



Seminar Hora Informaticae

Institute of Computer Science, Prague

Tuesday, November 8, 2022, 14.00 – 15.30 (2 - 3:30 PM) CET,

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

ZOOM:

<https://cesnet.zoom.us/j/95478234977?pwd=dXoyekFHbDJ0MkNrTjVVS3F2STZqUT09>

Meeting ID: 954 7823 4977 , Passcode: 712564

Jiří Wiedermann - Autonomous machines and machine minds

Autonomous machines, like robots and self-driving cars, are machines that perform behaviors or tasks with a high degree of autonomy (without, or with minimal external influence). Such machines operate and cooperate in the real world while coping with unpredictable external phenomena and error-prone technology. We model any autonomous machine as a cyber-physical system controlled by universal processors implementing so-called minimal machine consciousness, minimal machine experience, and minimal machine understanding.

In the talk, we present in more detail the respective model, define and justify our machine versions of consciousness, experience, and understanding and indicate their computational realization. We argue that endowing the systems with all these cognitive feats leads to more trustworthy, safer, and more reliable systems, by increasing their behavioral flexibility and autonomy in varying environments and by strengthening their resilience to operation or cooperation failures of their components or as a whole. The talk is based on joint research with Jan van Leeuwen, Utrecht, NL.

References:

Wiedermann, J., van Leeuwen, J. (2019): Finite State Machines with Feedback: An Architecture Supporting Minimal Machine Consciousness. In: Manea F., et al., Eds. *Computability in Europe: Computing with Foresight and Industry (CiE 2019)*, Lecture Notes in Computer Science, Vol. 11558, Springer, 2019, pp. 286-297, HTTP: [//www.springerprofessional.de/en/finite-state-machines-with-feedback-an-architecture-supporting-m/16879858](http://www.springerprofessional.de/en/finite-state-machines-with-feedback-an-architecture-supporting-m/16879858)

Wiedermann, J., van Leeuwen, J. (2021) Towards minimally conscious finite-state controlled cyber-physical systems: A manifesto. In: T. Bureš et al. (Eds.), *SOFSEM 2021: Theory and Practice of Computer Science*, Lecture Notes in Computer Science Vol 12607, Springer-Verlag, pp. 43-5,5 <https://www.cs.uu.nl/groups/AD/UU-PCS-2020-02.pdf>

Wiedermann, J., van Leeuwen J. (2021) Autonomous vehicles that cooperate and understand: Intelligent algorithms under the hood. TR UU-PCS-2021-01, Center for Philosophy of Computer Science, Department of Information and Computing Sciences, Utrecht University, <https://www.cs.uu.nl/groups/AD/UU-PCS-2021-01.pdf>

van Leeuwen J., Wiedermann J. (2021): Impossibility Results for the Online Verification of Ethical and Legal Behaviour of Robots, *TR UU-PCS-2021-02*, Center for Philosophy of Computer Science, Dept. of Information and Computing Science, Utrecht University, Utrecht, 2021, <https://www.cs.uu.nl/groups/AD/UU-PCS-2021-02.pdf>

Wiedermann, J., van Leeuwen, J. (2022). Validating Non-trivial Semantic Properties of Autonomous Robots. In Vincent C. Muller (ed.), *Philosophy and theory of artificial intelligence 2021*. Proceedings PT-AI 2021, SAPERE, Berlin, Springer. To appear. <https://www.cs.uu.nl/groups/AD/UU-PCS-2022-01.pdf>

Jiří Wiedermann belongs to the first generation of pioneers in computer science in former Czechoslovakia. Between 2000-2012 he served as the director of the Institute of Computer Science of the Czech Academy of Sciences. He is a member of the Learned Society of the Czech Republic and Academia Europaea. In recent years he focuses mainly on algorithms and models for AI inspired by modeling human higher-level cognitive abilities such as machine consciousness, experience, understanding, and other semantic properties of AI systems. For more information, cf. his homepage at www.cs.cas.cz/~wieder and <https://webpace.science.uu.nl/~leeuw112/list-vLW.html>

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>