PROF. DR. PAUL LUKOWICZ

Prof. Dr. Paul Lukowicz (born in 1967) has been Scientific Director and Head of the Research Unit “Embedded Intelligence” at the German Research Center for Artificial Intelligence (DFKI) in Kaiserslautern since 2011.

Paul Lukowicz studied computer science (Dipl. Inform.) and physics (Dipl. Phys.) at the University of Karlsruhe, where he received his PhD in 1999 on opto-electronic computer architectures. During his subsequent work at ETH Zurich, he established the Wearable Computer Group at the Institute of Electronics between 1999 and 2004. In October 2003 he was appointed Professor of Computer Science at the Private University for Health Sciences, Medical Informatics and Technology in Hall in Tyrol (UMIT). Three years later he accepted an appointment at the University of Passau, where he headed the Chair of Applied Computer Science with a focus on Embedded Systems and Pervasive Computing. In 2011 Paul Lukowicz accepted the call to the TU Kaiserslautern and since then he has been head of the newly founded research area “Embedded Intelligence” at DFKI. Since 2014 he also manages the DFKI SmartCity Living Lab.

His research focuses on cyber-physical systems, pervasive computing and social interactive systems.

The Challenge of Human Centric AI

The notion of “Human Centric AI” increasingly dominates public AI debate in Europe¹. It postulates a “European brand” of AI beneficial to humans on both individual and social level that is characterized by a focus on supporting and empowering humans as well as incorporating “by design” adherence to appropriate ethical standards and values such as privacy protection, autonomy (human in control), and non-discrimination. Stated this way (which is how it mostly appears in the political debate) it may seem more like a broad, vague wish list than a tangible scientific/technological concept. Yet, on a second glance, it turns out that it is closely connected to some of the most fundamental challenges of AI [1].

First of all consider the notion of “supporting and empowering” humans in the context of privacy protection and autonomy. Today a broad range of AI assistants exists: from smart loudspeakers through recommender systems, intelligent search engines and personalized news services to professional assistance systems e.g. for industrial applications. All of them struggle with privacy concern as most of the underlying Machine Learning (ML) techniques critically depend on having as much training data as possible. Thus, the ability to learn from as little data as possible, just as humans do, which is a core fundamental research question of ML [5], is also an essential component of the vision of Human Centric AI. Related is the problem of distributed, collaborative learning that does not require centralized collection of large amounts of possibly sensitive data.

Beyond privacy today’s AI assistants also run into fundamental limits with respect to the notions of “empowerment” and autonomy. Empowerment and autonomy imply that a system should help users make more informed decisions, pursue their own agendas more efficiently, and build up their own differentiated opinions. In other words systems should be able to truly, constructively elaborate and explore issues

with human users. By contrast today’s AI systems mostly provide the users with a limited set of recommendations to choose from, or collect and filter information, or try to prevent users from doing what the system considers as mistakes. In most cases the system reaches its decision in a “black box” like manner which the user has no way of understanding or arguing with. Thus, in a way, today’s systems are largely prescriptive and autonomy constraining rather than empowering and truly supportive. To change this AI systems must gain the ability to develop a differentiated understanding of human lines of reasoning, relate to human motivations, emotions, moral assumptions and implications in this reasoning, help human partners challenge their own assumptions as well as provide simulations with consequences and explain alternative “AI angle” on seeing the problem. They must be able to make their own reasoning transparent to the user and anchor it within complex differentiated world models grounded in both the physical reality and the user’s subjective perception of reality.

Transparency and explainability of ML systems together with the ability to reason within complex, differentiated world models are also core concerns when it comes to the adherence to ethical standards, fairness, and non discrimination. Thus, AI systems increasingly support or even make decision that have grave personal and/or social consequences. Examples are judges, doctors, policy-makers or managers who more and more rely on AI decision support or even decision making systems. The ability to challenge such decisions when they have impact on a person’s life is a fundamental ethical concern that can not be satisfied when the decisions are influenced or even directly made by a “black box” like AI system. Instead the AI systems must be able to translate their computation into an explanation that is accessible to a non expert. Such translation between the complex AI model

and related computation and simple non expert mental model anchored within a user's subject world view goes far beyond current state of the art in explainable ML [4].

Fairness and non discrimination are further issues where AI systems need to relate their own computation models to complex world models and the way humans perceive and judge real world situations. The problem is that, in most cases, discrimination and bias do not arise as a result of any objective errors in the respective AI algorithms. Instead, the systems do what they were designed to: build models based on patterns contained in the training data. Unfortunately training data often reflect social bias and unfairness of the real world which the system incorporates and often amplifies. To avoid this the systems must be able to go beyond mere statistical analysis and numerical optimization and relate the data to world models that reflect human ethical and moral values [6]. In a way it requires that the systems do not discover the actual statistical properties of the data but adjust what they discover according to what is desirable from an ethical/social point of view. Give the vagueness and fluidity of such an adjustment (we do not want the system to fully ignore the data and just produce outcomes that we like) this will in general not always be a fully automated process. Instead humans must be able to accompany and guide the learning process and specify high level boundary conditions and optimization goals. This goes beyond explainable ML towards the notion of interactive ML [2,3] where the learning process is a "co-creation" between the user and the AI.

In summary the notion of Human Centric AI should not

be seen as a potential regulatory roadblock to AI research but rather as a challenge involving basic open AI problems such as:

1. Comprehensive world models that, in their scope and level of sophistication, should strive for human-like world understanding
2. "Interactive AI" that allows humans to not just understand and follow the learning and reasoning process, but also to seamlessly interact with it and guide it.
3. Understanding and naturally interacting with humans and complex social settings within dynamic open world environments.
4. Reflexivity and expectation management

The above research challenge is at the core of the Humane AI initiative² which combines nearly 50 European AI labs of high renown in the effort to make the vision of Human Centric AI a reality.

References: [1] Nowak, Andrzej, Paul Lukowicz, and Pawel Horodecki. „Assessing Artificial Intelligence for Humanity: Will AI be the Our Biggest Ever Advance? or the Biggest Threat [Opinion].“ IEEE Technology and Society Magazine 37.4 (2018): 26-34. [2] Amershi, S., Cakmak, M., Knox, W. B., & Kulesza, T. (2014). Power to the people: The role of humans in interactive machine learning. AI Magazine, 35(4), 105-120. [3] Abdul, Ashraf, et al. „Trends and trajectories for explainable, accountable and intelligible systems: An hci research agenda.“ Proceedings of the 2018 CHI conference on human factors in computing systems. ACM, 2018. [4] Samek, Wojciech, Thomas Wiegand, and Klaus-Robert Müller. „Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models.“ arXiv preprint arXiv:1708.08296 (2017). [5] Snell, Jake, Kevin Swersky, and Richard Zemel. „Prototypical networks for few-shot learning.“ Advances in Neural Information Processing Systems. 2017. [6] Moor, James H. „The nature, importance, and difficulty of machine ethics.“ IEEE intelligent systems 21, no. 4 (2006): 18-21
¹ <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>
² <https://www.humane-ai.eu/>

Fotos: Privat

DIE IS4IT GRUPPE

EINE STARKE IT-MANNSCHAFT AUF EINEM SPIELFELD VOLLER MÖGLICHKEITEN!

BERATUNG

CLOUD

RECHENZENTRUM
& INFRASTRUKTUR

INFORMATIONSSICHERHEIT

ANWENDER-
SUPPORT

WORKPLACE
MANAGEMENT

IS4IT

IS4IT GmbH
Grünwalder Weg 28b
82041 Oberhaching
Deutschland
telefon +49 89 6389848-0
info@is4it.de

IS4IT
KRITIS

IS4IT KRITIS GmbH
Kraftwerkstraße 1
74847 Obrigheim
Deutschland
telefon +49 6261 31944-0
info@is4it-kritis.de

cyDIS CYBER DEFENSE
AND INFORMATION SECURITY

CyDIS Cyber Defense and
Information Security GmbH
Grünwalder Weg 28b
82041 Oberhaching
telefon +49 89 2488207-41
info@cydis.de

itsa 2019
08. – 10. Oktober 2019

Halle 10
Stand 311

kick IT
DAS TOP-TEAM IN DER SECURITY-LIGA