UNIVERSITÉ DE FRIBOURG
UNIVERSITÄT FREIBURG

DEPARTEMENT D'INFORMATIQUE
DEPARTEMENT FÜR INFORMATIK

INFORMATICS COLLOQUIUM

Speaker:

Asst Prof. Jie Yang, TU Delft, Netherlands

ARCH: Know What Your Machine Doesn't Know

Abstract:
Despite their impressive performance, machine learning systems remain prohibitively unreliable in safety-, trust-, and ethically sensitive domains. Recent discussions in different sub-fields of AI have reached the consensus of knowledge need in machine learning; few discussions have touched upon the diagnosis of what knowledge is needed. In this talk, I will present our ongoing work on ARCH, a knowledge-driven, human-centered, and reasoning-based tool, for diagnosing the unknowns of a machine learning system. ARCH leverages human intelligence to create domain knowledge required for a given task and to describe the internal behavior of a machine learning system; it infers the missing or incorrect knowledge of the system with the built-in probabilistic, abductive reasoning engine. ARCH is a generic tool that can be applied to machine learning in different contexts. In the talk, I will present several applications in which ARCH is currently being developed and tested, including health, finance, transport, and e-commerce.

Bio:
Jie Yang is Assistant Professor in the Web Information Systems (WIS) group at TU Delft. He co-leads the Kappa research line on Crowd Computing & Human-Centered AI at the WIS group and the Delft AI Lab Design@Scale. Before, he was a machine learning scientist at Alexa Shopping, Amazon Research (Seattle, US), and a senior researcher at the eXascale Infolab, University of Fribourg (Switzerland). He received his PhD from TU Delft in 2017, MSc from TU Eindhoven in 2013, and BEng from Zhejiang University in 2011. During his master program, he also spent some time at Philips Research.
Jie works on human-in-the-loop approaches for reliable and trustworthy machine learning. The underlying assumption is that to create AI that really serves the purpose of people, it is of key importance to involve stakeholders in the design and development of new technologies for every stage of the machine learning lifecycle. His research contributes a new set of human-in-the-loop methods and tools for the development and evaluation of, and the interaction with, machine learning systems. With such efforts, his utimate goal is to transform machine learning into an engineering discipline that gives humans the full control of AI such that it can be reliably and safely used in all kinds of contexts.

Date and time: *Wednesday May 18th, 2022, 4.00pm*
Location : *Pérolles 21, room C130, Bd de Pérolles 90, Fribourg*
Contact person: *Prof. Philippe Cudré-Mauroux*

The colloquium is free and open to the public.