The BMVA Technical Meeting “Cognitively inspired perception eXplainable perception-based AI” was held at the BCS headquarters in London, 7th February 2018. The workshop, organized and chaired by Assoc. Prof. Serge Thill (University of Plymouth, UK) and Assoc. Prof. Maria Riveiro (University of Skövde, Sweden) included three keynote speakers, seven presentations (including both theoretical and practical aspects of eXplainable AI) and concluded with a panel discussion.

Explainable AI (XAI) refers to AI systems that behave or provide necessary information so that their inner workings become comprehensible to the human user. Due to the fact that AI systems are increasingly present in everyday society, from simple computer systems to agents such as autonomous vehicles or social robots, several researchers have noted that it is critical to understand how human users perceive such systems - in particular, the degree to which they understand how the system works, and what mental models they build of the underlying algorithms. For this meeting, we focused on AI systems that operate at least somewhat autonomously based on real-world sensory data (in particular, based on machine vision). Therefore, the main goal of this highly interdisciplinary one-day BMVA workshop was to bring together contributions from the fields of robotics, AI, cognitive science, machine vision and HCI to outline the current state-of-the-art of XAI and talk about future challenges.

This workshop has been supported by The British Machine Vision Association and Society for Pattern Recognition (BMVA) as well as the research projects Dreams4Cars (H2020 grant agreement no 731593) and BIDAF (A Big Data Analytics Framework for a Smart Society, The Swedish Knowledge Foundation).

KEYNOTES

We were delighted to have three amazing keynotes speakers in this workshop, Dr. Alessandra Sciutti (Istituto Italiano di Tecnologia, Italy), Assist. Prof. Brad Hayes (University of Colorado Boulder, USA) and Prof. Yiannis Demiris (Imperial College London, UK) who provided their view on XAI from a robotics and human-machine collaboration perspective.

Alessandra Sciutti started the workshop presenting her and hers group current research aimed at defining which features of human and robot motion allow for natural mutual understanding, referencing to low-level kinematics properties (as biological motion) and higher-level cognitive aspects (as intention reading and attribution). Alessandra reflected on the fact that we are still far away from fully interactive robots, but we are moving from isolated to robots in society, but many challenges need to be solved. Alessandra elaborated on many relevant issues for XAI, like intuitiveness, implicit signals...
revealing internal states of collaborating partners, prospective signals, predictability and communication of intent. Brad Hayes started the afternoon session talking about robot collaboration and teamwork. He addressed relevant aspects of his research with supportive robots for XAI, such as communication of shared expectations, interpretability of models, reasonable behavior, learning by demonstration, etc. Another interesting discussion brought up by Brad was, e.g., how do we make the human to have a theory of mind about the robot? Yiannis Demiris closed the speakers’ sessions talking about personalized robot assistance and human modeling in areas like robotic wheelchair assistance and learning. He focuses on robots that help and provide support, not capability replacement, e.g., for children with disabilities or elderly. He suggested that explanations will be more understandable if presented closer to the internal models of the user; thus, we need to create personalized mental models of the users in each case. He coincided with Alessandra and Brad in bringing important aspects like intention recognition.

THEORETICAL ASPECTS OF XAI

Tarek Besold (City, University of London) characterizes four notions of explainable AI, focusing on the systems: (1) opaque systems (offer no insight into their algorithmic mechanisms), (2) interpretable systems (users can mathematically analyze the algorithmic mechanisms), (3) comprehensible systems (emit symbols enabling user-driven explanations of how a conclusion is reached) and finally, (4) truly explainable systems (where automated reasoning is central to output crafted explanations without requiring human post-processing as final step of the generative process). Moreover, Tarek presented some considerations on what a set of minimal requirements might be that an AI system’s output has to meet to meaningfully be considered "an explanation" in the context of human-machine interaction and collaboration. Henrik Svensson (University of Skövde) suggested the application of the situated and distributed cognition perspective to XAI. He provided several examples from the area of autonomous cars, and his & colleagues’ research in the Dream4Cars project. Henrik also suggested that more collaboration of traditional fields like AI with UXD designers is needed for the advancement of this area in the future. Tove Helldin (University of Skövde) provided a brief introduction to evaluation and metrics for explainable AI, providing lessons learned from other areas like HCI, robotics, information visualization, recommender systems, etc. She presented examples from her own research and other relevant examples from the literature.

GENERAL APPLICATIONS OF XAI

Elin A. Topp (University of Lund) started the morning session after Alessandra talking about how to generate meaningful clarification requests. She presented her research with industrial robots and her work on, e.g., extracting knowledge from experts to teach robots to do certain tasks. She discussed relevant issues for XAI like the need for the robot to understand what people actually mean (not say), action representation, etc. During her research, they could show that a close monitoring of the overall interaction context, including user behavior, gestures, robot positioning, and a short-term history over the interaction, allows to identify patterns in the interaction that can be used to detect ambiguities between what is said explicitly and shown implicitly, which can be exploited to
generate meaningful clarification requests. **Szonya Durant** (Royal Holloway, University of London) presented her and Prof. Kostas Stathis’ research on integrated cognitive user assistance for multiple-display systems. The aims are to reduce information overload due to the number of displays, the rapidly changing information in individual displays and the dependence of information between displays, which challenge human attention limits. For that, they focus on methods based on eye-tracking, where location and duration of fixations are used to measure locus and timing of user attention. Szonya presented the potential of creating multi-display interfaces that alert users and check if they have paid attention, using cognitive human-like agents able to control display parameters and detect eye movements to model the joint activity of the user and the display environment. **Sylvester Kacmarek** (Imperial College) presented a series of works and studies carried out with several companies in the field of applied AI focusing on transparency, trust, reinforcement learning, etc. Finally, **Swen Gaudl** (Metamakers Institute, Falmouth) brought to the workshop a “games” perspective and talked about XAI, games, genetic programming and human mimicry. Swen presented an agent framework approach using genetic programming to derive new reasoning agents that can be interpreted by human users.

**PANEL DISCUSSION AND FINAL REMARKS**

The meeting concluded with a panel discussion by the three keynote speakers. Several issues were discussed, such as why XAI? What does the user want to know from the system (wants to be able to predict the system behavior?) or lessons learned from other areas, such as decision support systems.

From our point of view, it was a very interesting meeting that engaged fully the participants in interactive discussion with the presenters, bringing together researchers and industry representatives from very disparate areas such as Robotics, AI, HCI, Computer Vision, Cognitive Science, etc. It is evident that there are many exciting challenges ahead for XAI, which need interdisciplinary solutions and the cooperation of several scientific areas.