

Application for a 2-year TDE Research Fellowship: Mutual understanding of humans and machines

Prof F. Cuzzolin (Visual AI Lab), Dr M. Rolf (Robotics) and Dr A. Bradley (Autonomous Driving).

Significance. The project concerns the design and development of novel ways for robots and autonomous machines to interact with humans in a variety of emerging scenarios, including: human-robot interaction, autonomous driving, personal (virtual or robotic) assistants. In particular, we believe novel, disruptive applications of AI require much more sophisticated forms of communication between humans and machines, something that goes far beyond conventional explicit and linguistic exchange of information towards implicit non-verbal communication and understanding of each other's behaviour.

Within this scenario, we wish more specifically to address the following **Research Issues**:

- How can machines detect a human communication attempt;
- What effective ways may a machine have to initiate communication with humans;
- How can machines put themselves in a human agent's shoes, in order to understand their reasoning and predict their future behaviour (an ability called 'theory of mind') in order to be able to promptly react;
- How to use the above abilities to develop a new class of empathic, moral artificial agents.

Motivation. For example, smart cars need to understand that children and construction workers have different reasoning processes that lead to very different observable behaviour, in order to blend in with the road as a human-centered environment. Empathic machines have the potential to revolutionise healthcare, by providing better care catering for the psychological needs of patients. Morally and socially appropriate behaviour is key in all such scenarios, to build trust and lead to acceptance from the public.

Exciting research is currently going on in moral robotics and AI, including moral development (how a robot can learn moral principles), fairness and bias in, for instance, AI-assisted recruitment. As smart cars head towards real world deployment, the field is shifting from mere perception (e.g. SLAM) to higher-level cognition tasks, starting from the automated detection of road events. Holographic AI is going to revolutionise the field of personal assistants, but needs effective communication interfaces.

We believe this to fit **School/Faculty objectives**, as it is placed in the context of current strong efforts (the list is not exhaustive):

- Our recent outline application (PI Cuzzolin) for a UKRI funded CDT in Human-centric AI and follow-up initiatives;
- Recently obtained HEFCE funding (Robotics) for the exploration of industry networks in AI implications;
- The Cognitive Robotics' group effort towards developing socially and morally appropriate robots, involving one PhD student and one upcoming MSc one;
- The Autonomous Driving research group's recent Innovate UK application (PI Bradley, CI Ball)
- Ongoing research work on human-vehicle interaction between Engineering and HLS (Bradley, Prof Dawes)
- A bid in preparation for ESRC on ethical AI interview in collaboration with Marilena Kyriakidou (Rolf).

Proposed Methods. Cuzzolin is exploring the design and implementation of a machine theory of mind model based on a simulation approach, in which input stimuli drive an agent-specific simulation of their mental states. Simulations are implemented as reconfigurable deep neural networks, learned by reinforcement learning. Closely related to this, Rolf is investigating socially-originated rewards for reinforcement learning, including pre-linguistic cues such as face detection, synchrony and contingency, as well as investigating robotic moral issues. Both research directions are directly applicable to autonomous driving – the Visual AI Lab is currently providing road event and agent activity annotation for the Oxford RobotCar dataset (<http://robotcar-dataset.robots.ox.ac.uk/>) which is bound to have a significant impact on the field, as the first such benchmark. The benchmark will be released in October 2018. In the first year of the project, the Fellow would implement reinforcement learning based machine theory of mind models and test them on the new data to provide a proof of concept.

Bradley has been working in the area of vehicle simulation for many years, and also upon driver behaviour analysis using a driving simulator (with Prof Helen Dawes). Bradley is currently working with Dr Peter Ball on areas of modelling autonomous vehicle behaviour, resulting in a recent Innovate UK application for Connected and Autonomous Vehicle (CAV) simulation.

External Funding Targets. One of the key objectives of the project is for the work of the Fellow to significantly support and facilitate our effort to gather external funding on this topic. We will target the following funding streams (with dates and size):

1. A Leverhulme Trust Research Grant application by Cuzzolin as the PI, with Co-I Prof Barbara Sahakian of the Neuroscience group of Cambridge University, entitled "Theory of mind at the interface of neuroscience and AI". The funding requested is £260,000 over 30 months. The grant would allow us to hire an additional Level 8 postdoc to work on the topic in an interdisciplinary context. Deadline for submission: September 2018. Possible start date: June-September 2019.
2. Sahakian is proposing a follow-up applications on moral AI agents in the healthcare setting (see above), to be submitted to the Wellcome Trust. This is still in early stages though.
3. An Innovate UK application to the Open grant funding competition: round 2 led by Cortexica Vision Systems, and with the participation of Cisco Systems, is in preparation with target submission date September 12 2018. The topic is efficient event detection on board, and as such could have significant impact in the autonomous driving context. Budget is being discussed. A first draft should be ready by August 23.
4. The Visual AI Lab's is engaged with University of Naples Federico II to seek funding on 'Event recognition for autonomous driving'. We currently have 2 MSc students involved, and a pending PhD studentship bid with Fiat Chrysler Automobiles whose outcome should be known by October 2018;
5. A bid is getting ready to be submitted to the EPSRC "Healthcare Technologies" theme by Cuzzolin (PI) with Prof Helen Dawes and Dr Derick Wade (CoIs) and a number of clinical partners working at 4 different sites on the intelligent audio-visual monitoring of patients in a coma. All documents are in place, internal reviews have been gathered and actioned upon. Budget: around £1,500,000. Duration: 3 years. Submission deadline: October 2018.
6. Cuzzolin will submit an EPSRC Fellowship application on the topic of machine theory of mind by February 2019. This was indeed the topic of his Professorial Lecture in January 2018, and is a key part of his medium term research plan. The bid has

received internal reviews already, and is being revised as we speak. Budget: around £2,000,000 (four PDRAs for 5 years). Deadline: February 2019.

7. A Horizon 2020 bid, entitled “Smart Data Analytics” (SMADA), led by KTH Sweden (Prof Mihhail Matskin) with a consortium of 12 partners. This was submitted in April 2018 and received a score of 10/15. Core components of this application are a Machine Reasoning and an Explainability components which leverage on the concepts of human-centric AI and machine theory of mind. Budget: 6M euros (Brookes’ share: 650,000). Deadline: April 2019. Possible start date: Jan 2020.
8. An additional Horizon 2020 application for the Artificial Intelligence funding stream is being discussed with Rita Cucchiara, head of the Italian AI Institute, which covers almost exactly the aims of this ECM-funded project. We are aiming for an April 2019 submission, as we are now forming a consortium.

People and Management. Our ambition is for the School to take the lead in these emerging research avenues (moral and empathic machines, machine theory of mind), by leveraging the existing pockets of excellence and laying the necessary groundwork in terms of both cutting-edge research and of engagement with relevant external partners and the public. Indeed, Cuzzolin is making this the focus of his five-year research plan in an upcoming EPSRC Fellowship application.

The proposers have extensive experience of successfully managing research projects (including various H2020, EPSRC, Innovate UK grants). The recently-formed Autonomous Driving research group aims to bring together the complementary expertise of Computing and Engineering to form a new area of strength in the areas of autonomous vehicles, ensuring the continued success of the Automotive and Motorsport degree programmes. The Visual AI Lab is leader in deep learning for action detection, has been awarded 1.3M in external funding in the last 12 months (including the SARAS H2020 project, £280,000 from Huawei Technologies, a KTP and a PhD studentship funded by Cortexica), and is projected to comprise around 20 people in 2019.

The Research Fellow will be part of a wider coherent effort in this sense, in close coordination with all four groups, as a first step towards a permanent cluster. Project plans and progress will be monitored on a weekly basis in meetings with the group leaders, and reported at regular group meetings and presentations.

Dissemination and Impact. Beneficiaries range from patients in healthcare, to drivers and users of our roads, to the general public. We already are in contact with a number of potential industry partners, including Roborace, StreetDrone, BMW, Zeta Automotive, Cortexica, Huawei, Prophesee. The work will also allow us to engage academic partners for external funding application (beyond those listed above), including Oxford University, Bielefeld University (DE), Osaka University (JP), the Italian Institute of Technology (IT), the University of Malta. The group are already working with Oxford County Council on the development of Smart Cities and CAV infrastructure, which provides an excellent opportunity for local development funding. The applicants have contacts with the Oxford NHS Trust that may be exploited for further work on empathic healthcare, possibly in the context of the current VC-lead Expanding Excellence application.

Finance

Staff. We envisage a Research Fellow Grade 8 (Starting salary: £30,688), for total employment costs over 26 months of £89,537.

Supervision time will be provided by the applicants as part of their research allowance, at no extra cost.

Travel. Of the remaining 10K, travel for conference attendance and networking is requested for around £5,000 (two international symposia and 1-2 European trips for conference or networking over the 2 years).

Equipment. £2,000 will be used to purchase a standard laptop. £3,000 pounds will be employed to purchase LIDaR equipment to explore the fusion of visual and 3D information while understanding the scene and its actors/activities, as standard practice in autonomous driving.

The Visual AI Lab will provide access to the necessary computational power in the form of several GPU-endowed workstations (of which two with 4 GPUs). The Robotics group can provide a number of humanoid robotic platforms that will be instrumental to the project. The Autonomous Driving group will provide access to vehicle simulation facilities, and a Driver-in-Loop simulator.