**BU PhD STUDENTSHIPS 2018**

**PROJECT DESCRIPTION**

<table>
<thead>
<tr>
<th>PROJECT DETAILS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Project Title</strong></td>
</tr>
<tr>
<td>Towards Explainable Artificial Intelligence (XAI): software design and development on top of AI algorithms to enable explainability of AI decisions</td>
</tr>
<tr>
<td><strong>Project Summary</strong></td>
</tr>
<tr>
<td>In collaboration with British Telecom (BT) Adastral park which is BT’s Global Research and Development Headquarters in Ipswich, this project will investigate the design and development of appropriate software on top of AI algorithms to ensure AI “explainability” to a range of audiences from AI experts to business managers.</td>
</tr>
</tbody>
</table>

BT’s research team has strong interests and experience in autonomic and self-adaptive algorithms in different application scenarios that are ranging from networks, medical applications, business processes, and personalisation in several application sectors. Lately, they are interested in the area of making all these algorithms and their results more understandable, accessible and therefore “explainable” both to AI experts teams and business managers, internally and externally. This is particularly important from a strategic point of view, as AI algorithms and their adoption heavily depends on the confidence they can instil.

Explainable AI (XAI) is an initiative originated by DARPA in 2016. The main goal identified was the production of machine learning models which, using appropriate explanation techniques, will enable users to better understand and ultimately trust the model’s process and results. Two types of explainability are generally recognised in the literature: ante-hoc explainability that is embedded inside the operation of the AI algorithms leading to explainability by design, this applies to simpler AI algorithms with less accurate results; and post-hoc explainability that is performed after the end of the AI algorithms and is used for complex neural-networks type AI algorithms with more accurate results. In this project, a hybrid approach will be followed that will combine the strength of both the above techniques. We will mainly aim for explainability by design, however, maintaining the accurate results of complex AI algorithms where this is appropriate and inevitable.

The method we will use is the design and development of “software for adding explainability” on top of a range of AI algorithms, utilising systems engineering methods such as abstraction, model-based development using SysML, patterns and framework development.

The proposed project will focus on the emerging and challenging research area of software development for AI explainability. The supervisory team is carefully selected in order to combine expertise in the above areas (BT-autonomics, AI, BU – systems and abstraction, software engineering and software development) and all areas are equally important for the “successful” completion of the project.

Therefore, the proposed project has the following objectives:

1. Research related to AI explainability. Both ante-hoc and post-hoc explainability will be considered.
2. Analysis of BT’s autonomic applications with emphasis on their explainability requirements.
3. Software design and development to enable XAI.
4. Assurance of explainability properties through appropriate Verification & Validation techniques.

The project will be mainly driven by the industrial partner’s requirements. BT will set specific application scenarios in the course of the project execution thus providing "real-world" case studies as well as dissemination of the project's outcome. BT’s autonomies team, business managers and external collaborators will be also involved for testing the results of the project.
This project aims at innovative and timely work inline with EU societal challenges for explaining AI decisions (new European General Data Protection Regulation (GDPR 2016/679 and ISO/IEC 27001) entering into force on May 25th 2018), making XAI a necessity for all future AI-based systems.

**Academic Impact**

The last decade has seen a significant increase in autonomies/AI research and issues such as trust in the decisions made by AI algorithms are becoming hot topic both in industry and academia. AI research results’ adoption mainly depend on its ability to inspire confidence to its potential users/adopters. On the other hand, systems and software engineering have made huge advancements in abstraction and embedding several aspects such as security, trust at the design level. Explainability by design would be a promising next step for this research. Therefore, this PhD research work is timely and innovative dealing with important research questions. There are at least four types of academic impact that will be achieved through this project: (1) Scientific output should be published in high-impact journals; (2) Networking: The candidate will collaborate with researchers from BU, the BT’s research headquarters in UK; (3) Prototype: The project will deliver fully functional software on top of AI algorithms that will provide AI explainability. (4) Participation in BU-BT bids for XAI.

**Societal Impact**

XAI research has huge societal impact as it directly affects confidence and trust in AI decisions. Legal and privacy aspects depicted in the new European General Data Protection Regulation (GDPR 2016/679 and ISO/IEC 27001) entering into force on May 25th, 2018, will make black-box approaches difficult to use in business, if they are not able (on demand) to explain why a decision has been made. Therefore, XAI will be required by all AI algorithms in all application areas that AI is being utilised. For example, BT’s autonomic team is currently applying autonomic algorithms in several application areas from medical diagnosis and prediction to business processes and business decision making. Transparency, confidence, trust through explainability in the algorithms’ decisions and steps are of vital importance to the final adoption of these algorithms. Wider and safer adoption of AI can contribute tremendously to the society of the future where AI and emerging technologies play a vital, transparent and accountable role.

**Training Opportunities**

This project offers very interesting learning and training opportunities for the candidate. Different subjects are involved in the project like autonomies & AI, interface design and software design & development. XAI is a very “trendy” and recent area of research that interests several industries. As such, it is anticipated that XAI will be part of research proposals from BU and BT where the candidate will be able to contribute and participate. The candidate will be exposed to the academic and industrial environments having access to periodic secondments at BT. While at BU, he/she will receive research-based training, at BT, he/she will get familiar with research and development in the industrial setting and will be given insights into the exploitation of the results of this project. Additionally, the candidate will be able to participate in periodic conferences/events organised by BT’s Tommy Flowers Institute which is a cross-industry initiative initiated by BT that brings together many companies, government bodies and academia to produce research leaders of the future. Moreover, the candidate will have the opportunity to interact with local and international researchers through supervisory team’s contacts. Last but not least, he/she will be expected to attend international conferences and relevant summer schools.

**SUPERVISORY TEAM**

<table>
<thead>
<tr>
<th>First Supervisor</th>
<th>Dr Sofia Meacham</th>
</tr>
</thead>
<tbody>
<tr>
<td>Additional Supervisors</td>
<td>Prof. Keith Phalp, Dr Angelos Stefanidis, BU Visiting Prof. Detlef Nauck</td>
</tr>
</tbody>
</table>


INFORMAL ENQUIRIES

To discuss this opportunity further, please contact Dr Sofia Meacham via email: smeacham@bournemouth.ac.uk

ELIGIBILITY CRITERIA

The PhD Studentships are open to UK, EU and International students. Candidates for a PhD Studentship should demonstrate outstanding qualities and be motivated to complete a PhD in 4 years and must demonstrate:

- outstanding academic potential as measured by either a 1st class honours degree or a Master’s degree with distinction or equivalent Grade Point Average (GPA)
- an IELTS (Academic) score of 6.5 minimum (with a minimum 6.0 in each component) for candidates for whom English is not their first language

In addition to satisfying minimum entry criteria, BU will look closely at the qualities, skills and background of each candidate and what they can bring to their chosen research project in order to ensure successful completion.

Applicants will be asked to submit an online application form and a proposal (c. 1500 words) outlining their understanding of the project for which they are applying, the approach they would envisage taking and what qualities they will bring to the research community.

Please note:

- Current BU Doctoral students are not eligible to apply for a Studentship
- Current MRes/MPhil students can apply, subject to satisfactory completion of their Research Degree prior to being able to take up the award
- PhD Studentships cannot be used to support BU staff to complete doctoral programmes

HOW TO APPLY

Please complete the online application form by Sunday 1 July 2018. Further information on the application process can be found at: www.bournemouth.ac.uk/studentships