Two Post-Docs in Operationalizing AI Ethics at TU Delft

Artificial Intelligence is used more and more in society, from healthcare to government decisions and recruitment. Along with the rapid increase of AI adoption come increased concerns about the inherent shortcomings of such technologies (e.g., robustness) and the social, and ethical implications. To create AI systems that can properly serve humans, it is crucial to put humans at the centre of the process such that the outcome system behaves in a way that fits the values and needs of people. This poses new challenges both to philosophical work on values and to (responsible) technological development. What (ethical) requirements should these systems adhere to? What does such 'adherence' mean and how can we demonstrate and validate claims regarding adherence? How to build AI systems that can be understood by humans and that can align their behaviour with human values? Tackling these challenges requires bridging the gap between philosophy and computer science within AI Ethics.
TU Delft has world-leading expertise in the operationalization of AI Ethics and now looks to strengthen this by hiring two three-year post-docs, one in philosophy and one in computer science, that will be closely collaborating on topics within AI Ethics of their choice.

The philosophy post-doc is embedded in the TU Delft Digital Ethics Centre which focuses on translating philosophical values into design requirements. Here, research is done on the methodology of embedding values, as well as on the specification of moral and epistemic values: tudelft.nl/digitalethics.
The computer science post-doc is embedded in the Web Information Systems (WIS) group, a world-class research group on data management, human-centered AI, information retrieval, and user modeling. The post-doc will work independently on research that can be well embedded in WIS; they will also contribute to the management of AI projects.
Both post-docs are also connected to the activities of the TU Delft AI Initiative and the national Digital Society program.

There are strong interdisciplinary connections already present, with joint supervision and publications already taking place. The joint research of the post-docs will furthermore be supported by dr. Jie Yang (computer science) and dr. Stefan Buijsman (philosophy) as well as Prof. dr. ir. Geert-Jan Houben (computer science) and Prof. dr. Jeroen van den Hoven (philosophy).

Practical Information:
2 Post-docs for 3 years
Salary €2.960,00 - € 4.670,00
Requirements: a PhD in a field related to the topic of the post-doc, so either related to philosophy or to computer science
The willingness to work in interdisciplinary teams

Apply by sending your CV and a motivation letter that briefly describes why you apply for this position, your research interests, and your fit to the two groups.
For the computer science position, send this to Dr. Jie Yang: j.yang-3@tudelft.nl
For the philosophy position, send this to Dr. Stefan Buijsman: s.n.r.buijsman@tudelft.nl

Computer Science Post-Doc: Knowledge and Language in Human-Centered AI

Job Description

Artificial Intelligence is used more and more in society, from healthcare to government decisions and recruitment. Along with the rapid increase of AI adoption comes increased concerns about the inherent shortcomings of such technologies (e.g., robustness) and the social, and ethical implications. To create AI systems that can properly serve humans, it is crucial to put humans at the center of the process such that the outcome system behaves in a way that fits the values and needs of people. This poses new challenges to technological development: how to build AI systems that can be understood by humans and that can align their behaviour with human values? Tackling these challenges requires new ways of looking at AI systems, e.g., machine learning models as knowledge bases and as autonomous agents that people can query, interact with, and influence.

This post-doc position focuses on the knowledge and language aspects in human-centered AI. It is up to the candidate to choose specific challenges to work on, provided that there is a clear component of knowledge or language and that the technical work connects with ethical requirements.

As support in this work, the post-doc is embedded in the Web Information Systems (WIS) group, a world-class research group on data management, human-centered AI, information retrieval, and user modeling. The post-doc will work independently on research that can be well embedded in WIS; they will also contribute to the management of AI projects. In particular, the post-doc will work in close collaboration with a post-doc in philosophy on interdisciplinary projects, doing joint research and co-supervising master students.

There are strong interdisciplinary connections already present and the joint research will be supported by dr. Jie Yang (computer science) and dr. Stefan Buijsman (philosophy) as well as Prof. dr. ir. Geert-Jan Houben (computer science) and Prof. dr. Jeroen van den Hoven (philosophy).

Job Requirements

The successful postdoc candidate should have:

- A Phd degree in related AI/CS fields such as databases, knowledge representation, machine learning, natural language processing, or equivalent;
- Strong conceptual and analytical skills;
- Excellent research, academic writing, and presentation skills;
- Strong communication, collaboration, and coordination skills;
- The interest and capacity to work in interdisciplinary teams with philosophers;
- Excellent command of the English language, both spoken and written.

Philosophy Post-Doc: Operationalizing AI Ethics

Job Description

Artificial Intelligence is used more and more in society, from healthcare to government decision and recruitment. The enormous benefits it can bring, however, also come with ethical risks. Machine learning systems can lead to discrimination challenging us to ensure fairness, are opaque and thus require new explainability tools, and their autonomy complicates the goal of meaningful human control over these systems. These values, and other ethical values we have for AI systems, now need to be expressed into design requirements that can be prototyped and tested to ensure that resulting AI systems are developed and used responsibly. This Post-doc position focuses on the philosophical side of the operationalization of these ethical values. It is up to the candidate to choose specific challenges in AI Ethics to work on, provided that there is a clear component of operationalizing these values and connecting the philosophical work to technical implementations.

As support in this work, the post-doc is embedded in the TU Delft Digital Ethics Centre, a world-leading centre of excellence on the operationalization of ethical values. Here, research is done on the methodology of embedding values, as well as on the specification of moral and epistemic values: tudelft.nl/digitalethics.

Additionally, the post-doc will work in close collaboration with a post-doc in computer science on interdisciplinary projects, doing joint research and co-supervising master students. There are strong interdisciplinary connections already present and the joint research will be supported by dr. Stefan Buijsman (philosophy) and dr. Jie Yang (computer science) as well as Prof. dr. Jeroen van den Hoven (philosophy) and Prof. dr. ir. Geert-Jan Houben (computer science).

Job Requirements

You will have completed, prior to appointment, a PhD degree in philosophy. In addition, knowledge of computer science and the ability to collaborate with computer scientists on the operationalization of AI Ethics is a core part of the position. We therefore look for candidates with:

- Strong conceptual and analytical skills
- Strong communication, collaboration, and coordination skills;
- The interest and capacity to work in interdisciplinary teams with computer scientists.
- Strong research, academic writing and presentation skills.
- Strong command of the English language, both spoken and written