

Responsible Digital Transformations – Call for Proposals 2022

Responsible Digital Transformations (RDT) is one of the four themes in the 2021-2026 strategic plan of the University of Amsterdam (UvA), focusing on technology and ethics of regulation, algorithms, systems and platforms; their impact on trust, dependence and equality in society. The first year of this theme will be devoted to the following three activities: setting the foundation of the moonshot project *Towards an AI4Society Sandbox*, establishing an UvA-wide platform to make the multitude of initiatives within the UvA in this domain visible, developing a UvA-wide call for proposals that can strengthen inter-faculty collaborations at the interface of existing initiatives with an eye to collaboration with external partners.

Towards an AI4Society Sandbox

As new AI systems are developed, it is often not immediately clear how they will function within their domain, not to mention the implications that they may have for society and economy more broadly. The same is true for interventions, be that legal, economic, or organizational ones, with the goal of steering technology development into a particular direction. With the AI4Society Sandbox moonshot project, the UvA steering group RDT will start setting-up a scenario-modeling platform that will allow UvA researchers, in cooperation with societal stakeholders, to develop, discuss, and test different scenarios of technology development and forms of interventions.

Originally stemming from software development and the cybersecurity field, sandboxing is a way to emulate a safe production or user environment to test applications upon its security, compatibility with a system, undesired side-effects, or malfunctions. We will extend the sandboxing method to develop future scenarios, probe technological and regulatory solutions discursively for their legal, societal, and ethical implications. This will help create a space for co-designing novel solutions and approaches. Researchers from different faculties are asked to investigate which technology, social intervention, regulation, etc. are needed to realize possible scenarios in a responsible manner in order to address challenges such as inequality or lack of opportunities. How can we adopt these technologies for the widest benefit of society and what happens if we lay the core responsibility/accountability for a certain technology with one particular actor or selected set of few? Can we arrive at a more intelligent division of responsibility?

The AI4Society Sandbox moonshot project aims to build on relevant research initiatives within the UvA. The Institute of Advanced Studies, for example, already has experience with modeling and simulating future scenarios through virtual sandboxes and programmable futures, in which researchers, policy-makers, and others can evaluate interventions ex-ante to assess the likely outcomes of interventions and the various uncertainties associated with those different interventions. Because of the strong societal relevance, the UvA steering group RDT encourages cooperation with external organizations such as the City of Amsterdam, and connection to existing initiatives in the city as well as to various societal actors. This demonstrates the broad relevance of the moonshot project, and further test the platform/sandbox/modeling environment upon its robustness.

Call for junior researchers 4 x 0,3fte (July-December 2022)

The UvA steering group RDT seeks 4 early-career UvA researchers (postdoc or UD level) to set the foundations for the AI4Society Sandbox. The researchers are ideally from various UvA

faculties, expected to form an interdisciplinary team and set-up the sandbox using specific topics such as:

- Digital infrastructure is an important topic that has generated broad interest in the research community, which has developed new methodologies to study the importance of infrastructure as a whole and begun to define its own needs of research infrastructures, which have on the national and European level received large amounts of funding. Despite these efforts, the Covid pandemic has shown how much researchers still depend on the critical infrastructure provided by big Internet companies, which generates new kinds of dependencies. The UvA's rector has therefore called for regulatory steps to protect 'independent and public knowledge' and reduce the dependence on large tech platforms. We are looking for a sandbox to understand current limitations to define infrastructures by researchers for researchers and/or future participatory infrastructures for the society as a whole.
- 2022 is the year that the European Union will start working together with its member states on its AI regulation following on to its earlier work on the General Data Protection Regulation. The question of whether and how we should regulate AI systems raises many socially and economically important questions but also research questions at the intersection of computer science, law, humanities, economics, society, and behavioral sciences. The regulation includes many assumptions on how such technologies can be best organized for the benefit of society. One example is the new right to authenticity that the regulation is introducing, the right to know whether you are communicating with a bot. This case study uses the sandbox/modeling platform to critically examine the philosophical underpinnings of such a right, avenues for its technical realizations, scenarios of how the algorithmic world would look with/without such a right and what kind of societal/economic/behavioral changes it can cause.
- Interactions in online social media are now ubiquitous, with a growing number of citizens and organizations using these platforms to share contents, debate and get access to information. At the same time, these platforms are increasingly impacted by AI-driven applications such as news filtering algorithms, link-recommendation algorithms, bots, or algorithmic content moderation. Here, we plan to focus on developing a sandbox to test the long-term impacts of these algorithms on opinion polarization, spread of misinformation, and the viability of human collective action to solve socio-economic and cultural challenges. Sandboxing will allow to test future scenarios where algorithms are adapted (e.g., link-recommendation algorithms following different heuristics), interventions are considered (e.g., signaling contents that (re-)shared too many times) or new regulation is adopted (e.g., to counter bias and inequality in online representation and visibility of different population groups).

The 4 junior researchers will devote 0.30 fte research time (teaching-release) between July – December 2022 to set up the platform/sandbox using the above or other suitable topics as a starting point, and with the goal to deliver:

- An inventory of existing sandboxing practices at UvA and beyond (qualitative, quantitative, and mixed) approaches, infrastructures, tools, and other (computing) resources, as well as needs, for modeling digital transformations and simulating future scenarios.
- Design requirements for a collaborative cross-faculty AI4Society Sandbox platform that balances, calls for innovation, disruption, and technological change with participation, deliberation, and legitimacy.

- Establish a first network and community of stakeholders (researchers, students, citizens, civil-servants, policy-makers, etc.), through workshops and other events, for example organized at and with support of the IAS.
- Completion and implementation of three test cases for AI4Society Sandbox to probe future scenarios, frameworks, interventions, and their legal, societal, and ethical implications.

Application information

Interested UvA researchers can apply by sending a short CV and a 1-page motivation letter as one PDF-file to subsidie-fgw@uva.nl before May 3th. Please, clearly describe the Sandbox topic you are planning to work on.

Ideal starting date is July 1th.

The UvA steering group RDT will select candidates based on proven affinity with the topic, experience in setting up and conducting interdisciplinarity and affinity with the subject matter, interdisciplinary research, ability to work independently and in a team.

The selected candidates will be asked to report and provide the lightest possible accountability to the Steering Group RDT before the end of the calendar year, which provides insight into the use of funds and substantive progress.