**Workshop proposal for PHTR 2020:**

**How can we integrate ethical deliberation in the design of Artificial Intelligence?**

Given the enormous actual and potential impacts of Artificial Intelligence (AI) on people's daily lives and on societies, the need for ethical deliberation is clear (Van Belkom 2019a, b). There are, however, numerous practical questions regarding the *how* (Dignum 2019, Mittelstadt 2019, Morley et al. 2019): *How* can we integrate ethical deliberation into the process of designing, developing, deploying and using AI systems? In this workshop we will playfully engage with this serious question.

We will work with a group of approx. 20 people; ideally people with mixed backgrounds, e.g., philosophy, technology and social sciences. We will create 4 multidisciplinary teams and give them the same assignment: to sketch and design a trustworthy AI-system (to be defined), using various tools. The teams can choose from pre-selected tools, such as Jet Gipsen's *Ethics for Designers* and Mario Alberto Sosa Hidalgo's *Ethical Toolkit*. Now, how can you *win* this assignment? Not by making beautiful sketches. The design is not the goal, but a means to expose ethical issues. We view co-design as way to support ethical deliberation (Steen 2013). Ethical issues the teams come up with can engage with principles like: respect for human autonomy, e.g., an appropriate allocation of agency between people and the system (Steen 2011); prevention of harm; fairness; and explicability (High-Level Expert Group on Artificial Intelligence 2019). The team that presents the best approach for integrating these principles into their AI design wins.

- Organizers: Marc Steen (TNO) and Rudy van Belkom (STT)
- Duration: 90 minutes
- Participants: 20 people
- Requirements: flipchart, post its, markers

**References**

Dignum, Virginia. 2019. *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*. Cham, Switzerland: Springer Nature.

High-Level Expert Group on Artificial Intelligence. 2019. *Ethics Guidelines for Trustworthy AI*. Brussels: European Commission.

Mittelstadt, Brent. 2019. "Principles alone cannot guarantee ethical AI." *Nature Machine Intelligence* 1 (11):501-507. doi: 10.1038/s42256-019-0114-4.

Morley, Jessica, Luciano Floridi, Libby Kinsey, and Anat Elhalal. 2019. "From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices." *Science and Engineering Ethics*. doi: 10.1007/s11948-019-00165-5.

Steen, Marc. 2011. "Tensions in human-centred design." *CoDesign* 7 (1):45-60.

Steen, Marc. 2013. "Co-design as a process of joint inquiry and imagination." *Design Issues* 29 (2):16-29.

Van Belkom, Rudy. 2019a. *De computer zegt nee: Scenario's over onze toekomst met AI*. The Hague, The Netherlands: STT (Netherlands Study Centre for Technology Trends).

Van Belkom, Rudy. 2019b. *Duikboten zwemmen niet: De zoektocht naar intelligente machines*. The Hague, The Netherlands: STT (Netherlands Study Centre for Technology Trends).