

## **Taming crowds: How practitioners address uncertainty when crowdsourcing data sets**

Crowdsourcing is a common approach to annotate a data set to be analysed directly or used for 'Artificial Intelligence' ('AI') applications. An initiator distributes tasks to crowd workers, who then annotate the data point. Turning to crowdsourcing exposes the initiator to multiple sources of uncertainty: How the task should be designed, who is part of the crowd, how to best make use of the annotations, and how to know if the crowd's work is any good are causes for concern. From a Science and Technology Studies perspective, this study investigates how practitioners that crowdsource data sets address uncertainties during this process. Based on qualitative interviews with practitioners it draws on Actor-Network Theory as a theoretical lens and Situational Analysis as method to analyse the empirical material.

This study identifies shared strategies to address uncertainty, and analyses how these strategies can fail. It shows that it is not straightforward to make crowdsourcing work but takes tremendous effort, labour that often remains invisible and hidden. At the same time, I show how epistemic approaches, whether the initiators consider the data as disputed and how this gets acknowledged, informs the structure of crowdsourcing processes.