# OVERVIEW AND GOAL OF THE FRAMEWORK

- One, quite philosophical approach, in applying this framework would be providing help for stating and investigating such assumptions about reality. Considering that AI-systems might greatly magnify the impact and reach of its deployers assumptions this phenomenological exercise could perhaps be very warranted. At least, knowledge should be held if one's own assumptions and those of the system engineers are the same; otherwise there seems to be a risk of mission drift without any other fault than differing human interpretations of the world we live in.

## 1) CONTEXT

- After surveying literature on AI ethics, one emerging challenge seem to be the risk of creating epistocratic structures where knowledge of what is done and what results means are opaque to many prospective users. I share the identification of crucial principles and the classification above.

- *Business model: for-profit use, non-profit use or public service [optional criteria]: Should perhaps be a core criteria since business model is an essential aspect of analysing the incentives and prospective impact of the deployment and further development of a certain technology*

## 2) DATA AND INPUT

- *D. Data quality and appropriateness [optional criteria]*: Approaching this subject from the "human" side, and from a context of using systems as a for profit service in private settings where deep expertise is often lacking, data quality or lack thereof seem to be a critical point in assuring accurate results and avoiding false belief in the systems. It could be an important "nudge" to make this a core criteria and reinforce the frameworks ability to get the policy makers to assess their potential assumptions and biases. The implications of this criteria is to extend the principles of robustness and transparency to the data-gathering and aggregation to the sources of information. Ensuring trust and reliability in the data seem to become a critical security and strategic issue for robustness. Point 41 below seem to capture the essential steps to keep in mind.

- This cumulative way of assessing risk would be very helpful to conduct analysis, it seem straight forward enough to be deployed in a non-AI expert group, given that the appropriate research is done and communicated good enough. Further work in recommendation on how to weigh together impacts in different areas would be helpful. For example how to weigh risks that are external to the operators direct control – what level of risk on impact outside the own ecosystem can be justified seem to be a crucial societal question which is not exhausted. This framework seem to be a step in the right direction.