

✓ Privacy, Surveillance

✓ Manipulation of Behaviors

✓ Opacity of AI systems

There are several technical activities that aim at “explainable AI”, starting with (Van Lent, Fisher, and Mancuso 1999; Lomas et al. 2012) and, more recently, a DARPA programme (Gunning 2017 [OIR]). More broadly, the demand for

a mechanism for elucidating and articulating the power structures, biases, and influences that computational artefacts exercise in society (Diakopoulos 2015: 398)

is sometimes called “algorithmic accountability reporting”. This does not mean that we expect an AI to “explain its reasoning”—doing so would require far more serious moral autonomy than we currently attribute to AI systems (see below §2.10).

✓ Bias in Decision Systems

→ social phenomenon of learned bias

→ human cognitive bias

→ statistical bias

✓ Human-Robot Interaction

✓ Automation and Employment

✓ Autonomous Systems

"Verifiable AI" for such

safety-critical systems

and for "security applications".

✓ Machine Ethics
(Subjects)

✓ Artificial Moral Agents

✓ Singularity & Superintelligence