

「人工知能の倫理」？

土屋俊

大学改革支援・学位授与機構

2016年6月6日、人工知能学会公開討論会にて

なぜ、人工知能研究者は(まだ)人工知能の倫理を語ってはいけないのか

- I 人工知能研究者が、その研究成果について社会に対してどのような責任をもつのかを決めていないから
⇒ 学会倫理綱領が不可欠

- II 人工知能研究者は、人工知能が存在する社会の全体的景観のなかで一要素にすぎないから
⇒ (しかし、一応市民であるので、市民としての倫理を理解していることを示しつつ、専門的な知識を理解可能な形で示す姿勢をもっていることを示す) 学会倫理綱領が不可欠

(カルト集団ではないことを自ら宣言することが必要 - 「ヒポクラテスの誓い」)

人工知能研究者は(今)何を決めるべきであるのか

- これらの二つの理由を前提として、あらためて、倫理綱領の最大の目的は、人工知能の研究開発が社会に埋め込まれた形で行われていることを示すこと
- 「市民としての」倫理と整合的な倫理規範に従って行動していることを示す ⇒ 草案の8(他者危害則), 7(法令遵守), 9(プライバシーの尊重), 3(公正性)が相当
- (人工知能の)「専門家としての」倫理規範に従って行動することを示す ⇒ 草案の2(誠実な行動・契約遵守), 5(「検証と警鐘」), 6(啓蒙), 10(説明責任), 4(不断の自己研鑽)が相当(ただし、重複、混乱の整理が必要)
- さらに、いわば「メタ条項」(この綱領を遵守し、遵守させる)が必要
- 人工知能技術の特別な性格を考慮して整理し直す余地はある ⇒ 行動規範と製品の特異性、the machine question
- できないことを宣言しない(「行動指針」で具体化)