AIに主体性を帰属させること：学際的アプローチの中間報告

宮原　克典（Katsunori Miyahara）

北海道大学　人間知・脳・AI研究教育センター（CHAIN）

近年、ビッグデータと深層学習を主な原動力として、人工知能（AI）技術の開発と実用化が急速に進んでいる。こうした現状は、近い将来、最新の情報処理技術としてのAIの利活用がますます進むだけでなく、AIが人間に類するようなある種の「主体」として人間と関わりあう場面も増えてくることを予想させる。本WSでは、このように単なる道具ではなく、人間がそれをある種の主体として経験することを誘発するようなAIシステムを「人工主体（Artificial Subject）」と呼ぶ。

人工主体の開発と社会普及が進行すれば、主体とモノ／客体の区別が比較的に明らかであった従来の社会では生じなかった新たな問題や緊張が生じてくるだろう。そして、これは遠い未来の出来事ではない。人間がAIにある種の主体性を見出してしまう事案はすでに現実に起き始めている。たとえば、2022年6月、Google社のエンジニア Blake Lemoine 氏は、同社で開発していた対話用の大規模言語モデル LaMDA（Language Model for Dialogue Applications）は意識と人格をもつ主張した。Google社の上層部はLaMDAが意識を持つと信じる十分な証拠はないと判断し、最終的に、社内情報を一般に漏洩したことを理由にLemoine氏に休職処分を下した。こうした事例は、人間と人工主体が共存するような社会は、遠い未来やサイエンスフィクションのものではなく、すでに現実に立ち現れつつあることを示唆する。

本グループでは、トヨタ財団〈特定課題〉「先端技術と共創する人間社会」の助成対象プロジェクト「人間と人工主体の共存のあるべき姿を学際的に問うための新たな枠組み「人工主体学」の構築に向けて」（D21-ST-0012）の資金を得て、今後、人間と人工主体はどのような形で共存すべきかを学際的に探究している。特に、人間がAIに主体性を帰属させる実践をめぐる三つのテーマに注目して考察を進めている。第一のテーマは、この実践そのものの本性の解明である。人間は、特定の具体的なコンテキストのなかでAIと相互作用するなかで、それを高度な情報処理技術として経験したり、ときには主体性や意識をもつものとして経験したりする。人間はいかに人工主体と付き合っていくべきかを有意味な仕方で議論するためには、その基本的な前提として、こうした場面で生じる経験を記述的に分析し、また、そうした経験の生成条件を実験的に解明する必要がある。第二のテーマは、AIに意識や主体性を帰属させることの倫理的意味を明らかにすることである。ここでは、人間と人工主体の理想的な共存を実現するためには、AIに意識や主体性を帰属させることが、どのようなときには倫理的に許容できるのか、どのようなときは倫理的に好ましくないことなのかといった点を考えておくことが課題となる。第三のテーマは、この実践の社会的意味を明らかにすることである。AIを一種の主体として見ることは、人間の社会的な実践や制度にさまざまな帰結をもたらしうる。人間と人工主体が調和した社会を実現するためには、AI

に主体性を帰属させる実践の定着によって生じうる帰結を予見し、それに付随する問題への対処を事前に考察しておく必要がある。

本 WS では、人間が AI に主体性を帰属させる実践をめぐる諸問題について、哲学、倫理学、科学技術社会論の観点から議論を提起する。長坂は、哲学史における惑星運動と不動の動者をめぐる議論を手掛かりに、身体的及び非身体的な人工物が、いかなる条件ものとで、我々にとって主体として現れるようになるのかを考察する。新川は、先端科学技術によって生み出された人工的存在者がどのような意味で「主体」とみなされうるのかを整理し、その倫理的な位置づけについて論じる。竹下は、人工主体と人間の関係だけでなく、そこに非ヒト動物を含めた三者の相互関係という観点から、人工主体との関係のあり方について論じる。特に人工主体と非ヒト動物（特にイヌ）のあいだの友情関係の可能性について検討する。池原は、人間と非人間を区別しないことで知られるブルーノ・ラトゥールのアクターネットワーク理論の観点の人工主体への応用可能性について論じる。濱田は、text2image 技術の進展によって、AI が従来は人間主体に特権的だと見なされてきた創造的な役割を担いつつあることをふまえて、人類のデータを使って学習する人工主体が人間社会に与える影響について論じる。