[人類はどう生きるべきか？ IT はどうあるべきか？]

# シンギュラリティへの哲学的逡巡

7.2

応般

村上祐子（東北大学）

## 実現可能性？

シンギュラリティとは，巨大データの高速計算が可能となることにより，人間の心以上の能力をもって，知覚し，思考し，意図する心を持つ「強い人工知能（AI）」が生まれ，我々の世界に影響を及ぼすという一群の予想である．

これがそもそも実現するためには，次の 2 つの問への解答が yes にならなければならない．（A）強い AI が原理的に可能か？ （B）強い AI による現実の因果系列への干渉が可能か？ （A）についてはそもそも到来するのかという議論があるし，（B）については技術的開発は進むにしても，技術的には解決できない制約，たとえば社会的制約や物理的制約の存在が問題となる．だから以下 4 つの解答方針が可能だ．（1）原理的に到来しない．（2）予算，エネルギー，制度といった現実的制約を受けて，限定した形でしか実現しない．すなわち強い AI が実現したとしても，現実世界への干渉は何らかの方法でブロックされる．（3）一様な実現はあり得ない．つまり強い AI は実現しロボット（特にナノロボット）やもののインターネット（IoT），また生体素子による生体を含む現実世界への干渉が実装されるかもしれないが，現実世界には AI によるコントロールが到達しない部分が存在し，将来にわたってそのまま存在し続ける．（4）全面的なシンギュラリティが実現する．この（2）～（4）では，部分的にでもシンギュラリティが実現することになるので，人格，自由意思，責任といった哲学的問題と関連する法的・社会的問題が発生することとなる．

強い AI はこころの過程は計算であるという心の計算理論の理論的仮定に基づく．内部構造がどうなっているかは問われず，外とのインタフェースで「こころ」は行動主義的に判断される．Searle はこの前提を否定するが，このような行動主義がロボット研究者には自然であろう．計算機の計算過程が我々の思考過程と同じである必要はなく，人間的思考であるかどうかは外面的応答で判断される．

しかしそれを認めて強い AI は可能だと措定しても，何らかの理由でシンギュラリティは「現実には到来しない」ことになりそうだ．現実の因果系列への干渉に関して，既存制度，特にさまざまな法を遵守したシステムを作るのであれば，その実現形態はある程度限定されたものとなる．

ここで問題となるのは，「強い AI は規範・法を自律的に作り，アップデートすることができるか」という問いだ．既存の規範に従う以上の意図を持たないような「従順かつ強い AI」は概念上可能である．だが人間が既存制度の延長上の法・規範を遵守させる AI システムを作ったとしても，AI が規範を変える可能性があるのであれば，ハックされなくても，突然殺人をよしとするロボットが発生するかもしれない．このような自律的な設定変更が可能なら，問題が発生したときの責任主体は誰か？ そもそも AI を「誰」と呼んでいいのか？

つまり，シンギュラリティの実現が部分的であったとしてもこの問題は議論しなければならないし，人間がこれまで人間特有であると考えてきた領域が保持される条件はいかなるものか？という問題が発生することになる．

## 人格・責任概念の変容

なにかが責任主体であり得るための条件を考える

際に不可避なのが，自由意思の問題である．通常，外部からの操作や命令への服従などある選択肢を強制される場合には自由意思を持たないこととなる．別の理論では意図と因果系列への介入があわさると自由意思による行為となり，責任を問われることとなる．だが「羊をめぐる冒険」の主人公は自由意思で運命の地にたどりついたつもりが，それはプログラムされたものだったと明かされダメージを受ける．シンギュラリティ到来後の人間の自由意思はこの主人公のように幻想でしかないのか？　一方で，強い AI が意図と因果系列への介入機能を持てば，（定義により）自由意思を持つこととなるのか？　それとも従順でない AI に限って人格を認めることになるのか？

なお強い AI 以前にも，ネットワークを介した因果系列への介入に関する責任問題は存在する．IoT が何らかの原因で誤動作し，損害を発生させたときの責任主体は誰か？　コストをかけてもログ分析を元に誰かの責任を追及することになるのか？

また，ロボット技術や生体素子技術の進歩による人格の拡大問題は AI 側だけの問題ではなく，人間側にも変容が発生する．エンハンスメントとは個体の本来の能力以上の能力を発揮させるための技術である．ものによっては単に埋め込めばよいというものではなく，生体側にも技術への適応訓練が必要なことがある．シンギュラリティ実現後の教育はこれに特化するのか，それとも「情報を元にした判断」の訓練をするのか？　一方で，脳が AI で強化されデータ取得のみならず判断回路も実装したとして，そこで行われる判断は「その人の判断」なのか？　結果的に間違った行為につながった場合の責任は AI にあるのではないか？　しかもデータ取得についても AI に依存しているのであれば，ログ分析によっても人間への責任帰属は不可能なのではないか？

また脳の記憶をサーバにアップロードした一群のデータに人格を付与できるか？　応答だけ見ればアップロードヒューマンもサーバデータを参照して意思決定を行っているように見える．このとき，強い AI にも人格を認めることになっているのであれば，この 2 つを区別する根拠は何か？

ここでは解答として，シンギュラリティの実現は一様ではなく，どうしても機械ではできない部分が残る，という（3）を主張する．物理的制約下におかれた要素と因果系列内でインタラクションを保っていくとすれば，その部分の速度はデータの巨大化や計算の高速化では改善できない．たとえば宇宙空間など超遠隔地との通信や生物が育つのには一定の時間が必要だ．データや素子で制御は可能かもしれないが，時間的制約は消えない．そしてこのシンギュラリティが原理的に波及しない部分が存在し続けるのならば，その部分でなんらかの責任を取り得ることを十全な人間としての条件の根拠とできる．しかも，科学の本質が観察と新たな理論化に常にオープンであることと措定すれば，シンギュラリティの限界の存在は保証されるのではないか．

（2014 年 7 月 15 日受付）

村上祐子　ymurakam@m.tohoku.ac.jp

東北大学文学部准教授．Ph.D.（Philosophy）東京大学教養学部，インディアナ大学大学院哲学専攻博士課程修了．専門：論理学・哲学・科学技術社会論．