

アイザック・アシモフのコンテキストにおける AI とロボットの知的財産権と倫理

金子 格†

概要：本稿ではアイザック・アシモフのコンテキストにおける AI とロボットの知的財産権と倫理について検討を加える。

キーワード：アイザック・アシモフ, AI, ロボット, 知的財産権, 倫理

Intellectual Property and ethical issues with AI and Robots in the context of Issac Asimov

Itaru Kaneko

1. はじめに

本稿ではアイザック・アシモフのコンテキストにおける AI とロボットの知的財産権と倫理について検討を加える。

アイザック・アシモフ(1920-1992)[1]は言うまでもなく非常に多くの SF・ミステリー・科学読み物ロボットを著した著述家であるが、特にロボットをテーマとした多くの SF 作品を著したことで有名である。近年知能を持つロボットの実現が現実的な可能性となるなか、アシモフ氏の著作に含まれる人間に近い知能を持つロボットに関する多くの見解が示唆に富むものとして再注目されつつある。本稿では、アシモフ氏の物語のコンテキストにおける AI とロボットの知的財産権と倫理に関する議論について、作品を確認しながら検討を加える。アシモフのロボット関連作品においては「ロボット工学三原則」という前程がそのほとんどの作品の前程となっている。アイザック・アシモフのコンテキストとはそのロボット工学の三原則を中心とした作品世界を指す。また本稿において、本稿の筆者の筆者を「筆者」、アシモフ自身は「アシモフ」、その他の作品の著者は、「著者」と表記する。

本稿の性質上アシモフ作品の物語の内容を最小限度記述しないわけにはいかない。できるだけ最小限度にとどめるよう努力するが、結末や論点

を先にしてしまうことで娯楽としての楽しみを損なうことを完全に避けるわけにはいかないことを、あらかじめお断わりしておく。

2. アシモフ以前のロボット

まずアシモフ以前のロボット作品について簡単にまとめておく。

ロボットを人造人間の一種ととらえれば、その暗い側面を最初に表したのは、イギリスの小説家メアリー・シェリーの小説(1818)に登場する「フランケンシュタイン」(原題は『フランケンシュタイン、あるいは現代のプロメテウス』)だと思われる。死体をつなぎあわせて再生したもので完全な合成物ではないが、一旦死体となったものを部品として作り上げ、材料とは異なる意識さえ持たせた、人造人間である。この人造人間は自らの存在について悩み、自らを創造した創造主に復讐を果たす。このストーリーもその後の多くのロボットに関する物語に強い影響を残した。

コンピュータ業界に身を置く者としては、詩人のバイロンの提案によりメアリー・シェリーこの作品を一気に書き上げ[2]、バイロンの娘エイダがコンピュータ技術の基礎を築き結果的に現代に人工知能を登場させることになった[3]ということに、偶然以上の不思議な因縁を感じる。

ロボットという言葉はチェコの劇作家カレル・チャペックの戯曲、「RUR(ロッサム万能ロボット)」

† 東京工芸大学
Tokyo Polytechnic University

で使われた言葉が普及したものだ。チャペックのロボットは電子機械的なものではなく、有機物に電気刺激を与えて生命を与えることで製造される。これは当時の有機化学と神経の電気的性質の研究が急速に進歩していたことを考えれば自然で現代のロボットの機構と大きく異なることは本論のテーマにおいては重要ではないと考える。

チャペックのロボットも人間と対立するが、怪物としてではなく人間と同様の知性を持つ社会的な存在として描かれていることが大きな特徴である。チャペックの戯曲でもロボットは人間を滅ぼすが、ロボットは集団として人間に敵対し、当時台頭しつつある新しい政府や国家と人間の関係をテーマとしていると筆者は感じる。物語で提起されるさまざまな論点や、物語の結末は、現代に至っても示唆に富むものである。

アシモフはこうした「人間より強力な人造人間が人間を、あるいは人類を滅ぼす」という固定概念を打破することを創作の動機とした。強力な装置を作る際に2重3重の安全策を講じないことはあり得ない。そして有名なロボット工学の三原則を生み出した。

「第1条 ロボットは人間に危害を加えてはならない。また、その危険を看過することによって人間に危害を加えてはならない。

第2条 ロボットは人間に与えられた命令に服従しなければならない。ただし与えられた命令が、第一条に反する場合は、この限りではない。

第3条 ロボットは、前掲第一条、および第二条に反する恐れのない限り、自己を守らなければならない。」 [6]

ロボット工学(Robotics)はアシモフの造語とされるが、以後実際のロボット工学を表すことばとしても用いられることになった。しかし現実のロボット工学でアシモフのこの三原則が実用的な意味で用いられることは、長年なかっただろうと筆者は確信する。これらの原則は初期の実用的なロボットの実際的设计に必要な具体的で機械的な設計と比べると、あまりに概念的で抽象的であり、これらの「原則」が工学的な設計に役立つ局面は想像しにくい。

しかし今日、急速に人工知能が発達をとげ、状況は変わりつつある。現実には人間をはるかに上回る知識や情報を元に人工知能に判断をゆだねる時代が目前である。現代では「ロボット法」といった議論も活発であるが、AIが人間の知能では計り知れない膨大な知識や情報に基づいて判断をする場合に、もはや個別の具体的機械的な制約の集合でロボットの動作を制約するのでは間に合わないと考える者は多い。アシモフが想定したの

とは全く異なる方法によるかもしれないが、何らかの形で具体的な状況を高所から俯瞰する「倫理的原則」を人工知能に組み込む必要性が生じていると筆者は考える。

もちろん、40年前にエンタテインメントを目的として書かれた原則が、現実のロボット工学に役立つとは考えにくい。現実のロボット工学で採用されるのは全くことなる形のものになる可能性が高いだろう。しかしそうした抽象度の高い目標設定が、具体的な動作に展開される場合に生じる様々な問題点について、アシモフが創作活動を通じて行った膨大な思考実験から、なんらかの示唆は得られるのではないだろうか。

アシモフの作品はSFでありながら推理作品でもあり、提示された問題が上記三原則によって単純に解決されるわけではない。三原則は非常に上位の目標設定、制約条件であって、その他に具体的な複雑な物語上の前程条件が細かく提示され、そこから導きださせる唯一の結論を読者に推理させるという様式をとっている。具体的な局面にあてはめると三原則と状況の間に複雑な相互作用が発生する。たとえば有名な「トロッコ問題」[7]にロボットが直面した場合三原則はどう機能するかといったことがテーマとなる。実際のアシモフ作品においてはさらに複雑な前程を用意し、三原則から導き出される帰結を推理してみせるのである。本稿ではこの後アシモフ作品の中でロボットの知的財産と倫理がどう扱われているか、そしてそれに三原則がどう影響するかについて検討を加える。

3. ロボットの知的財産

3.1. うそつき-高い知性の副作用

ロボットやAIの知的能力を人間に近づけていけば当然知的成果物が生じ知的財産の問題が生じる。今日では現実の問題として話題になっているがアシモフ作品ではどのように論じられていたのだろうか。

知的活動で人間以上の能力を有するロボットは「うそつき」(文献[6]に収録)に登場する。本作品に登場するロボットは人間を上回る知的能力を持つ上、人間の感情を読み取る能力、一種のテレパシーを有している。ここでアシモフはこうした能力を持つロボットが、第一条により人間を「心理的に傷つけない」という制約にしばられる可能性を示唆している。ロボットに人間への絶対服従を強く動機づけた結果、ロボットは「真実」ではなく「利き手が望む回答」を回答せざるを得なくなる。その可能性とそれが引き起こすより大きな危険性がこの作品のテーマである。

今日我々は、ロボット以外の組織運営において、同様の問題に直面する。ある組織やその構成員を、失敗にたいしては極端な罰則を与え、成功に対して極端な報償を与えていくと、虚偽の報告を受け取るようになる可能性が高い。賞罰の厳しさと報告の信頼性には反比例の関係がある。また、他人の感情に高い共感力を持つと精神的ストレスが高まり正常な人間関係を構築できない、という現象もある。

アシモフが実社会におけるこうした事例から着想を得たかどうかはわからない。ロボットに「人間が満足する」という強い動機を与えると、同様の動機を与えられた人間や組織同様「うそつき」の症状が現れる可能性がある、というのは興味深い着想である。

3.2. 心かけられたる者 - 創造性の条件

アシモフはある時期から、ロボットの「創作活動」に高い関心を持ち多くの作品を残した。文献[8]によると、「いつの日か」(1956)、「笑えぬ話」(1956)、「女の直感」(1969)、「バイセンテニアル・マン」(1976)[10]をあげている。

「心かけられたる者」(1972)[9]はロボットが重要な発明を行うという物語である。本作品ではロボット製造会社がロボット開発にロボットを利用することを決断する。ロボットの知的能力が高まれば当然ロボット自体にロボットの助けを求めめることは自然な帰結である。しかしこの試みが意外な結果を生み出す。

まず、アシモフの設定としてはこの時代ロボットは人類に対する脅威とみなされ、人間社会での利用が厳しく規制されている。そのためロボットの利用はおろかロボットを研究所から外へ出すこともできない。そのことが(地上で利用するロボットを販売できないことが)ロボット製造会社が苦境に陥った原因である。そうした中で、ロボット製造会社は会社に再び利益をもたらす新製品を、ロボットに行わせようとする。これまでにない高性能な人工知能を備えたロボットを開発し新製品開発のためのあらゆる情報をロボットに与え、新製品開発という課題を課すのである。

このように人間を上回る「創造性」を持つように作られたロボットはこれまでのロボットよりもはるかに人間的な要素を持つだろうとアシモフは想像したのではないか。そして物語でロボットは、会社の期待にこたえるには2つのことが必要であると監督者に要求する。1つは話し相手の提供である。もう一台の同性能のロボットを自分の思考を助ける話し相手として提供するよう求めるのである。もう一つは実体験である。知識だ

けでなく実際の「世界」と直接接することが必要であるとして要求するのである。この要求は受け入れられ、やがて2台のロボットはお互いに意見を交えつつ実体験に基づいて画期的な新製品を生み出す。

この物語の結末はネタバレになるので省くが、ロボット工学という前程を存分に使った一編となっており、また人間が持つ創造性のメカニズムをアシモフ自身がどう意識していたかを伺い知ることができる作品となっている。

3.3. バイセンテニアル・マン(知的所有権)

バイセンテニアル・マン[10]ではさらに人間に近いロボットが登場する。本作は後に長編化され[11]、ロビン・ウィリアムス主演で映画化[12]されたので、一般にもよく知られた作品のひとつではなかろうか。

本作の主人公アンドリューは、人間の家庭に買われたロボットであるが、偶然他のロボットと異なる高い創造性を有している。アンドリューの芸術的創造性は、購入者の娘によって発見される。持ち主はその価値が非常に高いことをすぐに察するが、ロボットが生み出す知的財産の価値をどう扱うかがストーリーの最初のポイントになる。ロボットの持ち主は「権利の半分はアンドリューに帰属する」と判断する。

これには法的義務はない。しかし必ずしも持ち主にとって不都合とは限らない。アンドリュー自体の所有権は持ち主にあるからけっきょく所有者がすべてを所有していることはかわらない。一種の法人化と考えることもでき、こうした扱いは今日でも可能かもしれない。そうすることでアンドリューに関する費用と生み出す利益を相殺し設備投資や、減価償却ができるなど会計上の利便性もあると考えられる。その意味では知能を持ったロボットには人間と同じ所有権と財産処分権を与える方が合理的であると筆者は考える。

しかし、物語においてアンドリューの持ち主はこうした功利的な理由ではなく、アンドリューに対する敬意から知的財産権を与えているようだ。にもかかわらずアンドリューが所有という関係から脱し完全な独立を求めた時には所有者はショックを受けるのである。

アンドリューはその後も高い創造性を発揮し、その能力は平均的な人間を上回っている。必然的に巨万の富を築くようになる。こうした中で当然アンドリュー自身の財産所有権の正当性を否定し資産を奪い取ろうとする様々な攻撃がある。しかしロボットの知的能力は高いから、的確な法的対応でこれらの試みをことごとく打ち破って

いく。

将来ロボットに知的所有権を初めとした財産権を与えることが考えられるが、ロボットの能力は高いから、必然的に市場競争の中で人類資産のほとんどがロボット所有となっていくことが予想される。この「人間とロボットの格差問題」もロボットに財産権を与えるということから派生する問題の一つであると思われる。

4. ロボットと倫理

4.1. バイセンテニアル・マン(人間願望)

ここからはロボットと倫理についてアシモフの設定したコンテクストの中で論じていくことにする。

まず先のバイセンテニアル・マン[10]のストーリーの続きから論じることにする。ロボットは最後に「人間と認めてもらう」というつよい願望を持つことになる。まずロボットの創造性とロボットの人間願望について少し考察したい。このロボットのように人間社会の知識を持ち、人間と同じ役割を人間社会の中で果たす。そのような知的存在にとって、「自分も人間でありたい」という強い願望を持つことは、はたして必然だろうか？このような命題にアシモフが強く魅かれ、彼が出した結論は Yes だったのではないかと筆者は推測する。人間を深く理解する知性が自らも人間でありたいという願望を持つことが必然であると筆者も強く確信するからだ。

4.2. バイセンテニアル・マン (人間の条件)

アシモフ作品においてロボット工学三原則が適用されるとき、人間とは何か、という問いがくりかえされる。ロボット工学三原則は人間がロボットの主人であることを保証するものだが、ロボットが人間をどう定義するかという問いは、けっきょく人間はなにを人間と定義するかという問題に帰結する。これにはいろいろな論点があるが、もっとも単純な論点は「知性」、「人間性」そして「倫理感」の関係である。人間は高い知性を持つ人間をより人間的であると感じてしまう。また倫理感のない人間は「人でなし」ということばで表すように人間的でないこととみなしてしまう。そこでアシモフ自身も何度か指摘しているが、この特徴でロボット工学の三原則を人間とロボットに適用するとロボットはロボットを人間と、そして人間をロボットと判断してしまいかねない。この論点には後で再度戻ることにする。

バイセンテニアル・マンの結末では「人間である」ことの別の条件が示されるが、それは生物としての寿命である。確かに唯一無二の自我である

ことと、寿命があることは人間とロボットを区別する明確な差異である。一方で、もし人間の権利を得るために寿命があることを条件とするならば、人間である限り不老不死や無制限の再生治療をあきらめ、有限の寿命を受け入れなければならなくなる。近い将来の人間社会における倫理上の課題の一つであると考えられる。

4.3. 逃避 - 死の定義

次にとりあげるのは「死」の定義である。「われはロボット」[6]に収録された「逃避」という作品ではロボットが「人間の死」を扱うという困難に直面する。

ロボット工学の三原則第1条によりロボットは人間への危害を看過できないから人間の死は何者にも優先して防止しなければならない。しかし心臓移植手術の際には人間は仮死状態になるが、ロボットが手術を妨害しては手術がだいなしだ。妊娠・出産はかなりのリスクを伴うがロボットがその妨害をしてもらっては困る。人間活動においてあえて死亡や重症を負うリスクを受け入れるべき場面は多々あるが、それをどう扱うかはロボットだけでなく現代の人間社会においてもしばしば問題になる事である。

4.4. 夜明けのロボット - 危害のインテンシテイ

本作品は 1983 年に発表され同シリーズ前作から 25 年のブランクを経ての発表であることが話題になった。また本作品は人工知能の父であるマーヴィン・ミンスキと産業用ロボットの父といわれるジョセフ・F・エンゲルバーガーにさげられている。

1950 年代にアシモフは代表作となる長編シリーズを 2 つ世に出した。その一つ「ファウンデーション」シリーズは人類文明が恒星間に広がった世界の物語でありロボットは全く登場しない。もう一つが「ロボットもの」とアシモフが呼ぶ作品群で、きわめて高い知性を持ち人間と区別がつかないほど精巧な人型ロボット、オリヴォー・R・ダニールが活躍するシリーズ(以下「R シリーズ」と省略する)である。これらの作品群について「ロボット小説の舞台裏」でアシモフ自身が経緯や意図を説明している[13]。

R シリーズではロボットの力を借りて宇宙に進出したスペーサーという種族とロボットを忌避して地球にとどまり過密と汚染に苦しむ地球人との文明が対立している。その中でスペーサーに属するロボットのダニールと地球出身の刑事イライジャ・ベイリが共同で 1 作品につき一件の非常にこみいった事件の捜査を行う。2 人(1 人と一

体)はお互いの文明や種族が持つ違いを乗り越えて見事なチームワークで事件を解決する。

R シリーズは 1956 年に「はだかの太陽」が書かれたあと 25 年後にこの「夜明けのロボット」が発表された。この間起こったこととしては、ソ連による初の人工衛星打ち上げ、コンピュータの急速な進歩と実用化、冷戦と核軍拡、ベトナム戦争の敗北、産業用ロボットとマイクロプロセッサの実用化だった。ロボットと宇宙に関係した現実があまりに急速に進展し、アシモフだけでなくすべての SF 作家にとって近未来を描くことがきわめて難しい時期だったと考える。同時にこの作品では、以前の作品の設定を活かしつつ、この間に実現したすべての政治的変化や科学技術の発達をふまえて作品世界をさらに発達させている。そして物語では現代に通じる多くの倫理問題が提示されている。

夜明けのロボットでダニールとイライジャが取り組むのは「殺ロボット事件」である。ストーリーでは実に多くの倫理問題が提示される。主なものを以下に上げる。

- (1) ロボット殺害の倫理性
- (2) Sex ロボット(ロボット側の動機、人間の社会の反応)
- (3) 精神的危害
- (4) 危害のインテンシティ問題(トロッコ問題の一般化)
- (5) 波及効果による危害

たとえば参考文献の下巻 p34 では第 1 条において「真実を曲げることによる加害」と「虚偽を述べることによる加害」が拮抗する場合の行動について尋ねるベイリ(人間)に対し、ロボットが説明する。

「第一条がこういっていてもか。<ロボットは人間に危害を加えてはならない。またその危険を看過することによって、人間に危害を及ぼしてはならない。>この第一条の前半と後半は合い等しいといえるだろうか? やったという罪の方がやらないという罪より君は言うんだね。」

… 中略 …

「通常は。二者に及ぼす危害のインテンシティがほぼ相等しければ、非真実より真実を常に選ぶということです。つまり通常の場合。」

このように第一条、第二条、第三条とも対象となる目標が複数対立する可能性がある以上(複数の人間を助ける、あるいは複数の矛盾する人間の命令を受ける可能性がある以上)、必然的にそれぞれの強度、インテンシティの評価という問題を受け入れざるを得ない。そして必然的にロボット自身に「命の相対的な重さ」の判断を求めることに

ならざるを得ないことになる。その中には精神的な苦痛や波及効果による加害も含める必要がある。これを進めると人間がロボットに多数の人間に対する生殺与奪権を与えることを認めざるを得なくなってしまう。

このような問題を生じるのはロボット工学三原則第一条の後半、<またその危険を看過することによって、人間に危害を及ぼしてはならない。>という項目があるからで、ロボットに対するこの強制は「危険」の推測方法や防止すべき水準によってはロボットを危険な動作に導く可能性を持っているといえる。

5. まとめ

本稿はアイザック・アシモフ作品のコンテキストにおいて知的財産権問題と倫理問題がどのように描かれているかを主要な作品を参照しながら検討した。

アシモフのロボット関連作品は、よく知られているようにロボット工学の三原則という前程がほとんどすべてのストーリーでの骨格となっている。ロボット工学の三原則についてはもともと工学的な応用を目的としたものではなく、それを元に面白いエンタテインメント作品を作る目的で創作されたものであるから、その工学的な意味や妥当性を細かく論じることはあまり意味がないと筆者は考える。したがって本稿で論じたかった問題はむしろアシモフがロボットに関するどのような問題に注目したかという点である。

再度まとめると以下のような様々な問題にアシモフは言及していることがわかる。

表 1 言及項目と作品

言及項目	作品
知性の副作用	うそつき
創造性に必要なもの	心かけられたる者
知的所有権	バイセンテニアル・マン
人間願望	バイセンテニアル・マン
人間とみなす条件	バイセンテニアル・マン
死とみなす条件	逃避
危害のインテンシティ問題	夜明けのロボット
Sex ロボット	夜明けのロボット
精神的危害	夜明けのロボット
波及効果による危害	夜明けのロボット

くりかえしになるがアシモフはロボットや人

工知能の実現方法についての技術的知見を持って作品を書いたわけではない。これらの作品中の展開における技術的側面は、肯定的に評価するにしても否定的に評価するとしてもあまり重要な意味を持たないと筆者は考える。

アシモフ作品から読み取るべきはロボット及びロボット社会が人間社会に及ぼす影響、および人間社会の反応である。アシモフはロボットを排した社会、1人の人間が無数のロボットを使う社会。ロボットとだけ接し人間同心の接触が絶たれた社会など様々な可能性を描いている。

ロボットのどのような側面に人間は価値を見出し、恐怖を感じるか。そういったロボットという今までにない存在に直面した時の人間社会や人間の反応については、非常に示唆に富む作品群であると評価できる。

最後に作品から以下の部分を引用したい。最も初期の「われはロボット」におけるロボット工学者スーザン・キャルビンとそのインタビューアの会話である([6], p403)。アシモフはロボット SF をエンタテインメントとして作りつつも、やがて人間をはるかに上回る能力を有するロボットが登場する時代がやってくることは必然であると考えただろう。この会話を、人類の悲観的な未来を表すものと受け止める人も多い。しかし筆者は逆ではないかと思う。その後の人類社会のありかたについて思考実験を重ね、そこに一つの光を見出したことを表しているのではないかと筆者は考える。

以下、ハヤカワ文庫版の p403 より引用。

「人類の究極的な幸福が何を必要とするか、わたしたちにわかるかしら？…中略(以下同様)…もっと文明度の低い、人口の少ない、田舎生活風の、あるいは田園生活風の文明の方がよいのかもしれない。…あるいは、完全な都会化か、あるいは完全に鑄型にはめ込まれた社会か、または完全な無政府状態か、いずれかが答えとなるでしょう。わたしたちにはわからない。」

「…それじゃあ人類は未来に対して自らの発言権を失ってしまったのだと？」

「いまだかつて発言権があった試しはありませんよ。」

参考文献

- [1] アイザック・アシモフ,
<https://ja.wikipedia.org/wiki/アイザック・アシモフ>, (UTC 2016年8月6日(土)08:29)
- [2] デイオダディ荘の怪奇談義,
<https://ja.wikipedia.org/wiki/デイオダディ荘の怪奇談義>(UTC 2016年3月17日(木)21:46)

- [3] エイダ・ラブレス,<https://ja.wikipedia.org/wiki/エイダ・ラブレス> (UTC 2016年6月22日(水)03:30)
- [4] カレル・チャペック, RUR(1920), 深町真理子(訳), グーテンベルク 2 1,(2013)
- [5] カレル・チャペック, RUR(1920), 同書の青空文庫版,
http://www.aozora.gr.jp/cards/001236/files/46345_23174.html (2003)
- [6] アイザック・アシモフ, われはロボット(1950), 小尾 英佐(訳) 早川書房, ハヤカワ文庫 SF, (2004)
- [7] トロッコ問題,<https://ja.wikipedia.org/wiki/トロッコ問題> (2015年12月4日(金)03:51)
- [8] 瀬名秀明, 「ロボット学」の新たな世紀へ/アシモフ<ロボット工学の三原則>の受容と発展(ハヤカワ文庫「われはロボット」後書として収録), われはロボット, 早川書房, ハヤカワ文庫 SF, (2004)
- [9] アイザック・アシモフ(訳 池 央耿), 心にかけてきたる者(1974), 創元 SF 文庫, 「アイザック・アシモフ 聖者の行進」収録 (1979)
- [10] アイザック・アシモフ(訳 池 央耿), バイセンテニアル・マン(1976), 創元 SF 文庫, 「アイザック・アシモフ 聖者の行進」収録 (1979)
- [11] アイザック・アシモフ(訳 池 央耿), アンドリュウ NDR114(1992), 創元 SF 文庫, (2000)
- [12] クリス・コロンバス監督, ロビン・ウィリアムス主演, アイザック・アシモフ原作, 映画「アンドリュウ NDR114」, ソニーピクチャーズ (1999)
- [13] アイザック・アシモフ, ロボット小説の舞台裏(「はだかの太陽」の序文), 早川書房, 小尾 英佐(訳), ハヤカワ文庫 SF (2015)
- [14] アイザック・アシモフ, 夜明けのロボット(1983), 早川書房, 小尾 英佐(訳), ハヤカワ文庫 SF (1994)