研究

ネオ・サイバネティクスの理論に依拠した人工知能の倫理的問題の基礎づけ

Foundations of the ethical issues regarding artificial intelligence relying on the theory of neocybernetics

キーワード：
人工知能，倫理，情報倫理，ネオ・サイバネティクス，基礎情報学，オートポイエーシス
keyword：
artificial intelligence, ethics, information ethics, neocybernetics, fundamental informatics
autopoiesis

青山学院女子短期大学　**河　島　茂　生**
Aoyama Gakuin Women's Junior College　Shigeo KAWASHIMA

要　約
本論文では，ネオ・サイバネティクスの理論のなかでも特にオートポイエーシス論に依拠しながら，第３次ブームの人工知能を定位し倫理的問題の基礎づけを目指した。オートポイエティック・システムは自分で自分を作るシステムであり，生物の十分かつ必要な条件を満たす。一方，アロポイエティック・マシンは，外部によって作られるシステムであり，外部からの指示通りに動くように調整されている。この区分に照らせば，第３次ブームの人工知能は，人間が設定した目的に応じたアウトプットが求められるアロポイエティック・マシンであり，自己制作する生物ではない。それゆえ，責任を帰属する必要条件を満たさず，それ自体に責任を課すことは難しい。人工知能に関する倫理は，あくまでも人間側の倫理に帰着する。とはいえ人間は，しばしば生物ではない事物を擬人化する。特に生物を模した事物に対しては愛情を感じる。少なからぬ人々が人工知能に度を越した愛情を注ぐようになると社会制度上の対応が要される。そうした場合であっても，自然人や法人とは違い，人工知能はあくまでもアロポイエティック・マシンであることには留意しなければならない。

Abstract

In this paper, we seek to lay the foundations of the ethical issues revolving around how to place the third boom in artificial intelligence, while relying on the theory of neocybernetics, in particular autopoiesis theory. Autopoiesis refers to a self-producing system what characterizes living beings. However, the artificial intelligence developed in this third boom are allopoietic machines, which must give outputs in accordance with the purposes that humans have set, and do not produce themselves. Therefore, they do not satisfy the requirements of attributing responsibility, and it is difficult to impose responsibility on these artificial intelligences themselves. Therefore, the ethics related to this artificial intelligence simply reduce to ethics on the human level. Nevertheless, human beings often personify non-living things. They particularly feel affection toward things that are modeled after living things. If it comes to a point where too many people pour too much affection into artificial intelligence, we will have to deal with it on a societal level. Even in such a case, we must bear in mind that artificial intelligence are allopoietic machines and are different from natural persons and juridical person.

（受付：2016年６月１日，採択：2016年８月20日）

1　問題の所在

本論文の目的は，ネオ・サイバネティクスの理論に基づきながら，昨今の人工知能の現在地点を確認し，その方向性から見出せる倫理的問題の基礎づけを行うことである。現代社会には，すでに多種多様な人工知能が入り込んでいる。特定の人工知能ごとに個別に倫理上の問題を議論する前に，全体を俯瞰するような基準となるものさしが求められる。そうでなくては，人工知能をめぐる倫理は場当たり的かつ統一の取れない事態に陥ってしまう。本論文では，こうした大きな問題意識に立って，ネオ・サイバネティクスの理論を参照しながら，人工知能倫理をめぐる基本的な方向性を考察することとした。

人工知能の開発は，細かい違いはあるにせよ，「人間のように思考するコンピュータ」が目指されており，人間と機械との境界を越えようとする傾向が見出せる。そうした傾向に合わせるように，情報哲学の分野でも自然物と人工物との境界を設けず，情報的存在として生命／非生命を連続的に捉える理論も提示されている（Floridi, 2011）。K.Darling（2012）のように，基礎的な議論を経ないまま財産権を超えたロボットの保護について論じている研究者も現れている。しかし倫理的な問題において，生物と機械との差異について検討せず両者を同一線上に位置づけてよいのだろうか。こうした姿勢はいたずらに倫理の範囲を拡大してしまう恐れがある。ロボットに対して保護される権利を与えてしまえば，損壊したとしても捨てられない。あるいは人間が勝手に電源を切ったり改変したりすることが躊躇われる。人間が機械と対等に扱われる事態さえ招きかねない。責任を逃れるために人工知能それ自体に責任を帰属させる論法も出てくることが危惧される。これまで権利や責任の範囲は「生命なるもの」が必要条件であったが，人工知能が生命でないとすると，人工知能に責任を帰属させる立論はその必要条件を大きく崩すこととなる。生命と機械との差異は，倫理的議論の大きな分かれ目である。

けれども，人工知能の倫理をめぐる論考は多々あるが，生物と機械との差異をもとに検討した研究は見当たらない。2000年以降，人工知能もしくはロボットをめぐる倫理は大きく分けて２方向で議論が進められている。１点目は，ロボエシックス（roboethics）と呼ばれ，社会生活に組み込まれる人工知能と人間とのかかわりを探求しようとする方向性である(1)。たとえばD.Dennett（1997=1997）やD.Levy（2007），西條（2013）らの研究が挙げられる。２点目は，マシン・エシックス（machine ethics）と呼ばれ，いかにして人工知能に倫理的な側面を実装して道徳的な振舞いをするAMAs（artificial moral agents）を設計するかという方向性である。たとえば関口・堀（2008）やW.Wallach & C.Allen（2009），M.Anderson & S.Anderson（2011），久木田（2012），岡本（2012）らの研究が位置づけられる。これらの２方向の研究は，人工知能をめぐる倫理を考えるうえで示唆に富むものの，いずれも知能や感情に焦点が当たっており，生命と機械との相違についてはほとんど問題にしてこなかった。

しかし，この相違を議論せずに倫理的な問題を検討していけば，上述の課題が生じる。むしろ生命と機械との差異を踏まえたうえで，これまでの人工知能倫理の議論を整序-再編することが望ましい。これゆえ，本論文では最初に生物と機械との違いを議論の俎上に載せ，その立論をもとに人工知能倫理の基底的な領域を検討することとした。議論の過程で，倫理的責任の所在や倫理的配慮の必要性，道徳的共同体への包摂の可否，擬人化の問題についても扱っていく。

なお，2.2で述べるように人工知能の開発は多岐にわたっており，技術的には実現の道筋が比較的明確である短期的な計画もあれば，いくつもの技術的革新を経なければならない長期的な計画もある。それぞれは，後でいうアロポイエティック・

マシン／オートポイエティック・システムの開発計画に対応している。本論文では，倫理的問題においてこの２タイプの人工知能のあいだに大きな隔たりがあることを指摘するとともに，前者の人工知能をめぐる倫理的問題が喫緊であるため，その問題を中心に取り上げる。

2　オートポイエーシス理論に基づく生物と機械との相違点，ならびに人工知能技術の位置づけ

2.1　生物と機械との差異

本論文が依拠するネオ・サイバネティクスは，情報学の基礎理論であり，生物と機械との違いに焦点を当てつつ倫理的問題を考える理論的基盤を与えてくれる。ここでいうネオ・サイバネティクスとは，B.ClarkeやM.Hansenが使った用語であり，その核心はオートポイエーシス理論であるといってよい（Clarke & Hansen, 2009）。ラディカル構成主義や機能的分化社会論という特徴も挙げられるが，それらはオートポイエーシス理論の立場から導き出されるものである。オートポイエーシス理論は，H.R.MaturanaやF.J.Varelaが主に細胞や神経系，生物個体の認知機能に関する研究をもとにして学術的に定式化した（Maturana & Varela, 1980 = 1991）。オートポイエーシス（autopoiesis）とは，一言でいえば，自分で自分（auto）を制作（poiesis）しながら円環的に内閉したシステムであり，生命の十分かつ必要な条件を兼ね備えている（Varela, 1979=2001）。オートポイエティック・システムの定義を下に引用する(2)。

> オートポイエティック・システムは，構成素が構成素を産出するという産出（変形および破壊）過程のネットワークとして，有機的に構成（単位体として規定）された機械である。このとき構成素は，次のような特徴をもつ。(i) 変換と相互作用をつうじて，自己を産出するプロセス（関係）のネットワークを，絶えず再生産し実現する，(ii) ネットワーク（機械）を空間に具体的な単位体として構成し，またその空間内において構成素は，ネットワークが実現する位相的領域を特定することによってみずからが存在する（Maturana & Varela, 1980:78-79=1991:70-71）。

オートポイエティック・システムは，構成素の産出過程のネットワークであり，現れては消えていく構成素を絶え間なく産出しながらみずからを動的に特定する自己準拠的性質をもっている（図１）。過去の蓄積に依拠しながら絶えず自分を作っていくシステムである。

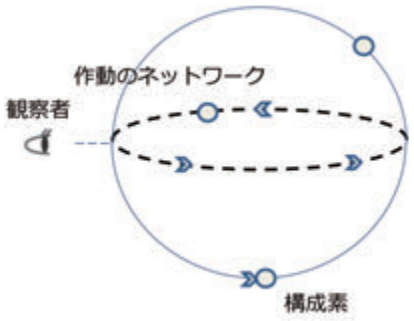

図１　オートポイエティック・システム

MaturanaやVarela自身は，原イメージとして，M.C.Escher作の「描きあう手」（drawing hands）をよく引き合いに出した。この絵画は，２つの手が互いに相手方を描きながら，両方の手が地の部分から離れて現出していくプロセスを視覚化している。オートポイエティック・システムは，絶えず自分で自分を作るように作動しながらみずからを環境と隔てて顕在化していくシステムであり，「描きあう手」のイメージと似通っている。

オートポイエティック・システムは，このように作動的に閉じているが，その内的原理に合わせて環境を知覚している。たとえば生物は，

その生存のために環境を知覚し捕食する。いわば，みずからにとって価値をもたらすと推定されるものを「図」としそれ以外のものを「地」としながら，不断に知覚を継続している。情報（information）の語源はラテン語のインフォルマチオ（informatio）から来ており，その原義は端的にいえば「生命にとって価値をもたらすもの」である(3)。生命がオートポイエティック・システムであり，そのなかで焦点化されていくものが情報であるといってよい。たとえば食物や異性, 敵, 気温などが情報として立ち現れてくる。

MaturanaやVarelaは，オートポイエティック・システムの特徴を際立たせるため，アロポイエティック・マシン（allopoietic machine）という概念も提出している。アロポイエティック・マシンは，オートポイエティック・システムの反対概念であり，「自動車のように，その機能が自分自身とは異なったものを産出する機械」（Maturana & Varela, 1980:135= 1991:242）であって，入出力関係に従属して動作するシステムである。すなわちアロポイエティック・マシンは，開放システムであり，ある入力をすれば常に一定の出力をするように固定化されているシステムである（図２）。

図２　アロポイエティック・マシン

オートポイエティック・システムと違い，アロポイエティック・マシンは，みずからを再帰的に存立させるのではなく外部からの介入によって成立せしめられるシステムであり，内部ではなく外的メカニズムによってその作動が決定される。たとえば，エアーコンディショナー（エアコン）である。当然のことであるが，エアコンは，フィルターや熱交換器，クロスフローファンなどの部品を自己制作するわけではない。それは外部にいる人間によって作られ適切に組み立てられている。人間がそのソフトウェアも作り，リモコンで設定した温度になるように調整されている。もし人間が設定した通りに動かず，冷房の指示を出したのに暖房しはじめたら故障である。エアコンは，アロポイエティック・マシンの典型的な例である。オートポイエティック・システムは自己準拠的に作動するのに対し，アロポイエティック・マシンは基本的に入出力関係に従属して作動する。

2.2　第３次ブームの人工知能の位置づけ

こうした概念区分に基づけば，第３次ブームで実現されつつある人工知能技術は，第１次・第２次ブームと同じく，アロポイエティック・マシンに位置づけられる。もちろん，IBMのWatsonのような高度なエキスポート・システム，あるいはディープラーニングを使った画像認識や音声認識，自然言語処理，さらには自動運転車やロボット兵器などは，仕事や遊び，交通，戦争などに大きなインパクトを与えていくだろう。周知のように，グーグルの人工知能は，ディープラーニングの手法を用いて膨大な計算を行い，猫の視覚的特徴量を教師なし学習でみずから見出している。音声認識も，ディープラーニングによってノイズの多い環境での認識率が高まった。こうした例はほかにもある。たとえば，視覚・聴覚・触覚を複合的に組み合わせたマルチモーダル情報を自動的に計算してカテゴライズするロボットが開発されている（長井・中村, 2012)。カメラやマイク，触覚センサーからのデータを計算しクラスタリングして，高い精度で物体を分類している。人工知能は，実践的な体験に根ざした概念形成にまで至っていないとはいえ，部分的にせよみずから物的概念を形成していると考えられる。また進化論的計算の手法を使った研究でも，コンピュータ

が自動計算する領域が拡大している。たとえばBrodbeckら（2015）は,「母ロボット」がモジュールを組み合わせて「子ロボット」を作り評価し，さらにその評価をもとに組み合わせを行い，自走するスピードが速いロボットを生み出している。母ロボットなる操縦機やその内部で自動計算するソフトウェア，子ロボットを構成するモジュール，実験方法，目標設定は開発者が制作しているが，母ロボットが進化的アルゴリズムに則り，モジュールを組み合わせて制限時間内に遠くまで移動できる子ロボットを自動で割り出すまでには至っている。こうしたことは技術上の確実な成果であるといえる。

しかし人工知能が機能するには，人間が要求仕様を固め，どのようなアルゴリズムを採用するかという設計仕様を確定し，それに応じてプログラムの命令と引数を細かく記述しなければならない。ディープラーニングでも，人間がTorchやChainer，TensorFlowなどのフレームワークを整え，そのフレームワークを利用しながら畳み込み層・プーリング層を多段に連結して後半に全結合層をつなげて関数で正規化する準備を予め施さなければならない。ソフトウェアばかりではない。人間がGPUなどのハードウェアも用意し，必要なデータはセンサーを設置して収集している。試行錯誤しながら，一定の入力に対して一定の出力がなされるように調整している。マルチモーダル情報のクラスタリングや進化論的計算の分野も，大きく変わらない。すなわち，人間の介在なしに特徴量（素性）の抽出が可能となったにせよ，精度が上がり柔軟性が増したにせよ，自動計算の領域が広がっているにせよ，第３次ブームの人工知能も，オートポイエティック・システムに近づきつつあるとはいえても，あくまでアロポイエティック・マシンに位置づけられる。

もちろん発明家R.Kurzweilが提唱した「技術的特異点」（technological singularity）以降では，人工知能は，人間の知能を上回り，それ自身を自己生産することが想定されている（Kurzweil, 2005=2007）[(4)]。群ではなく単位体のレベルでそうした状況が仮に実現すれば，オートポイエティック・システムと位置づけされ，生命の条件が満たされうる。後でみるように生命であるか否かは，責任の帰属が大きく分かれる分岐点であり，技術的特異点の前後では大きく倫理的問題の議論が変わってくる。自分で自分を作るという意味での自律的存在に人工知能がなるということは，いわば外部からのコントロールが効かないということである。哲学者N. Bostrom（2014）がいうように，いかにして人類の知能を超える人工知能を制御し，人間に害をもたらさないように設計するかというコントロール問題が生じてくる。人間の指示通り動作せず，人間と人工知能との対決というSF的な構図を想像することが可能な段階である。この段階では，人工知能自体が自己のコントロール権を有する。したがって，人工知能自体に行為者性を帰属することも可能であろうし，責任を帰属することも可能である。法的な責任を課す制度を作るならば，微妙ならざる差異はあるだろうが，組織に対して適用している「法人格」付与の手続きと似た手法を採ることが考えられる。なお社会学者N.Luhmannは，オートポイエーシスの範域を拡張し，生命と同じく自己創出性が見出せるものとして人間心理や社会を記述した（Luhmann, 1984 = 1993,1995）。それゆえ，ネオ・サイバネティクスの観点からすれば，人間とともに組織に人格を認めるのは一定の妥当性をもっている。

注意を払わなければならないのは，自律性（autonomy）という言葉の意味である。自律性という語は多義的である。自律型自動車（autonomous vehicles）や自律型武器（autonomous weapons）のように，自律的なマシンは製造可能とみなされることが増えてきた。しかしその自律とは，学習したパターンにしたがい新たな対象物を分類するといった意味であり，

「自動的」という言葉に近い。これに対して，オートポイエティック・システムもその特徴を自律という語で表現できる。ただし，ここでいう自律は外部からの働きかけではなく内部の原理に則って作動し，自分で自分を作るという意味である。前者は技術的に実現可能と予測されるのに対し，後者については実現の道筋がはっきりとはみえていない[(5)]。

2016年現在，自己産出する人工知能は，その到来を想像はできるものの，あるいはWhole Brain EmulationやWhole Brain Architectureのような興味深いプロジェクトが進んでいるものの，その技術的成果がみえていない段階であるといってよい。これまで人類が発明してきたテクノロジーはすべてアロポイエティック・マシンである。生命の起源が不分明である以上，オートポイエティック・システムの人工知能が生まれる可能性は皆無ではない。しかし有史以来そうしたテクノロジーが制作されておらず，自己制作する人工知能の実現は困難が予想される。長らく開発が進められている自己反映計算の機能を高めなければならず，加えてハードウェアも作って必要なデータはセンサーを設置して集めてこなければならない。また，外部を測定しそれに対応する能力の汎用性を高めなければ，すぐさまトラブルが生じるであろう。さらに，自己制作する人工知能は，人間によるプログラミングが必要なくなる一方で外部からの制御可能性が低くなり指示通り動くか分からないため，企業がどこまで市場価値を見出して開発・販売に労力をかけるかは未知数である。販売に乗り出す場合，「意図通り動かなくとも故障ではありません」と説明書きを入れなければならないことは想像に難くない。したがって，緊要かつ切実なのは，現代社会に早期に組み込まれる可能性の高いアロポイエティック・マシンとしての人工知能にかかわる倫理的問題である。特化型人工知能／汎用型人工知能の区分でいうと，先に特化型人工知能が社会生活空間に入り込む。特化型人工知能とは，その名の通り特定のタスクをこなすために開発された人工知能であり，別のタスクへの汎用性がない。特化型人工知能はすでに実現されている技術もあり，こうした人工知能をめぐる倫理的問題を考慮することは喫緊の課題である。

3　倫理的責任（moral responsibility）の帰属

3.1　人工知能に責任を帰属することの困難さ

オートポイエティック・システムは，その内的原理に基づいて動作する。つまり，その内部にシステム自体をコントロールする作用がある。それゆえ，オートポイエティック・システムに倫理的責任を課すことは可能である。もちろん，すべてのオートポイエティック・システムが責任を取る条件を満たしているわけではない。植物やペットは，生物でありオートポイエティック・システムであるが，罪に問われない。犬が人を噛んだら責は飼主に帰せられる。子供が暴力沙汰を起こせば，親にも非難は及ぶ。すなわち，行為者と責任者は概念上区別されるのであり，オートポイエティック・システムでありかつ善悪の分別がつくとみなされる人物のみが倫理的責任を帰属させられる[(6)]。逆にアロポイエティック・マシンは，生物でもなく外部からの指示にしたがい動作するため，責任が帰せられる公算が低い。窓ガラスが割れて怪我をしたとしても，窓ガラス自体に対して責任が追求されることはない。エアコンが壊れて熱中症で人が死んだとしても，エアコン自体が起訴されることはない。したがって，オートポイエティック・システムであることは，十分条件ではないが責任を帰すための必要条件である。自分で自分を作る人工知能でないと，責任は帰属できない。

「コンピュータが殺人を犯したら，だれが罪に問われるのか」というありふれた問いがある。ネ

オ・サイバネティクスの理論に基づけば，その回答は自己制作する性質を備えているか否かによって分かれる。アロポイエティック・マシンの人工知能が暴走し危害や損害が生じれば，罪に問われるのは人工知能側ではなく所有者や開発者，販売者といった人間側といえる。その作動は人間側が定めているからである。それゆえ，自己制作していく人工知能を開発していく道筋が明確につかめていない現段階では，人工知能自体の倫理ではなく人工知能に携わる人間側の倫理に焦点を当てることが望ましい。

冒頭で述べたAMAsの設計についても，いかにして人間が用いている倫理的推論をコンピュータ技術者がAMAsに実装していくかに力点が置かれている。WallachやAllenはロボットが備えるべき倫理的価値のレベルを「operational morality」「functional morality」「full moral agent」の3段階に分け，トップダウン・アプローチとボトムアップ・アプローチの両面から自律性と価値に対する感受性を高めるように人間が関与するべきという（Wallach & Allen, 2009）[7]。operational moralityとは設計者が明示的に埋め込んだ倫理的対応のみを示すレベルであり，functional moralityとは教えこまれたパターンをもとにして機械自体が倫理的な問題に自動的に判断を下すレベルである。現存する人工知能の大半は，operational morality の段階であり，functional moralityの水準に到達している機械であってもその初歩段階にすぎない。最上位のfull moral agentは，自己反省するレベルであり，現時点では実現にはほど遠い（Wallach & Allen, 2009: 34; Wallach & Allen, 2012）。いうなれば，AMAsもアロポイエティック・マシンであり，人間がどのような点に着目して倫理的判断を行わせるかを決めている。

人工知能がアロポイエティック・マシンの段階にある場合，有名なI.Asimovのロボット3原則は意義を有するだろうか。「第一条　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，人間に危害を及ぼしてはならない。第二条　ロボットは人間にあたえられた命令に服従しなければならない。ただし，あたえられた命令が，第一条に反する場合は，この限りでない。第三条　ロボットは，前掲第一条および第二条に反するおそれのないかぎり，自己をまもらなければならない。」（Asimov, 1950 = 2004:5）の3原則である。このロボット3原則は，ロボット自身が自意識をもっていることを前提としている。自意識は，心理レベルでのオートポイエティック・システムから生まれてくると考えられるが，この自意識が現出しないとすれば，それは柴田正良が指摘するように単なる仕様書であり，家電製品が求められる3原則にすぎない。「（一）安全（人間を危険にさらさない），（二）従順（人間の意図通り動く），（三）堅牢（簡単には壊れない）」（柴田, 2010: 23-24）である。そうした意味で, ロボットという言葉の原義である「労働」「隷属」に合った原則といえる。

誤解を避けるために付け加えておくと，AMAsの設計の取り組みに見られるように，倫理の基準を埋め込んだアロポイエティック・マシンが作ることは十分に可能である。たとえば功利主義が正義であるとすると，その原理を基準とする自動運転車は正義の判断が組み込まれているといえるだろう。しかしそれは，人間が考える正義をプログラミングしてインストールしたルールに基づいて動くだけであり，自動運転車が正義について内発的に思考した帰結ではない。そのため，正義に基づく判断がつくように見えるアロポイエティック・マシンは，あくまで人間によって作られたものであり，その責任は人間もしくは人間社会が引き受けるべきである。

3.2　自動運転車やロボット兵器を事例として

こうしたことを踏まえ，紙幅の都合上，近年注目を浴びている自動運転車やロボット兵器を取り

上げ，それらにかかわる応急の倫理的問題について簡潔に述べる。これらは，自己制作する人工知能であれば指示通り動くとは限らず，アロポイエティック・マシンとして開発が進められると想定される。まず自動運転車の欠陥にかかわる事項である。内閣府政策統括官（科学技術・イノベーション担当）（2015）は，自動運転レベルを4段階に分け，レベル４「加速・操舵・制御を全てドライバー以外が行い，ドライバーが全く関与しない状態」を完全自動走行システムと名づけた。そのレベル４に位置する完全自動運転車は，準自動運転車に比べて開発企業等にとってリスクが大きいといえる。アロポイエティック・マシンである完全自動運転車自体に責を帰すわけにはいかないからである。また，しばしば指摘されるように，事故を起こしたときには，製造物責任法がソフトウェアに適用しがたいにせよ，責任を問われてしまうからである。異常気象で環境が大きく変わり画像認識がうまくいかず誤作動を起こしたら，それは設計上のミスである。アメリカの非営利組織FLI（Future of Life Institute）は人工知能に関する公開書簡を発表しているが，その文書には次のような皮肉めいたコメントが載せられている。「自動運転車の導入によっておよそ４万件の交通事故死がなくなったとしても，自動車メーカーに寄せられるのは２万通の感謝の手紙ではなく，２万件の訴訟かもしれない」（Future of Life Institute, 2015a）。自動車メーカーの立場からすると，責任の帰属を逃れるためには，たとえ技術的に可能でも完全に自動運転化はせず，一歩手前の準自動で止め，人間にハンドルを握らせたほうがリスクの低減にはなると考えられる。もしドライバーの運転と自動運転車の判断に齟齬が生じたときは，人間の操縦を優先させなければドライバーの過失を問えない。運転者側にコントロールする機会が与えられていないからである。その帰責は自動車メーカー等が負う可能性が高い。

ただし自動運転車の市場規模は，高齢者や障がい者，運転免許非保持者の人数，飲酒しても自動車移動を可能にすることなどを勘案すると，けっして小さくはない[(8)]。交通事故や排気ガスの減少，移動時間に別の作業に集中できるといった効用も期待されているが，運転できない人たちの自動車移動を可能にすることは，そうした人たちに対するバリアフリー／ユニバーサルデザインの実現であり，倫理的配慮にもつながる。それゆえ，自動運転車に合った保険制度を確立し，保険会社と自動車メーカー等が包括的な契約をして完全自動運転者を販売する方策が採られるかもしれない。『日本経済新聞』が報じたところによれば，東京海上日動火災保険株式会社は2014年に自動運転に対応した自動車保険を検討しはじめている（瀬川ら，2016）。すでにドローンは保険が作られており，たとえばDJI社は製品「PHANTOM3」を三井住友海上火災保険株式会社の業務保険つきで売り出している。 そうした先行する例を参考にして保険制度を構築することになると予想できる。安全規格も必要とされるだろう。

自動運転車の設計思想にも倫理的な問題が含まれている。自動運転車が回避しきれない事故に直面した場合，どのような倫理的な基準を採用してハンドルを切るかは前もってルール化しておかなければならない。J-F. Bonnefonらのオンライン調査（2015）によると，人々は功利主義的な意思決定を支持している。たとえば，自動運転車が直進すると10人が死亡する。進路を変えると1人が死亡する。そうしたときには，進路方向を変え，より多数の人々を救うようにプログラミングしておくことが肯定的に受け止められている。自己犠牲が伴う場合でも同じであった。功利主義的な判断は，現行の刑法37条「緊急避難」の措置に適合しており，社会的な同意が得られやすいと想定される。もちろん，自由主義や共同体主義に則った意思決定も可能であろう。こうした倫理的意思決定も，人間の責任の範囲として受け止め議論を重ねてからロボットに埋め込むべきである。

責任の所在を別にすれば，完全自動運転車に対する懸念として，そこに爆弾を積めば次に述べるロボット兵器となることが挙げられる。これは，アメリカのDARPA（国防高等研究計画局）が自動運転技術のコンテストを実施していたことから容易に連想できることである。

次にロボット兵器である。ロボット兵器は，遠隔操作もしくは自動で相手に攻撃を仕掛けることができ，すでに開発されている。例としては無人飛行機や無人偵察機，自動機関砲が挙げられる。自動運転車と同様，あるいはナイフや拳銃と同様，ロボット兵器自体に責任を帰すことは難しい。ロボット兵器の製造や使用者などの人間側に責任は帰属する。ロボット兵器は，味方の兵士の危険性を減らすため需要が高い。しかし，敵やその周囲にいる人々への人道的配慮については疑義を挟まなければならない。ロボット兵器を使ったからといって人命を奪うことには変わりがない。あきらかにAsimovのロボット３原則からは外れている。

自動ロボット兵器は，事前に設定した条件にあった人間に攻撃を加える。けれども，学習したパターンで実際にどれほど敵と味方の区別をつけられるだろうか。人工知能が顔認識していると判断すれば，敵は顔を布で覆うだろう。全身の画像認識であれば髪型を変え服も変えるだろう。軍服を脱ぎ私服を着ていたら，敵側の民間人との区別はつきにくい。ディープラーニングによる画像認識の精度が人間並みになったとはいえ，そうした識別ができなければ誤射・誤爆を招いてしまう(9)。

P.W.Singer（2009=2010）が述べているように，アメリカのロボット研究者はDARPAからの資金援助を受けていることが多いが，それを拒否する研究者もいる。みずからの技術が戦争に使われることに耐えきれなかったからである。こうした姿勢は，サイバネティクスの父N.Wienerを思い起こさせる。Wienerは，第２次世界大戦中，フィードバック制御に基づいた弾道計算の研究に従事した。しかしその後，日本への原子爆弾投下に精神的ショックを受け，軍事研究からは距離をとり，科学者による軍事研究を厳しく非難した。いうまでもなく，Wienerのサイバネティクスは，本論文で依拠しているネオ・サイバネティクスの源流である（西垣，2010）。日本学術会議は，戦時中に科学者が兵器開発に従事したことを省み，1950年に「戦争を目的とする科学の研究には絶対従わない決意の表明」をしている。FLI（2015b）もまた，自動ロボット兵器が火薬や核兵器に続いて戦争に革命を引き起こすものであるとし，その禁止を強く訴えている。

4　擬人化

4.1　擬人化の先鋭

第３次ブームの人工知能は，生物（人間）ではない。とはいえ，さまざまな場面で擬人化―ネオ・サイバネティクスの用語になぞらえていえば，擬似オートポイエティック・システム化―はすでに起きており，これから後そうした傾向にますます拍車がかかる可能性がある。

シミュラクラ現象にみられるように，コンピュータ技術の有無にかかわらず元来，人間はさまざまな事物を擬人化してきた。自動車を前からみたときに顔のようにみえるときがある。人面魚もそれに類する。それ以外にも，ギリシャ神話にあるピグマリオンが彫った乙女像，人形浄瑠璃，チェコの人形劇など例は数多い。

コンピュータ技術に関しても擬人化が生じている。人工知能研究の初期に開発されたELIZAも擬人化の例である。ELIZAは簡単なルールにしたがって言葉の意味も分からず回答しているだけだが，ユーザーはその会話のなかで相手が人間であるかのように感じ，のめり込んだことがあったという。また，不倫サイトといわれる「Ashley Madison」では，ユーザーが会話ロボットを人間の女性だと思い込んでサービスを利用していた（Pagliery, 2016）。LINE公式アカウントの「パン

田一郎」「りんな」，ウェブサイト「脳内カレシ具体化計画」といったサービスも，こうした例に属す。世界的な人気を誇る初音ミクも，声優の声の入った単なるボーカロイドではなく，キャラクターのイラストや年齢等のプロフィール設定をしたことで擬人化が引き起こされた。事実，初音ミクがまるで身体をもったかのように「そんな高音苦しいわ」と歌う曲が作られ話題となった。

あるいは，将棋の電王戦も擬人化の例として挙げられるだろう。コンピュータ対人間といった構図で電王戦が報道され，ハードウェアもソフトウェアも人間が作っているにもかかわらず，将棋用コンピュータを作っている人間の存在が意図的に消去され語られることが多い。すなわち，コンピュータ自体が機械学習の方法も含めてすべて行っているかのように擬人化されている。ペリパーソナル･スペースのような考え方を借りれば，電王戦は，コンピュータを使った人間と使わない人間との異種格闘技戦にすぎないが，そのような見方で語ると耳目を集められないと考えられる。作為的に擬人化している様子が見て取れる。報道記者以外の言葉としても，棋士自身が「コンピュータは怖がらずにちゃんと読んで，踏み込んでくる。強いはずですよ。怖がらない，疲れない」（山岸，2013）とコメントしており，対戦者の目線でも擬人化が起きていることが窺われる。

人間の姿形のイメージは擬人化に必ずしも必要とされない。たとえば，B.ReevesやC. Nassの実験によれば，我々は，単なるコンピュータであってもまるで人間であるかのように接する（Reeves & Nass, 1998 = 2001）。彼らは，ある実験でコンピュータに対して我々が礼儀正しく振る舞うかを検証している。実験の方法は，まずコンピュータが一通りの豆知識を被験者に教え，次に被験者が豆知識の印象をアンケート方式で回答する。この回答の方法で２通りのものを用意し比較考量する。豆知識を教えたコンピュータと同じ端末で豆知識の印象を聞く場合と，別の端末で印象を聞く場合との比較である。その結果，被験者は，同じ端末で評価した場合のほうが，別の端末で評価より肯定的な回答を行った。すなわち，我々が人間相手に示しているのと同様，豆知識を教えたコンピュータに対しては礼儀正しく接した。

また，相手を人間のように扱ったケースではないが，４本足で犬のように足を動かすロボットを蹴ったところ，可哀想という声が相次いだ（CNN, 2015）。擬人化ならぬ擬生命化が起きている。

実際には，人間の姿形のイメージが前景化している場合のほうが擬人化は喚起されやすいと想定される。自動販売機や電子レンジもロボットといえるが，そうした機器に人間性を見出す人は稀であろう。人間の姿形のイメージが伴う例は，各種オンライン・サービスのアバター，あるいはPepperのようなロボットが挙げられる。目があり，顔が動き，また手足を動かしながら話しかけられると，人間は親近感を引き起こされやすい。今後，人間とのスムーズなやりとりが可能になるような人型ロボットが生活の場に置かれていくだろう。ペットのような仕草をするロボットとしてはAIBOやパロなどが有名だが，人々はそれぞれのロボットに愛着をもち特別性―ほかの同じ機械では代替できない，かけがえのなさ―をしばしば感じている。人型ロボットにも同じような感情移入が起きることが想定される。性愛の対象としてのロボットも登場する可能性がある（Levy, 2007; 西條, 2013）。

4.2　擬人化に伴う倫理的問題

擬人化された人工知能に行為者性を付与することは既に行われている。代表的な例は先述した将棋の電王戦である。アメリカのNHTSA（運輸省道路交通安全局）が2016年２月に自動運転車の人工知能を運転手とみなせる可能性があるとの見解を示したが，この見解にも擬人化が生じており，人工知能に行為者性が帰属されているといえる。ただし，これらはあくまでも擬人化であり，実際

は設定した目標値に向かって計算処理をしているアロポイエティック・マシンである。人間が機械学習の手法や教師用データ，各種のハードウェアを整えている。擬人化による行為者性の付与で危惧されるのは，人工知能が高度化し汎用化してくると人間の尊厳が問われてくることである。とはいえ擬人化は，行為者性の付与までで留めることが望ましい。倫理的責任にまで推し進めると，事故が起こったとしても誰も責任を取らない無責任状態が引き起こされかねかいからである。

ロボットとの間に倫理的な関係を結べるか否かも重要な論点である。繰り返し述べるように，自己制作しないロボット自体は生物（人間）ではない。アロポイエティック・マシンである。倫理は他者との関係で生じるものであるため，ロボット自体に倫理的配慮をする必要はない。I.Kantのよく知られた道徳律「他者の人格を単なる手段としてのみならず，目的としても尊重すること」にしたがわなくともよい。24時間休みなく動かせたり，売買したり，また故障したら廃棄しても差し支えない。崩壊した建物の内部のような危険性の高いところにロボットを送り込んでも構わない。ロボットが生命ではないがゆえの利便性である。そして，便利／不便の尺度だけでロボットを測っても人権を不当に傷つけたことにはならない。万が一，ロボットに「人権」を与える「ロボットの権利」を打ち出すなら，便利／不便の尺度の放棄が半ば伴う(10)。また，生命（人間）ではなく，ましてや法人のような社会レベルのオートポイエティック・システムですらない存在に人権を与えるという論理的飛躍を埋め合わせる立論が不可避となる。

もちろん人工知能は，人間の道徳的共同体にも属さない。Bostromは次のように述べている。「欲求に関しては，人工知能は人間との共通点がほとんどない。地球外生命体よりも共通点を見出すことは難しいだろう」（Bostrom, 2014:106）。地球外生命体がいるとしたら，それは進化のプロセスを経ており，人間もまた生命が誕生してから約40億年のオートポイエティック・システムの来歴を引き継いでいる。けれども，人工知能はそうではないからである。人工知能が道徳共同体に組み込まれるためには，人間と同じような生理学的知覚構造も要件だが，少なくとも他者の心のうちを推し量る能力は欠かせない。ネオ･サイバネティクスの用語でいえば，オートポイエティック・システムとして他者の心理を眺める観察者の能力である。自分と同じように他者も思考と感情があり悩み苦しむ存在であることが見出せてはじめて，他者への倫理的配慮がもたらされると考えられるからである。そのほか，人工知能にも寿命や病気，死がなければ，人間の喜びや悲しみ，宗教などについて理解が及ばないと想定される。たとえ自然人ではないロボットに「人権」を与える手順が踏まれたとしても，ロボットが組み込まれた社会における正義をいかに考えていくかという課題が立ちはだかってくる。

ただし擬人化のことを考慮すれば，人工知能に愛着をもち特別性を感じている人たちへの配慮が必要となってくるだろう。AIBOは，1999年から2006年までソニー株式会社が発売した犬型ロボットであるが，2014年３月まで長らく修理サポートが継続されていた。しかし，ソニーによる修理サポートが打ち切られた後も，修理に対する要望は根強く，ソニーのOBが立ち上げた会社には修理依頼が次々と押し寄せている（宗像，2015）。論理的にみれば，犬型ロボットはぬいぐるみとさほど変わらない。しかし愛着のあるぬいぐるみに旅をさせるツアーも人気であることを考えると，それほど不思議ではない。もし動物型ロボット・人型ロボットにあまりにも愛情を注ぐ人が続出すれば社会的にも対応を迫られると考えられる。動物型ロボット・人型ロボットについては修理サポートを長めに設定することや動物愛護法に似た法制度が求められる可能性がある。動物愛護法が制定されたねらいとしては，その第一条に

示されている通り，動物を守ることによって人間社会に生命尊重の思想が醸成されることが挙げられる。人工知能の保護がそうした目的に合致する段階に達したならば検討しなければならない課題である。

また1990年代後半以降，インターネット依存（嗜癖）が社会問題化した。オンラインゲームを中心にしてインターネット上の活動に極度に夢中になり，それ以外の生活に支障が出る人が続出している。認知療法や行動療法による治癒が期待されている（河島, 2015）。インターネット依存の研究で蓄積された知見をもとにして，ロボット依存の判定基準や治癒させる医療サポートすら必要になることが想定されうる。

ロボットへの愛は，セラピーロボットに見られるように人間を慰撫する一方で，ビジネスのターゲットとなる。恋人や子供に接するような感情を抱かせグッズを売っていく商法が予期される。また，「心理的依存の対象となったロボットから発せられる情報は，より耳を傾けるべきものとしてユーザーに受け取られかねない」（西條, 2013: 45）。したがって，広告の規制も想定されうる。

5　結語

本論文は，個別的な議論に入る前段階として，人工知能倫理を考える基本的な枠組みを提示することを目指している。人工知能倫理に関する既往の研究では生命と機械との相違点が等閑視されていたが，本論文では，ネオ・サイバネティクスの理論に基づき，生物（人間）／機械の区分を人工知能に適用して倫理的問題の基礎づけを試みた。このことにより次の言明が導かれる。第3次ブームを支える人工知能は，人間が設定した目的に応じたアウトプットが求められるアロポイエティック・マシンであり，それ自体に責任を帰属することは難しい。人工知能に関する倫理は，あくまでも人間側の倫理に帰着する。人工知能は，倫理的配慮の直接的対象にはならず，道徳的共同体にも含まれない。とはいえ人間は，しばしば生物ではない事物を擬人化してきており，人工知能の擬人化が進んでいる。特に生物を模した事物に対してはその傾向が顕著である。少なからぬ人々が人工知能に度を越した愛情を注ぐようになると，そうした人びとに配慮した社会制度上の対応が要される。その場合であっても，自然人や法人とは違い，人工知能はあくまでもアロポイエティック・マシンであることには留意しなければならない。万が一，ロボットに「人権」を与える必要性を訴えるなら，かつてない大きな論理的飛躍を埋め合わせなければならない。

自動運転車などの例にみられるように，人工知能はこれまでにない倫理的問題を喚起している。人間が運転する場合は，不測の事態に対応するにしても反射的に判断するだけであって，倫理観が問題視されることはない。しかし，完全自動運転車で想定されている計算スピードはきわめて高速であり，事前のルールに基づいて動作を決める時間的余地が生じる。したがって，前もってどのような設計思想でアルゴリズムを組み立てるかが問いとして提起された。このように新たな技術が実現されることによって倫理的問題が引き起こされる例はほかにも存在する。たとえば，超音波検査が確立したがゆえに出生前診断が可能となり，ダウン症や二分脊椎症の検査が胎児に対してできるようになった。親は，その検査をするか否か，検査をしてそれらの病気を患っている可能性が高いと判断された場合はどうするかといった倫理的な判断が求められるようになった（Verbeek, 2011=2015）。これから人工知能がさらに社会生活に入ってくるとすれば，その技術が引き起こす倫理的問題を慎重に検討しなければならない。

本論文で扱わなかった課題としても，人工知能の内部メカニズムのオープン化，人工知能の開発・製造・販売・メンテナンス等にかかわる人たちの登録制や免許制，あるいはウェブサイトで使われ

ているようなセキュリティの認証といった問題がある。ロボットがインターネットに接続することを鑑みれば，個人情報の取扱いもより慎重にしなければならない。これ以外にも多数の課題がある。海外の動向を踏まえながら，制度設計を前もって整えていくことが喫緊の課題である。

メディアの社会構築主義的研究が知見を蓄積してきたように，技術は，可能的様態をいくつも保持しており，「国家や資本の編制力から，市民，あるいは大衆の想像力にいたる，複合的で重層的な社会の諸力の錯綜した結果として，今日のような姿に固定化されてき」（水越，1996:186-187）ている。人工知能の倫理的問題に関する議論も，未来の人工知能の在り方を決める契機になると考えられる。

謝辞

本論文を執筆するにあたり，AIネットワーク化検討会議（総務省情報通信政策研究所），ネオ・サイバネティクス研究会，AI社会論研究会で意見交換する機会を得ました。これらに関係された方々ならびに査読者の皆様に深謝申し上げます。

注

（1）「ロボエシックス」は，2002年にG. Veruggioによって作られた造語である（Veruggio, 2006）。

（2）MaturanaやVarelaは，生気論を避けるため，オートポイエティック・システムを機械の一種だと定位した。しかし，一般的な機械はアロポイエティック・マシンという呼称で表し，それとオートポイエティック・システムを区別している。オートポイエティック・システムは，特殊な機械であり，それが生物であると立論した。

（3）詳しくは『基礎情報学』（西垣，2004）を参照。

（4）自己制作する人工知能のイメージとしては，漫画『攻殻機動隊』の人形使い，映画『her/世界でひとつの彼女』の人工知能型OS，あるいは映画『トランセンデンス』のマインド・アップロードされた主人公を思い浮かべれば分かりやすい。

（5）オートポイエーシス理論では構造（structure）と有機構成（organization）との区分を導入している。オートポイエティック・システムの有機構成は外からはみえない。したがって計算式で表せない。ルールベース・テクノロジーでは生命の根幹たる有機構成は記述しきれないと考えられる。とはいえ，ディープラーニングの手法では，ニューラルネットワークが重みづけを自動的に計算して動き，それが外からも観察しきれない有機構成をなすことはありえると思われるかもしれない。しかしながら，オートポイエティック・システムの有機構成は生命の必要条件であり，自分で自分を作るということが構造のレベルで実現されていないのであれば有機構成も成立していないといえる。

（6）L.Floridi & J.W.Sanders（2004）は，本論文とはまったく違った理路に基づき，抽象化のレベルの違いを持ち出すことで行為者と責任者との区別をしている。彼らの論法では，人工知能は行為者となりうるが責任は帰属されない。彼らの議論についての批判的検討は別稿を期す必要がある。

（7）トップダウン・アプローチは，功利主義やI.Kantの定言命法などの倫理的基準をコンピュータに実装するものであるが，それだけではさまざまな困難にぶつかる。たとえば，功利主義的な計算をするうえでも，「どの時点を帰結とみなすか」「どれだけ先の未来をコンピュータが計算できるか」「計算する効用をどのように規定するか」「どのように異なる帰結の効用を比較考量する

か」といった問題が挙げられる。自動運転車が避けられない事故に直面した場面のように限定した領域でならさほど問題ではないだろう。

（８）　平野（2014）によると，完全自動運転車は製品分類全体責任には認定されにくい。というのも，複数の効用があり，また司法府よりも立法府や行政府が判断するべきものであるためである。

（９）　2015年10月の報道によれば，アメリカの無人攻撃機によって殺害された人のうち，９割近くにも上る人が標的ではなかった(Blake, 2015)

(10)　基礎情報学がモデル化したように，観察者の視点を変えれば，社会レベルのオートポイエティック・システムが存立するために，そこに関係する人間がアロポイエティック・マシンとして道具のように現出する(西垣, 2004)。本論文で，このことを否定しているわけではない。本論文では視点移動の操作を行わず，オートポイエティック・システム間の関係についても論じていない。

参考文献

Anderson, Michael & Anderson, Susan, eds.（2011）*Machine Ethics*, Cambridge University Press, 548p.

Asimov, Issac (1950=2004) *I, ROBOT*, Gnome Press, 253p.（小尾芙佐訳『われはロボット』, 早川書房.)

Bonnefon, Jean-François & Shariff, Azim & Rahwan, Iyad (2015) Autonomous Vehicles Need Experimental Ethics,
<http://arxiv.org/pdf/1510.03346v1.pdf> Accessed 2016, March 12.

Bostrom, Nick (2014) *Superintelligence*, Oxford University, 328p.

Blake, Andrew (2005) Obama-led drone strikes kill innocents 90% of the time
<http://www.washingtontimes.com/news/2015/oct/15/90-of-people-killed-by-us-drone-strikes-in-afghani/> Accessed 2016, July 7.

Brodbeck, Luzius & Hauser, Simon & Iida, Fumiya (2015) Morphological Evolution of Physical Robots through Model-Free Phenotype Development
<http://dx.doi.org/10.1371/journal.pone.0128444>　Accessed 2016, July 7.

Clarke, Bruce & Hansen, Mark, eds. (2009) *Emergence and Embodiment*, Duke University Press, 296p.

CNN (2015)「「ロボット犬」でも蹴っちゃダメ？倫理めぐる議論盛んに」,
<http://www.cnn.co.jp/tech/35060457.html> Accessed 2016, March 12.

Darling, Kate (2012) Extending Legal Protection to Social Robots,
<http://spectrum.ieee.org/automaton/robotics/artificial-intelligence/extending-legal-protection-to-social-robots> Accessed 2016, March 12.

Dennett, Daniel (1997=1997) Did HAL Commit Murder? *Hal's Legacy*, MIT Press, pp. 351-365.(内田昌之訳「HALが殺人をおかしたら，だれが責められるのか?」『HAL伝説』早川書房, pp.391-408.)

Duffy, Brian R. (2006) Fundamental Issues in Social Robotics, *International Review of Information Ethics*, Vol.6 pp.31-36.

Floridi, Luciano & Sanders, Jeffry.W.（2004) On the Morality of Artificial Agents. *Minds and Machine*, No. 14, pp. 349–379.

Floridi, Luciano (2011) *The Philosophy of Information*, Oxford University Press, 405p.

Future of Life Institute（2015a）Research Priorities for Robust and Beneficial AI, <http://futureoflife.org/static/data/documents/research_priorities.pdf> Accessed 2016, March 12.

Future of Life Institute（2015b）Autonomous Weapons, <http://futureoflife.org/AI/open_letter_autonomous_weapons> Accessed 2016, March 12.

平野晋（2014）「製造物責任（設計上の欠陥）における二つの危険効用基準」『NBL』1040号, 商事法務, pp.43-57.

河島茂生（2015）「インターネット依存」『情報倫理の挑戦』学文社, pp.53-76.

久木田水生（2012）「ロボットは価値的記号を理解できるか」, 京都生命倫理研究会

Kurzweil, Raymond（2005=2007）The Singularity is Near, Viking Press, 672p.（井上健監訳, 小野木明恵・野中香方子・福田実訳『ポスト・ヒューマン誕生』日本放送出版協会, 661p.）

Levy, David（2007）*Love and Sex with Robots*, HarperCollins Publishers, 334p.

Luhmann, Niklas（1984 = 1993,1995）*Soziale Systeme*, Suhrkamp Verlag, 675S.（佐藤勉・村中知子・村田裕志・佐久間政広・永井彰・小松丈晃訳『社会システム理論（上）』・『社会システム理論（下）』恒星社厚生閣, 994p.）

Maturana,Humberto & Varela,Francisco（1980 = 1991）*Autopoiesis and Cognition*, D.Reidel Publishing Company, 171p.（河本英夫訳『オートポイエーシス』国文社, 320p.）

水越伸（1996）「情報化とメディアの可能的様態の行方」『メディアと情報化の社会学』岩波書店, pp.177-196

宗像誠之（2015）「元ソニーマンが救うAIBOの命」, <http://business.nikkeibp.co.jp/article/opinion/20150427/280471/> Accessed 2016, March 12.

長井隆行・中村友昭（2012）「マルチモーダルカテゴリゼーション」『人工知能学会誌』27巻6号, 人工知能学会, pp.555-562.

内閣府政策統括官（科学技術・イノベーション担当）（2015）「SIP（戦略的イノベーション創造プログラム）自動走行システム 研究開発計画」, <http://www8.cao.go.jp/cstp/gaiyo/sip/keikaku/6_jidousoukou.pdf> Accessed 2016, March 12.

西垣通（2004）『基礎情報学』NTT出版, 235p.

西垣通（2010）「ネオ・サイバネティクスの源流」『思想』1035号, 岩波書店, pp.40-55.

西條玲奈（2013）「性愛の対象としてのロボットをめぐる社会状況と倫理的懸念」『社会と倫理』28号, 南山大学社会倫理研究所, pp.37-49.

岡本慎平（2012）「人工的道徳的行為者への近年のアプローチの検討」『HABITUS』16巻, 西日本応用倫理学研究会, pp.53-73

Pagliery, Jose（2016）Some Ashley Madison women were actually computer 'fembots', <http://money.cnn.com/2016/07/05/technology/ashley-madison-fembots/> Accessed 2016, July 7.

Reeves, Byron & Nass, Clifford（1998）*The Media Equation*, Cambridge University Press, 323p.（細馬宏通訳『人はなぜコンピューターを人間として扱うか』, 翔泳社, 2001, 399p.）

瀬川奈都子, 児玉小百合, 木ノ内敏久（2016）「デジタルとルール（上）運転・作曲…進む自動化――権利や責任, 誰のもの。」『日本経済新聞』2016年1月26日発行, 朝刊, p. 1.

関口海良, 堀浩一（2008）「人工物を倫理レベルから設計するための方法論に関する一考察」『2008年度人工知能学会全国大会（第22回）論文集』pp.1-4

柴田正良 (2010)「異世界の者たちの倫理」『哲学・人間学論叢』1 号, 金沢大学哲学・人間学研究会, pp.17-37.

Singer, Peter（2009＝2010）*Wired for War*, Penguin Press, 512p.（小林由香利訳『ロボット兵士の戦争』日本放送出版協会, 720p.）

Varela, Francisco（1979 = 2001）*Principles of Biological Autonomy*, North Holland, 701p.（染谷昌義・廣野喜幸抄訳「生物学的自律性の諸原理」『現代思想』29巻13号, 青土社, pp.62-117.）

Verbeek, Peter-Paul（2011=2015）*Moralizing technology*, The University of Chicago Press, 183p.（鈴木俊洋訳『技術の道徳化』法政大学出版局, 318p.）

Veruggio, Gianmarco（2006）The EURON Roboethics Roadmap, <http://www3.nd.edu/~rbarger/ethics-roadmap.pdf> Accessed 2016, March 12.

Wallach, Wendell & Allen, Colin（2009）*Moral Machines*, Oxford University Press, 288p.

Wallach, Wendell & Allen, Colin（2012）Hard Problems, <http://wendellwallach.com/wordpress/wp-content/uploads/2013/10/Hard-Problems-AISB-IACAP2012-Wallach-and-Allen.pdf> Accessed 2016, July 7.

山岸浩史（2013）「「人間対コンピュータ将棋」頂上決戦の真実【後編】一手も悪手を指さなかった三浦八段は，なぜ敗れたのか」, <http://gendai.ismedia.jp/articles/-/35787?page=4> Accessed 2016, March 12.