

# 管理するコンピュータ：それは合理的な

## 「権限の委譲」か、無責任な「自律の放棄」か？

A. カフリック

出典：Arthur Kuflik, “Computers in control: Rational transfer of authority or irresponsible abdication of autonomy?” in *Ethics and Information technology*, vol.1, No.3, 1999, pp.173-184.

キーワード：

責任(responsibility)、自立した道徳的行為者(autonomous moral agent)、背後からの監督責任(background oversight responsibility)、管理するコンピュータ(computers in control (コンピュータによる管理 computer-dominion))、道徳的作業の分業(division of moral labor)

### はじめに

従来人が行って来たような仕事（経済分析や医療診断等）を、現在ではコンピュータや機械が引き受けるようになりつつある。それに従って、我々ほどの程度まで自らの責任(responsibility)をコンピュータに委譲・放棄する事ができるのかという問が生じる。カフリックはこの論文で、責任の概念を6つに分け、それらのどの意味でコンピュータは責任を持ちうるかを論じている。更に彼は、人間が決して放棄する事の出来ない種類の責任があるということを主張する。

### 概略

コンピュータによる管理に反対する代表的な意見として、カフリックは次の2つを挙げる。

- (i) 人間の能力には限界があり、我々は誤りを免れない。故に、人間が複雑なシステムをつくって自らの欠陥を補うというのにも、限界がある。すなわち、いかに信頼性の高いコンピュータを人間が造ろうとも、必ず間違いが起こる。
- (ii) いかにコンピュータが「人間より正しい答えを出す」としても、仕事の中には代行されえない種類のものがある。

これらを考察してカフリックが導き出した主要な結論は以下の3つである。

- (1) 人がどれほどコンピュータの性能向上に努めても、人間の「誤りやすさ」という欠陥は克服しきれない（性能向上自体が「誤りやすい人間」の行為の産物であるから）。またコンピュータに頼り切ることによって、プログラムデザインの基礎となっている人間の経験や直観とのつながりを我々が失ってしまったら、人間の非信頼性を低減するどころか増幅することになる。
- (2) 自律した道徳的行為者(autonomous moral agents)が専門的知識を持つ者(人であれ機械であれ)に従うことは理に適っている。しかし、従うからといって「背後からの監視責任」(background-oversight responsibility)まで放棄することは適当でない。
- (3) 他者(人でも機械でも)に自分の人生を代行してもらうのには限度がある。人生を意味と目的を持って生きるには、自らの技能を用い、他者と共感的につながる事が不可欠だからである。

## 導入

技術が進歩するにしたがって「コンピュータが人よりも特定の分野で優れている場合、コンピュータに管理させるのは道徳的に正しいか？」という問題が現実的になってきた。現在、また将来的に行われるであろうコンピュータ管理は2種類ある。一つは工場など技術的装置を制御するようプログラムされたメタ機械 (meta-machines)であり、もう一つは、医療診察・経済分析等、特定の問題に対する決定を下すようプログラムされたエキスパート・システム (expert systems)である。そして後者の下す決定は、人間の行動に何らかの指示や影響を与えると考えられる。

後者のシステムが十分な専門知識を持つようになれば（専門家にその知識故に決定を下す役割を任せるのと同じように）人がエキスパートシステムに頼ることも起こりうる。カフリックはそのような時代の来る前に、コンピュータに決定を任せる事に関しての道徳的考察すべきであると考へた。そして彼はこの論文で、人間の自律と責任の概念について再考する。

### 1. 「意思決定」「責任」という語の明確化

まず「コンピュータが（単に命令を実行するだけでなく）意思決定する」という言葉の意味を明らかにしておかなければならない。現在のコンピュータは、プログラムの設計者が定めた意思決定の為の基礎的決定事項に従う。その意味で現在のコンピュータが下す意思決定は二次的、下位のものである。それゆえ結果に対する究極的道德責任があるのはプログラムを実行し意思決定したコンピュータではなく、基礎的決定を行いプログラムした人間である。

それにもかかわらず「新しい技術がコンピュータに意思決定において更なる責任を負わせる」  
“New technologies are making computers responsible for more and more decisions”  
という言い方に違和感を感じないのはなぜだろうか。カフリックは「責任」という語の多義性が原因であると考え、責任概念を6種類に分類することでそれを説明する。

**1 因果責任 (causal-responsibility)** 結果に近接した原因。例文としては「the hurricane-force winds are responsible for the falling the oak tree (台風の風が木を倒した)」など。

**2 機能的役割としての責任 (functional-role responsibility)** 機能を持つシステムの中である役割を果たしているという意味での責任。例文としては「the heart is responsible for pumping blood through the circulatory of a human being (心臓は人間の循環器系を通じて血液を送り込む働きをする)」

(1、2は道徳的な意味合いのない責任である。)

**3 道徳的説明責任 (moral accountability responsibility)** 道徳的行為者は、人に何らかの影響をもたらすような行為を行った場合に、その意思決定や振る舞いに対しての事実説明ではなく、道徳的にもっともな説明や言い訳、また謝罪や訂正を要求される。そのように、道徳的に応答するよう責められる対象である場合に負う責任。機能として道徳的応答をすることができるなら（機械の良心が痛むかどうかという事は問題ではなく）機械であっても人と同じく道徳的行為者である。

**4 榮譽ある責任感 (honorific sense of "responsible")** 3の意味で道徳的に応答できる行為者であるだけでなく、自分の行動の影響を適切な道徳的見地から判断した上で行動することができる行為者が負っている責任。日常会話で「責任感のある人」という時の責任はこの4の意味。この responsible (責任) の対義語は irresponsible (無責任) である。

**5 役割責任 (role-responsibility)** ある役割を与えられた人が、その役割において求められ、期待されている仕事を果たすよう求められる事。(2、3の結合型)

**6 監督責任 (oversight-responsibility)** 役割責任の特別な形。仕事を委譲した相手がきちんと役割を果たしているかどうかを監視し、場合に応じて適切な処置(役割の補助介入や代行)を行う事。(例えば、車のパワーステアリングが壊れたとしても、手動でハンドル操作を続ける事はできる。そのように故障の際、介入や停止をする選択肢を人が保持している場合に負っているような責任)

この論文の中心的問題を、件の責任概念の分類を用いて述べるならば次のようなものになる。(括弧内の数字はそれぞれ分類の番号に対応する)

**「監督責任(6)」を放棄することなしに、どれほどの責任(2又は5)を「責任を負う(3)人間」は「責任を持って(4)」コンピュータに割り当てることができるのか。**

カフリックは、道徳的責任を持つ行為者(3又は4)である人間は完全に監督責任(6)を放棄する事はできない、と主張する。そして以降はその主張を擁護するための議論を展開する。

## 2. 問題：中心的なものと同付随的なもの

この論文で扱うのは道徳哲学的な問いであり、次のようなものである。我々はコンピュータによる管理に、どのくらい、どのように、頼るべきか?更にカフリックはこの問を次の二つに分ける。

- A コンピュータに意思決定の役割を負わせるかどうかを決める際に、道徳的問題として考慮に入れなければならないものは何か?(結果の最終的な説明責任はコンピュータではなく人間にあると前提する)
- B コンピュータに管理を委譲するとして、どのような条件のもとでそうすべきか?合理的で責任能力のある人間として、コンピュータの管理に対してどの程度の命令権を保持しておくべきか?

## 3. 考慮すべき要因1 能力と信頼性

コンピュータは処理速度やデータ保存、持久力においては人よりも格段に優れた性能を持つ。その点でコンピュータを使う事には大きなメリットがある。しかし一方で、安全性が非常に重視される場合には「複雑な機械に頼りきるのは危険だ」という心配がある。確かに、大きなソフトのプログラムはテストするのも、エラーを直すのも難しい。また仮にソフトが高い割合で正常に働くとされてもその安全性について絶対的に信頼がおけるといふ事にはならない。そのような問題に対して二つの対応の仕方があるとカフリックは言う。

- (1) コンピュータによる管理を安全性がそれほど重要でない分野に限る。
- (2) 安全性が重要になる分野でも、全てを人が管理するのは更に信頼性が低いことになりかねないので、基本的にはコンピュータに管理をゆだね、人がそのコンピュータ管理を無効にしたり補助したりすることを選択できる決定権を持つ。

コンピュータの能力と信頼性に対する考慮からカフリックが導いた結論は次のようなものである。人は自分の不完全性をコンピュータの性能を利用する事で、改善したり克服したりできる。しかしまた、自分の欠陥を(欠陥を持つ)自分自身がつくり出した複雑な機械で乗り越えようという試みには、限界がある。更にコンピュータに頼り切ることによって、プログラムデザインの基礎となっている人間の経験や直観とのつながりを我々が失ってしまう恐れもある。そうすると人間

の非信頼性を低減させるどころか増幅させることになるだろう。

#### 4. 考慮すべき要因2 能力と信頼性の他に

人の生死や健康に強く関わる仕事でない場合や、仕事自体が退屈でやりがいの無いものである場合は、多少信頼性が低くともコンピュータに管理を任せることは、人間的なやり方であるといえるだろう。しかし逆に、仮にコンピュータの方が人よりも上手く目的を達成するとしても、機械化を拒む種類の事柄がある。

例えば贈り物を送りたい相手の名前をうちこむだけで、完璧なプレゼントを選び、送り主の銀行口座から代金の引き落としができる店を選びだし、商品の購入をし、更に贈り物を送る手筈まで全て自動的に整えてくれる「Gift Perfect」なる機械があるとする。だが、その機械は贈り物を送ることが持つ意義の大部分を損なってしまうだろう。贈り物は相手にとって貰って嬉しいものである事も重要だが、同時に、送り主がどれほど贈るという行為に関わったかということも大切なのである。

カフリックはその事から次のような事を指摘する。人は他者（人でも機械でも）によって代行された生では意味を持って生きられない。意味と目的を持って生きるには、自らの技能を用い、他者とつながり共感しあうことが必要だからである。ノージックの思考実験（望み通りの体験をあらかじめ提供する「経験機械」によって、仮想の体験を通して満足した人生を送るのと、苦しいことが多くとも現実の生を生きるのとどちらがよいか）が示唆するように、我々は真の活動と経験から満足を得る事を非常に重視している。

#### 5. 管理するコンピュータ：どんな条件付きで？

カフリックは「道徳的に責任を持つ人間は、自身の生活と周囲とが持つ道徳の質に対して、高位から制御する権限を保持し続けるべきだと私は信じる」と言う。そしてこれを「自律的に道徳が機能すること（autonomous moral functioning）」と呼び、次のように説明している。

- 1 それは終わりのなき熟考をすることではない。それ以上考えたり議論する必要がないとわかれば、合理的な自己管理のもとで行動に移る。
- 2 複雑な社会においては、ある人が全ての道徳に関わる問題について熟知することは不可能である。よって道徳的自律の理想と「道徳的作業の分業（division of moral labor）」とは両立する。すなわち、道徳的に自律した人とは「時には、他人の方がある物事の道徳的側面を良く見ていて、それらに対して適度に共感し利己心を離れた判断を下せる」ということを認める用意がある人である。
- 3 道徳的に自律した人とは、その人が何を行うのであれ、道徳的にもっともな見地からその行動を正当化する準備がある人である。

そのような道徳的自律を保つ為の方法として、カフリックは次の事を提案している。

- A 常に（しかし耽溺しない程度に）人生があるべき方向に進んでいるかどうかを確認し、自分が間違っている事を示す徴候に敏感であること。
- B 時には一歩引いてより高い見地から自分の人生をよく考えてみる。つまり人生を一生という大きな単位で見直し、自分の生きる際の原理と人生とを照らしあわせてみる。

このようにすれば、人は自律した人間として、自らの人生の道徳的質の高さを、過度な反省・熟考に陥ることなく評価し、保ち続けることができるはずであるとカフリックは言う。

次にカフリックはこの「自律した人間」という観点を「コンピュータによる意思決定」の文脈にあてはめる。「統計的に見て人間の下す決定よりコンピュータの下す決定の方がより確かだから、人間による介入という選択肢をなくして、人の持っている責任をコンピュータに譲るべきだ」という言葉は反感を招くだろう。そのばあい、この言葉は「人は道徳的責任をコンピュータの為に放棄するべきだ」という事を意味する。しかしそれはカフリックによると事態を正しく見ていない。統計上どれほど人より信頼性の高いコンピュータであっても、無謬ではない。信頼性に欠ける我々人間が一体どうやって、その（道徳の基盤に関わる決定をする）コンピュータが「絶対正しい」と信じられる十分な理由を持ちうるというのか。このように彼はコンピュータが間違いを犯すということを主張する。

次に彼は人間のコンピュータへの介入の仕方を分類する。

i/a 人間がコンピュータに介入する無制限の特権をもつ。

i/b 特別な偶発事故の時に介入する選択肢を、十分な技術と道徳的責任を持つ人間が持つ。

さらに ii/a(i/a、i/b 両者を含む)と ii/b を区別すべきと主張する。

ii/a 状況に応じて人間の判断で介入できる。(situational-override) (人間の判断は広い裁量権があるものでも、特別な事態における裁量権を認めるものでもよい。)

ii/b 背後からの監督責任 (background-oversight responsibility) を持つ。定期的にシステムが正常に機能しているかを点検し、必要とあらばコンピュータに与えた役割を撤回することのできる監督権を持つ。

そして彼は、仮に ii/a の「人間がコンピュータへ状況に応じた介入をする」ことを放棄するのが賢明な選択であるという場合であっても、ii/b 「背後からの監督責任」を保持しなければならないと主張する。

カフリックによると、このようなコンピュータとの関わり方の問題はわれわれにとって新しい問題ではない。カフリックは例として、医師-患者関係における、我々は専門家である医者にどのように頼るべきか、という問いを挙げ、さらに次のように言う。医師がどんなに医学的に信頼できても、患者が自分の決定権・責任を放棄するのは正しいことではないであろう。「専門家に頼る」というのは「道徳的責任を放棄すること」や「自律的道徳判断を譲渡すること」とは違うのである。

この哲学における最も古い誤りの一つである「ひとたび最高の知識を知ったなら、それに無条件に従わなければならない」という考えが既にプラトンの『国家論』に見られる。絶対的知識は危険なものであり、専門知識でさえも適当な監視とバランス調節機構のもとに置かれるべきであるというのは、道徳的政治的知恵の初歩である、とカフリックは主張する。

## 6. 一見すると対立している主張

J.Moor は"Are There Decisions Computers Should Never Make"という論文のなかで、「コンピュータが人間より優れた意思決定を行えるのであれば、人間はコンピュータに頼るべきである」と主張している。これはカフリックの主張と一見対立しているように見える。しかしカフリックはムーアと自分の意見は完全に対立しているのではないと考える。

ムーアはコンピュータによる意思決定の範囲が広くなりすぎると、人間性を損なう

(dehumanize) と警告している。カフリックによれば、このムーアの警告は、贈り物の例でカフリックが示そうとしたのと同じ主旨である。またムーアは「コンピュータは人間の基礎的な目的や価値を定める事はできないし、またすべきではなく、コンピュータの意思決定が問題になる時には、初めの責任だけでなく、「継続的な責任」を負う」とも言っている。これはカフリックの監督責任を放棄すべきでないという主張と主旨は同じであるとカフリックは言う。

## まとめ

以下はカフリックによるの自身の主張のまとめである。

人々の安全を脅かすようなエラーをコンピュータが犯したことが明らかの場合に、適切な介入を（適切な）人が出来るようなシステムを設計するのは理にかなっている。この場合、コンピュータによる意志決定の仕組みが人のそれより優れているかどうかは問題ではない。人の安全が問題になる時、コンピュータの処理速度や持久力は、人間の道徳的判断力・パターン認識力・柔軟性など結びつけて利用されるべきである。「偶発事故が起こったが人の介入が適当でないとき」でさえ、人は「背後からの監督責任」を放棄する事はできない。その責任を保持出来ないという事は、すなわち、自律した道徳的行為者としての責任を果たす事を断念することだからである。

## エピローグ：未来像ーコンピュータによる究極的管理

SF ではコンピュータという被造物が創造者である人間のコントロールを越えて動き出す未来を夢想している。その場合、コンピュータの未来は次の二種類が考えられるだろう。

- 1 彼らが単なる超知的に洗練されたロボットである場合。
- 2 彼らが考え、望み、恐れ、愛し、気遣い、人間よりも技術的に優れているだけでなく、道徳的に優れた行為を為す存在である場合。

1 のロボットはスーパー機械であり、人間の意思を反映するものである。その場合、人は従来通りの生を生きるだろう。

2 の場合は、我々が何を為すべきか考えねばならない。全ての責任放棄をするのか、または今までの「人間の意思を実行する機械」という関係を逆転させ、道徳的に正しいコンピュータの意思に従う為に誠実な努力をするようになるのか。

この問題は人の道徳性 (morality) とは何かをよく反映している。道徳的に適切な人としての我々の役割は、道徳的に完璧なモノに仕えることではないのだ。道徳的に適切な人の役割とは、そのように単純なものではなく、自身の人生を正直に、正しく導くために最善のことをなし、誰もがしているように良心的に生きるよう努力する事なのである。

## コメント

この論文の意義をあげるとするなら、「コンピュータによる管理反対派」の論点をまとめた事と、「responsibility の分類」をして機械も道徳的行為者 3 として扱われうると示した事であろう。

(自律の概念の再考察をと言っているが、5 の部分はそうなのだろうか。私には、カフリックは自分が理想とする自律の項目を挙げているだけのように思えるのだが。)

カフリックの議論の根幹の一つに、人間は欠陥を持つ存在である、ということがある。欠陥を

持つ我々は決して間違えないコンピュータをつくることはもちろん、作ってもそうだと正しく知ることができないはずだ、というのである。それに基づいて「コンピュータは誤る」という判断がカフリックの中にある。だからこそ彼は「患者-医師関係」を「人-コンピュータ関係」のアナロジーとしてとらえた。しかし、我々が医師に全ての決定を委ねない理由は、医師も同じ人間であり完全な知識と判断力を持っているわけではないと知っているからである。とするとコンピュータが仮に人間以上の（あるいは神並の）知恵を持つとした場合、このアナロジーは成り立たなくなる。その点について、カフリックのこの議論では、究極的知識を承認する人々を説得することはできないだろうという気がする。

また、カフリックの主張は少しおおまかすぎると感じる。彼の主張をエキスパートシステムによる管理の文脈にあてはめれば、「背後からの監督責任」さえ守ればよいということになるが、そんな大雑把な指示では現実はどうするのがいいかわからないだろう。どこまでエキスパートシステムに責任を譲るのが適当かを判断する「基準」を示すことをこの論文はしていないのだ。初めに、コンピュータ（エキスパートシステム）に決定を委ねることの道徳的是非を問うとしながら、このような大雑把な結論になったのは何故だろうか。その理由は彼の議論の中で「コンピュータ」の意味するところが一貫していないことにあると思われる。

カフリックが導入部で「コンピュータ」という場合、医療診断システムなどのエキスパートシステムを念頭においた議論である。しかし後半になり議論が進むにつれ、人間の道徳や生を管理し指示するコンピュータ（1984年や火の鳥未来編などに出てくる、都市を支配するマザーコンピュータが彷彿とされる）を指していると思われる部分が散見される。本当は彼はこの論文で、エキスパートシステムの是非を仔細に論じた現実的な議論よりも、エピローグに触れられた究極的コンピュータ支配について書きたかったのだろう、と想像された。

(北口景子)