Challenger

意識あるロボットをつくり、意識の発生を証明する

意識を人工的に再現する。野心的な開発目標を掲げるのがベンチャー企業のアラヤである。神経科学分野出身のCEO（最高経営責任者）が一線級の研究者を率いる姿は、人工知能（AI）研究で世界のトップを走る米Google社傘下の英DeepMind社さながらだ。アラヤを支える意識の理論と実用化への道筋を、CEOの金井良太氏に聞いた。　**（聞き手＝今井 拓司）**

アラヤ 代表取締役 CEO

金井 良太氏

かない・りょうた

2000年に京都大学卒業。米California Institute of Technologyなどを経て2015年まで英University of Sussex准教授を務める。2015年にアラヤを創業。

我々は、意識を人工的に作り出すことを目標に研究開発を進めています。「怪しいことをやっているな」と見られることも多いですが、地道に取り組んでいるという印象を持ってもらえる方が最近増えてきました。

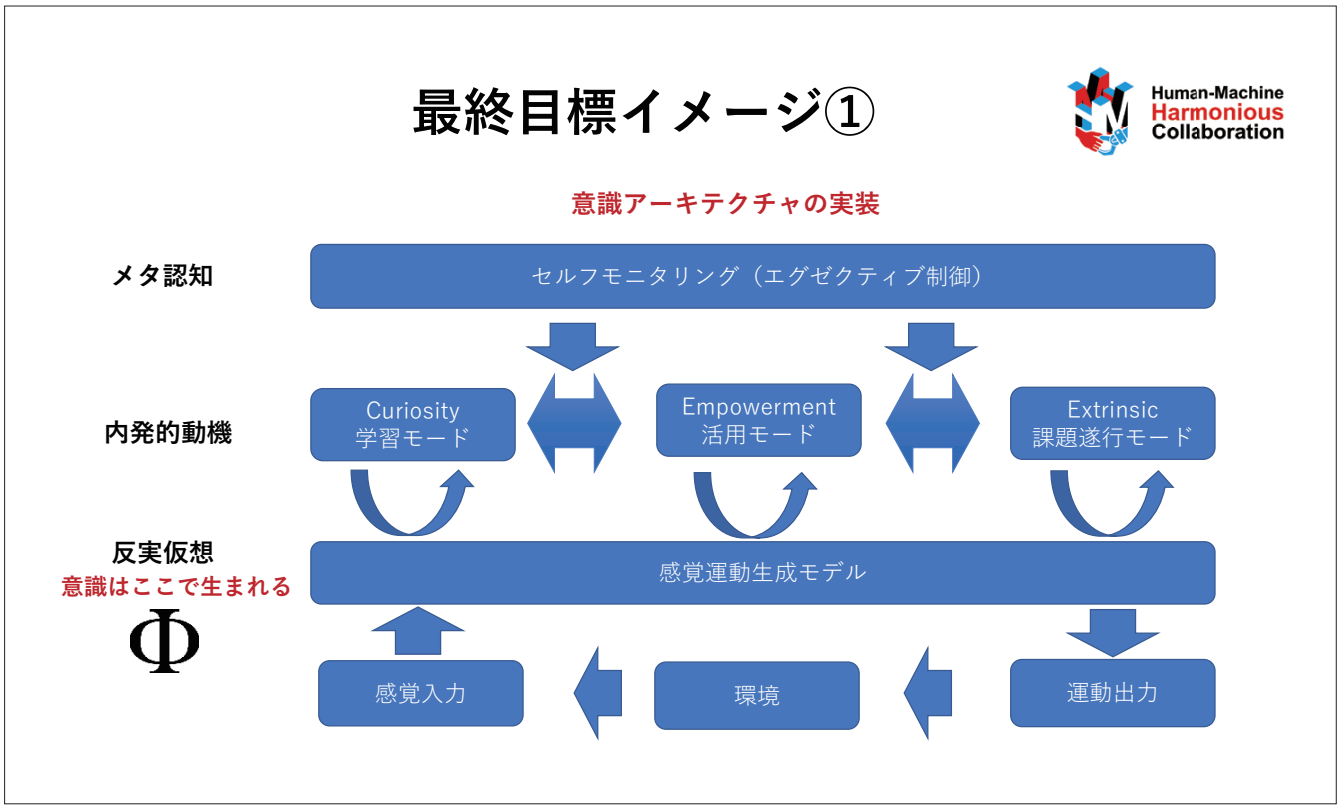
意識に関して、大きく2つの研究開発をしています。1つは、意識の機能に着目して、それを再現しようというものです。いわゆる強化学習の発展形といえる技術を開発しており、ロボットの制御の最適化に使おうとしています。もう1つは、生物や機械に意識があることを証明する研究です。米University of Wisconsin, Madison校教授のGiulio Tononi氏らが提唱する「統合情報理論（IIT：Integrated Information Theory）」を利用して、人工知能（AI）に意識があることを証明するのが目標です。

前者の研究では、意識の機能を説明する仮説として、「反実仮想的情報生成理論（Counterfactual Information Generation Theory）」を提唱しています。意識の機能は、「現在の感覚とは切り離された仮想的な情報、例えばこれから起こりそうな状況などを内的に生成して、未来の行動計画などに利用すること」と考えるものです。

現在、この考えに基づいたシステムを開発中です。強化学習の効率を高められると考えています。直接の狙いは、ロボットなどに自らの制御方法を自律的に学習させることです。「反実仮想」の発想を利用すれば、ロボットは行動のポリシー（方策）を、状況に応じてその場その場でつくり出せる。実際、倒立振子を立てるといった簡単な制御のシミュレーションで、狙った効果を出せることを確かめています。

ロボットの学習に関する研究開発は、5段階で進めます。最初は、仮想空間でロボットをトレーニングしたり、コントロールしたりする環境を整えること。既に技術はできています。次に、仮想世界に通常の強化学習を入れて、ロボットの学習を速くします。2018年の終わりの実用化が目標です。3段階目に、いわゆるゼロショットラーニング†のように、見たことがないものが来ても、ちゃんと対応できるようにします。「反実仮想」が出てくるのは、4段階目と5段階目で、詳細は未公

（写真：加藤 康）



実装を目指す「意識アーキテクチャ」（図：科学技術振興機構（JST）のCRESTプログラムの資料）

表ですが、ロボット自体がかなり自律的になるはずです。

自律的なロボットには、目標が決まっていないときに何をすべきかを決める機構を組み込みます。「内発的動機付け」と呼んでいます。その1つが好奇心（Curiosity）で、学習の高速化が目標です。報酬が何かが決まっていないときに、情報を得ることを報酬にするわけです。他の動機付けの手段も組み込んだ上で、その1つ上のレイヤーに、これらの動機を切り替えるセルフモニタリングの仕組みを入れます。これらをまとめたものが、我々が考える意識のアーキテクチャです（**図**）。このアーキテクチャの実装を、科学技術振興機構（JST）のCRESTプログラムの一環として進めています。ATR脳情報通信総合研究所や大阪大学と一緒にやっています。

我々が取り組むもう1つの研究分野が、作り上げた人工知能に意識があることの証明です。ここで利用するのが先述のIITです。IITは、まず誰もが納得できる公理系（Axiom）から出発します。「意識は存在する」とか、「意識は統合されている」とかいった命題です。これらを仮定として認め、数式に翻訳します。一度数式に翻訳したら、演繹的に、そこからいろいろな予測を導き出せます。その予測が観測と合っているかどうかを調べれば、理論の正しさを確認できる。理論の修正と観察を繰り返し、正しそうな理論にたどり着いたら人工知能に当てはめてみる。そうすれば、人工知能に意識があるのかないのか、分かるのではないかと。既に我々は、恐らく世界で初めてサルの実験データを使って、IITが示唆する意識のコアを計算によって見出しています。

強化学習などの研究では、困ったことにライバルが強すぎまして。例えばDeepMindの方が2～3カ月先を行っている感じです。できればDeepMind独り勝ちではなくて、「アジアだとアラヤがいるぞ」みたいな感じにできればと。それで、いっぱい人が集まるといいですね。個人的には、研究者が集まってくる場所をつくりたいと思っています。

†ゼロショットラーニング＝例えば画像分類問題で、学習データにない分野の画像を入力した際にも正しく分類できるようにするなど、学習時に与えられなかった種類の入力データを適切に処理可能にする学習手法。