

人工知能は精神疾患に罹りうるか

信原 幸弘 (Yukihiro Nobuhara)

東京大学大学院総合文化研究科

昨今の人工知能の発展は目覚ましい。人間並みの知能や、それどころか人間を超える知能すら、実現されそうな勢いである。しかし、人工知能とは何か。それは人間の知能と根本的に異なるのか。それとも、人間の知能と同種のものか。この問題にたいして、人工知能も精神疾患に罹りうるかといういささか特殊な観点から考察を試みる。

人工知能が人間並みの知能をもつようになれば、当然、人工知能も精神疾患に罹りうるとまずは思われよう。人間の知能が精神疾患に罹りうるのだから、人間並みの人工知能もそうであるはずだ。敵兵を殺害したロボットが激しい罪悪感に苛まれ、戦闘意欲をなくしてうつ状態になることは十分考えられる。そのようなことが起こりえないとすれば、それはまだそのロボットが人間並みの知能に達していないというだけのことであろう。

人間並みの知能をもつ人工知能は、そうであるがゆえに精神疾患に罹りうるようにみえる。しかし、本当にそうであろうか。話はそう単純ではないように思われる。本発表では、心に関する物理主義的な見方を前提にして話を進めるが、物理主義のもとでは、人間の心の変調である精神疾患は、人間の心を実現する人間の脳の変調にほかならない。そして人間の脳がどのような変調を来すかは、それを構成する物質的素材に決定的に左右される。したがって、あるものが精神疾患に罹りうるかどうかはそのものの脳が人間の脳のような素材で出来ているかどうか決定的に依存することになる。しかし、人工知能の脳は人間とはまったく異なる素材で出来ているから、それが起こす変調も人間の脳が起こす変調とは根本的に異なるだろう。そうだとすれば、人工知能は精神疾患には罹りえないことになる。正常なときには人間のような知能を示す人工知能であっても、その素材が人間と異なるために、変調を来したときにはその変調のあり方が根本的に異なるのである。そのような変調もある種の精神疾患だと言えるかもしれないが、少なくとも人間の罹る精神疾患とは異なるであろう。

また、これと関連するが、精神疾患はたんに精神に変調を来しただけではなく、その変調に何とか対処するために精神を再編成した結果でもある。そのような変調を来した精神がどのように再編成されるかは、やはり脳がどんな素材で出来ているかということに大きく規定される。そうだとすれば、やはりこの点でも、人間と素材が異なる人工知能が人間のような精神疾患に罹る可能性は低いように思われる。

しかし、ここで次のような反論がなされるかもしれない。たとえ素材が異なっても、人工知能が人間のような知能を示すようにプログラムされうるとすれば、人工知能は人間の精神疾患の症状を示すようにプログラムされることも可能ではないだろうか。たしかにそれは可能であろう。しかし、そのようにプログラムされて生じるような状態は、はたして精神疾患と言えるであろうか。それは精神疾患のシミュレーションではあっても、本物の精神疾患ではないように思われる。

ただし、人間でも、遺伝子の異常によって精神疾患に罹ることがあり、そのような場

合には、精神疾患が遺伝子的にプログラムされていたと言えそうである。しかし、そうだとすると、人工知能のプログラムが精神疾患の症状を示すように設定されたのにたいし、人間の遺伝子プログラムはそうではなく、何らかの異常によってそうなったのであり、その結果、精神疾患の症状を示すことになったにすぎない。この違いのゆえに、人間は本物の精神疾患に罹りうるのにたいし、人工知能はシミュレートされた精神疾患に罹りうるにすぎないと言えるように思われる。

このように見てくると、人工知能が人間並みの知能をもったとしても、人間のような精神疾患に罹ることはむしろありえなさそうである。しかし、そうだとすると、人工知能がそれ独自の精神疾患に罹ることはありうるだろうし、人間のような精神疾患の症状を示すこともありえよう。そうだとすれば、ここで次のような問いが浮かんでくるかもしれない。すなわち、いかなる精神疾患にも罹らず、いかなる精神疾患の症状も示さないような高度な人工知能を設計することは可能だろうか。人間は精神疾患に罹らないような強靱な精神をもつことに憧れるが、そのような強靱な精神をもつ人工知能を作ることにはできるだろうか。

この問題に対する鍵となるのは、精神疾患に罹るような人間の心の脆弱性はじつは人間のような複雑で柔軟な知能を実現するために不可欠かもしれないということである。人間の知能は膨大な数の諸要素の統合的な働きによって成り立っているが、そのような諸要素の統合は何らかの原因・理由によって崩壊する危険性を本質的に秘めている。そのような統合の崩壊が精神疾患だとすれば、人間の知能は精神疾患に罹りうるような脆弱性を本質的に孕んでいる。人工知能も、たとえ人間と素材が異なり、プログラムされたものであるとはいえ、そのような統合の崩壊の危険性を本質的に孕んでいるとすれば、人間とは異なる精神疾患かもしれないとはいえ、精神疾患に罹る可能性を本質的に秘めていると言える。

なお、現在、開発されている人工知能はまだ人間の知能の域に達していないが、そのような未熟な人工知能はそうであるがゆえに、すでに精神疾患に罹っていると言われることがある。しかし、それはおそらく誤りであろう。精神疾患は人間のような知能をもちうるものが何らかの理由・原因で精神に変調を来すことによって生じるものであり、そもそも人間のような知能をもたない未熟なものに生じる状態ではないだろう。