

随想

AI殺人兵器とアシモフのロボット三原則

人間こそ悪の根源と機械が認知しかねない大国の自儘

（株）PPQC研究所 加藤 宏光

アシモフのロボット三原則というものがある。

① ロボットは人に危害を加えてはならない。また、人に危害が加えられているのを見過ごしてはならない

② ロボットは人の与えた命令に従わなくてはならない。ただし、その命令が①に違反するときはこの限りではない

③ ロボットは①、②の原則に違反しない限りにおいて、自分を守らなければならない

（二〇五八年ロボット工学ハンドブック、第五六版、われはロボット、から）

著者がこれを知ったのは、今は亡き小松左京氏のSF短編小説が最初であった。この原則は、

ロボットが進化したあげく、人間がロボットに抑圧されるリスクを防ぐための必須の条件として覚えている。

思い起こせば松本零士氏の《銀河鉄道999》のテーマも、機械の知能が人のそれを超え、機械が人を不要のものとして全滅させようとする世界での哲郎の機械世界へのレジスタンスであり、世界の人間は人類のみならず全宇宙の知的生物すべてが味方、というスケールの話であった。人間が作り上げた機械文明が人間の敵になるとするのは、SF（サイエンス・フィクション）の興味深いテーマとしていろいろな面から取り上げられている。

先に挙げたロボットの三原則

二年に、武力行使のハードルが下がり、テロリストに拡散する危険もあると警告。国際的関心が高まった。

規制の議論は一四年から特定通常兵器使用禁止制限条約（CWC）の枠組みで始まり、今回が四度目の政府専門家による協議。（中略）法的規制には、米

国やロシア、イスラエル等AI技術先進国が開発の制約に繋がることを懸念して反対（後略）。

同じ記事の囲みコラムでは、有識者として国際会議に出席した福井康人・広島市立大広島平和研究所准教授の話として《政府担当者が積極的参加を》のタイトルで、今回はほとんど議論が進まなかったが、十一月のCWC締約国会議で（方向性を示す）「政治宣言」は出せるかもしれない。ただ宣言を出すことにロシアや米国、豪州が反対に回った。日本の立場も難しくなる。オーストリアやチリは規制を主張している。会議には武器製造関係者も出席し「安全で皆のために技術だ」と宣伝し

をSF作家として名の知れた故小松左京氏が取り上げていた（著者が《ロボットの三原則》というモノを知ったのは、四〇年以上も以前に書かれた小松左京氏のこの小説による）。

あらずじをまとめると以下のようになる。

ヒト型ロボット（今はアンドロイドという呼称が一般的かもしれない）が人間生活に欠かせない伴侶となった未来に、そのロボットを製造する工場に雷が落ちた。

それからしばらくして《人型ロボットが突然人を襲って殺す》という事件が頻発する。さまざまな怪奇なストーリーをたどったあげく、先の落雷で、AIに組み込まれる《ロボットの

三原則が逆に組み込まれた》ことよって起きていることが想定された、というものである。ちなみに三原則を逆の定義に設定してみよう。

① ロボットは人に危害を加えてはならない。また、人に危害が加えられているのを見過ごさなければいけない

② ロボットは人の与えた命令に従ってはいけない。ただし、その命令が①に違反するときはこの限りではない

③ ロボットは①、②の原則に違反しない限りにおいて、自分に危害を加えなければいけないということになる。

この事実を突き止めた主人公は定義に含まれるパラドックスにも気づき、そのパラドックス

というのである。

ヒトを認知すると、危害を加えないように自動停止する、と聞いて、先のロボット三原則を思い出した。

小松左京氏の小説のように、AIの定義が逆に埋め込まれていけば《ヒトを認知すると、危害を加えるために暴走する》となる。これはいみじくも《殺人ロボット》そのものではないか!? ましてそれが自動運転されているとすれば、《無人殺人ロボット》であろう。

そうなつてはならないからこそ、人知を結集させてヒトの知恵で暴走を防ごうとする会議そのものが、大国の自儘で前へ進まない、という事実と直面すると、松本零士氏になる《銀河鉄道999》にあるように、《機械人間》が、生きている人間の存在自体を悪の根源と認知し、《ヒト型無人殺人ロボット》が人間を見つけては殺戮する、という恐ろしいSFが近い将来現実のものとなるような錯覚に襲われる。

スを利用してくだんのロボットを駆逐するというのが小説のラストである。

四月九日の東京新聞二六面に《AI殺人兵器規制で溝》とタイトル付けされた記事がイラスト付きで掲載されていた。内容はかつての未来世界の話が現実になっているような深刻さを伴っているように感じられた。

以下に大略を紹介する。

人工知能（AI）を持ち人間の関与なしに自律的に目標を攻撃する「殺人ロボット兵器」の規制について議論する国際会議が先月末、スイス・ジュネーブで開かれた。日本政府はとりま

とめに向け、規制対象となる殺人ロボは「人間の関与が必須だ」とする作業文書を初めて提出したが、規制推進派と慎重派の隔たりは縮まらなかつた。政府は来月の非公式会合にも出席し、議論を継続する。（大杉はるか）

殺人ロボは「自立型致死兵器システム（LAWSS）」と呼ばれる。無人攻撃機等の技術開発を背景に、国際NGOが二〇一