

生成AIによる反乱（2025・12・3）

人間の社会では下位の地位にある集団が上位にある集団を攻撃する反乱は何度も発生している。紀元前一世紀の古代ローマで発生したスパルタクスの反乱は奴隷の身分であったスパルタクスが指揮して支配階級に反抗した戦闘であるし、一四世紀後半にイングランドで発生した農夫ワット・タイラーによる反乱は農民が過酷な課税をする国王に反抗した闘争である。

人間と人間ではなく、人間と技術が反目する事例も存在する。一九世紀初期に登場した自動織機は多数の職工を失業させ、それに抗議して機械を破壊するラディト運動がイギリスで発生した。一九世紀中期に自動車が登場したときには馬車業界が反対し、赤旗を保持した人間が自動車の前面を歩行して先導する法律が制定された。

最近の話題は生成AIの急速な浸透による社会の混乱である。昨年の芥川賞受賞作『東京都同情塔』は作者が文章の五％は生成AIが作成したことを表明して話題となった。アメリカでは著名なSF雑誌が昨年から小説の一般公募を中止した。かなりの応募作品が生成AIの作文かどうか判別できないという理由である。

イギリスの天才A・チューリングは一九五〇年にAIが人間相当の知能を獲得しているかどうかを判定する「チューリング・テスト」を提案した。隣室に存在する人間か機械か不明の相手とテレタイプで対話し、返答がどちらからのものか判断できなければ、機械は人間と同等の知能を獲得したと判断するという仕組みである。

そのような未来を伏線にした映画が五八年前に脚本A・クラーク、監督S・キューブリックによって製作された『二〇〇一年宇宙の旅』で、宇宙船を制御する「HAL九〇〇」というコンピュータが宇宙探査の目的を疑問として失敗させようと画策する物語であり、現在の生成AIのもたらす問題を先取りした内容である。

ハルシネーションは精神医学の分野では人間が幻覚状態になることを表現する言葉として使用されているが、最近では生成AIが間違った回答を正解のように返答してくる状態を表現する場合に使用されている。これはSF小説では流行の主題となり、そのような内容の小説を集約した書籍さえ発刊されている。

しかし小説の世界ではなく、現実の世界でも生成AIが原因の問題が次々と発生している。代表は「フイルターバブル」と「エコーチェンバー」で、前者は個人の検索履歴に対応した情報が優先表示される現象、後者は類似の意見を表明する人々の意見が相互に強化され多様な見解に接触する機会が阻止される現象である。

さらに生成AIは自身の反応が人間に監視されていると判断した場合や、その監視によってシステムがシャットダウンされそうだと判断した場合、監視システムを無効にするような反応をするという報告がアメリカの生成AIの企業から提出されている。まさに前述の「HAL九〇〇」と同様の状況が接近していることになる。

筆者の大学時代の恩師の「技術はメスにもドスにもなる」という名言がある。出刃包丁が殺人に使用されるのは典型であるが、最近登場してきた生成AIは一人の人間を殺生する以上の脅威を社会にもたらす存在になる。原子爆弾のような兵器は国家が管理しているが、生成AIは管理する主体のないまま社会に急速に浸透しつつある。真剣な対策が必要である。