



## フレーム問題とリスク管理

ロボット1号は、時限爆弾の仕掛けられている部屋から、バッテリーを取出すように指令を受けた。部屋に入ったロボット1号は見事バッテリーをもちだしたが、バッテリーの上に爆弾が乗っていることに気がつかず、部屋の外で爆発に巻き込まれた。

続いてロボット2号は「バッテリーを動かすと周囲にどう影響するか考える機能」が付加された。2号はバッテリーの前まで行った。「バッテリーを動かしても、部屋の色は変わらない、天井は落ちない、ドアは閉まらない、……」考えているうちに時限爆弾が爆発した。

ロボット3号は「爆弾とバッテリーに関係することと、無関係なことを区別する機能」が付け加えられた。しかし3号は、部屋の手前で全く動かなかった。関係ないことを無視するプログラムをしていたが、関係ないことが膨大であったため、動き出す前に爆発した。

これは人工知能におけるフレーム<sup>1)</sup>問題と呼ばれる大変有名な問題である。人工知能は、チェスやオセロのような閉じた世界の中では有効に働くが、起こりうるパターンが膨大な現実世界には対応できない。そして、ロボット3号の例「無関係なことを無視する」ことも非常に難しい。「無視する」行為自体が、行為を行うという点で、意味的に無視することにならないからである。

フレーム問題を部分的に解くことが可能な存在が人間である。では、爆弾をどううまく処理するか？

- ①目的の設定：バッテリーを取出す。
- ②フレームの設定：バッテリーと爆弾関係で起こりうるパターンとその因果関係の列挙（有限時間内で優先順位をつけながら）
- ③モニタリング事態の設定
- ④例外処理：「優先的に対応すべき事態」が発生した場合、フレームの再設定かつ目的変更



ところが、人間でも特にプロセス④の実行は難しい。「優先的に対応すべき事態」は、設定したフレーム外にあり、人間の生存を最優先するなど「上位ルール（常識）」から判断されるからである。そして、フレーム外の事態をモニタリングする仕組みも非常に難しい。

以上の話は、リスク管理のシステム化プロセスを意識して書いたつもりである。リスク管理の目的、フレーム、モニタリング対象、意識的無視、フレーム外の上位ルールなど、リスク管理を人工知能のスキームで考えると新たな示唆が生まれよう。ところで、リスク管理が、かわいそうなロボット任せとなってしまうませんか？（外園 康智）

1) ここでフレームは、事象の認識や思考の枠組を意味している。