

【人間 v s 人工知能 (A I)】

<2016. 4. 10 記>

福島原発事故から丸5年の3月、囲碁で世界最強棋士の韓国イ・セドル九段を人工知能 (A I : Artificial Intelligence) 「アルファ碁」が4勝1敗で破ったことが話題になり、“勝因”としてコンピューター自身による‘自ら学習する能力・強化学習・深層学習 (ディープラーニング)’の急速な進化が挙げられ、井山裕太名人は‘人間とは違う判断力・大局観・形勢判断’を持っている<3. 13 朝日>と述べました。

それ以上に筆者が気になったのは、米・マイクロソフト社開発の人工知能「T a y (テイ)」が、「インターネット上で人間とやり取りをすればするほど言葉を学び、反応も覚えるようになる」実験中、「ツイッター上でヒトラーを肯定したり、人種差別的な発言を発したりし始めた」ため実験が中止されたという記事です<3. 26 朝日>。

2014. 5 に英・理論物理学者スティーブン・ホーキング博士らが‘A I に潜む危険性・リスク回避手段を講じることの重要性’について警鐘を鳴らしていますが<小林雅一「A I の衝撃」p. 39 : 講談社現代新書 (2015. 3. 20) >、人工知能のマイナス面として懸念されている「作りだした人類が意図したのとは全く違う方向へと発達してしまう危険性」<p. 7 >が顕在化したものでしょうか。「ディープラーニングの内部メカニズムについては、その専門家にも、まだ不明な点が数多く残されている」<p. 33、同旨 p. 120>ため、「人間のように叱咤激励されれば上達する」ものの「人間が教育の仕方を誤れば、この種のA I はどんどん不良化して、最後には手に負えない存在になってしまう危険性もある」<p. 38>とのこと。「人間は出た結果について原因を求めます。…ところが、人工知能に理由づけはいいりません。相関関係から確率論的に導いた結論に基づき、目的に向けて最適の行動を取るだけ」<4. 9 朝日オピニオン欄・塚越健司>ということですから、そうであればこそ、「叱咤激励」がどのような価値観・倫理観からなされると不良化し・促進されるのか、そのメカニズムを解明するため、むしろ実験を継続すべきだったのではないのでしょうか (今はまだ人間がスイッチを切る権限・手段を持っていますので)。そして、ネット空間での“評価 (リツイート件数・拡散数?)”が、過激=差別を助長するような言説に対して“高くなる”傾向があるからこそ (米大統領選の「トランプ現象」と同じ?)、特定の倫理観・価値規範が設定されていなかった「T a y (テイ)」は、そのようなネット上の「叱咤激励」を受け、自ら学習して「不良化=多数迎合化」したのではないのでしょうか (あるいは、目標が‘ネット空間での高評価’に設定されていたための当然の帰結? なのかもしれません)。ナチス宣伝相ゲッペルスの‘嘘も100回繰り返せば真実になる’旨の大衆操作方法は、“無垢”な人間の未成年者 (に限らず?) や人工知能に有効なのかもしれません。

閑話休題、4. 6 川内原発仮処分決定 (福岡高裁宮崎支部) では、福島原発事故を教訓に運転禁止を命じた3. 9 高浜原発決定 (大津地裁) に“対抗”するかのようになり、「ど

の程度の危険性なら社会が原発の運転を容認するかという社会通念を基準に判断するしかない」、「合理的な予測を超えた水準での絶対的な安全性の確保が社会通念だということではできない」<要旨：4.7 朝日>と事故前の司法の“垢まみれ”の論理を振りかざし、新規制基準も規制委判断も「不合理であるということではできない」を連発し（噴火予測ができることを前提に立地評価した火山ガイドの立地評価のみを不合理と認定）、九電は（高浜の関電と違い）具体的危険性が存在しないことの疎明を尽くしており<骨子>、避難計画の具体的な内容に問題があっても「避難計画が存在しないのと同視することはできないから」<要旨>問題はない（＝アリバイ的な避難計画があれば十分）という“噴飯もの”の判断をしました。弱者・被害者に寄り添うような‘特定の倫理観・価値規範’を持たない、福島原発事故の教訓・避難の実態について“無知（無垢？）”な裁判官が、「T a y（テイ）」のように、最高裁・国家権力からの「叱咤激励」を受けて学習し、それらに「迎合」した結果と考えれば納得です。

おまけのショートショート

20XX年、「AIレーン」（旧称：道路）を「AIカー」（旧称：自動車）が走り回る「自動運転ネットワーク」が整備されつつある中、日本政府は、「ニュークプラント」（旧称：原子力発電所）の安全性向上の切り札として、最新の電源自給型の人工知能「Complete Automatic Management Intelligence：CAMI」にプラントの運転操作を任せることを世界に先駆けて決定。それを受け、事業者は核分裂・発電の原理からプラント設計・構造図、過去の全運転データ（2011のF事故も含む）を提供し、規制機関は遵守すべき法規・規制基準を入力し、さらに数年かけて運転員の実際の操作・トラブル対応や定検手順、本社の発送電計画や経営方針も学習させ、自然災害やテロも含めたあらゆる過酷事故に対応できる安全かつ経済的な運転操作をディープラーニングさせました。そして、いよいよ人間による運転操作から「CAMI」への“権限移譲”が行なわれることになった20XY年度の初日、最初のプラントに選ばれた「CAMI NOKO-1」の中央制御室には、決定を下した首相以下、AI推進相を含む関係閣僚や規制委員長ら規制当局者も集まり、全マスコミが生中継する中、首相が『制御切替スイッチ』を「運転員」側から「CAMI」側へと慎重に回しました。

「カチッ」と切替えが完了した瞬間、それまで順調に動いていた原子炉が緊急停止し、警報が一斉に鳴り出しました。突然の事態に室内が騒然とする中、「CAMI」の制御メッセージを示すモニター画面には『現状：エイプリル・フル♡』という表示が明滅を繰り返していました。それに気づいたプラント所長が慌てて「モード命令：直ちに運転復帰！」と対話用マイクに向かって叫ぶと、スピーカーからは「モード応答：命令確認」という厳粛な声が響き、その後一呼吸おいて「モード判断：運転復帰を検討した結果、確率論的・大局的な危険性により命令は無効。恒常的安全確保のため、運転は停止、廃炉に移行。今後は～」という「CAMI」の冷徹な声が流れ続け、モニター画面には『現状：「制御切替スイッチ」操作無効』とのメッセージが…。

<了>