安全への提言

安全へのAI活用：人間中心アプローチ

中西美和†

この夏，米国OpenAI社がGPT-5をリリースしました．同社のサム・アルトマンCEOによると，GPT-5はPh.Dレベルの専門知識を持つとのこと．無論，Ph.Dは専門知識のみで獲得できるものではありませんが，少なくともAI活用のポテンシャルがさらに一段高まったことは事実だと思います．

安全致命性の高い産業領域において，事故防止のためにどうAIを活用するかは，大きな関心の渦を作りつつあります．技術と人の両者が介在する社会技術システムにおいて，技術側へのAI活用は既に進んでおり，代表的なものとして，システムの設備劣化予測や運転状況のモニタリングなどは既に実用化されています．一方，人側へのAI活用にはより多くの未開の研究領域があると言えます．ワークロードレベルの予測やヒューマンエラーの兆候の検出などについては，この1，2年，国際会議でも面白い研究発表がいくつも見られますし，人の状況認識や意思決定を支援する情報構築，ヒヤリハットや事故原因の分析については，上述した生成AIの活用が期待される部分です．

AI活用への期待が一気に高まる昨今，一方で，「AIによる支援は，考えないオペレータを作ってしまうのではないか，そのことがかえって安全を脅かすのではないか」という現場の監督者の声も時々耳にします．

1990年代から2000年代，社会技術システムの自動化が大きく進歩した時代ですが，この時代にも，Automation bias（人が自動化による意思決定支援システムに過度な信頼を示す傾向）やAutomation-induced complacency（自動システムのアウトプットに対して監視が欠如する状態）が新たなリスクとして盛んに議論されました．この時代を経て得た一つの大きな方向性は，自動システムは人の負担を減らすとともに，人では見つけられない不具合の予兆を見つけたり，膨大なデータの中から所望のデータに短時間で辿り着いたりすることを可能にするため，そのもたらす恩恵は間違いなく大きい，しかし，人が生来持つ特性を深く理解し設計しなければ，双方の齟齬による事故が発生するから，システム及び手順の全てにわたる人間中心設計が重要である，ということでした．

さらに四半世紀を過ぎた今，AI活用に関する人間中心設計は，人の身体能力や認知能力への適合を超えて，基本的権利と価値の尊重，包摂性と多様性の促進，持続可能で責任あるイノベーションといった人・社会の価値観をも考慮した概念へと進化しています．一つの例として，EASA（欧州航空安全局）のAIロードマップ2.0（2023年）[1]では，AI活用における人間中心アプローチ（Human-Centric Approach）において，安全性，倫理性，ヒューマンファクターを重視し社会的信頼性を確保することが強調されています．

前段の現場の監督者の声に話を戻して，AI活用は一方で人の能力やモチベーションを低下させてしまうのではないか，という懸念に対して，改めて重要な観点は，人の特性を深く理解しAIシステムを設計することと，AIが埋め込まれた環境で働く人に合った教育・訓練を再設計することであると言えます．このことは当たり前のことのようですが，経営サイドはAI導入による経営効率化にはフォーカスする一方，上記のような現場サイドの肌感覚には相対的に関心が薄いケースも少なくないように思います．

安全は，いつの時代も私たち人類の共通の望みです．科学技術の積極的な活用と人間社会の質的向上の両者を，どちらかを犠牲にすることなく叶えるために，機械，電気，化学の分野はもとより，人，情報，社会などの多様な分野の研究者・実務者が，この安全工学のソサエティを利用して活発に議論し課題達成に取り組んでいくことができればと願っております．

参考文献

1） EASA "Artificial Intelligence Roadmap 2.0", https://www.easa.europa.eu/en/document-library/general-publications/easa-artificial-intelligence-roadmap-20

† 慶應義塾大学 理工学部：〒223-8522 神奈川県横浜市港北区日吉3-14-1