

原子力分野の AI 品質に係る基本的な考え方の開発 (3) AI のリスク回避性に関する考え方と今後の課題

Development of basic concept of quality management of the use of AI/ML in nuclear power plants

(3) Views on Risk Avoidance of AI/ML in basic concept related to AI quality and future challenges

*河合 理城¹, 小野寺 将規¹, 江藤 淳二¹, 板倉 豊和¹, 柴垣 和広¹, 八木 伸太郎¹,
河野 尚幸², 樫又 恒一², 荒金 賢二³, 出町 和之⁴

¹三菱総合研究所, ²日立 GE ニュークリア・エナジー, ³原子力エンジニアリング, ⁴東京大学

革新炉や既存炉への機械学習技術の導入促進を目的に、AI 技術のうち特に機械学習技術に着目した、原子力分野の信頼性評価の基本的な考え方の草案を開発している。特に、プラント保安分野 AI 信頼性評価ガイドラインをベースとし、原子力特有の変更点としてリスク回避性の指標の簡易評価手法を検討した。

キーワード : AI, 信頼性評価, 機械学習, リスク

1. 緒言

石油化学プラントの保安分野を対象として AI (特に機械学習) の信頼性評価に関して整備された、「プラント保安分野 AI 信頼性評価ガイドライン」^[1]をベースとし、原子力分野における AI の信頼性評価を検討した。ベースとしたガイドラインでは、機械学習の品質を「利用時品質」、「外部品質」、「内部品質」の 3 階層に分け、それらの達成を通じて機械学習利用システムの信頼性を管理する。そのうち、「外部品質」が機械学習要素の満たすべきことを指し、「リスク回避性」と「パフォーマンス」に分けられる。このリスク回避性のレベルについて、原子力分野でどのように評価すべきかを検討した。

2. AISL (リスク回避性のレベル) の考え方

機械学習要素の誤判断による悪影響に対する回避の要求の強さに応じてレベルを設定し、これを AISL と呼ぶ。原子力分野においては、取り扱うリスクの内容、リスクレベルが一般産業等と必ずしも一致しないことを踏まえ、右図のように AISL の評価方法の変更を検討した。フロー中に記載の AISL 表は、機械学習要素の誤判断に起因する人的または経済的な影響の強度と、誤判断を人間が回避する可能性の程度で、レベルを簡易評価するものである。今後、具体の機械学習要素を有する AI 技術で評価を実施し、課題抽出や改善等を実施する。

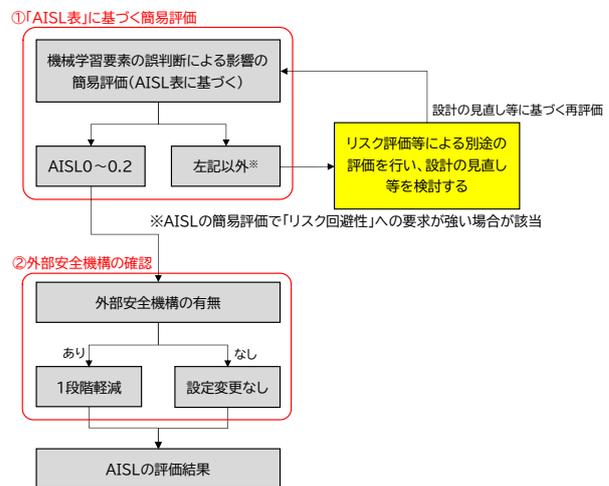


図. AISL の設定フロー

謝辞

本調査研究は、経済産業省資源エネルギー庁の「令和 5 年度原子力産業基盤強化事業」として行われたものである。本稿は著者らの意見を表明したものであり、資源エネルギー庁の見解を反映するものではない。

参考文献

[1] 石油化学コンビナート等災害防止 3 省連絡会議, "プラント保安分野 AI 信頼性評価ガイドライン", 2020.

*Masaki Kawai¹, Masaki Onodera¹, Junji Etoh¹, Toyokazu Itakura¹, Kazuhiro Shibagaki¹, Shintaro Yagi¹, Naoyuki Kono², Koichi Kashimata², Kenji Aragane³, Kazuyuki Demachi⁴

¹MRI, ²HGNE, ³NEL, ⁴UTokyo