

AI システムの利用者視点からのトラスト構築の考察 Consideration on building trust form the viewpoint of AI system users

島 成佳 * Shigeyoshi Shima 小川 隆一 † Ryuichi Ogawa 佐川 陽一 † Youichi Sagawa

キーワード AI システム, 誤判断リスク, 利用時品質, トラスト構築

あらまし

機械学習を用いた AI 技術は実用化が進み, AI 学習モデル (以下, AI) を搭載した製品やサービスが提供され, 応用も広がりを見せている. AI を搭載し, AI の判断に基づき制御等を行う IT システム (以下, AI システム) の開発では, 従来の演繹的なソフトウェア開発に加え, 学習データに基づき, 探索的・反復的に学習モデルを構築する必要がある. 学習モデルは, 学習を網羅的に実施することが困難であり, 精度は 100%にならない[1]. このため, AI システムには, AI の誤判断が存在する.

従来 AI の誤判断リスクの削減は, 学習モデルの精度向上の観点で検討されてきた. 背景には, 精度向上が AI の重要な品質要件とされたことが考えられる. これは, AI を直接利用する場合の要件としては合理的である.

一方, AI システムにおける誤判断リスクの影響は, AI システムを構成するソフトウェアやハードウェアの機能で低減することも可能である. すなわち, AI の誤判断リスクは, 学習モデルを直接使う利用者には重視すべき品質要件であるが, AI システムの利用者にとってはそうとは限らない. むしろ AI システムの利用者には, システム全体の安全性や有用性等が喫緊の品質要件になるとされている.

AI システムでは, 安全性や有用性等の品質要件を満たすことによって, 利用者と提供者の間にトラストが構築される. このため, AI の品質要件として重視されている精度が, AI システムのトラストを構築するためにどの程度重要であるのかは精査が必要であり, ここに筆者らは注目した.

筆者らは, まず AI システムの構成とその利用を含むトラスト階層モデルを定義 (図 1) し, 各層の品質要件を検討した. また, AI システムに組み込まれる学習モデルの精度 (誤判断の程度) は AI システムのトラスト (品質要件) に基づいて決定されるべきであると考え, 各層ごとの品質要件と品質の伝搬に関してモデル上で検討した. このモデルに基づけば, AI システムのトラスト構築に AI の誤判断リスクの影響が必ずしも大きくないことが示される. 本論文では上記の検討についての詳細を述べる.

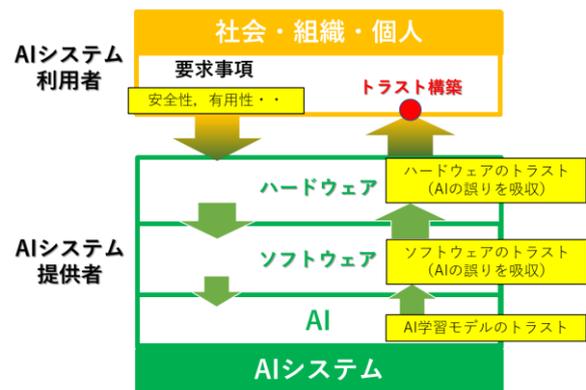


図 1 : トラスト階層モデル

参考文献

[1] AI プロダクト品質保証コンソーシアム (QA4AI), “AI プロダクト品質保証ガイドライン 2020.08 版,” <http://www.qa4ai.jp/QA4AI.Guideline.202008.pdf>

* 長崎県立大学, 〒851-2195 長崎県西彼杵郡長与町まなび野 1-1-1, University of NAGASKI, Manabino 1-1-1, Nagayochi, Nishisonogi-Gun, Nagasaki, Japan 851-2195

† 独立行政法人情報処理推進機構, 〒113-6591 東京都文京区本駒込二丁目 28 番 8 号, Information-technology Promotion Agency, Japan (IPA), 2-28-8 Honkomagome, Bunkyo-ku, Tokyo, Japan 113-6591