Hu(man) vs. Machine

Wen Li CHAN[1]* and Hsin-Vonn SEOW[1]

[1]Nottingham University Business School, University of Nottingham Malaysia

**Correspondence: Wen Li Chan, Nottingham University Business School, University of Nottingham Malaysia, Jalan Broga, 43500 Semenyih, Selangor, Malaysia.*
Email: wenli.chan@nottingham.edu.my

Abstract

With the continued development of decision-support systems based on artificial intelligence (AI) / machine learning (ML) techniques that display superior performance and prediction capabilities over traditional logistic regression models, an inherent inability of AI-generated output to be 'explainable' due to ML models being associative rather than causal, has become a crucial issue in credit scoring particularly in the area of equal credit opportunity.

From the consumer's point of view, the inability to explain the reason for a credit decision (e.g. reason for a reject decision) can translate to issues of public trust in AI, which could hinder adoption.

From the lender's point of view: We have previously suggested that the non-causality of credit scoring models could pave the way for the inclusion of protected variables (that have traditionally been deemed 'discriminatory' in the legal sense) to improve the predictive accuracy of models, while recognizing the challenge of explainability to consumers of such practice (Chan and Seow, 2013). Subsequent studies have since provided empirical evidence that the effect of a protected variable is not avoided by its exclusion in a model due because its influence remains present through association with 'surrogate' variables; and further that the prohibition of protected variables in current practice could have the unintended consequence of perpetuating inequality in respect of disadvantaged groups.

It appears therefore that there may increasingly be a case for considering the inclusion of protected variables in credit scoring models, although this would face the challenge of explainability head-on – in order for seemingly 'politically incorrect' requests for protected variables to be deemed acceptable.

We recognize that various technical mechanisms have and are being developed for explainable AI (xAI) in credit scoring, and that there may be trade-offs between explainability and model performance. We also note the UK government's recent White Paper proposing a context-specific regulatory framework for AI (as opposed to creating a new cross-industry regulatory body). Building on this and to bridge the AI trust deficit in tandem with the development of xAI models, we consider the option of having an oversight body – either within- or cross-industry, to bridge the gap.

Keywords: Credit scoring models; Anti-discrimination laws; Explainable AI.