

AI Risk Management Framework RFI Responses

Accrete AI – Prashant Bhuyan, CEO and Founder

NIST RFI 1: The greatest challenges in improving how AI actors manage AI related risks- where “manage” means identify, assess, prioritize, respond to, communicate those risks.

Accrete Response: One of the greatest challenges in improving how AI actors manage AI related risks is to translate AI performance in terms of accuracy, reliability, and explain-ability into actionable business value. The reason that this is of critical importance is that it forces the user of AI to quantify the business value of current employees that don’t use AI. In particular, users must first establish benchmarks for performance that elucidate the true impact and risk of error in real world contexts. Using such benchmarks set in accordance with the performance of human domain experts, the AI’s performance can be held accountable.

For example, we’ve created an AI that automates the grading of baseball cards and trading cards for a major collectibles grading company. This company has been experiencing incredible growth due to the demand for collectibles in a post covid world. The grading company is experiencing an unprecedented backlog of cards that are waiting to be graded. The reason that there is a growing backlog of cards is that the company cannot find, hire and train enough graders to keep up with the demand. As such, the longer it takes to turn around a graded card to the customer, the greater the risk of losing market share to competitors.

Our AI is able to grade baseball cards with junior grader accuracy. However, the accuracy itself is meaningless without a benchmark. We worked with the card company to establish a benchmark against which we could evaluate our AI’s performance. The grading company established this performance benchmark by quantifying the true monetary and reputational cost of misclassification errors incurred by human graders. Once we understood those costs, we were able to superimpose business rules over AI outputs that dictated when it was appropriate for the human using our AI to take the risk and allow the AI to autonomously grade a card and when it was too risky to allow the AI to grade a card without manual review.

Finally, AI explain-ability and continual learning is the key to AI risk management because in situations where it is too risky to allow the AI to operate autonomously, it is critical for the human reviewer to understand the AI’s reasoning. If the AI’s reasoning is flawed, it’s important for the human to correct AI bias by providing conceptual feedback and for the AI to incorporate that feedback, learn and improve. In essence, explain-ability and continual learning enables the machine to correct human bias and, conversely, for the human to correct machine bias. As time goes on, a continually learning AI’s performance will improve with decreasing human feedback and the AI will be able to perform increasingly risky tasks with increased autonomy.