Rational Intelligence Manifesto

The Rational Intelligence (RI) Lab was established to address emerging mathematical and practical challenges that arise at the interface between intelligent systems and society.

Artificial intelligence (AI) breakthroughs owe their triumph in part to machine learning whereby computers learn from past data to tackle intricate tasks. Rosenblatt's Perceptron algorithm (1957) was initial proof of a versatile learning rule, enabling computers not just to discern patterns from data but also to extrapolate to new scenarios. The domain has since seen a surge of more complex models and learning algorithms, from kernel machines to deep neural networks. Empowered by Vapnik's pioneering statistical guarantees (1995), this learning approach has revolutionised how we address scientific and engineering challenges.

In the wake of scaling computational power, the efficacy of the learning-based approach has surged, leading to the widespread adoption of modern AI systems across diverse platforms. Non-technical users now harness these capabilities, prompting the integration of AI into an extensive array of applications. However, this evolving landscape has given rise to new challenges that drive researchers to scrutinise the very foundation of machine learning. Notably, the efficacy of established techniques can stem from superficial data patterns vulnerable to real-world fluctuations. Moreover, the flourishing diversity of real-world datasets introduces prospects for collective learning and digital democratisation, yet imposes limitations on inter-environment data sharing due to critical concerns like privacy, security, and equity. The frailty of prevailing learning methods in grappling with distributional shifts and adversarial attacks underscores their inability to deal with uncertainties beyond the purview of classical probability theory (Kolmogorov 1933). Lastly, the need for novel solution concepts, for example, in cooperative and non-cooperative game theory contexts (Nash 1950; Shapley 1951), becomes pronounced in the face of feedback loops and strategic manipulation. These trials collectively underscore the formidable gap that remains in our quest to engender machines with robust generalisation capabilities.

The RI Lab's research embraces and navigates these emerging complexities and, in order to address them, we envision an expanded machine learning landscape, spanning not only algorithmic design and architectural considerations, but also data collection and model deployment. Simultaneously, we recognize the imperative to redefine machine learning and generalisation in our modern context. On the one hand, knowledge of the data collection process illuminates a guiding beacon, empowering autonomous agents to see their own blind spots as well as to ascend the ladder of causation (Pearl 2018), not only foreseeing intervention outcomes but also engaging in counterfactual reasoning. On the other hand, reliable real-world model deployment mandates anticipation of novel uncertainties and intrinsic heterogeneity, particularly those rooted in human preferences. The shortcomings of classical probability theory in capturing broader uncertainties, highlighted by scholars like Walley (1991), Shafer and Vovk (2019), among others, form the driving force behind our research. Last but not least, modern learning systems encompass more than minimising risk function; they entail eliciting, aggregating, open-sourcing, and trading algorithmic models. This reshapes the conventional communication model between AI engineers and its users. Amid these complexities, we recognize the need to establish rationality (von Neumann and Morgenstern 1944; Arrow 1951) as a theoretical foundation for modern learning algorithms.

Our solutions thus demand an interdisciplinary fusion of computer science, statistics, machine learning, and economics. Through this synergy, our primary objective is to maximise the value derived from data, inference, and decisions. In essence, our research will establish a strong foundation for creating systems that exhibit not only intelligence, but also rationality, capable of engaging in interactions with complex environments. In turn, this effort will further foster the accountable integration of these systems into society.

Dr. Krikamol Muandet
15th August 2023, Songkhla, Thailand