Enabling AI Adoption through Assurance

Jaganmohan Chandrasekaran[1], Feras A. Batarseh[1], Laura Freeman[1], D. Richard Kuhn[2], M S Raunak[2], Raghu N. Kacker[2]

[1]Virginia Tech, [2]National Institute of Standards and Technology (NIST)
{jagan, batarseh, laura.freeman}@vt.edu, {d.kuhn, ms.raunak, raghu.kacker}@nist.gov

Abstract

The wide-scale adoption of AI will require that AI engineers and developers can provide assurances to the user base that an algorithm will perform as intended and without failures. AI assurance is the safety valve for reliable, dependable, explainable, and fair intelligent systems. It provides the necessary tools to enable AI adoption into applications, software, hardware, and complex systems. This interactive tutorial will provide an overview of AI assurance, introduce a new set of assurance goals for intelligent systems, discuss the open challenges in assurance, and present recommendations to overcome its drawbacks.

Full Description

AI assurance involves quantifying capabilities and associating risks across deployments including data quality to include inherent biases, algorithm performance, statistical errors, and algorithm trustworthiness and security. Data, algorithmic, and context/domain-specific factors may change over time and impact the ability of AI systems in delivering accurate outcomes (Freeman, Rahman, and Batarseh 2021). In this tutorial, we present the importance and different angles of AI assurance and introduce a general 6-point framework (Batarseh, Freeman, and Huang 2021) that addresses its challenges.

This tutorial covers the major aspects of AI assurance; it will be organized into three parts. First, we introduce the challenges in testing and evaluating AI systems. In part 2, we present assurance metrics for AI systems such as explainability, fairness, and trustworthiness, followed by a discussion and involvement by the audience. In the third part of the session, the attendees will be encouraged to participate in open polls and discussions set up to exchange ideas and challenges in AI assurance based on their own experiences. We aim to cover different angles of assurance; for instance, ones that arise due to the domain in which an AI system is deployed, the architecture of the AI system itself, and ones related to government and policy regulations. We conclude the tutorial with a discussion on the future of AI assurance and a lessons-learned poll from the participants.

Outline of Tutorial

The tutorial commences by illustrating the need for AI assurance. We discuss assurance challenges that arise due to the data-intensive nature of AI systems and present a roadmap on how to address such challenges. The following topics are covered:

- Validation and Verification (V&V): Traditional Software vs. AI-based Software
- Data vs. Algorithmic Assurance
- Testing beyond correctness: the need for new assurance methods

The remainder of the session will be *interactive*. Next, we introduce the six AI assurance goals; afterwards the audience will be encouraged to interact and discuss the goals using polls and exercises; followed by an open discussion. The 6 assurance goals are:

- Explainable AI and Trustworthy AI
- Safe AI and Secure AI
- Fair AI and Ethical AI

In the third part of the tutorial, we present assurance challenges that arise from three angles: *domain*, *model*, and *policy*; followed by a team interaction segment where each team picks one of the categories for elaboration. Finally, an open discussion regarding the future of AI is held.

References

Batarseh, F. A., Freeman, L., and Huang, C. H. (2021). A survey on artificial intelligence assurance. Journal of Big Data, 8(1), 1-30.

Freeman, L., Rahman, A., and Batarseh, F. A. (2021). Enabling Artificial Intelligence Adoption through Assurance. *Social Sciences*, *10*(9), 322.