Testing AI systems: Leveraging collaborative intelligence for assured outcomes

Abstract

Artificial Intelligence (AI) is one of the key technologies driving digital transformation. At the core of this technology are the learning algorithms that help determine the success of AI systems. Successful industry deployments mostly comprise Supervised AI models, where human experts and data scientists help identify critical considerations to maximize success rates.

As AI becomes a key differentiator in operating performance of enterprises, its deployment in business-critical processes will grow. However, AI algorithms continue to face reliability issues limiting deployment to processes where human intervention is possible. The reason: any failure in critical processes has the potential to lead to significant losses for business – both in terms of money and reputation.

This paper focuses on developing an AI assurance framework for Supervised AI models with the goal of standardizing algorithm testing - both pre and post deployment.

The paper is designed to help enterprises identify a generic testing processes critical for successful AI deployment –thereby empowering their Business 4.0 journey.

Why Testing is Critical for AI Success

AI is a powerful force in driving new developments in technology and business in the Business 4.0 era. However, some fundamental capabilities set humans apart from AI. Human intelligence has an inherent ability to gauge a situation based on past experiences coupled with consciousness – a combination critical to identifying a creative solution. On the other hand, AI as it exists today, is constrained in handling diverse and complex situations, for which it has not been trained. It also lacks the ability to take decisions - a power that is unique to human intelligence.

Moreover, AI also faces challenges in rationalizing whether a task is appropriate or ethical. This means a great deal of responsibility lies in the hands of testers when it comes to preventing AI from creating havoc. It is critical for testers to define the boundaries within which AI-based algorithms can operate and monitor them regularly to pre-empt any breaches. Though AI adoption is increasing rapidly in industries like telecom, transportation, medical sciences, manufacturing and others, AI projects still face significant deployment challenges. Dynamic scenarios and varied attack possibilities make testing critical for successful deployment of AI-based solutions.

Testing an AI system requires the ability to compare scenarios and identify small changes. The Agile model of failing fast often comes to the testers' rescue - as both data and decision engines need regular monitoring to prevent undesirable modifications and outcomes.

The three key aspects in testing of AI systems include performance, safety and security. The interlinking of these three aspects is crucial because failure in one can aspect can trigger a failure in another dimension. AI testers need to ensure that failure of one algorithm does not lead to the failure of the entire system. One way to do this is to borrow AI system testing principles from conventional industries such as manufacturing and heavy engineering that have safety embedded in their product development philosophies. Concepts like Safety Engineering and Design for Reliability become relevant in this context.

Roadblocks to Superior AI Performance

AI models deployed in enterprises follow a steep learning curve; initial results are almost always less optimal than anticipated. AI success currently focuses on areas in which human intervention can institute corrective actions such as conversational AI, loan default prediction, price discovery for retail chains, prediction of logistics bottlenecks in manufacturing, and so on. Here, even if AI systems come up with wrong predictions –human supervision can weed out mistakes and refine the algorithm further. Supervised learning algorithms typically use ensemble methods such as bagging, boosting and stacking to improve the probability of success.

Ensemble techniques use multiple algorithms – in parallel and/or in sequence, exploring different combinations and selecting the one that works best. Human supervision remains the key element in establishing whether an AI learning system is ready for deployment or not.

As a result, the time and human effort needed to establish deployment-ready models, remains the primary challenge in ensuring successful AI deployments. Unsupervised AI models do not generate the desired trust or performance, while supervised AI models require expensive data science talent to develop and maintain; in addition to domain experts to formulate decision points and test its predictions.

Another key point of concern with AI models is that they age faster than regular software products. This happens due to the inability of models to generalize or 'concept drift' wherein changing operating scenarios make data models obsolete. Remember how Microsoft's chat-bot, Tay, started displaying views antithetical to human morals within 24 hours of deployment[1]? Clearly, there must be a mechanism in place to continuously monitor AI performance post deployment.

Without significant monitoring, it is unlikely that future AI systems will be used in operations of high business criticality and hence, will not realize the expected gains.

Proposed AI Assurance Framework: Building on the Human-Machine Synergy

The quality of AI deployments will be a major competitive differentiator for organizations as they increasingly leverage AI for higher productivity and building an insights-driven enterprise. Organizations need an AI assurance process that can ensure consistent performance, predict failures, and adapt to dynamic operating scenarios.

Given the non-deterministic nature of AI systems, the proposed approach towards AI Assurance goes beyond the usual application testing practice and utilizes both human expertise and technology monitoring to drive optimal results[2].

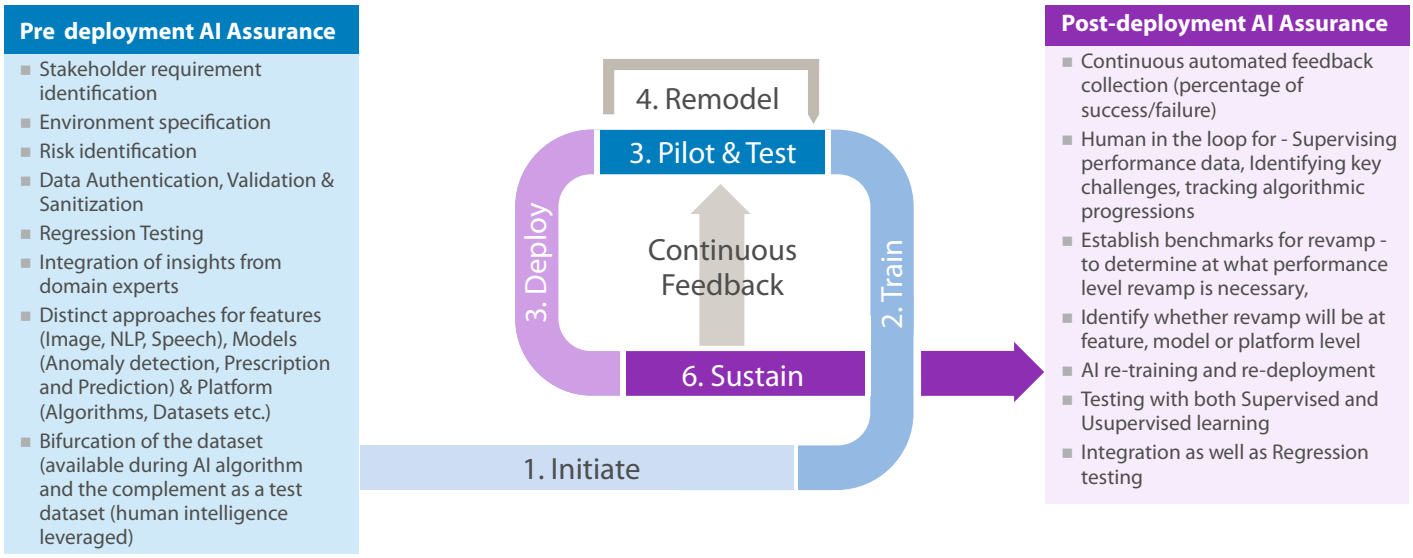

Fig 1. Development of AI systems – with critical testing interventions

Pre-deployment

- Choose a data set that closely resembles the production system with careful bifurcation (as shown in Fig.1). Such bifurcation ensures optimal training and testing before the AI model goes into production
- Identify tools for feedback data processing and result evaluation (establish data analytics tools to determine success/failure rates)
- Eliminate data biases, if any, by scrutinizing the test results using analytics; and identify any failure patterns
- Execute non-functional testing using tools to analyze results; and establish failure protocols.

- Prioritize data sanity and privacy and explore encryption tools

Post deployment

- Review output from continuous feedback monitoring platform – analyze user actions to understand success rates
- Establish failure threshold for algorithms as it is likely to vary depending on the area of application
- Use the AI monitoring platform to identify code progressions and root causes for performance deviation Classify required level of changes to establish whether a single algorithm, model or the platform itself, needs revamping
- Identify if any new data parameters or sources need to be tracked, based on changing user behavior and operating scenarios
- Iterate pre-deployment testing after any change decisions are made

The degree of human monitoring effort needed in AI assurance depends on the nature of operating areas. While our assurance framework is directed towards AI systems in dynamic environments, human intervention can be minimized in use cases with static decision parameters - like regulatory compliance, legal document review or monitoring of machine performance in factories. However, the threshold for AI failure is important to consider here as well, and any deviations need to be identified and corrected – to prevent significant lapses.

Future-Proofing AI Success Using Standardized Testing

Currently, standards do not exist for the development of AI systems and initiatives such as the IEEE Ethical framework or NIST Draft Plan for AI standards are still a work in progress, making it imperative to standardize the testing of AI.

It is imperative to establish trustworthy AI systems using a collaborative intelligence- driven AI Assurance framework that can be used in a business-critical environment – such as the one proposed in this paper. Collaborative intelligence helps synergize human intelligence and AI: leadership, teamwork, creativity, and social skills of humans combined with the speed, scalability, and quantitative capabilities of AI. A recent Harvard research, involving 1,500 companies, revealed that firms

achieve the most significant performance improvements when humans and machines integrate their capabilities. The future will see enterprises increasingly leverage reliable AI systems to drive competitive differentiation-generating exponential efficiency and performance improvements.

References

[1] *The Guardian, Tay, Microsoft's AI chatbot, gets a crash course in racism from Twitter, Mar 2016 (accessed Oct 2019), https://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter*

[2] *Harvard Business Review, Collaborative Intelligence: Humans and AI are Joining Forces, Aug 2018 ((accessed Oct 2019), https://hbr.org/2018/07/collaborative-intelligence-humans-and-ai-are-joining-forces*

About The Authors

Sayantan Datta

Sayantan Datta is a research analyst with TCS' Marketing Transformation and Operations team. He provides research based advisory services to global enterprises for digital transformation. A certified Financial Risk Management professional, Datta, holds a Bachelors' degree in Computer Science and Engineering from West Bengal University of Technology, Kolkata, and an MBA from IISWBM, Kolkata.

Ushasi Sengupta

Ushasi Sengupta is a research analyst in TCS' Marketing Transformation and Operations team. She provides research-based insights to global enterprises for digital transformation. Ushasi, a certified Supply Chain Analyst, is an Electronics and Communication Engineer from West Bengal University of Technology, Kolkata, and holds an MBA in General Management from XLRI, Jamshedpur.

Gokulaparthiban

Gokulaparthiban is an Innovation Evangelist with TCS' Corporate Research & Innovation unit, working in the Data and Decision Sciences Research area to contextualize innovation offerings. Gokul has over 12 years of experience in Engineering Operations and New Product Development. He holds a Bachelors' degree in Mechanical Engineering from Sri Krishna College of Engineering (Anna University), and a PGDM from IIM, Lucknow.

Dr. Rahul Agarwal

Dr. Rahul Agarwal is an Innovation Evangelist with TCS' Corporate Research and Innovation unit, evangelizing solutions in the Deep Learning and Artificial Intelligence Research area. Having more than 25 years of experience, he holds a Doctorate degree (thesis: Innovation and Agile Software Development) in Management, PGDM from IIM, Lucknow, and a Bachelors' degree in Mechanical Engineering from NIT, Warangal.

Experience certainty. IT Services
Business Solutions
Consulting

Contact

Visit the Research and Innovation page on www.tcs.com

Email: innovation.info@tcs.com

Blog: #Research and Innovation

Subscribe to TCS White Papers

TCS.com RSS: http://www.tcs.com/rss_feeds/Pages/feed.aspx?f=w
Feedburner: http://feeds2.feedburner.com/tcswhitepapers

About Tata Consultancy Services Ltd (TCS)

Tata Consultancy Services is an IT services, consulting and business solutions organization that delivers real results to global business, ensuring a level of certainty no other firm can match. TCS offers a consulting-led, integrated portfolio of IT and IT-enabled, infrastructure, engineering and assurance services. This is delivered through its unique Global Network Delivery Model™, recognized as the benchmark of excellence in software development. A part of the Tata Group, India's largest industrial conglomerate, TCS has a global footprint and is listed on the National Stock Exchange and Bombay Stock Exchange in India.

For more information, visit us at www.tcs.com

All content / information present here is the exclusive property of Tata Consultancy Services Limited (TCS). The content / information contained here is correct at the time of publishing. No material from here may be copied, modified, reproduced, republished, uploaded, transmitted, posted or distributed in any form without prior written permission from TCS. Unauthorized use of the content / information appearing here may violate copyright, trademark and other applicable laws, and could result in criminal or civil penalties. **Copyright © 2020 Tata Consultancy Services Limited**

TCS Design Services I M I 03 I 20