Developing Trustworthy AI Applications: A Methodology to Manage Risk from European Regulations

Robert Hobbs, PhD
Service Management
Crowley Maritime
Jacksonville, FL
ORCID: 0000-0001-9880-8195

***Abstract*—The European Union has taken a lead position in establishing standards of artificial intelligence with its Trustworthy AI regulations. Much like the predecessor General Data Protection Rights, the Trustworthy AI sets industry standards for a broad range of products operating in the European market. Trustworthy AI expands on the concepts of individual impact to consider factors such as the impact on society, overall accuracy, system dependability, and environmental impact. While the artificial intelligence components trigger the applicability of regulations, the regulations explicitly expand from the technical aspects of the AI model to encompass the entire scope of the application's implementations.**

The Methodology for Development of Trustworthy AI (MDTAI) proposes a model to address the myriad of needs across the full scope of development efforts required for an AI system implementation. The process reflects the EU requirements by starting with the corporate initiating goals and progressing through the data collection, AI Modeling, inferencing, quality management, and result applications. The Methodology includes the capture of intents and processes required from the development teams to demonstrate both actual and intended compliance. Managing activities to the development model can reduce the risk and potentially large liabilities from Trustworthy AI auditors.

Keywords—European Union, EU, Trustworthy AI, Explain AI, Compliance, Audit, Development, Models

I. Outline of Session

The walkthrough will address the current regulatory context and its implications for development requirements. Providing an understanding of how European Union regulations differ from other markets is essential to understanding the needs and risk for willful compliance with the Trustworthy AI (TAI) standards.

A. Trustworthy AI. The European Union has taken the General Data Protection Rights as a model to expand to address concerns over how individuals, organizations, company's, and countries employ artificial intelligence. GDPR focused primarily on the impact on the individual. Trustworthy AI goes beyond the immediate impact of electronic systems on the individual to consider how AI systems can impact individuals and societies. The EU regulations explicitly require companies to go beyond the individual to consider the impact of their system on democratic institutions and the environment.

B. Trustworthy Regulatory Risk. The EU has applied the general regulatory framework for GDPR to Trustworthy AI. The framework provides a centralized standard for companies to adhere to with decentralized regulators. Companies gain from having a single formal standard but face different regulatory bodies in each country. The risk is significant with fines of up to 30 million Euros per instance.

C. Processes and Standards. The Trustworthy AI regulations and standards set forth key processes, checkpoints, and documentation requirements for company compliance. Establishing the processes required to adhere to and capture the required information during the development process can reduce the risk of non-compliance as well as drop the overall development costs.

D. Model for Development of Trustworthy AI. The MDTAI proposes an overall process that begins with corporate goals for the AI project and applies standard development steps with the additional key processes required to comply with the Trustworthy AI Standards. The MDTAI steps through the phases and components of the multi-functional environment required for AI applications and adds the steps required to demonstrate the intent and actual compliance with Trustworthy regulations.

E. Initiation and Scoping. The TAI standard requires organizations to capture their intent and goals at the initiation of their AI projects. Documenting the intents upfront provides both guidance for subsequent development and an audit trail for future interrogatories.

F. Modeling. While AI modeling has frequently garnered publicity, the modeling team demonstrated the balance of multiple factors in the modeling effort. The need for explainability is but one driver that may cause the AI to use a model with lower objective accuracy but higher subjective value.

G. Developing Data collection. The legacy of data frequently gets overlooked. However, poorly managed data has caused expensive errors when implementations misinterpreted critical conditions.

H. Inference. Distributed architecture and IoT processing can lead to processing delays and failures that negate the trust in an otherwise valuable result.

I. Application. Once the inference engine provides an answer, the consuming application has to match the use of the answer to the business context.

J. Quality Processes. The application of quality monitoring and control processes is essential. TAI requires companies to establish upfront their quality standards and measures.

K. Production Validation. The most important factor of TAI is the end result of the application. Far beyond the question of does the model work in the lab is the answer to does the full application give a result that is accurate, dependable, and consistent with the values of the European Union. The MDAI outlines the processes and steps required to achieve and demonstrate compliance with the Trustworthy AI regulations.

II. Proposed Walkthrough

A. Proposed Length. The walkthrough for the development methodology is targeted for a two-hour session.

B. Resume

a) Robert Hobbs (M'03) from Mobile, AL, received a B.S. in Physics/E.E. United States Military Academy, an M.S. in Business Process Management from Colorado Technical University and a Ph.D. in Business Administration, Information Systems and Enterprise Resource Management from California Intercontinental University, CA. Robert is the Chair of the Jacksonville, FL Section and Computer Chapter. He is a voting member of the IEEE Standards Committee on Artificial Intelligence and hosts a recurring series of webinars on Machine Learning for the Jacksonville section.

He has an extensive history in developing enterprise applications to include Machine Learning solutions to identify maintenance issues from high-definition images and the development of autonomous control for robotic systems. He managed the development of multiple classes of applications for numerous regulated industries, including Transportation, Finance, Insurance, Healthcare, and DOD. He has performed multiple needs analyses for Enterprise applications, including Enterprise Resource Management and Maintenance Management systems. He has led and participated in hundreds of projects as a manager, project manager, Scrum Master, and Scrum Master for Scrum-of-Scrums. He has presented papers and training seminars at prior IEEE SouthCon and other local conferences on AI Ethics and effective Agile development practices.

C. Budget

a) The session is offered a no additional cost to SoutheastCon Attendees.

D. Pertinent Information

a) The author submitted a separate paper on European Union Trustworthy AI and accepted standards for Non-Functional Requirements: "European Union's Trustworthy AI: Functional and Non-Functional Requirements." The paper addresses the benefit of considering software quality metrics as part of the development concerns for Artificial Intelligence. This walkthrough addresses multiple factors of software development, placing them as well as non-functional requirements in a methodology that corporate reduces risks and liabilities.

Partial Bibliography

[1] N. Smuha, "Ethics guidelines for trustworthy AI," Shaping Europe's digital future - European Commission, Apr. 19, 2019. https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (accessed Oct. 15, 2020).

[2] L. Chazette and K. Schneider, "Explainability as a non-functional requirement: challenges and recommendations," Requir. Eng., vol. 25, Dec. 2020, doi: 10.1007/s00766-020-00333-1.

[3] "Europe fit for the Digital Age: Artificial Intelligence," European Commission - European Commission. https://ec.europa.eu/commission/presscorner/detail/en/ip_21_1682 (accessed Apr. 21, 2021).

[4] B. Shneiderman, "Bridging the Gap Between Ethics and Practice: Guidelines for Reliable, Safe, and Trustworthy Human-centered AI Systems," ACM Trans. Interact. Intell. Syst., vol. 10, no. 4, p. 26:1-26:31, Oct. 2020, doi: 10.1145/3419764.

[5] "Proposal for a Regulation on a European approach for Artificial Intelligence | Shaping Europe's digital future." European Commission, Apr. 21, 2021. Accessed: Apr. 21, 2021. [Online]. Available: https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-approach-artificial-intelligence

[6] C. Stix, "Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment | Shaping Europe's digital future." European Commission, Jul. 17, 2020. Accessed: Apr. 27, 2021. [Online]. Available: https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment

[7] J. Estdale and E. Georgiadou, "Applying the ISO/IEC 25010 quality models to software product," in European Conference on Software Process Improvement, 2018, pp. 492–503.

[8] "ISO 25010." https://iso25000.com/index.php/en/iso-25000-standards/iso-25010 (accessed Apr. 01, 2021).

[9] T. Hovorushchenko, "Methodology of evaluating the sufficiency of information for software quality assessment according to ISO 25010," J. Inf. Organ. Sci., vol. 42, no. 1, pp. 63–85, 2018.

[10] E. Hickman and M. Petrin, "Trustworthy AI and Corporate Governance: The EU's Ethics Guidelines for Trustworthy Artificial Intelligence from a Company Law Perspective," Eur. Bus. Organ. Law Rev., Oct. 2021, doi: 10.1007/s40804-021-00224-0.

[11] P. Santhanam, "Quality Management of Machine Learning Systems," ArXiv200609529 Cs, Jun. 2020, Accessed: May 06, 2021. [Online]. Available: http://arxiv.org/abs/2006.09529