2023 IEEE Conference on Artificial Intelligence (CAI) | 979-8-3503-3984-0/23/$31.00 ©2023 IEEE | DOI: 10.1109/CAI54212.2023.00155

AI Digital Tool Product Lifecycle Governance Framework through Ethics and Compliance by Design†

Eduardo Ortega⁎, Michelle Tran⁑, Grace Bandeen⁂
⁎Duke University, Pratt School of Engineering, Durham, NC, USA
⁑Duke University, The Graduate School, Durham, NC, USA
⁂Duke University, School of Law, Durham, NC, USA
{eduardo.ortega, michelle.tran, sydneygrace.bandeen}@duke.edu

Abstract— **The acceleration of Artificial Intelligence (AI) has brought forward new digital tools that have had a wide impact across society. However, AI digital tools (such as ChatGPT, midjourney, DALL-E 2) have brought forward legal and ethical concerns. – Internationally, public, and private leaders are introducing regulatory frameworks to address data governance for such ~~these~~ AI digital tools (i.e., Global Data Protection Regulation, the European AI Act, Blueprint for an AI Bill of Rights, NIST Risk Management Framework, etc.). We recognize that these AI digital tools are a vital aspect of future technological development, but they require input from various sectors in addressing ethics and compliance design. We survey the current landscape of published AI-specific regulatory frameworks and known engineering design process methods. Using a product lifecycle approach, we also introduce a trans-disciplinary framework to address AI ethics and compliance via design. This product lifecycle approach considers several principles: a Human-Centered Design for Risk Assessment, Functional Safety and Risk Management Standardization, and Continuous Governance throughout Product Lifecycle. Establishing risk management throughout AI product lifecycles can ensure accountability for AI product use cases. In addition, by utilizing previous Functional Safety considerations we can create safety mechanisms throughout the product lifecycle of AI digital tools. Finally, establishing in-field testing for continuous governance will enable the flexibility for new compliance standards and transparency. We establish this governance framework to aid in new compliance strategies for these emerging issues with AI digital tools.**

Keywords—Ethics, Compliance, AI, Risk Management, Human-Centered-Design, Engineering Design Thinking

I. Introduction and Motivations

Artificial Intelligence (AI) is a computer-science field that creates decision-making with intelligent behavior [1]. AI systems often make these decisions to influence the surrounding environment [2]. These decisions have real-world impact (i.e., autonomous vehicles), and the liability and compliance of these AI-made decisions are that of a public concern [4]. Given its new and unprecedented nature [4], new AI regulatory frameworks have been introduced to establish AI compliance standards [5][6][7]. However, there lies an inherent challenge to bridge the gap between law practitioners and computer scientists [3]. As new AI regulation emerges the greater the need for a trans-disciplinary solution that can support current legal statutes and the technological advancement of AI.

In this work, we are motivated to establish a trans-disciplinary governance framework to support compliance strategies and AI's advancement. By evaluating existing compliance trends, regulatory policies and initiatives, engineering standards, and known design processes (Human-Centered-Design), we assert that future AI ethics and compliance must engage the needs of multiple sectors across varying jurisdictions. Through cross-referencing each field of study, we deduct general principles and language that assesses ethics and compliance by design. As stressed by IEEE (Institute of Electrical and Electronic Engineers), ACM (Association for Computing Machinery), and OECD (Organization for Economic Co-operation and Development) [8][2][9], a shared motivation includes the need to create a culture of engagement with governance standards, risk assessment, accountability, transparency, monitoring, and iteration of AI. We design our governance framework from several principles: Human-Centered Design for Risk Assessment, Functional Safety and Risk Management Standardization, and Continuous Governance throughout Product Lifecycle. As these new regulatory frameworks are deployed, we employ an iterative governance and engineering design process to support a forward-thinking compliance strategy. The rest of this paper is as follows. Section II covers a preliminary literature review of recent data governance efforts, human-centered design, and functional safety. Section III covers the proposed trans-disciplinary governance framework. Then finally, we end the paper with our conclusion in Section IV.

II. Background

A. Current AI and Data Governance Developments

There are several AI-centric regulatory frameworks currently in circulation: the European Union (EU) Global Data Protection Regulation (GDPR) [22], the EU released draft versions of the EU AI Act [5], the National Institute of Standards

†We acknowledge the Ethical Technology Practicum course from the Duke Law School instructed by Lee Tiedrich J.D. This course informed our multi-disciplinary approach. In addition, we also acknowledge Krishnendu Chakrabarty Ph.D. for supporting this multi-disciplinary effort (Ph.D. advisor to Eduardo Ortega).

979-8-3503-3984-0/23/$31.00 ©2023 IEEE
DOI 10.1109/CAI54212.2023.00155

and Technology (NIST)Risk Management Framework (RMF) [7], and the White House Office of Science and Technology Blueprint for an AI Bill of Rights [6]. Each documents stresses key important considerations for AI regulatory frameworks.

The first and most notable of these frameworks is the EU GDPR [22]. The GDPR is a regulation in EU law that establishes new privacy and security requirements to protect individual control and rights over personal data. This includes the fair, accountable, and regulated processing of personal data (direct or indirect) as well as other components such as data processing consent regulation and default data protection. [22]. The breadth of privacy regulation bolsters overall data protection for all citizens of the EU. In addition, the GDPR addresses automated decision making (i.e., AI decision making, etc.) from said user data. We acknowledge that the GDPR is enforced per member state and thus can be left open to interpretation of enforcement. Thus, the implicit lack of harmonization between member states can make it difficult for compliance groups to support. However, GDPR is the first of its kind to ensure that those who use user data inappropriately are held accountable.

The EU AI Act introduces a risk-based approach to AI governance and is the first document to provide a comprehensive regulatory framework for AI. The EU AI Act classifies AI into several categories of risk: unacceptable (e.g., infringement on human rights), high risk (e.g., interaction with consumers with non-transparency), and low/minimal risk (e.g., transparent AI interaction with consumers or non-interactive) [5]. This document asserts high standards for data quality and AI system. However, as AI has progressed since the time of the last draft – most emerging AI models (i.e., generative AI, general-purpose AI) would be classified as high-risk. The EU has taken a more proactive approach in applying strong legal protections to product safety with a strong principle-based approach. This includes establishing strong data quality standards for emerging AI. The Act aims to create a proportionate regulatory system with a risk-based approach that allows for flexibility. While the EU has taken a strong principle-based approach aimed to optimize AI use case, there is limited information that asserts recommended technical standards and implementation strategies. It is unclear how the EU AI Act aims to protect fundamental and safety rights as such norms, laws, policies, and enforcement techniques continue to evolve.

The U.S. based NIST RMF improves upon the previous EU AI Act with its risk-based approaches to assessing AI governance [7]. The NIST RMF considers risk as the measure of a harmful event's probability of occurring and the impact of that event on either people, organizations, or the ecosystem [7]. In addition, risk management should be supported throughout AI system lifecycles (i.e., development, deployment, or in-field as standalone/integrated components) [7]. They also broaden the definition of risk compared to the EU AI Act. NIST acknowledges that the overall use cause and the measurability of risk should be considered when performing risk assessment [7]. They recommend creating specific user profiles for the use case and intended audience to aid in AI development with risk management (i.e., use-case profile, temporal profile, cross-sectoral profile, etc.) [7]. Overall, the NIST RMF is a voluntary framework intended to provide the horizontal groundwork for mapping AI risk [10]. The RMF intends to create a culture of trust and transparency behind AI and the public sector [10]. While the NIST RMF is forward-thinking, the document is still unclear on risk quantification and profile shaping [10]. The RMF looks to the greater community for insights into how to roadmap future iterations [10].

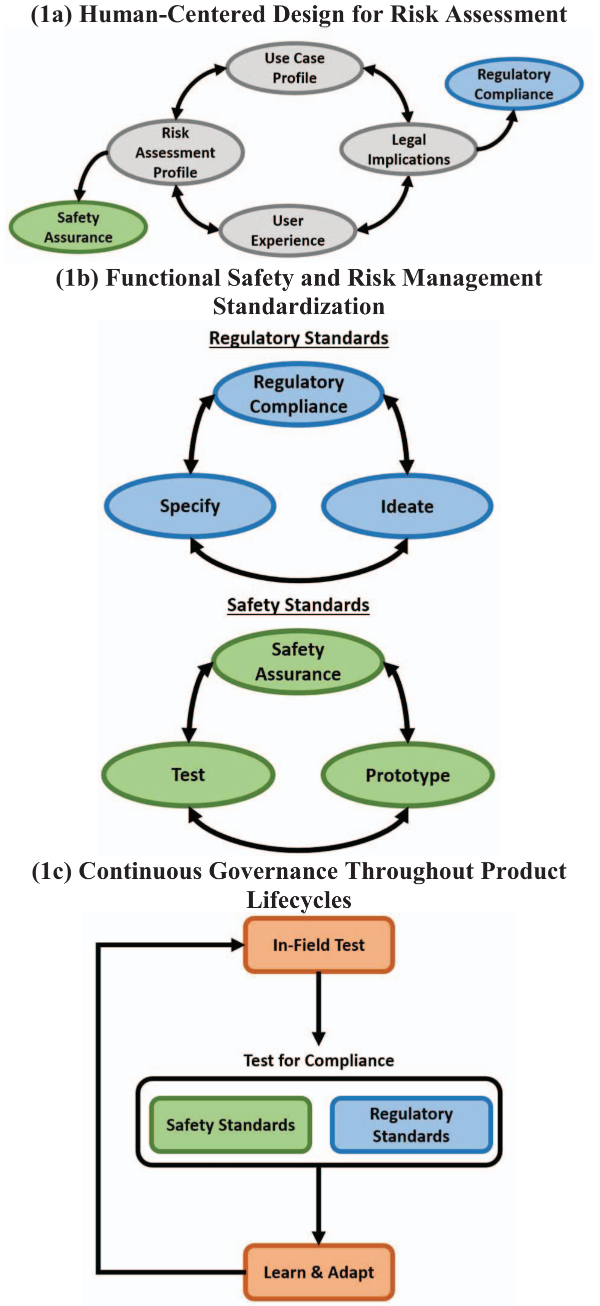
The Blueprint for an AI Bill of Rights illustrates the relationship between individuals and AI/automated systems [6]. This document includes the following rights of individuals: to be safe from ineffective systems, discrimination, abusive data practices, awareness of AI/automated systems, and human alternatives [6]. This document differs from the EU AI Act and the NIST RMF, as the Blueprint for an AI Bill of Rights is a white paper that outlines a set of principles to guide the design, use, and deployment of automated systems. The Blueprint for an AI Bill of Rights promotes democratic values and civil liberties among the deployment automated systems. It also justifies the possible implementation of automated systems in the contexts of national security, defense, or law enforcement. This may present normative challenges in a representative democracy. This guide on AI compliance is unique in the way that personal liberties are prioritized; such protections are only extended so far until AI use case is justified under defense related conditions. However, like the other frameworks, specificity is lacking in the Blueprint for an AI Bill of Rights. While it is understood that such recommendations are ambiguous to be futureproof, precise technical standards are needed to characterize AI use case and risk (or individual liberty infringement).

B. Human-Centered Design

A common engineering design methodology is human-center design (HCD). The broad philosophy of HCD is to empower end-users to aid in the engineering system design process (typically during requirement specification and usability testing) [11]. HCD engages the engineering design process as participatory design through a multi-disciplinary perspective [11]. HCD addresses the whole user experience and allows users to become, in essence, co-designers. For example, several HCD activities include use-case understanding, user requirement understanding, design solution/iteration, and evaluation methodologies. The designer's role is to facilitate these activities to fulfill several HCD principles. These HCD principles include defining task suitability, ease of use, user expectation conformity, suitability for learning, controllability, error tolerance, and suitability for individualization. [11]. In addition, HCD recognizes that some engineering systems will have a real-world impact [12]. The utility of these systems must be taken into consideration from a multi-stakeholder perspective (end-user group, business, environment, etc.). While HCD has been helpful for product usability, the application toward AI is still unclear as compliance standards are emerging. However, HCD can enable forward-thinking compliance tasks such as the NIST RMF use-case profiles [7].

C. Functional Safety

Standardization organizations have recently focused on functional safety (FuSa) guidelines for engineering systems.



(1a) Human-Centered Design for Risk Assessment

(1b) Functional Safety and Risk Management Standardization

(1c) Continuous Governance Throughout Product Lifecycles

Figure 1: AI Ethics and Compliance by Design Data Governance Framework.

The focus of these FuSa standards has been centric on automotive applications. However, newer developments in these standards have shifted toward general interoperability and AI applications [17][18][19][20][21]. The International Organization for Standardization (ISO) released an international standard for FuSa relative to automotive vehicles (ISO 26262) [16]. This standard has been influential for automotive vehicle development internationally. ISO 26262 addresses "possible hazards caused by malfunctioning behavior… including the interaction of these systems" [16]. In conjunction with ISO 26262, IEEE responded by forming its own standard to aid in specializing further FuSa guidance for engineering practitioners (IEEE P2851) [17]. The standard focuses on the interoperability of product lifecycles within the framework of FuSa. It is important to note that AI has become a key part of the activity for this IEEE standard's working group [17]. In addition, ISO has begun several working groups to address AI development and FuSa. These working groups include machine learning data quality measures [18], data quality process frameworks [19], AI system life-cycle processes [20], and AI-specific FuSa standards [21].

III. AI Ethics and Compliance by Design

We establish a trans-disciplinary viewpoint through current AI regulatory frameworks and human-centered design processes. Our trans-disciplinary viewpoint considers risk assessment, functional safety/risk management, and continuity as central principles for data governance. In addition, the key aspect of our governance framework involves collaboration between compliance entities and designers. In the following sections, we will cover each aspect of our proposed data governance framework: Human-Centered Design for Risk Assessment, Functional Safety and Risk Management Standardization, and Continuous Governance Throughout Product Lifecycles. We show our proposed trans-disciplinary governance framework in Fig. 1.

A. Human-Centered Design for Risk Assessment

We insert HCD into the initial stage of AI product design (Fig. 1a). We use the illustrated model to exemplify how regulatory compliance can inform the greater engineering design process. At the initial stage of design, we consider the protections and priorities of the end-user and associated groups. User feedback is vital to the HCD process as the system design models them as end-users. In addition, through HCD we can utilize NIST-supported risk-based profiles for essential risk assessment of the AI system. Legal frameworks are considered to exemplify the regulatory considerations of the use case and the users (i.e., privacy, data protection, preservation of personal liberties, etc.). In this stage, compliance teams can inform engineers of the legal implications of the use case and AI system(s) interoperability, to include the jurisdiction of the AI system itself. Hence, designers can be aware of the regulatory concerns from current frameworks [5][6][7][22]. From this initial stage, designers may consider the overall user experience from the use case, risk assessment, and legal perspectives. In addition to bolster traceability, these profiles are meant to aid in documenting the design process for quality assurance groups. The design process is aimed to inform the safety assurance and regulatory compliance requirements through product lifecycles.

B. Functional Safety and Risk Management Standardization

The previous legal implications and risk assessments inform the next stage of our proposed governance framework (Fig. 1b). The purpose of this governance stage aids in scoping the overall design of the AI system and strategizing in-field safety/risk-management mechanisms (Fig. 1b). We define these mechanisms as the regulatory and safety standards of an AI system. Legal and regulatory considerations inform the design process in this stage. In addition, we use the risk assessment and FuSa standards to inform safety standards [16][17][21]. These standards are met to inform the design considerations of the safety/risk-management mechanisms of AI digital tools. Collaboration between compliance groups and designers will enable the previously documented legal implications and risk assessment to be streamlined toward system design. We consider the output of this trans-disciplinary collaboration as the standards of the system. Please note that some of the AI-centric FuSa standards are in development [18][19][20][21]. However, utilizing previously documented FuSa practices can ensure safety assurance as part of the overall system design (design for error) [11]. These regulatory/safety standards are specific per engineering system and should be developed based on the previous human-centered design for the risk assessment process (Fig. 1a). We leave this part of our governance framework open to interpretation to enable flexibility between engineering and compliance groups. The design/proposed governance framework is supported as an iterative process. As new regulatory frameworks and FuSa standards are released, then these can be included to inform new specifications per use case.

C. Continuous Governance Throughout Product Lifecycles

The last stage of our proposed compliance by design process utilizes the previous system-specific standards for in-field testing for AI digital tools (Fig. 1c). We employ a cyclic but continuous execution and testing for AI products in-field. Mitigating AI drift problems is essential to [10]. ensuring strong test strategies and update mechanisms during the execution of in-field products will bolster overall user safety and compliance. We refer to this as Test for Compliance. In Test for Compliance, we utilize the previously documented regulatory/safety standards to establish tests per system design. Establishing a pathway for compliance groups and designers to create standards and system-level test strategies will enable accountability across disciplines. In addition, this will also provide a strategy to enforce monitoring, traceability, and accountability through AI governance. It is vital that collaboration between designers and compliance groups takes place in all stages of the governance's framework to ensure transparency from engineering design and legal compliance.

IV. Future Work

The proposed governance framework aims to provide a general design and compliance strategy for AI digital tool product lifecycles. AI digital tools have the potential to widely impact society (i.e., medical, automotive, security, etc.). As AI and related regulatory frameworks develop, it is important to give form to these next-generation design and compliance strategies. We acknowledge that AI development and the related regulatory landscape may change exponentially. Further investigation and case studies are needed to exemplify how ethics and compliance by design can occur as a collaboration between compliance groups and designers. Centering governance through ethics and compliance by design is how we may enable trustable and responsible AI digital tools. In addition to public frameworks for regulatory compliance, we urge actors to consider the evolving landscape of private AI law. Future investigation should be done into the patchwork of varying legal guidelines for AI design and implementation between private parties (e.g., third-party vendors and companies who are transitioning to AI-driven programming).

V. Conclusion

Overall, we have exemplified how regulatory frameworks can be used for AI ethics and compliance by design. Throughout the proposed governance framework, we include legal implications, risk assessment, and regulatory compliance. By utilizing insights from the design process, legal perspectives, and Functional Safety standards; compliance groups and engineering practitioners can inform safety and risk-management guidelines to ensure compliance. Establishing a transdisciplinary viewpoint through engineering and law establishes an insightful design and governance framework. In addition, establishing verification strategies from legal and Functional Safety perspectives will ensure Ethics and Compliance-by-Design.

References

[1] S. C. Shapiro, “Artificial intelligence (AI),” Encyclopedia of Computer Science, 2003, GBR, pp. 89–93.

[2] OECD, “Advancing Accountability in AI Governing and Managing Risks Throughout the Lifecycle for Trustworthy AI,” *OECD*, 2023.

[3] P. Hacker, et al., “AI compliance – challenges of bridging data science and law,” *Journal of Data and Information Quality*, vol. 14, no. 3, pp. 1–4, 2022.

[4] Maliha, “Who is liable when Ai Kills?,” *Scientific American*, 01-Mar-2023. [Online]. Available: https://www.scientificamerican.com/article/who-is-liable-when-ai-kills1

[5] Mauritz Kop, EU Artificial Intelligence Act: The European Approach to AI, Transatlantic Antitrust and IPR Developments, 2021.

[6] White House Office of Science and Technology Policy, 2022, pp. 1–73.

[7] E. Tabassi, “AI Risk Management Framework,” *National Institute of Standards and Technology*, 2023.

[8] Daniel Schiff, et al., “What's Next for AI Ethics, Policy, and Governance? A Global Overview,” *In Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 2020, pp. 153–158, https://doi.org/10.1145/3375627.3375804

[9] Aligned Design: A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems, Version 2. IEEE, 2017. http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html

[10] CREDO AI, 2023.

[11] ISO 9241-110, 2020.

[12] T. C. H. Wong, et al., "A Human Centered Design Framework to Support Product-service Systems," *2018 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM),* Bangkok, Thailand, 2018, pp. 545-549, doi: 10.1109/IEEM.2018.8607680.

[15] L. Tiedrich, “Data and emerging technology: The New Ethics and Compliance Frontier,” *Data and emerging technology: The new ethics and compliance frontier | COSMOS Compliance Universe*, Oct-2022. [Online]. Available: https://compliancecosmos.org/data-and-emerging-technology-new-ethics-and-compliance-frontier.

[16] ISO 26262-1, 2018.

[17] R. Mariani, et al., “Exchange/Interoperability Format for Safety Analysis and Safety Verification OF ip, SoC & Mixed Signal ICs,” IEEE P2851, 2019.

[18] ISO/IEC CD 5259-2, 2023.

[19] ISO/IEC CD 5259-4, 2022.

[20] ISO/IEC DIS 5338, 2023.

[21] ISO/IEC CD TR 5469, 2022.

[22] EU Global Data Protection Regulation, Official Journal of the European Union, 2018.