AI2X Co-evolution Guidebook and Case Study for Human-centered AI Framework

Akihisa Morikawa
IMAGINARY Corporation
Nagoya, Japan
morikawa@imaginary-inc.jp

Yutaka Matsubara
Nagoya University
Nagoya, Japan
yutaka@ertl.jp

Daichi Mizuguchi
Atelier Corporation
Tokyo, Japan
daichi.mizuguchi@ateliet-inc.com

Kiyoshi Fujiwara
National Institute of Advanced
Industrial Science and Technology
(AIST) Tsukuba, Japan
k-fujiwara@aist.go.jp

***Abstract*—As AI systems become more prevalent, concerns about their development, operation, and societal impact grow. Establishing ethical, social, and safety standards in response to advancing AI capabilities presents significant challenges. Global initiatives are currently formulating guidelines for AI development and deployment. Given the rise of collaborative human-AI tasks, adapting AI systems to meet evolving user and environmental needs is crucial. Failure to synchronize AI evolution with these changes could have ethical and safety implications. This paper updates on the progress of the AI2X co-evolution research project launched in 2022 to promote responsible AI development and deployment.**

***Index Terms*—AI, co-evolution, regulation, guidebook**

I. Introduction

AI drives rapid progress in sectors such as manufacturing, education, and healthcare, enhancing societal convenience by personalizing education, improving defect detection, and automating disease diagnosis. However, integrating AI into safety-critical systems like vehicles and robots raises safety and privacy concerns. International committees have developed AI risk and lifecycle management guidelines, such as ISO/IEC 23053 [1] and ISO 21448 [2]. Projects like SEAMS and TIGARS have advanced AI safety [3,4]. Generative AI, which creates new data from learned information, introduces risks like unintentional human guidance, bias, and copyright infringement. To address these issues, regulations like the EU's AI Act and the US's TAG Act have been enacted [5,6]. As AI becomes more pervasive, a new lifecycle management framework is essential. Recognizing this need, we launched the AI2X Co-evolution research project in 2022, aiming to facilitate the co-evolution of AI with stakeholders [7]. This paper introduces our guidebook with a case study approach for human-centered AI Framework.

II. Coevolution Accompanied by AI Evolution

AI has been integrated into various machines and services, impacting users, developers, maintainers, and organizations, potentially affecting societal norms and regulations. This paper introduces the concept of AI2X co-evolution, promoting mutual adaptation between AI and its stakeholders (X) to achieve their goals. Fig. 1 illustrates this as a continuous process involving individuals, companies, communities, and governments, with goals ranging from personal achievements to social objectives like SDGs.

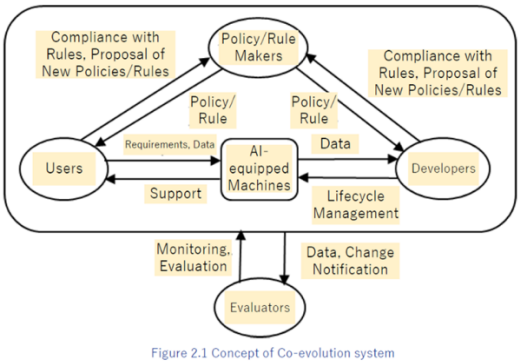

Fig. 1. Concept of AI2X Co-evolution

In AI2X co-evolution, AI systems and users evolve together. Systems adapt to users' changes, involving developers and maintainers, leading to collaborative changes among all stakeholders. For example, regulatory bodies may update policies in response to AI-user interactions. Evaluators monitor and evaluate these changes to manage new AI-related risks.

To support AI2X co-evolution, we have proposed:

- AI2X co-evolution guidebook, which assists in AI adoption, encourages stakeholder involvement, and provides a framework for regular evaluation.
- AI2X co-evolution infrastructure systems, which includes deployment software and monitoring systems to support stakeholders' goals.
- AI2X co-evolution use cases, which provide examples demonstrating the guidebook's effectiveness.

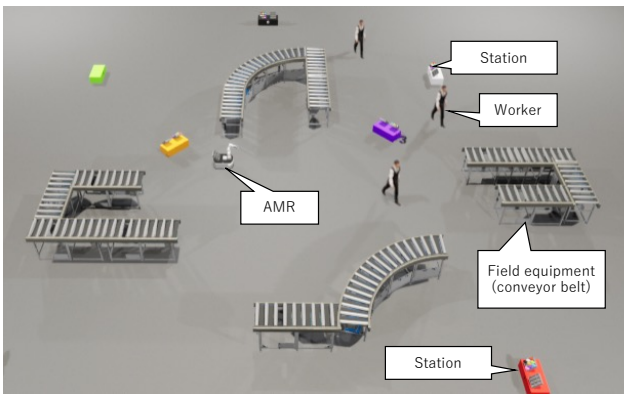

Fig.2 Overview of the target system running in virtual environment

TABLE I. Evaluation metrics and results with existing guidelines and studies.

Existing Guidelines or studies	Evaluation Metrics	Results			
		C	AE	AN	N/A
Guidelines for Human-AI Interaction [8]	Remember recent interactions		✓		
	Learn from user behavior	✓			
	Update and adapt cautiously		✓		
	Encourage granular feedback		✓		
	Convey the consequences of user actions		✓		
	Provide global controls		✓		
	Notify users about changes		✓		
Shared Control in Human Robot Teaming [9]	Novel situations		✓		
	Context-aware communication		✓		
	Performance vs preference		✓		
	Interruptions and cognitive burden		✓		
Human-Robot Collaboration and Machine Learning [10]	Precision of movement			✓	
	Robustness		✓		
	Proof of concept				✓
	Performance improvement	✓			
	Reduction of physical workload			✓	

By embracing AI2X co-evolution, we can foster harmonious advancement of AI and society, stimulating growth while mitigating potential risk

III. Enhancing Human-Robot Collaboration System: A Case Study Approach

We have developed a virtual human-robot collaboration system to evaluate its compliance with existing guidelines [8-10]. Fig. 2 shows an overview of the system, where workers and automated mobile robots (AMRs) collaborate. AMRs minimize wait times by pre-positioning components and halting transport to avoid collisions, adjusting actions based on workers' skills. Workers produce final products by processing components, moving between stations, and adapting based on AMR behaviors. The evaluation results are summarized in TABLE 1. 'C' indicates the system meets the requirement, 'AE' means the requirement is applicable and employed, 'AN' means it needs future reconsideration, and 'N/A' means it is not applicable. From these results, we identified functional requirements to enhance understanding between humans and robots and enable post-verification:

- Notifying each other of situational changes.
- Recording the history of predictions and actions.
- Facilitating functional updates.

Continuous adaptation of humans, robots, and the environment is crucial for reducing burdens and improving task efficiency in human-robot collaboration.

IV. AI2X CO-EVOLUTION GUIDEBOOK

We are updating the guidebook we developed last year based on the results of the case study. The guidebook organizes recommendations for each stakeholder for each phase of the lifecycle defined based on ISO/IEC 22989:2022 [11]. This guidebook aims to complement existing guidelines by organizing requirements for lifecycle management of humans, machines, and environments. The co-evolution guidebook draws from existing guidelines while incorporating additional requirements. This guidebook allows stakeholders to periodically adjust their behavior and organizational rules and evaluate whether the evolution of users is consistent with the expected trajectory.

V. Future Plan

Currently, we are in the process of developing the guidebook and applying the co-evolution guidebook to the target system to assess its effectiveness and refine its application. We are also investigating related papers [12-13] and sorting out metrics related to coevolution, which we would like to incorporate into the guidebook. In conclusion, our proposed AI2X co-evolution guidebook is a critical step towards navigating the rapidly evolving landscape of AI. By providing an evaluative framework for understanding the co-evolution of AI and human interaction, we aim to provide a roadmap for future development and implementation of AI technologies.

References

[1] ISO/IEC, ISO/IEC 23053 Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML) First Edition, June 2022.

[2] ISO, ISO 21448 Road vehicles — Safety of the intended functionality First Edition, Jun 2022.

[3] A. Morikawa et al. , Safety design concepts for statistical machine learning components toward accordance with functional safety standards, Aug 2020. https://doi.org/10.48550/arXiv.2008.01263,

[4] R. Bloomfield et al., Towards Identifying and closing Gaps in Assurance of autonomous Road vehicleS -- a collection of Technical Notes Part 1, Feb 2020. https://doi.org/10.48550/arXiv.2003.00789,

[5] EU, The AI Act, https://artificialintelligenceact.eu(Access 2023/07/03)

[6] US, Transparent Automated Governance Act, https://www.congress.gov/bill/118th-congress/senate-bill/1865/text (Access 2023/07/03)

[7] Y. Matsubara, et. al., Toward Human-centered AI Framework: An Introduction to AI2X Co-evolution Project. SAFECOMP 2023, Position Paper, Sep 2023. ⟨hal-04191518⟩

[8] S. Amershi, et. al., Guidelines for Human-AI Interaction, 2019.

[9] S. Matsumoto, et.al., Shared Control in Human Robot Teaming: Toward Context-Aware Communication, arXiv:2023.10218v1 [cs.RO], 2022.

[10] F. Semerato, et. al., Human-Robot Collaboration and Machine Learning: A Systematic Review of Recent Research, arXiv:2110.07448v1 [cs.RO], 2021.

[11] ISO/IEC, ISO/IEC 22989 Artificial intelligence — Artificial intelligence concepts and terminology First Edition, Jul 2022.

[12] Praveen Damacharla, et. al., Common Metrics to Benchmark Human-Machine Teams (HMT): A Review, IEEE, Jul 2018.

[13] Udara Ranasinghe, et. al., Resilience engineering: Resilience Engineering Indicators and Safety Management: A Systematic Review, ScienceDirect, Jun 2020.