Toward Human-centered AI Framework: An Introduction to AI2X Co-evolution Project

Yutaka Matsubara, Akihisa Morikawa, Daichi Mizuguchi, Kiyoshi Fujiwara

Toward Human-centered AI Framework: An Introduction to AI2X Co-evolution Project

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

Akihisa Morikawa
IMAGINARY Corporation
Nagoya, Japan
morikawa@imaginary-inc.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*—The rapid progress in Artificial Intelligence (AI) has significantly enhanced efficiency across numerous sectors, including manufacturing, education, and healthcare. However, as AI becomes more pervasive, safety and privacy concerns arise, necessitating the development of international AI standards. Numerous initiatives are underway to mitigate these concerns and enhance AI safety. Yet, the advent of generative AI introduces additional risk layers, necessitating more robust regulatory strategies. In response to this challenge, we launched a research project in 2022 titled "AI2X Co-evolution" to foster responsible development and use of AI. This paper outlines the project's objectives and provides an update on the progress achieved to date.**

***Index Terms*—AI, co-evolution, regulations, guidebooks**

I. Introduction

AI propels rapid evolution in sectors ranging from manufacturing to education and healthcare. It enriches societal convenience, personalizing primary education, bolstering defect detection in semiconductor factories, and automating disease diagnosis. However, the infusion of AI into safety-critical systems such as vehicles and robots necessitates consideration of safety and privacy. International standardization committees have responded by formulating AI risk and lifecycle management guidelines like ISO/IEC 23053 [1] and ISO 21448 [2]. Notably, our SEAMS and TIGARS projects have significantly advanced AI safety [3], [4]. Generative AI, capable of creating new data from learned information, poses unique risks. It may unintentionally guide human actions, display bias towards certain groups, or infringe on copyright and privacy laws. To counter these challenges, regulations like the EU's AI Act and the US's TAG Act have been enacted, safeguarding privacy, voting rights, and copyrights [5], [6].

As AI permeates society further, a new lifecycle management framework is essential for ensuring its beneficial integration. Recognizing this, we launched the research project named AI2X Co-evolution in 2022, aiming at the efficient co-evolution of AI with all stakeholders. This paper delves into the project's perspective and the surrounding ongoing discourse.

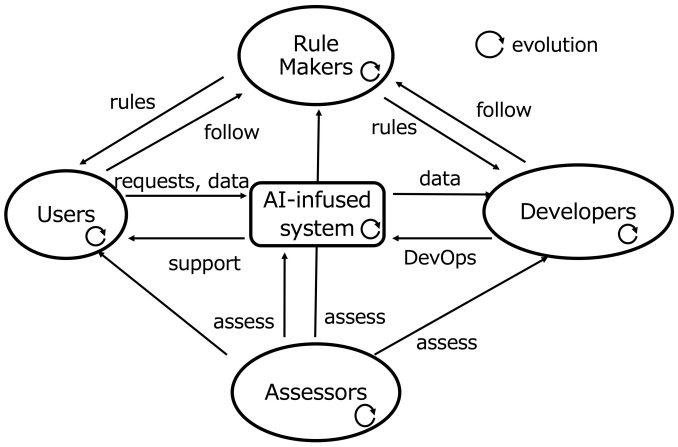

Fig. 1: Concept of AI2X co-evolution.

II. Co-evolution accompanied by AI evolution

AI has been infused to many kinds of machines and services, and affected to surrounding humans including users, developers, maintainers, and diverse organizations, with potential effects on societal norms and regulations. This paper presents the concept of AI2X co-evolution, promoting reciprocal evolutions between AI and its stakeholders (X) so that their goals are met. Fig. 1 depicts AI2X co-evolution as a continuous process striving for mutual adaptation and stakeholders' goal achievement within the AI-stakeholders interaction. Stakeholders can be ranging from individual users and families to companies, regional communities, and government. The their goals are also diverse from personal objectives, which are like getting desired jobs, improving sports skill and well being, to social objectives like Sustainable Development Goals (SDGs). In the context of AI2X co-evolution, AI-infused systems coexist with users. These users, guided by self-determined or externally defined goals, evolve as they strive towards their goals with AI's assistance. The AI-infused systems adapt to the user's changes, possibly involving developers and maintainers of the

Tab. I: Comparison of Human-AI related documents.

Contents		[7]	[8]	[9]	[10]	[11]	**Our guidebook**
Goal (policy and/or concept)	New risks		✓		✓		✓
	Human-centered policy	✓		✓	✓		✓
	New concepts	✓			✓		✓
Stakeholders	Users	✓			✓	✓	✓
	Developers	✓		✓		✓	✓
	Rule makers						✓
	Assessors				✓		✓
Life-cycle Management	Development	✓		✓	✓		✓
	Deployment						✓
	Operation	✓		✓	✓		✓
	Disposal						✓

systems, and then associated stakeholders also collaboratively change as this co-evolution occurs. For example, regulatory entities update policies in response to AI and users co-evolution. This co-evolution process, encompassing users, AI-infused systems, and other stakeholders, forms the core of AI2X co-evolution. The assessors are essential to monitor and evaluate these evolutions to manage new risks related to AI.

To support AI2X co-evolution, we propose:

- AI2X co-evolution guidebook: Assists users in introducing AI and encourages stakeholder involvement and rule updates. It provides a framework for regularly assessing users and AI evolution.
- AI2X co-evolution infrastructure systems: Includes helpful software for AI deployment, and a monitoring system to support stakeholders' goal realization.
- AI2X co-evolution use cases: Provide concrete examples to imply effectiveness of the proposed co-evolution guidebook and supportive systems.

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

III. AI2X CO-EVOLUTION GUIDEBOOK

AI technology is rapidly advancing, with great potential and unique and unknown risks. Existing articles and studies discuss the growth, applications, and potential hazards of AI, but there is a lack of in-depth analysis on the co-evolution of AI and stakeholders interaction. In Tab. I we compare existing Human-AI related documents with our proposed co-evolution guidebook. It highlights key principles of AI use, offering a human-centered approach, identifying anticipated risks, and exploring various applications. Our guidebook aligns with the human-centered AI utilization advocated by previous research [7], [10]. The potential risks of AI are clarified by [8], while references [7], [10], [11] provide valuable insight into user requirements for AI deployment and development. Additionally, studies [7], [9], [11] offer crucial guidance for AI developers. However, the current studies lacks a mechanism for assessing the evolution of AI and the resulting human adaptability. This gap emphasizes the need for a co-evolution guidebook to facilitate continuous AI management. The guidebook will enable stakeholders to periodically adjust their actions and organizational rules, and assess whether user evolution aligns with the expected trajectory. In [7], [9], the AI life-cycle focusing on development and operation are mentioned with providing valuable insights for AI development. We have been developing an initial draft of our proposed co-evolution guidebook to cover the unmet requirements.

IV. FUTURE PLANS

Our goal is to develop the first version of the AI2X co-evolution guidebook, supportive software systems, and related use-cases within the next two years. The guidebook will be constantly revised based on interactions with a variety of projects, thus helping to shape international standards. We always welcome collaborations and discussions with any researchers, developers and others who are interested in our project.

REFERENCES

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

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

[3] A. Morikawa and Y. Matsubara, "Safety design concepts for statistical machine learning components toward accordance with functional safety standards," arXiv preprint arXiv:2008.01263, 2020. [Online]. Available: 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," arXiv preprint arXiv:2003.00789, 2020. [Online]. Available: https://doi.org/10.48550/arXiv.2003.00789

[5] EU, "The AI Act," 2023. [Online]. Available: https://artificialintelligenceact.eu. [Accessed: July 3, 2023].

[6] US, "Transparent Automated Governance Act," 2023. [Online]. Available: https://www.congress.gov/bill/118th-congress/senate-bill/1865/text. [Accessed: July 3, 2023].

[7] H. J. Wilson and P. R. Daugherty, "Collaborative Intelligence: Humans and AI Are Joining Forces," Harvard Bus. Rev., pp. 114–123, Jul. 2018.

[8] C. Feijóo et al., "Harnessing artificial intelligence (AI) to increase well-being for all: The case for a new technology diplomacy," Telecommun. Policy, vol. 44, no. 6, Jul. 2020.

[9] S. Amershi et al., "Guidelines for Human-AI Interaction," in Proc. 2019 CHI Conf. Hum. Factors Comput. Syst., pp. 1–13, May 2019. [Online]. Available: https://doi.org/10.1145/3290605.3300233

[10] Z. Akata et al., "A Research Agenda for Hybrid Intelligence: Augmenting Human Intellect With Collaborative, Adaptive, Responsible, and Explainable Artificial Intelligence," IEEE Comput., vol. 53, no. 8, pp. 18-28, Aug. 2020. doi:10.1109/MC.2020.2996587

[11] A. Weiss and K. Spiel, "Robots beyond Science Fiction: mutual learning in human–robot interaction on the way to participatory approaches," AI & Society, vol. 37, pp. 501–515, Jun. 2022. [Online]. Available: https://doi.org/10.1007/s00146-021-01209-w