SafeComp2024 Position Paper

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

Akihisa Morikawa (IMAGINARY Corp., Japan)
Yutaka Matsubara (yutaka@ertl.jp, Nagoya Univ., Japan)
Daichi Mizuguchi (Atelier Corp., Japan)
Kiyoshi Fujiwara (AIST, Japan)

Negative impacts of robotaxis to traffic and social activities

https://gigazine.net/news/20230413-self-driving-vehicles-blocked-traffic/

<u>Reported cases in San Francisco</u>

- Stopped due to fog. *... This might be covered by SOTIF*
- Stopped for long time on a busy street.
- Encroached on fire scene and interfered with extinguishing the fire.

What's AI2X Co-evolution?

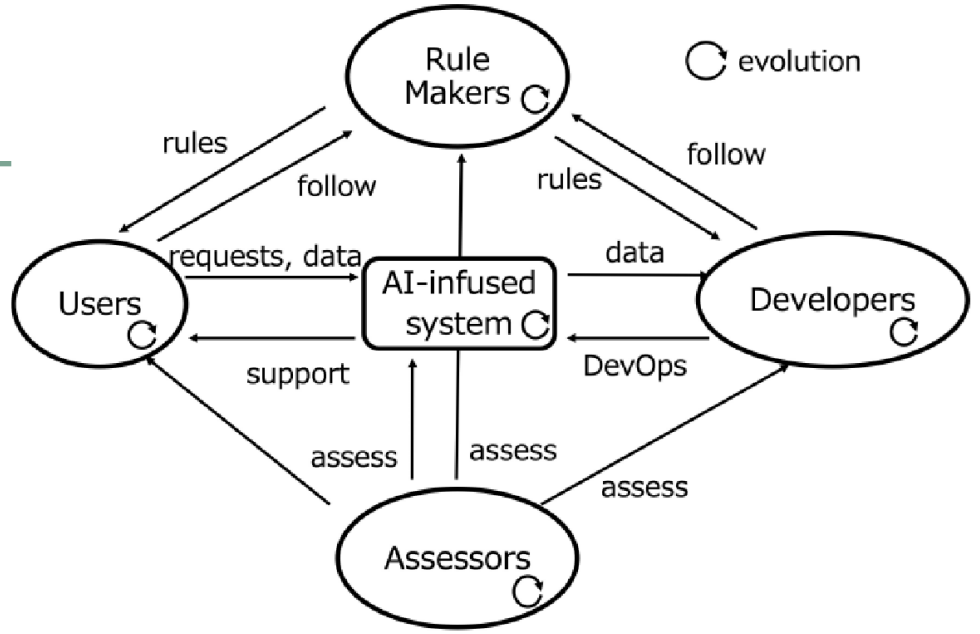


Fig. 1: Concept of AI2X co-evolution.

System safety

- Focus on how to develop and manage the AI(e.g., Statistical ML)-infused system.
- Several guidelines like ISO/IEC 23053 and ISO 21448 have been published.

System of Systems (SoS) safety

- To spread AI-infused systems with users/company/social acceptance, not only AI but also stakeholders (X) need to be continuously changed to improve each other. ***Co-evolution***
  - e.g., understanding to AI, education, rule making, assessment, ethics, …)
- RQ: How can we design a continuous framework and process striving for mutual adaptation and stakeholders' goal achievement within the AI-stakeholders (X) interaction?

We have started a new research project for AI2X Co-evolution in 2022.

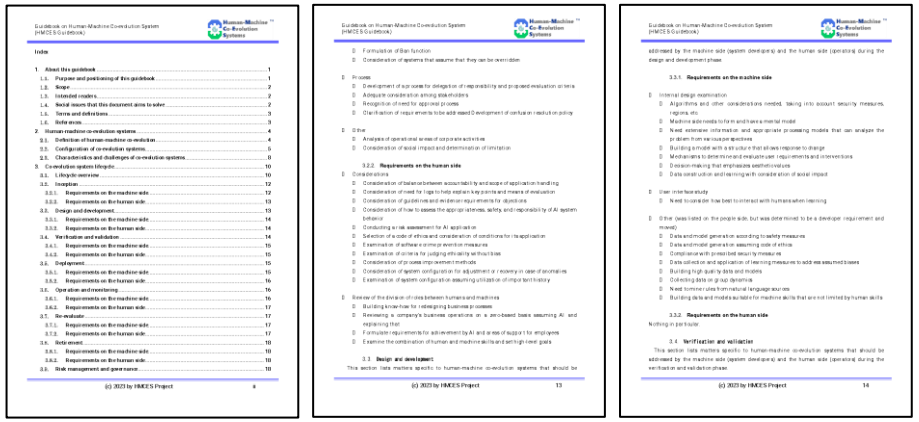
Overview of our project

	Structure for our project		Outcome
What	Goals	Improving human well-being in a society where humans and machines coexist (safety is one of the elements)	• Papers • **Guidebook**
	Issues	• Lack of (standard) mechanisms for people, machines, organizations, and society to work together • New risks due to the evolution of machines (AI)	
How	Concept	**Making structure for positive circle of mutual evolution among stakeholders** **= Co-evolution**	
	Approach	Co-evolution framework	• **Guidebook** • Supportive software tools
	Proof of concept, Demo	Development of case-study	• Case-study report

Case Study

Discussion points at our poster

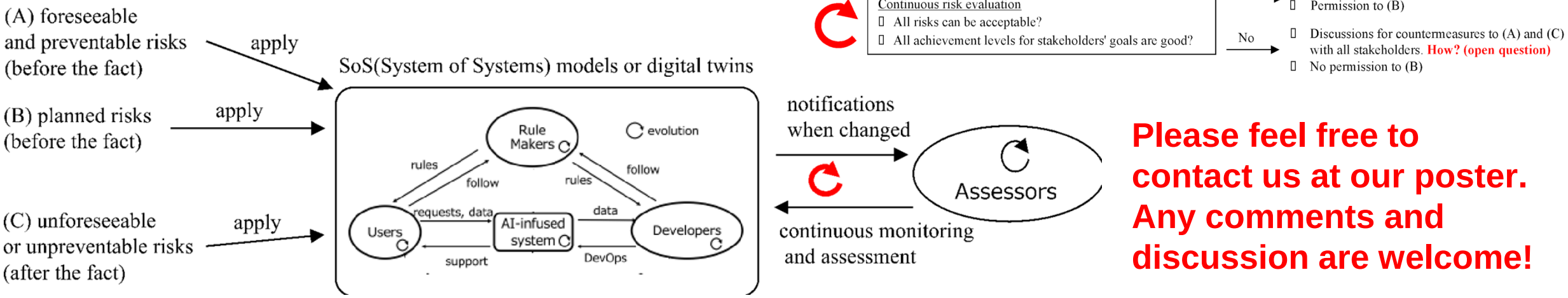
Guidebook (Preliminary version)



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						✓

Continuous dynamic risk assessment with multi-stakeholders



Please feel free to contact us at our poster. Any comments and discussion are welcome!