









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





Reported cases in San Francisco

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

- 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?

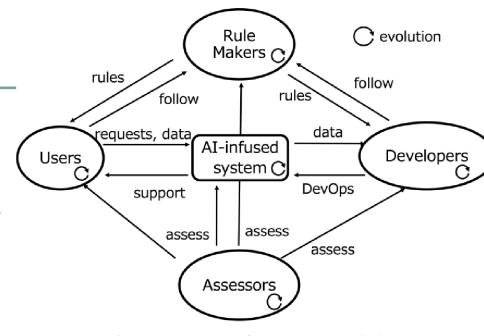
System safety

- 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

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

Case Study

Discussion points at our poster

Guidebook (Preliminary version)



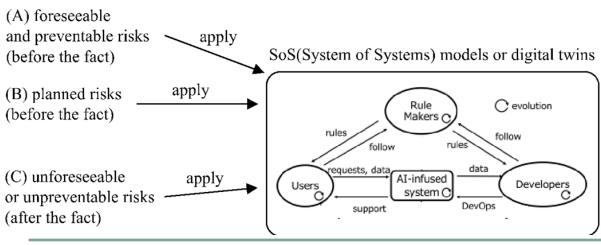
I fremente d'abliquement de l'acquire de l'	Ė		Systems				
Security							
Description of any seaso to single and respective of the proposal control of the season of the seaso		0	Consideration of systems that assume that they can be overridden				
A migray or mode that many periodics: A mapping of meeting to upon a power For expension of meeting and a meeting of meeting of meeting or meeting of meeting of meeting of meeting or meeting of meeting or meeting of meeting or m	ı	Process					
Interpret and restrict approxy powers Consideration restricts and the second powers ar			Development of agrocess for delegation of responsibility and proposed evaluation or terms				
Define the control of the control o		0	A dequate consideration among stokeholders				
Description of the control of the co		0	Recognition of need for approval process				
A Angular of an entire wear organization of a consistent of a		0	Clarification of requirements to be addressed D evelopment of confusion resolution policy				
Description 1.2. Application in the control of description of inflation. 1.2. Application in the control of inflation in the control of int	ò	01	be				
13.2 Avaignment in the human title. Construction Construction Construction Construction of the public beginning the symptomic beginning the sympto		0	Analysis of operational areas of corporate activities				
Content of the Conten		0	Consideration of social impact and determination of limitation				
Consistent of the anti-time is according to detail or equipment included. Consistent of the anti-time is according to the anti-time included. Consistent of the I take the according to the acco			3.2.2. Requirements on the human side				
Consistent of matter injury being with a primary agreement in mean or measure. Consequent of primary and even primary agreements and the primary agreement of the primary and even primary and e	D	0.0	ns derations				
Consistent of particular deliveraria animalità appoiata Consistenti on Particularia deliveraria animalità appoiata Consistenti on Particularia deliveraria animalità aggini deli Consistenti on Particularia animalità deliveraria animalità aggini deliveraria aggini deliveraria deliveraria deliveraria animalità aggini deliveraria animalità administrati aggini deliveraria animalità aggini delle a			Consideration of balance between accountability and scope of application handing				
Description of the transmission description are profressionally of Agent Johnson Leading Committee of the Section of the Sect		0	Consideration of need for logs to help explain key points and means of evaluation				
Information of the Continue of the Augment of the A		0	Consideration of guidelines and evidence requirements for objections				
Section of sold of the act of sold of sold of the act of sold of the act of sold of the act of sold of sold of the act of sold of sold of the act of sold of so		0					
Entered the date of address concept medicine measure. Licensists of address supplying managements are made to a concept and a c			Conducting or risk assessment for All application				
Entered and definition of pulpage relations by which less and considerable of pulpage and confidence of pulpage and confidence of pulpage and confidence of pulpage and analysis of pulpage analysis of pulpag		0	Selection of a code of ethics and consideration of conditions for its application				
Economics of processing in contrasting in contrasting in the contrasting in contr		0	Examination of software orime prevention measures				
Occasion of agent modigates for degeneral income an acertamental Extension of agent modigations among states of a second and agent modifications. Extension of agent acertamental modification and acertamental modification of a positive process of a centering not acceptance to acceptance and acertamental modification acceptance and acceptance accepta		0	Examination of criteria for judging ethicality without bias				
Examination of regions condigation assuming collabor of importance between Exempt of the collaboration and consequence of the collaboration of the collabor		0	Consideration of process improvement methods				
Election of the folders of closes between harmonic committees. So disting how from the relating to become processes. Reviewing a surprisely between operation as a proclosed base seasoning A and expension (see. If it is a processing the committee of the processing the committee operation of the abbreviate object to be of a page 11% and of your committee or abbreviate or abbre		0	Consideration of system configuration for adjustment or recovery in case of anomalies				
Building know how for referging biscones process Remarking a company's business operations in a service basis sessiming AF and explaining building building and explaining building buildi		0	Examination of system configuration assuming utilization of important history				
Removing a company's business operations on a pro-based basis instructing All and explaining the last of the second section of the second section of the second section of the end of the second section of the second section of the second section of the section of	D	Re					
esplanning that From laber requirements for althreament by All and areas of support for employees Examine the containation of fournary and modifies all than 6 settlings from goals 3.1 Besign and development							
Formulate requirements for attrievement by AI and areas of sugger 1 for employees Examine the combination of human and modifine skills and sethigh Ferel goals 3.3. Besign and development.		0					
Examine the continuation of human and modrine skills and set high level goals 3.1. Besign and development.							
3.1. Design and development							
		0	Examine the continuation of human and machine skills and sethigh level goals				
This section lists matters specific to human-machine co-evolution systems that should be							
		this	section lists matters specific to human-machine co-evolution systems that should be				

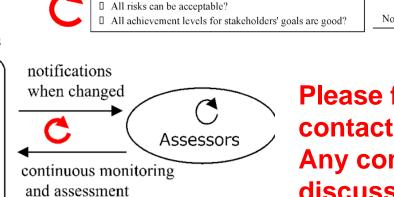


Tab. I: Comparison of Human-AI related documents.

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

Continuous dynamic risk assessment with multi-stakeholders





Continuous risk evaluation

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

□ No permission to (B)

Activation of countermeasures to (A) and (C)

Discussions for countermeasures to (A) and (C)

with all stakeholders. How? (open question)