International Symposium on AI Standards:
towards trustworthy AI through standards

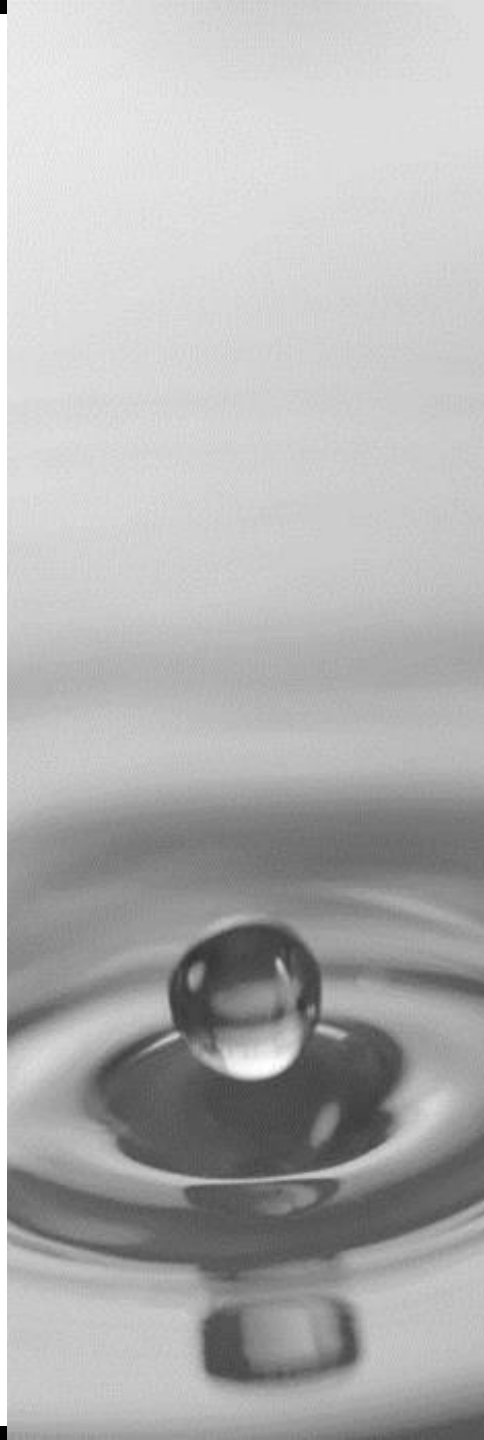
Human-Machine Co-Evolution System Guidebook

4th/July/2024

Akihisa Morikawa
Project Leader of HMCES Project
Director of Imaginary Corporation
Corporate Officer of Service Design Business Area, Witz Corporation

Human-Machine Co-Evolution Systems

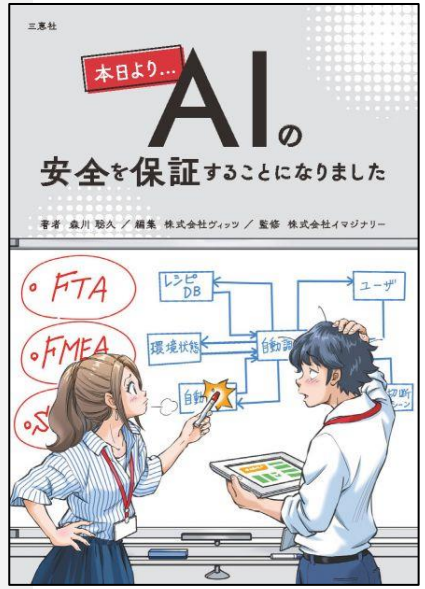
IMAGINARY



My self-introduction

I specialize in developing highly safety and reliable embedded systems.

- Biography
  - Mainly experienced in new development of embedded software for information home appliances, in-vehicle software PF development, etc.
  - Started functional safety development in 2006. Successful acquisition of **IEC 61508 process certification in 2010 for the first time in Japan.** Moreover **in 2012, ISO 26262 process certification is the first time in the world.**
  - Leading business centered on functional safety / product safety / AI safety (supporting more than 100 companies in Japan)
  - Organize methods for conforming to functional safety standards for AI and make international technical proposals (publish technical paper) https://arxiv.org/abs/2008.01263
  - **Contributed to the formulation of the AI functional safety standard (ISO/IEC TR 5469) at ISO/IEC JTC1/SC42 WG3**
  - November 2022: Publish a book that explains the points of safety assurance measures for AI systems in a story style



Published: 2022/11/14
(from Amazon)

Published: 2022/7/1
(only PDF from JSA)

INDUSTRY-ACADEMIA-GOVERNMENT JOINT RESEARCH:

RESEARCH AND DEVELOPMENT OF SAFETY CASE TECHNOLOGIES FOR AI SYSTEMS SUCH AS AUTONOMOUS DRIVING

SEAMS Project https://www.seams-p.jp/

Ministry of Economy, Trade and Industry (METI) Support Project (**2017 to 2019**)

<Main R&D results>

- Functional safety design/evaluation patterns for AI systems
  - "Safety design concepts for statistical machine learning components toward accordance with functional safety standards“
    - Akihisa Morikawa (Witz Corp.), Yutaka Matsubara (Nagoya Univ.), https://arxiv.org/abs/2008.01263
- How to quantify uncertainty in AI systems
  - based on IEC/TS 62998-1
- Building a reliable ML development process (procedures, templates and checklists)
  - based on 8.5 of UL 4600, Annex B of ISO/TR 4804, Automotive SPICE v4, etc
- Comprehensive verification method for complex conditions (using virtual simulation)
- Several concrete developments of AI systems

Agenda

1. Technical issues for the future society we envision
2. Our R&D Project “HMCES Project”
3. Human-Machine Co-Evolution System Guidebook
4. On-going tasks

The future society we envision

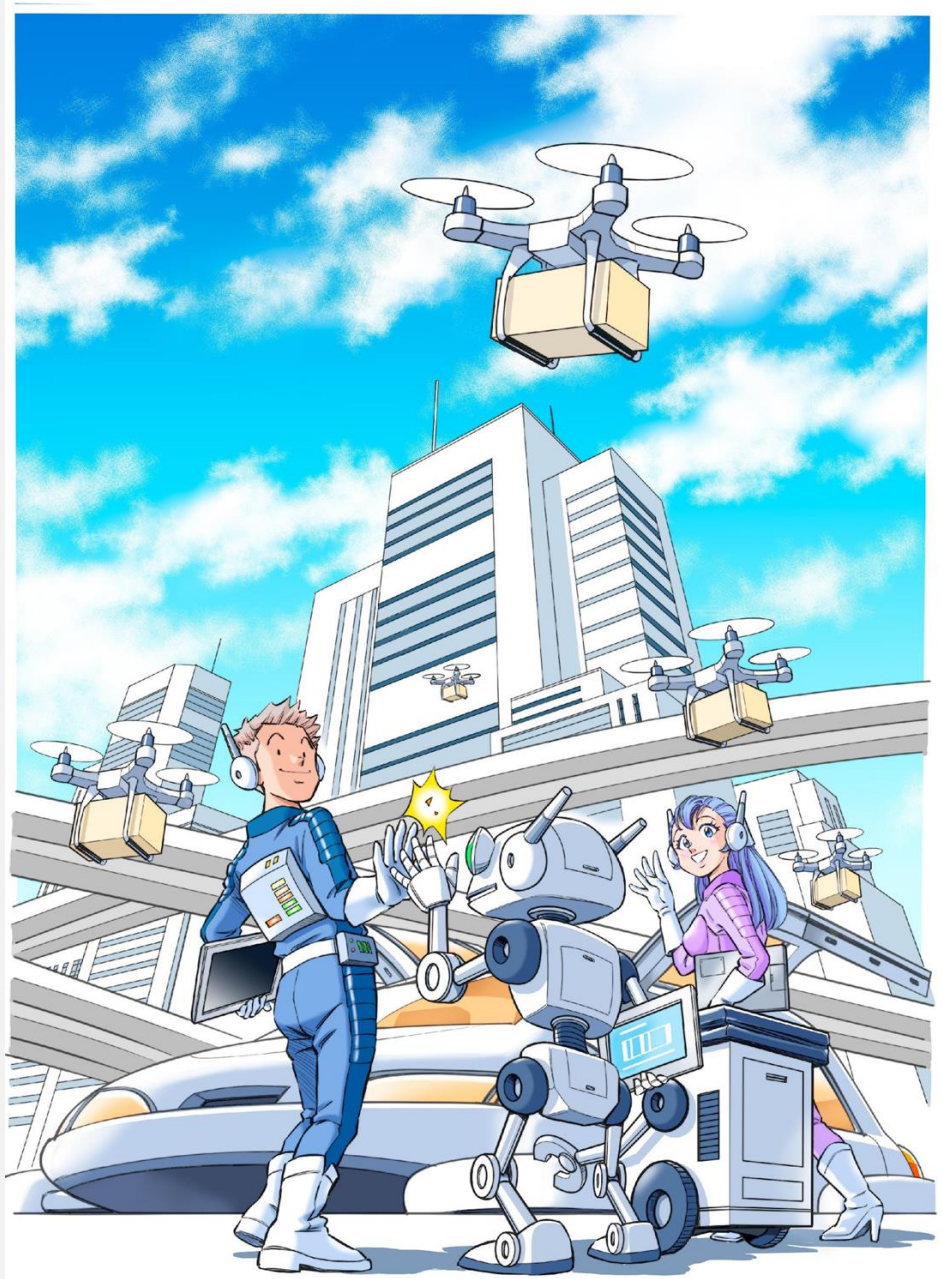
Serious problems in Japan

- **Decrease in productivity** (aging society, shortened working hours, etc.)
- **Low sense of happiness** (47th in world happiness ranking)

Improvement

Advanced symbiotic society of humans and machines (with AI)

- improves well-being by **optimizing for each individual**
- **continues to improve productivity, convenience, safety, and happiness**



"Co-Evolution" of humans, machines (AI), and society

1. Not only AI's improvement
2. Changes in human values and behavior
3. Changes in laws, infrastructure, lifestyles

Issues of continuous evolving social system

Technical Issues	Case of conventional systems	Case of co-evolving systems
(1) Verification timing of continuous changing systems	Thoroughly verify before deploying.	Since it is constantly changing, there is no timing for verification.
(2) Quality and safety assurance of unknown systems	Extract unknown risks to the assumed system as much as possible (using Guideword) and implement risk mitigation measures (SOTIF)	It is not convenient to detect deviations from the assumed system (concept drift) and shut down the system.
(3) Human tracking of machine and environmental changes	Compliance with safety operation rules for the assumed system	Humans also need to be flexible to keep up with change.

Guidebook against issues of continuous evolving social system

<Co-evolution Guidebook>

Supporting the Safety and Security of continuous evolving social system

- Establish new design and verification methods
- Guides to enable human flexibility and adaptability

Based on

< Key Existing Technologies >

- Human-Machine Teaming (HMT)
- Human behavior analysis
- Resilience Engineering (Safety II)
- Collaborative Safety
- Other several AI standards (by SC42)

Our guidebook vs Key existing practices

Key practices	Goals to be achieved through co-evolution			Monitored for "change" toward goal attainment				Uncertainty about "future change"			
	Safety Improvement	Achievement of project objectives (e.g., increased productivity)	Improved well-being of stakeholders	Entire system *The speed of change differs from each other. Machine (mechanical/ electronic systems/AI)	Surrounding environment (physical infrastructure)	Regulations and rules (logical infrastructure)	Stakeholders, including users	Known knowns (Knowing what I know)	Known Unknowns (Knowing what I don't know)	Unknown knowns (I don't know what I know)	Unkown Unkowns (I don't know what I don't know)
Functional Safety	✓	–	–	✓	–	–	–	✓	–	–	–
SOTIF	✓	–	–	✓	✓	–	–	✓	✓	✓	✓
Collaborative Safety	✓	✓	–	✓	✓	–	✓	✓	–	–	–
Human behavior analysis	–	–	✓	–	–	–	✓	✓	–	–	–
HMT	–	✓	–	✓	–	–	✓	✓	–	–	–
HMCES Guidebook	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Measures

Conventional issues：
- There is no comprehensive practice that takes into account other factors than safety.

Measures

Conventional Issues：
- No good practices exist to adapt to comprehensive changes in the overall system.
- Safety assurance methods have not yet been established for an overall system that changes frequently.

New risks：
- Complex changes increase risks that cannot be predicted or foreseen by humans.

Position of our HMCES Project

【Industries】
- AI system alone: Development is becoming more active. Interest in quality and safety is increasing.
- Co-Evolution/symbiotic society: unexamined

Application

Precondition

【Standardization】
- AI system alone: Compliant with international standards such as ISO/IEC JTC1/SC42, EU AI Act, NIST AI RMF, etc.
- Human-machine collaboration: Developing collaborative safety standards
- Co-Evolution/symbiotic society: Standard development not yet started

Contribution: Points to note/PF/Application examples
➡ Reducing costs/improving safety

References

Contributing to the standardization of guidebook

Human-Machine Co-Evolution Systems

HMCES Project https://www.hmces-p.jp/

「R&D of co-evolution guidelines to continuously improve the potential of machines」

Ministry of Economy, Trade and Industry **(METI) Support Project (2022 to 2025)**

References

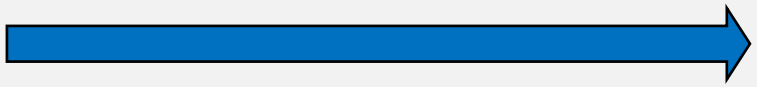
【Researchers】
- Raising various issues
- No specific solution presented (unestablished)

【Existing technologies】
- Resilience engineering (Safety II)
- Human-Machine Teaming(HMT) / Human-AI Teaming
- behavior analysis
- Human-Machine collaboration practices

SEAMS

Organization of HMCES Project

[R&D Budget]
Ministry of Economy, Trade and Industry (METI)



[Management]
The Public Foundation of Chubu Science and Technology Center (CSTC)

[R&D members]

- WITZ Group
  - Imaginary [Project Leader]
  - Witz
  - Atelier
- Nagoya University [Sub Leader]
- National Institute of Advanced Industrial Science and Technology (AIST)
- Gomes Company

[Advisors/Observers]

- Japanese members of ISO/IEC JTC1/SC42
- Big maker companies
  - Mitsubishi
  - Suzuki
  - Aisin
  - Kobelco
  - Meiden
  - Hitachi
- Japan Automobile Research Institute (JARI)
- Certification bodies
  - UL Japan
  - DNV

Requirements for human-machine cooperative system development for safety

[Theme-1] Arrangement of requirements to ensure the safety and security of human-machine cooperative systems

Human-Machine Co-Evolution Systems GUIDEBOOK

Contribution

Standardization activities (for AI system, Collaborative Safety, etc)

System operation requirements/social rule requirements for safety

Feedback

Guidelines / laws

[Theme-3]

Advanced Pilot APPLICATION

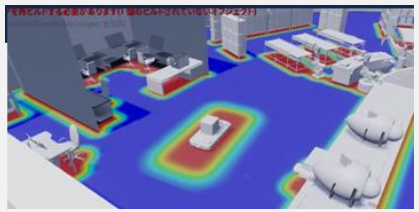
Development of integrated control robot for hands, feet, eyes, ears, and mouth

[Theme-2]

Human-Machine Co-Evolution Systems PLATFORM

social implementation

"Advanced symbiotic society" of humans and machines

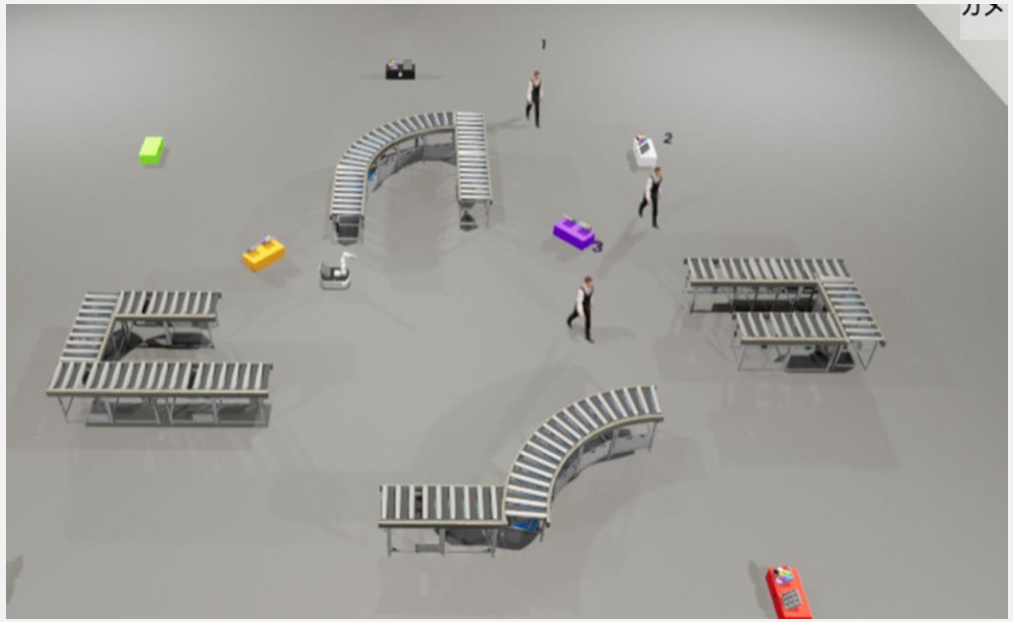


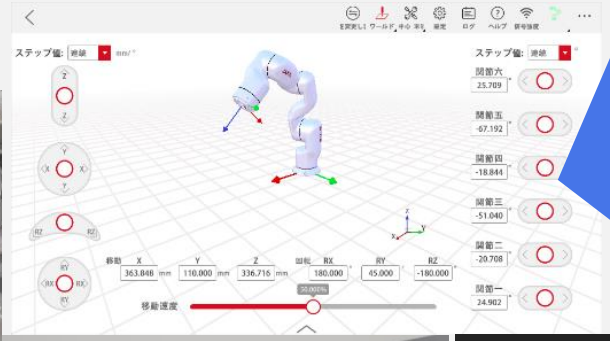
< Examples of application >

- Intelligent collaborative robot (in factory)
- Housekeeper robot (in-house)
- Secretary robot (in the office)
- etc

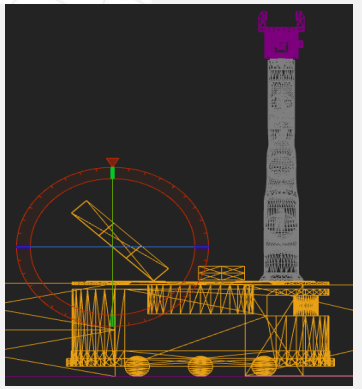
Co-evolutionary pilot system development and evaluation

- Coexistence/collaborative robot development
  - System configuration: Hands (arm robot), feet (AMR), eyes (camera), ears & mouth (AI concierge app)
- Humans and robots share tasks in a common space
- Each behavior evolves in response to mutual changes, such as improvements in human skills and robot performance.
- By building a similar system (including a human model) in virtual space, flexible operation verification is possible.

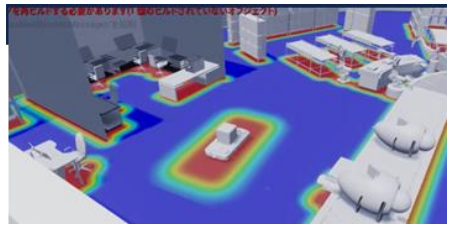




Virtualization as faithful as possible



< Real Target >



Positioning of co-evolution necessary for future society

GOAL
(Issues we want to solve for future society)

Well-being

Productivity

Safety

Addressing risks unresolved by existing guides for AI systems

Problem solving concept

Human-Machine Co-Evolution

New risks arising due to co-evolution

①Measures to promote co-evolution

②Measures to reduce new risks

Methodology
(How to appropriately realize co-evolution)

Human-Machine Co-Evolution system life cycle

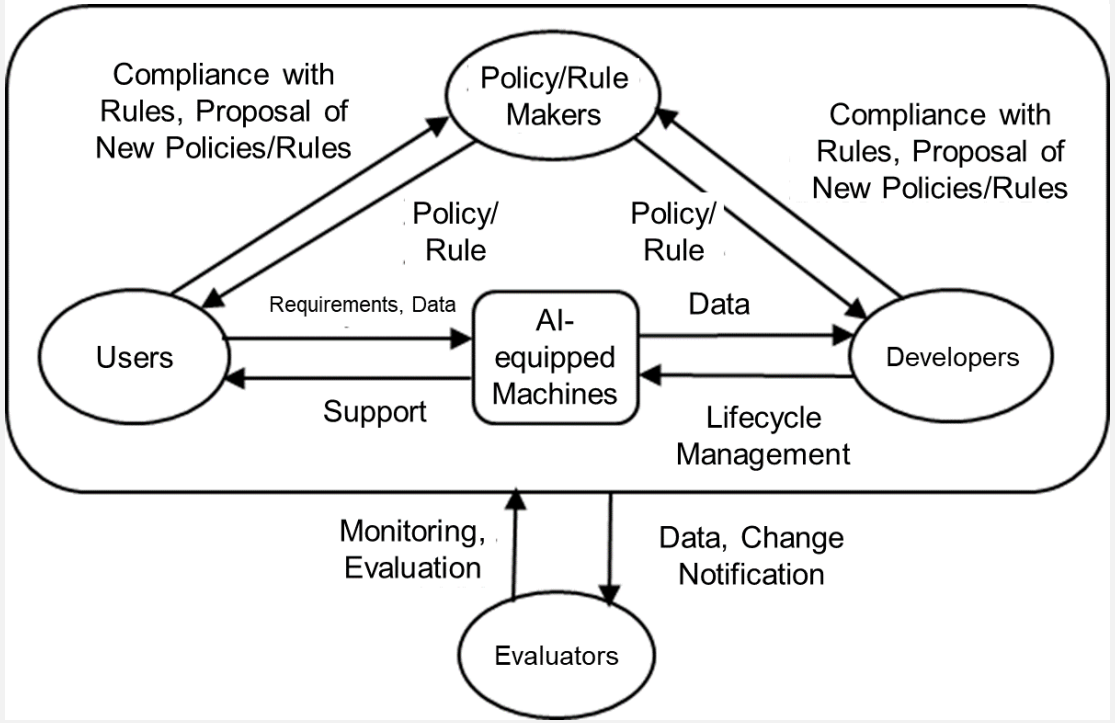
Human-Machine Co-Evolution Systems GUIDEBOOK ™

Table of contents of HMCES Guidebook

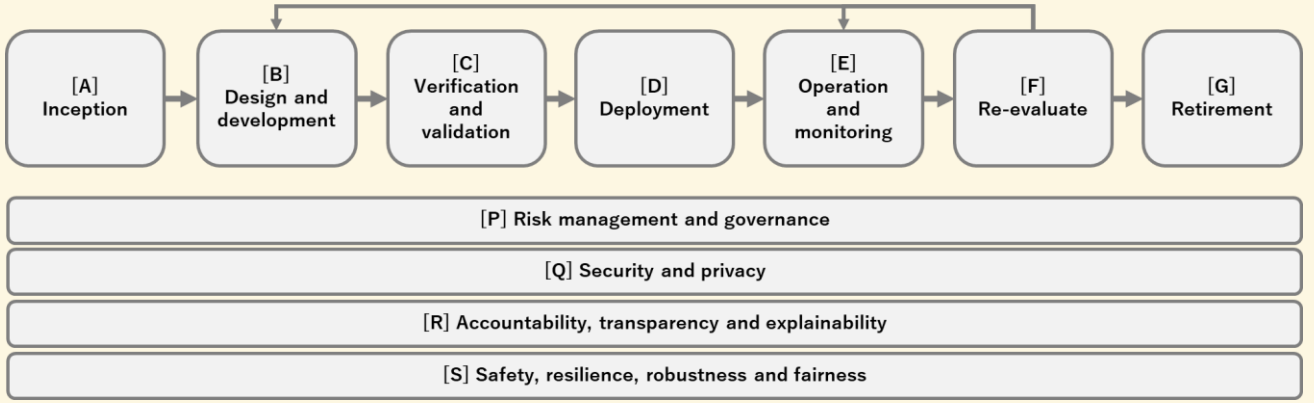
Summary of HMCES Guidebook (1/2)

- Defining Co-evolving Systems and Stakeholders
- Definition of the co-evolving system life cycle (based on Figure.3 of ISO/IEC 22989:2022)
- Organize the requirements for human-machine co-evolution for each phase

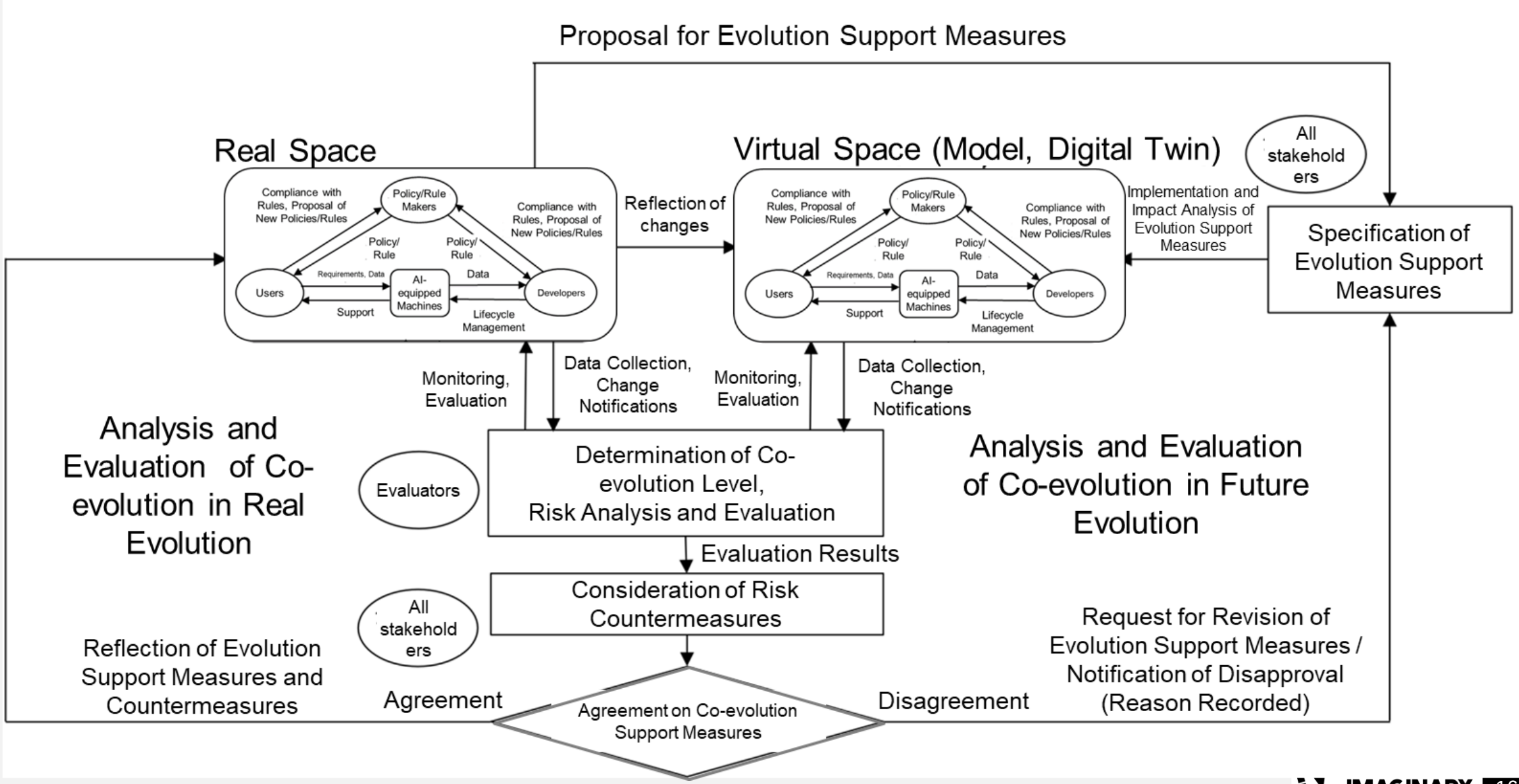
<Concept of Co-evolution system>



<Life cycle>



Summary of HMCES Guidebook (2/2)



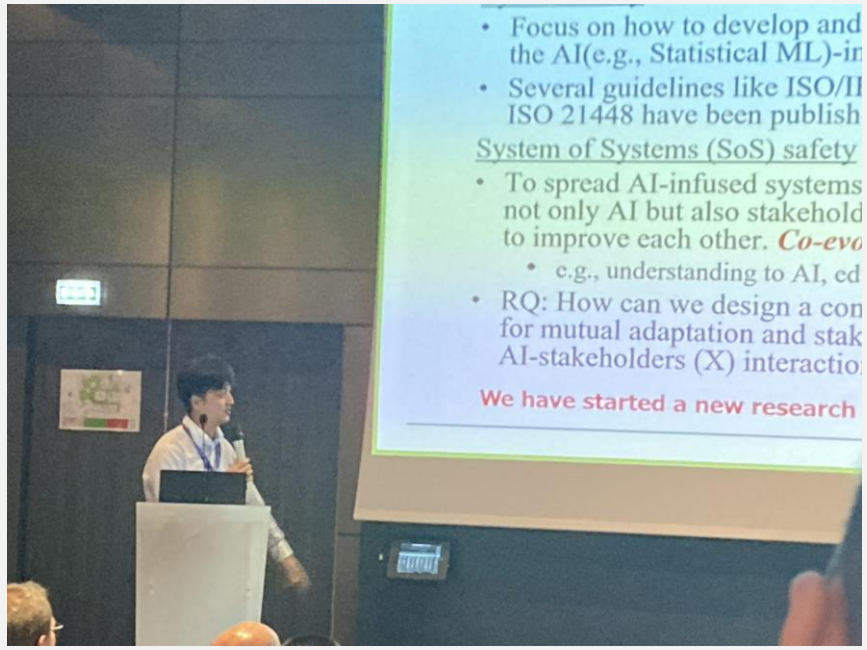
2 main roles of virtual space

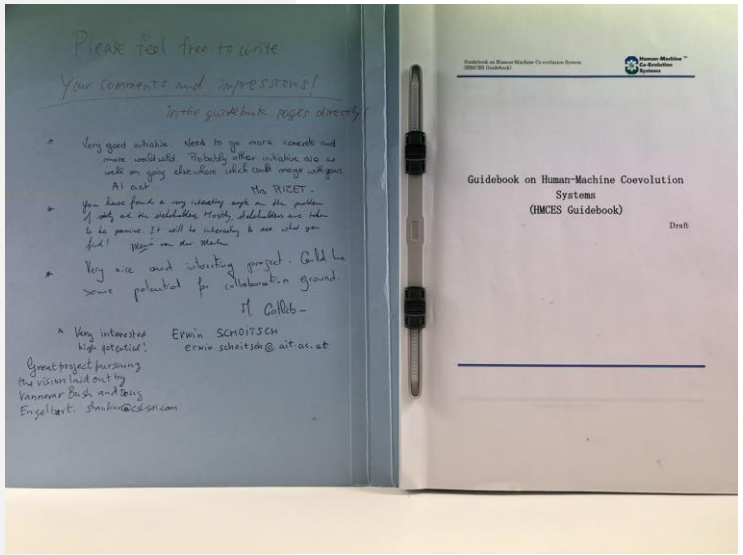
- Monitoring the real space
  - events and changes that are actually occurring in the field
  - **Known-known** events: sufficiently predictable
  - **Known-unknown** events: not easy but are somewhat predictable

- Simulation for unknown conditions
  - providing an opportunity to take countermeasures in advance
  - **Unknown-known** events: not clearly known at this time but could occur in the future
  - **Unknown-unknown** events: not expected at all

SafeComp2023 presentation and poster

- Presented the Co-Evolution Guidebook at an international conference (SafeComp2023, September 2023) with a presentation and poster booth and received some feedback comments from international experts.
- **Toward Human-centered AI Framework: An Introduction to AI2X Co-evolution Project**,
  - Yutaka Matsubara, Akihisa Morikawa, Daichi Mizuguchi and Kiyoshi Fujiwara
  - https://safecomp2023.cnrs.fr/position-papers/





Indicators related to human-machine cooperation

- Evaluation indicators are essential to promote, analyze, and improve the co-evolutionary system

< Base indicators >

- HMT: Common Metrics to Benchmark Human-Machine Teams (HMT): A Review (https://ieeexplore.ieee.org/document/8404030)
- Resilience engineering: Resilience Engineering Indicators and Safety Management: A Systematic Review (https://www.sciencedirect.com/science/article/pii/S2093791120302663)

138 indicators

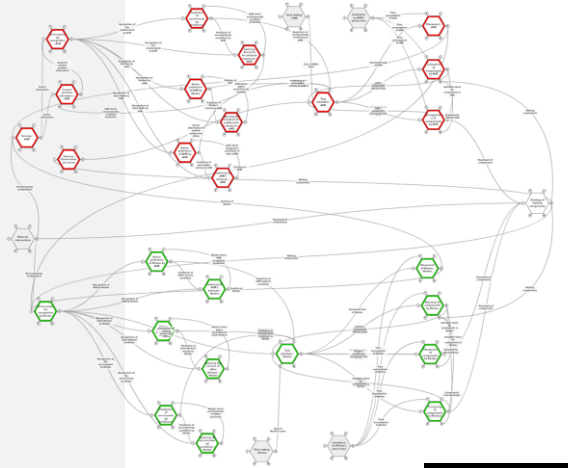
Human/Machine measurement

applicability

				計測可否			Stakeholders			
	Metrics/RE indicators	指標	指標詳細説明	Users側にしかわからない指標	システム側にしかわからない指標	全体統合しないとわからない指標（For 全体監視 あるいは ルール）	Users システム側に提示が望ましい	Developers ユーザ側に提示が望ましい	Rule Makers インフラとして計測が望ましい	Assessors 第３者による監視が望ましい
1	Adaptability	適応性	主観的指標(SM)は、人間の知覚に基づいて抽象的な品質を測定するために使用される。適応性は専門家による5段階の評価を用いて測定される。主観的な尺度と人間の属性または特性は、観察者の尺度に依存する。リアルタイムで測定するのは困難	–	–	○ 専門家による5段階の評価を用いて測定可能。 "Personality, adaptability, and performance: Performance on well-defined problem solving tasks," [102]	○ ユーザの使い勝手に大きな影響を与える指標と思われる。	–	–	○ 客観的立場でユーザが容易に使用できる操作性などの監視は必要を思われる。
2	Assertiveness	自己主張（自信に満ちた態度）	主観的指標(SM)。Rathus の自己主張尺度[103]、[104]に基づいて測定される。観察者の尺度に依存するため、リアルタイムで測定するのは困難。	–	–	○ Rathus の自己主張尺度[103]、[104]に基づいて測定可能。 "A 30-item schedule for assessing assertive behavior," [103] "The effects of crit ical team member assertiveness on team performance and satisfaction,' [104]	–	–	–	–
3	Composure	落ち着き（冷静さ）	人間のパフォーマンスの向上を通じてチームの結束を高めるための人間のパフォーマンス属性の一つ。19の異なる尺度を使用して測定される[105]、[106]。観察者の尺度に依存するため、リアルタイムで測定するのは困難。	–	–	○ 19の異なる尺度を使用して測定される[105]、[106]。 "Academic resilience and the four Cs: Confidence, control, composure, and commitment," [105] "A methodological review of resilience measurement scales,"[106]	○ ユーザの冷静な使用を考慮するのであればシステム側への提示は望ましい。	–	–	○ ユーザの使用に関して、客観的立場での監視は必要と思われる。

Development of verification method for Co-Evolution systems

- Technical study conducted with European safety experts in September 2023
- Examining methods for verifying the safety of ever-changing and unknown systems using specific applications.
- Acquisition technology
  - Tips on modeling, analysis, and evaluation methods for analyzing and evaluating Co-Evolution systems
  - Tips for Formal Verification of Safety of Co-Evolution systems



On-going tasks

- Brush up on the Guidebook
  - Reflecting comments from various experts
  - Feedback from the demonstration experiment (Theme 3)

- Technology building
  - How to use HMT indicators
  - Analysis, verification, and evaluation methods for systems that continue to change or behave in unknown ways
  - Human models on virtual space

- Publication of guidebooks
  - Contribution to standardization activities

IMAGINARY

THANKS FOR LISTENING!

Akihisa Morikawa
Director
Imaginary Corporation

+81 50 5211 5282
morikawa@imaginary-inc.jp
www.imaginary-inc.jp