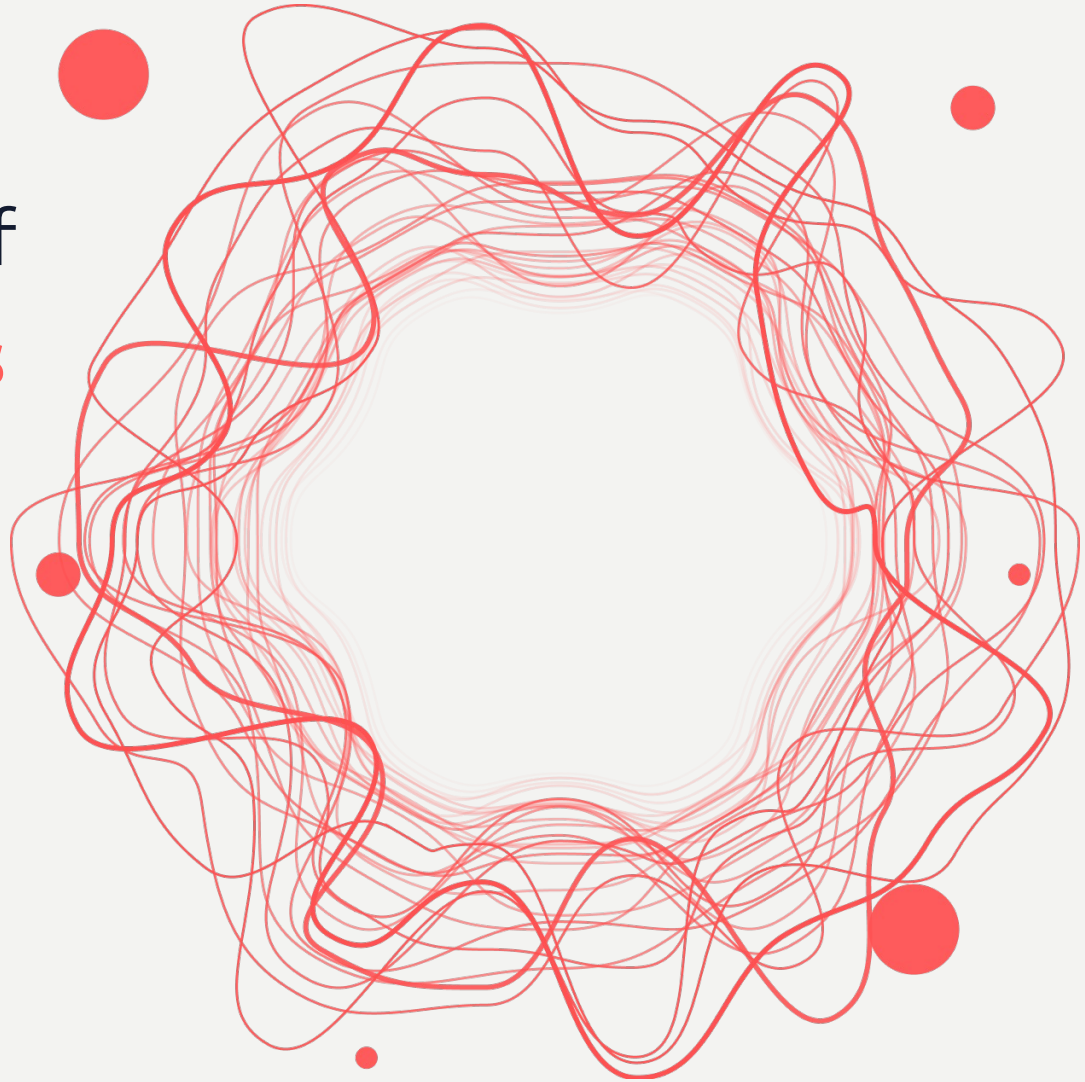
EQUMENOPOLIS

ITU Workshop - September 6, 2024

Standardization and innovation for multimedia and cable TV ecosystems

Towards the Standardization of **Conversational Competencies** for Digital Humans



Yoichi Matsuyama, Ph.D

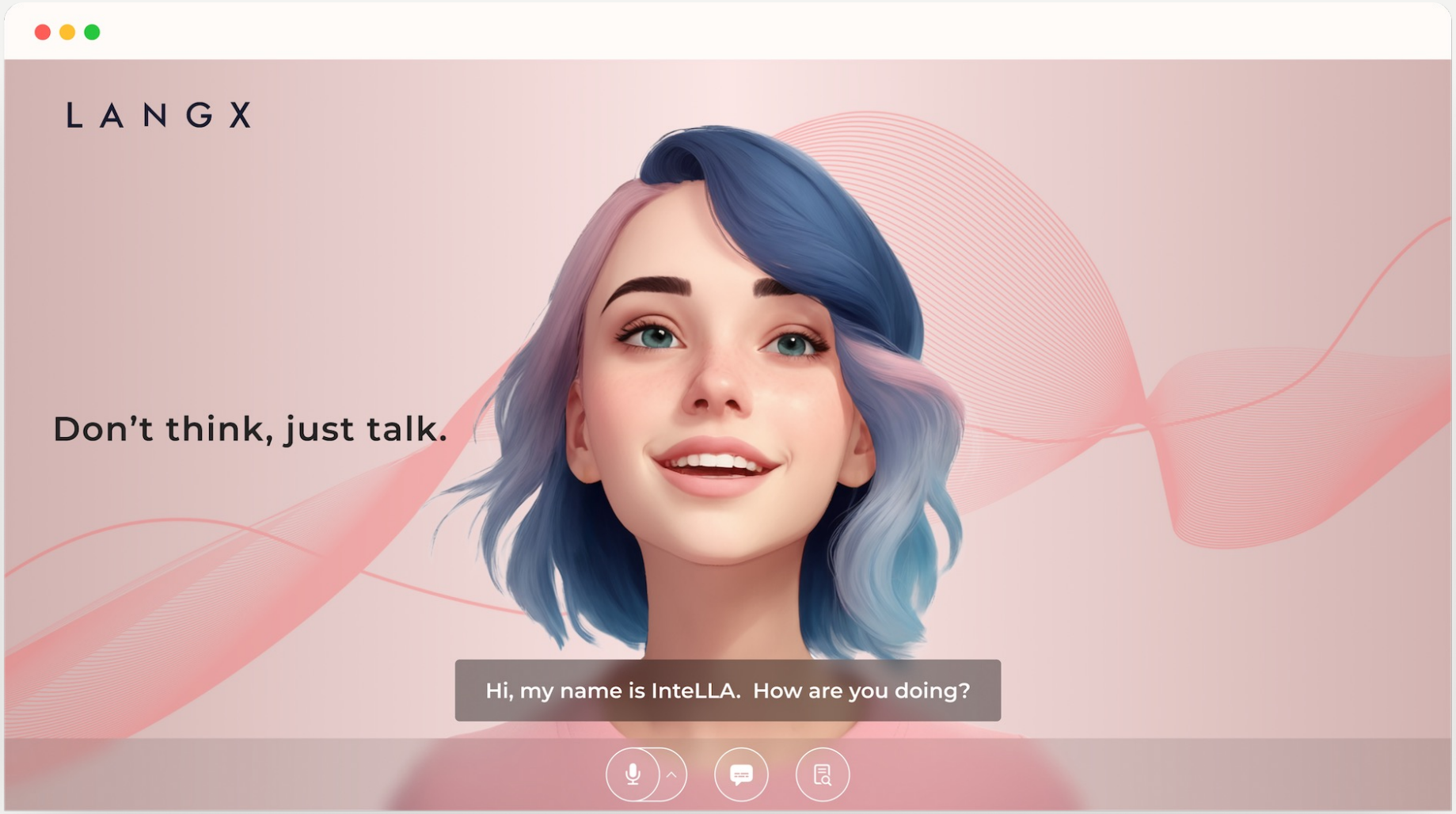
CEO, **Equmenopolis, Inc.**

Associate Research Professor, **Waseda University**

yoichim@equ.ai

Application | Online Language Learning

EQUMENOPOLIS



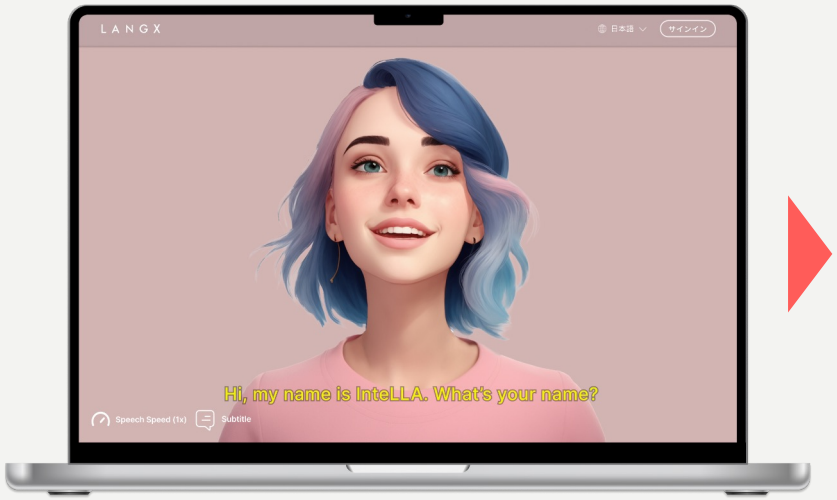
- SXSW EDU Launch 2023 Finalist
- Reimagine Education Award 2021
- Amazing Venture 100 Toyo Keizai 2023
- University Spin-our Venture Award 2024

BBC NEWS NIKKEI NHK news every. 早稲田大学 WASEDA University NEDO JST NICT Z-KAI OECD 文部科学省

Application | Online Language Learning

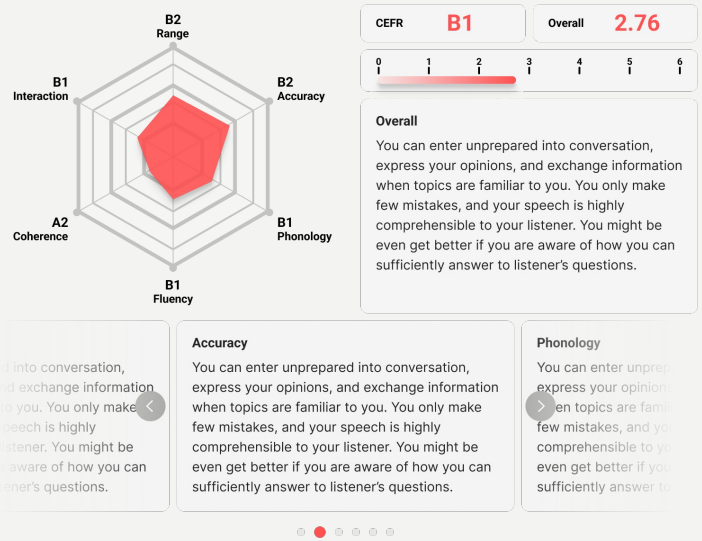
EQUMENOPOLIS

1. Natural Interactions
Eliciting latent skills and needs



Naturally draw out the **user's latent abilities and needs** through face-to-face conversations.

2. Diagnostic Feedback
promoting learner's growth



The assessment model analyzes the users's **strengths and weaknesses** of current skills and suggests the next steps to take.

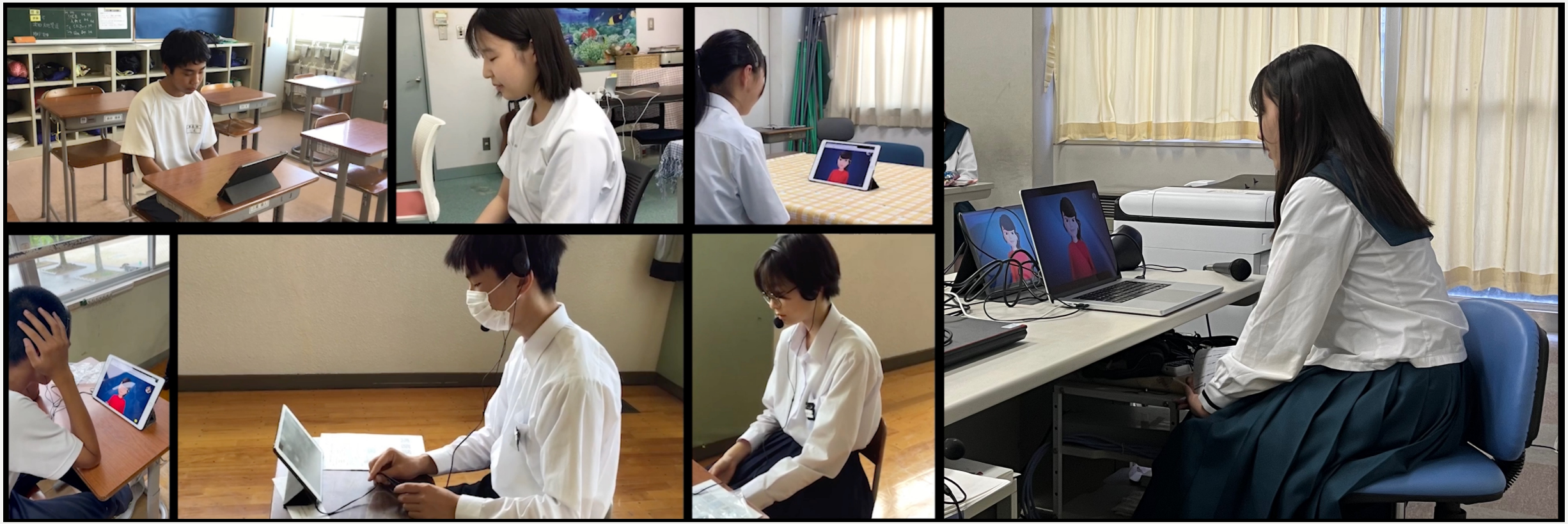
3. Curriculum Generation
based on diagnostic & learning goals

Based on diagnosis results, learning goals, and learning history, we provide **conversation curriculums tailored to daily life and business situations**.

Social Impact | Agent System Operations

EQUMENOPOLIS

Being used by significant number of **K-12** and **university students** across Japan



文部科学省
Ministry of Education, Culture, Sports, Science and Technology

OECD

岐阜市
City of Gifu

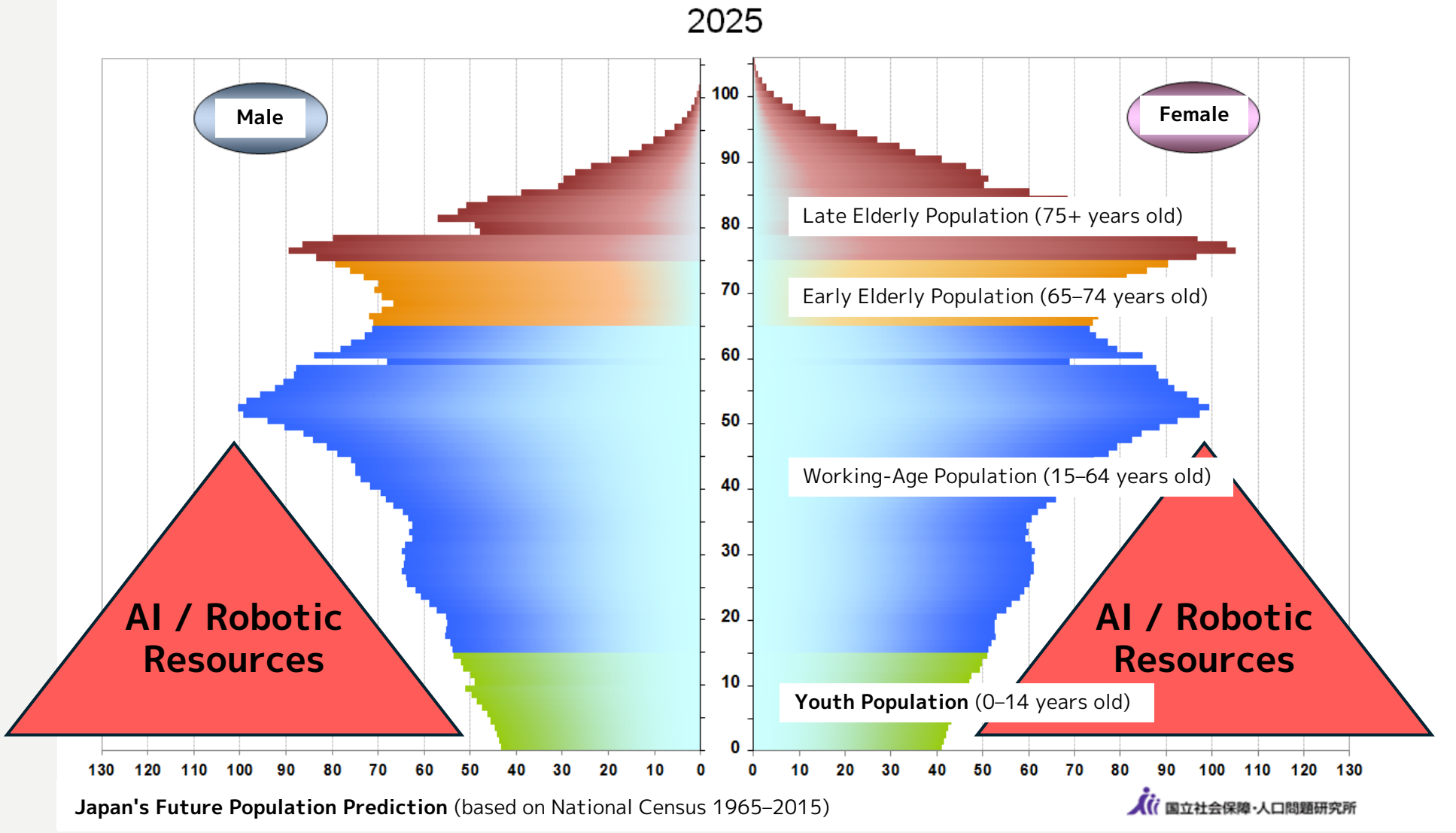
千葉県
chiba prefecture

早稲田大学
WASEDA University

Our Mission | Towards **Human-AI** Co-evolving Society

EQUMENOPOLIS

We dispatch **conversational AI agents** to schools and workplaces to improve creativity and productivity



How can **Conversational Competencies**

for **Digital Humans** be measured?

SAE Levels of Driving Automation™

EQUMENOPOLIS

	SAE LEVEL 0™	SAE LEVEL 1™	SAE LEVEL 2™	SAE LEVEL 3™	SAE LEVEL 4™	SAE LEVEL 5™
What does the human in the driver's seat have to do?	**You <u>are</u> driving** whenever these driver support features are engaged – even if your feet are off the pedals and you are not steering			**You <u>are not</u> driving** when these automated driving features are engaged – even if you are seated in "the driver's seat"		
	You must constantly supervise these support features; you must steer, brake or accelerate as needed to maintain safety			When the feature requests, you must drive	These automated driving features will not require you to take over driving	
	These are driver support features			**These are automated driving features**		
What do these features do?	These features are limited to providing warnings and momentary assistance	These features provide steering **OR** brake/ acceleration support to the driver	These features provide steering **AND** brake/ acceleration support to the driver	These features can drive the vehicle under limited conditions and will not operate unless all required conditions are met		This feature can drive the vehicle under all conditions
Example Features	• automatic emergency braking • blind spot warning • lane departure warning	• lane centering **OR** • adaptive cruise control	• lane centering **AND** • adaptive cruise control at the same time	• traffic jam chauffeur	• local driverless taxi • pedals/ steering wheel may or may not be installed	• same as level 4, but feature can drive everywhere in all conditions

ITU-T | SG16 “Digital Human”

EQUMENOPOLIS

Established as recommendations

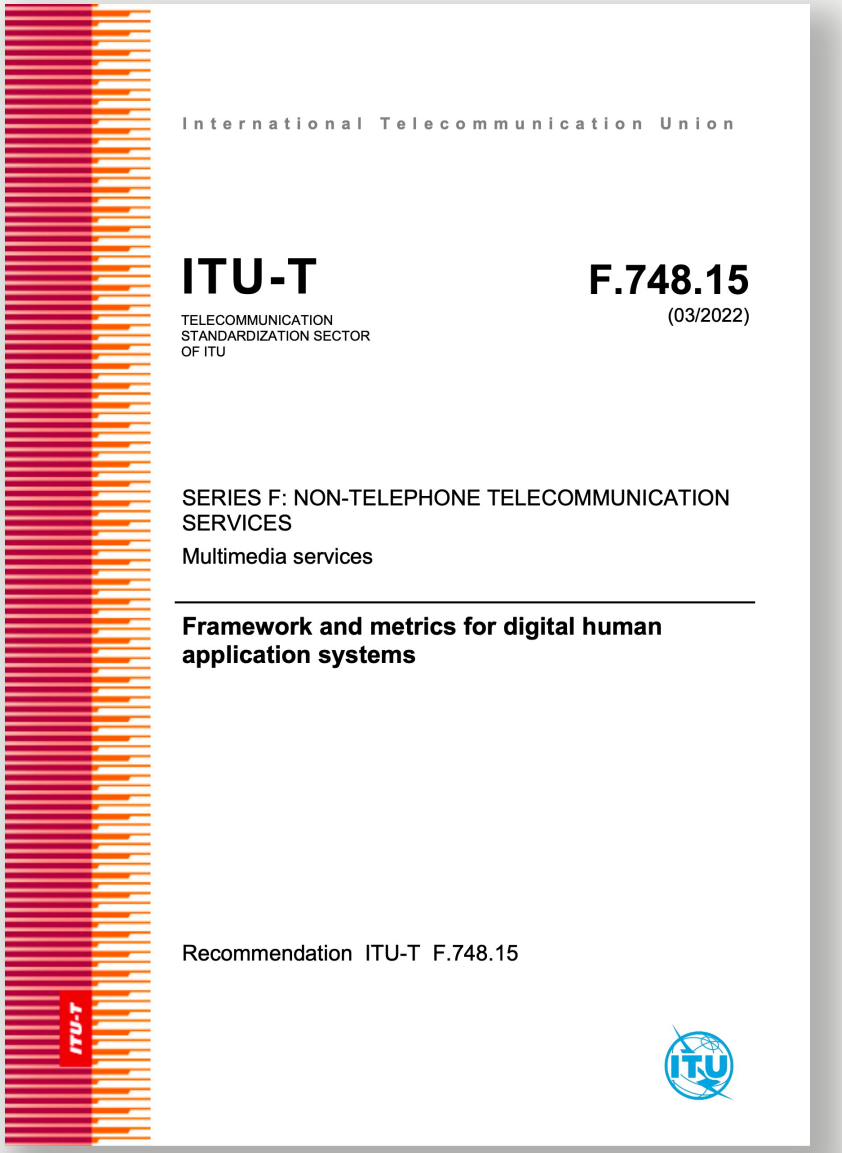
[Q5; 2022] **Framework and metrics for digital human application systems**

[Q5; 2022] **Requirements and evaluation methods of non-interactive 2D real-person digital human application systems**

[Q5; 2024] **Technical requirements and evaluation methods of 3D digital human system based on smart mobile devices**

In discussion

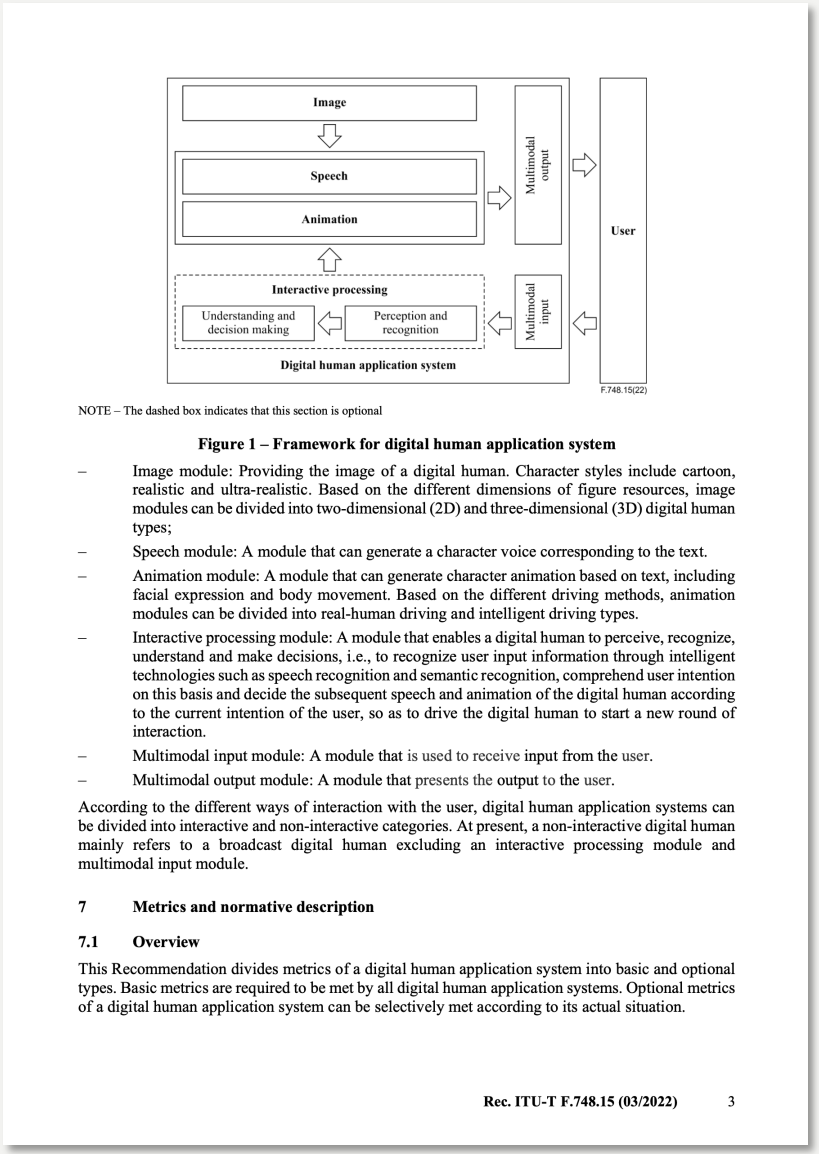
- [Q5] **Reference architecture of Large Model enhanced digital human system**
- [Q5] **Metrics and Evaluation Methods for the Presentation and Interactive Effect of 3D Digital Human**
- [Q5] **Framework and requirements of digital human access interfaces**
- [Q5] **Requirements and evaluation methods of digital human platform**
- [Q21] **Overview of digital human multimedia platform**
- [Q21] **Requirements and architectures of multimedia platform for digital human services using edge cloud**
- [Q24] **Framework for learnable digital humans through human machine interaction**
- [Q24] **Requirements of communication services for digital human**
- [Q24] **Factors and framework for digital human customization**



[Q5; 2022]

Framework and metrics for digital human application systems

- **Definitions**
- **Framework**
- **Metrics**
  - Image
  - Speech
  - Animation
  - Interactive processing
  - Multimodal input
  - Multimodal output



NOTE – The dashed box indicates that this section is optional

Figure 1 – Framework for digital human application system

– Image module: Providing the image of a digital human. Character styles include cartoon, realistic and ultra-realistic. Based on the different dimensions of figure resources, image modules can be divided into two-dimensional (2D) and three-dimensional (3D) digital human types;

– Speech module: A module that can generate a character voice corresponding to the text.

– Animation module: A module that can generate character animation based on text, including facial expression and body movement. Based on the different driving methods, animation modules can be divided into real-human driving and intelligent driving types.

– Interactive processing module: A module that enables a digital human to perceive, recognize, understand and make decisions, i.e., to recognize user input information through intelligent technologies such as speech recognition and semantic recognition, comprehend user intention on this basis and decide the subsequent speech and animation of the digital human according to the current intention of the user, so as to drive the digital human to start a new round of interaction.

– Multimodal input module: A module that is used to receive input from the user.

– Multimodal output module: A module that presents the output to the user.

According to the different ways of interaction with the user, digital human application systems can be divided into interactive and non-interactive categories. At present, a non-interactive digital human mainly refers to a broadcast digital human excluding an interactive processing module and multimodal input module.

7 Metrics and normative description

7.1 Overview

This Recommendation divides metrics of a digital human application system into basic and optional types. Basic metrics are required to be met by all digital human application systems. Optional metrics of a digital human application system can be selectively met according to its actual situation.

[Q5; 2022]

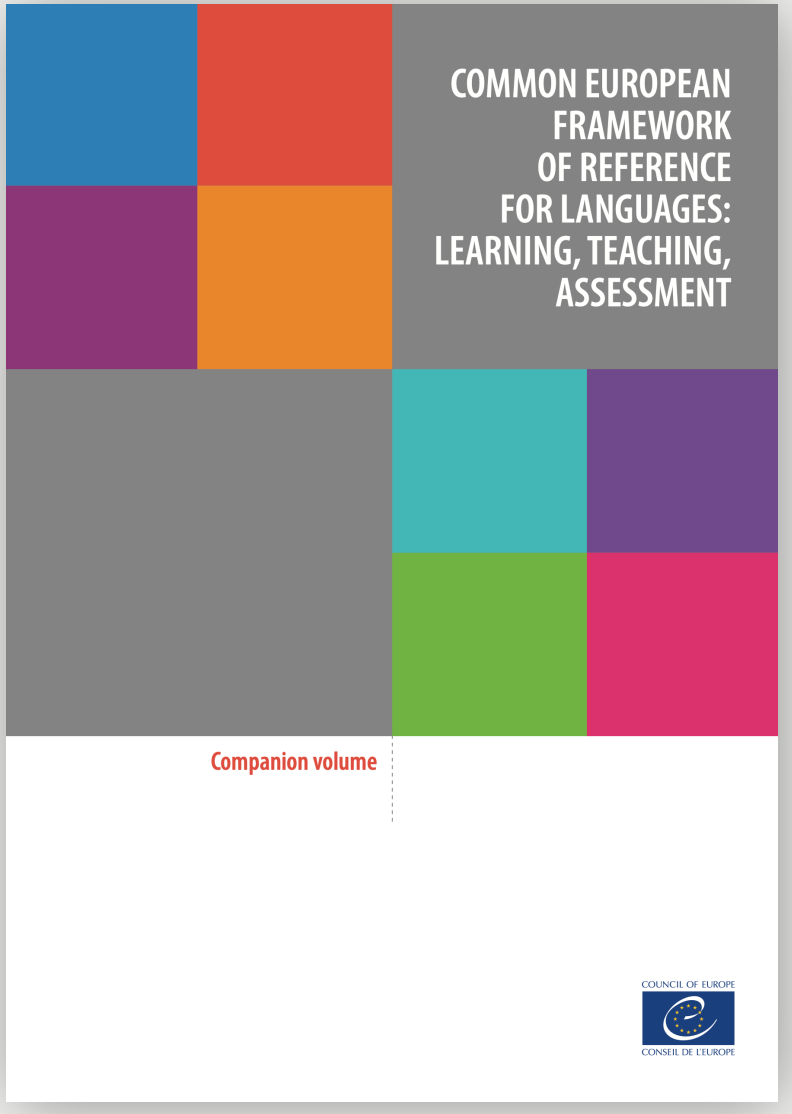
Framework and metrics for digital human application systems

- **Definitions**
- **Framework**
- **Metrics**
  - Image
  - Speech
  - Animation
  - **Interactive processing**
  - Multimodal input
  - Multimodal output

Need further discussions

Common European Framework of Reference for Languages

Council of Europe, 2001



A comprehensive assessment standard that views humans as socially situated beings (**Social Agents**) and organizes how they can **skillfully utilize fundamental abilities** to achieve specific **communication tasks**.

Socially Situated Conversational Tasks

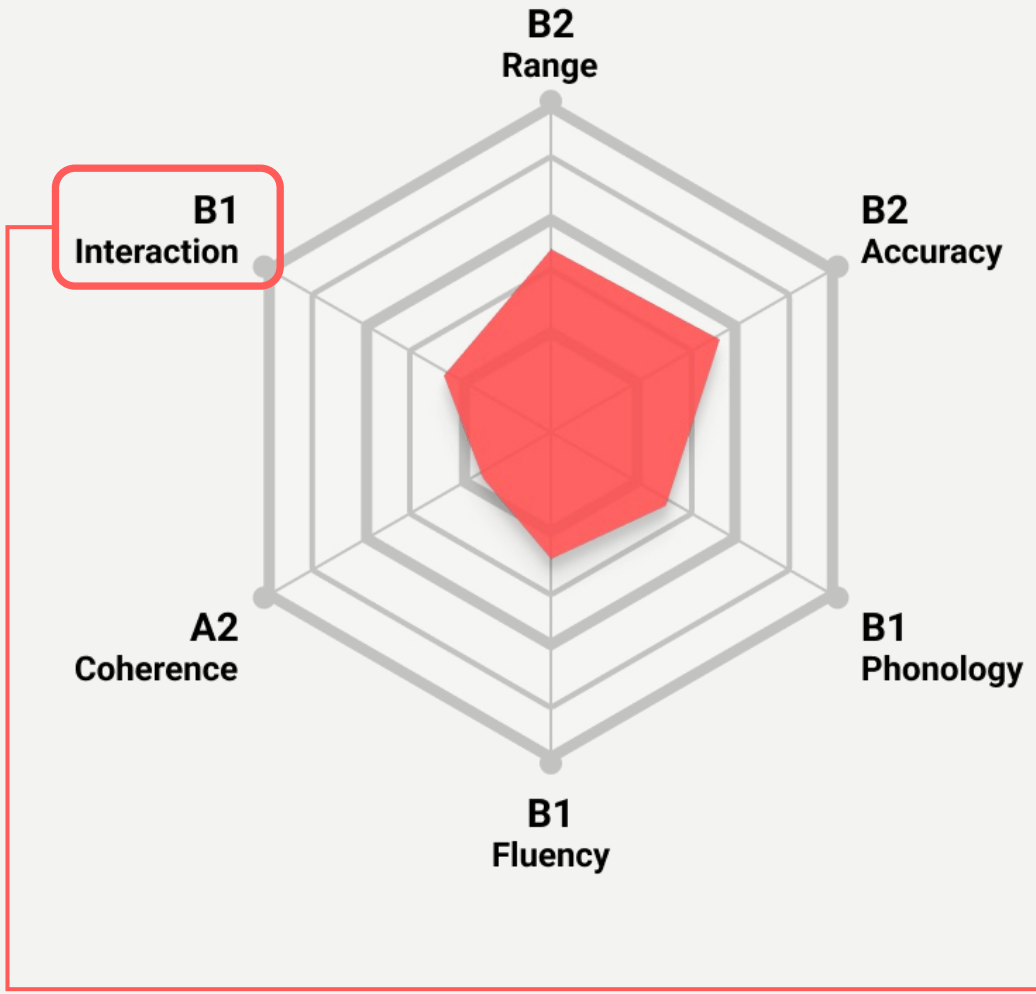
CEFR | Standardized Measurement of Language Use

EQUMENOPOLIS

	Proficiency Level	Descriptions
Proficient User	C2	**Mastery** : Fluent in any situation
	C1	**Advanced** : Effective and flexible use
Independent User	B2	**Upper-Intermediate** : Fluent on familiar topics
	B1	**Intermediate** : Handles daily interactions
Basic User	A2	**Pre-Intermediate** : Manages simple tasks
	A1	**Beginner** : Uses basic phrases

CEFR | Standardized Measurement of Language Use

EQUMENOPOLIS



Range	Ability to use a wide range of expressions and complex sentences
Accuracy	Ability to construct understandable and appropriate sentences
Phonology	Ability to speak clearly and audibly using articulation, intonation, and rhythm
Fluency	Ability to continue speaking at an appropriate speed, naturally and without stagnation
Coherence	Ability to connect and talk about a variety of information, providing specific examples and reasons
Interaction	**Ability to appropriately answer listeners' questions and collaboratively develop a conversation**

Interaction Competencies

EQUMENOPOLIS

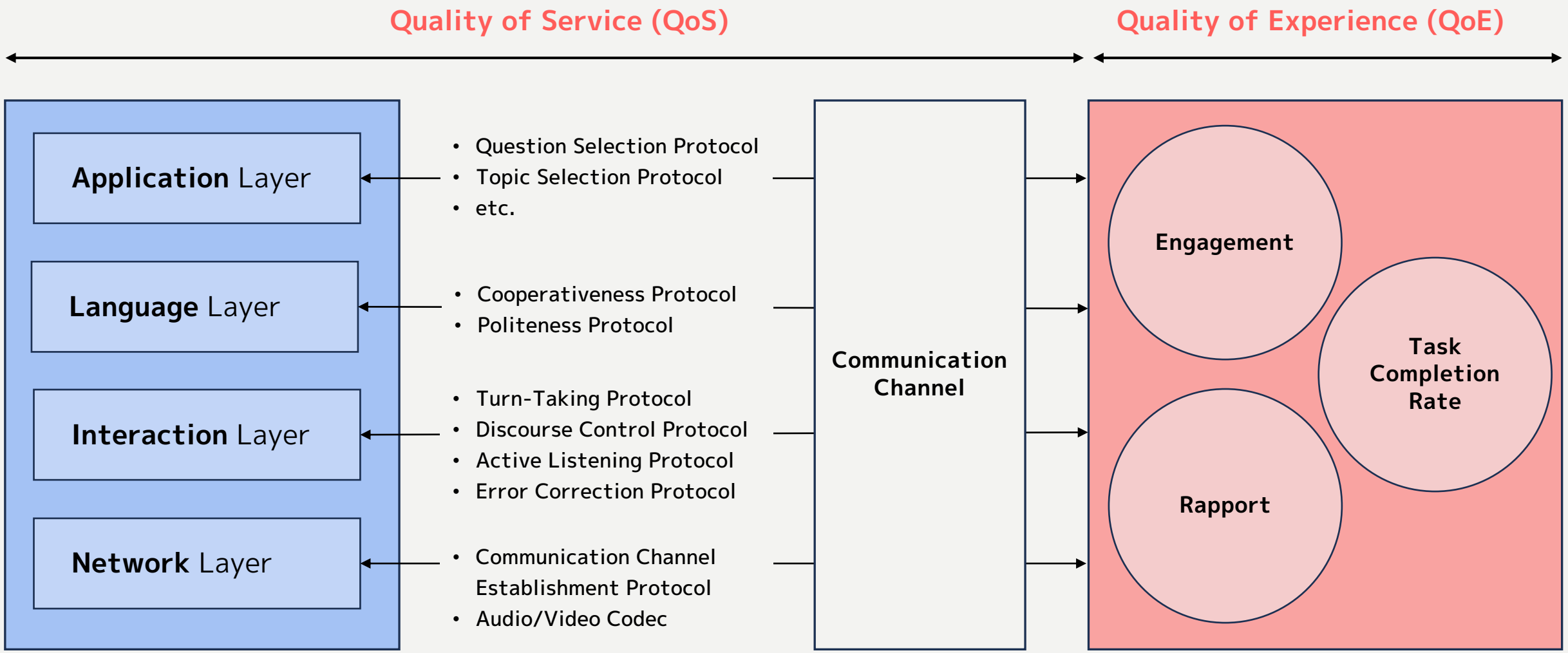
Our CEFR-based IC scale based on [Galaczi & Taylor 2018]

Turn Management	The ability to keep the conversation cooperative and collaborative through **appropriate turn-taking**.
Topic Management	The ability to **develop ideas collaboratively** towards the expected outcome of the interaction.
Interactive Listering	The ability to encourage the current speaker to keep their turn and continue talking by using **reactive actions** (e.g., backchanneling, acknowledgement).
Breakdown Repair	The ability to **repair breakdowns** in interaction proactively and preemptively.

[Galaczi & Taylor 2018] Evelina Galaczi, and Lynda Taylor, **Interactional competence: Conceptualisations, operationalisations, and outstanding questions**, Language Assessment Quarterly 15, no. 3 (2018): 219-236.

Evaluation Model | Conversation as Communication

EQUMENOPOLIS



Acknowledgement

Standardization Advisors

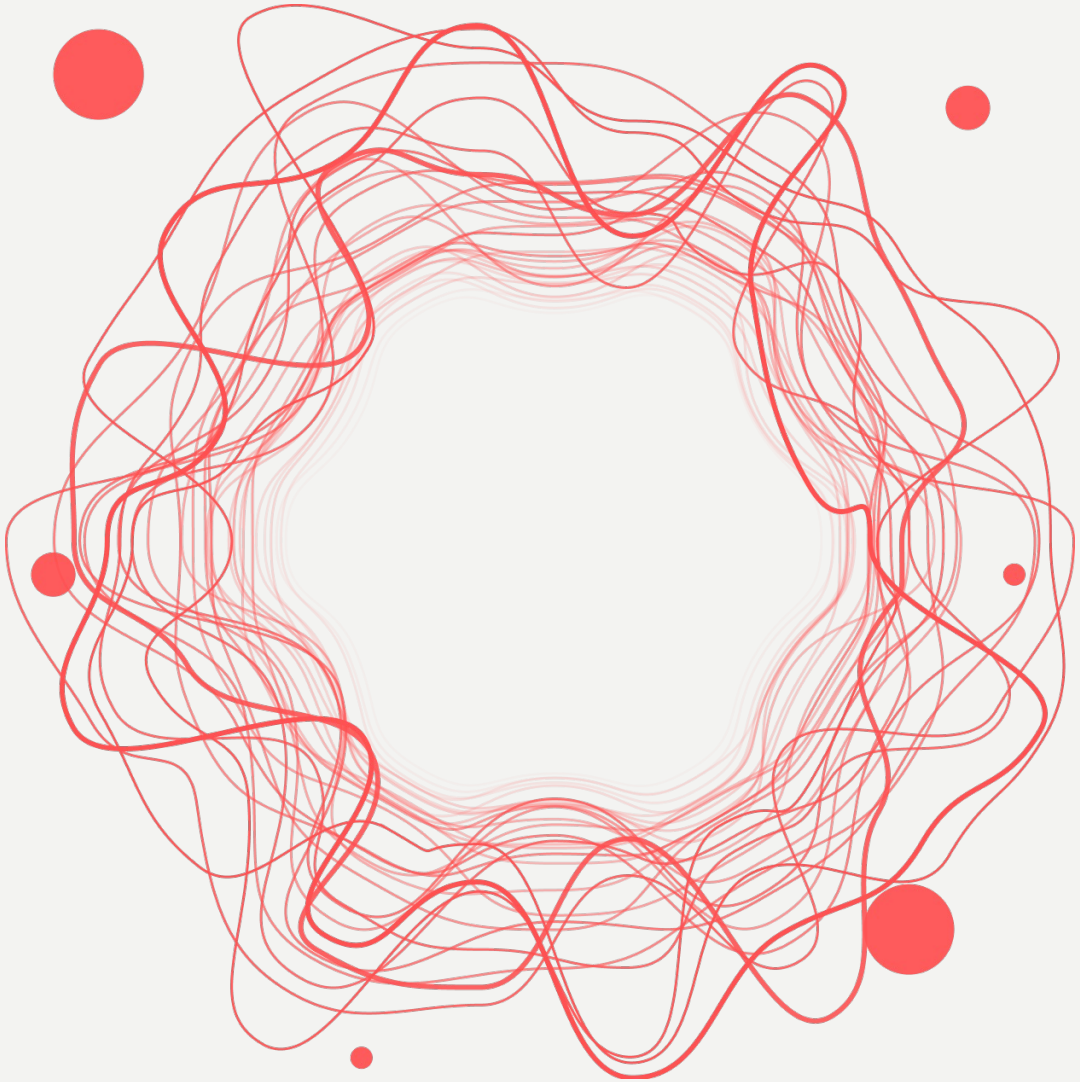
- Hideo IMANAKA (NICT)
- Yoshiaki SHIBATA (metaFrontier.jp, LLC)

Research Funding

- NICT Beyond 5G

  XR Communication Infrastructure for Realizing High-Immersion Interaction Experiences with Conversational AI Agents

  JPJ012368C06301



Yoichi Matsuyama, Ph.D

CEO, **Equmenopolis, Inc.**

Associate Research Professor, **Waseda University**

yoichim@equ.ai