

Trustworthy Generative AI from the perspective of industry

Shanshan YU

CognitiveGraph Core Project

Fujitsu Research

Fujitsu Limited



What kind of company is Fujitsu?



- A leading Japanese ICT company, offering a full range of technology products, solutions & services around the world.
- We use our experience and the power of ICT to shape the future of society with our customers.

The Supercomputer Fugaku FUJITSU

#1 for 9 consecutive terms in HPCG and Graph500

#1 Machine Learning Benchmark MLPerf HPC

#1

- HPCG
- Graph500
- TOP500 (2020.06-2021.11)
- HPL-AI (2020.06-2021.11)

High performance & efficiency

- A64FX (ARM)
- 5 PB memory
- 158,976 nodes
- 442 PFLOPS (HPL)
- 30M W

The supercomputer Fugaku has been jointly developed by RIKEN and Fujitsu



Release of "Fugaku-LLM" – a large language model trained on the supercomputer "Fugaku"

Enhanced Japanese language ability, for use in research and business

2024/05/10 Fugaku-LLM Press Conference

Tokyo Institute of Technology
Tohoku University
Fujitsu Limited
RIKEN National Research and Development Institute
National University Tokai National University Organization Nagoya University
Cyber Agent Co., Ltd
Kotoba Technologies Inc.

Who am I?

- Researcher on human centric technologies

Visibility
Enhancement

Human
Sensing

Affective
Computing

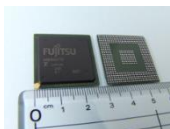
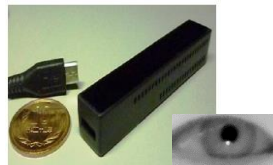


Image sensor



Car navigation



EyeExpert
(Gaze detection sensor)



Psychological
estimation

AI Trust

- Senior Research Manager of Cognitive Graph Core Project, Fujitsu
- Global research activities and collaboration to make AI trustworthy

Fujitsu Neuro-Symbolic Explainer

Extract validated explanations for image classifications made by AI

[link](#)

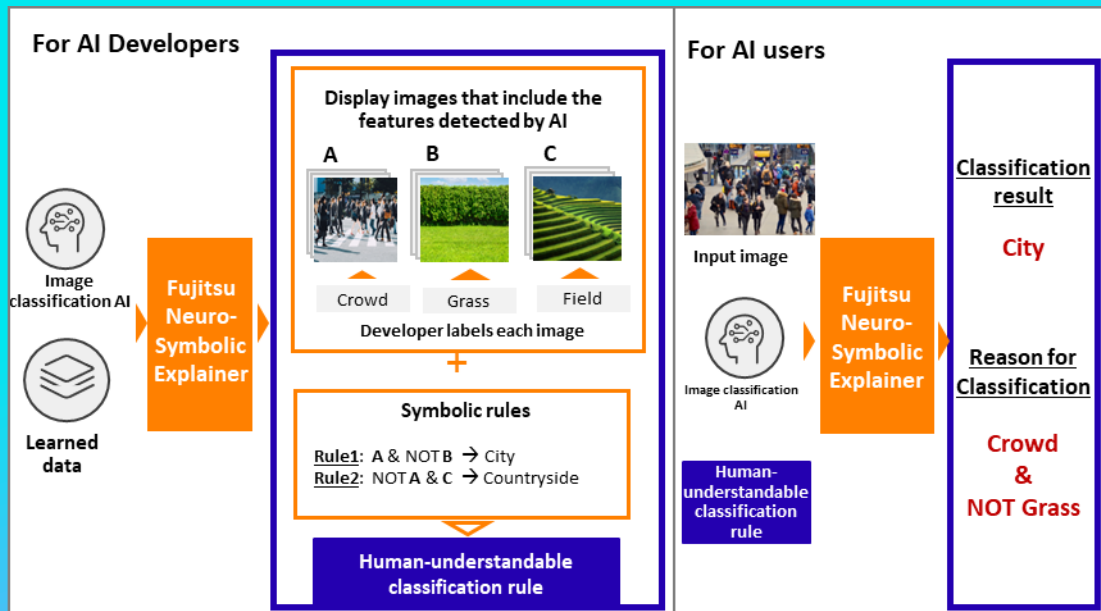
Values of the technology

【For AI Developers】

Generate reasons for image classification models

【For AI users】

Provides image classification results and reasons

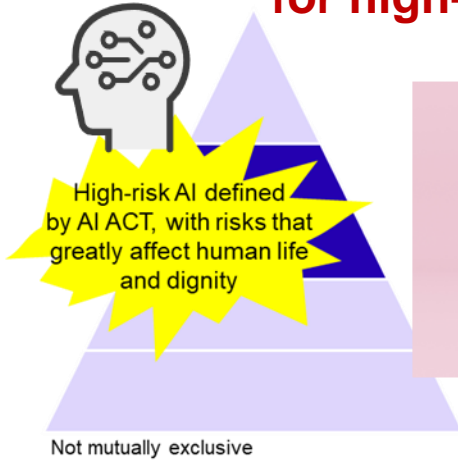


“Neuro-Symbolic” is an approach to AI that combines the learning capabilities of neural networks with the explainability properties of symbolic AI (rule-based AI)

Examples of use cases for medical AI

For medical AI developers:

Be ready for regulation/audit for high-risk AI



Utilize the explanation as the basis of the medical AI diagnosis, to **meet the requirements for regulation/audit for medical AI**

For medical AI users(doctors)

Be ready for the doctors to cooperate with medical AI



Utilize the explanation to provide accountability to patients, which leads to **better medical service**

What happened in the past year?

[Asia Pacific Research Integrity \(APRI\)](#)
[2023 Tokyo, Japan](#)
21st, March, 2023

[GPT-4 released by OpenAI](#)
14th, March, 2023

Video (Language: English)

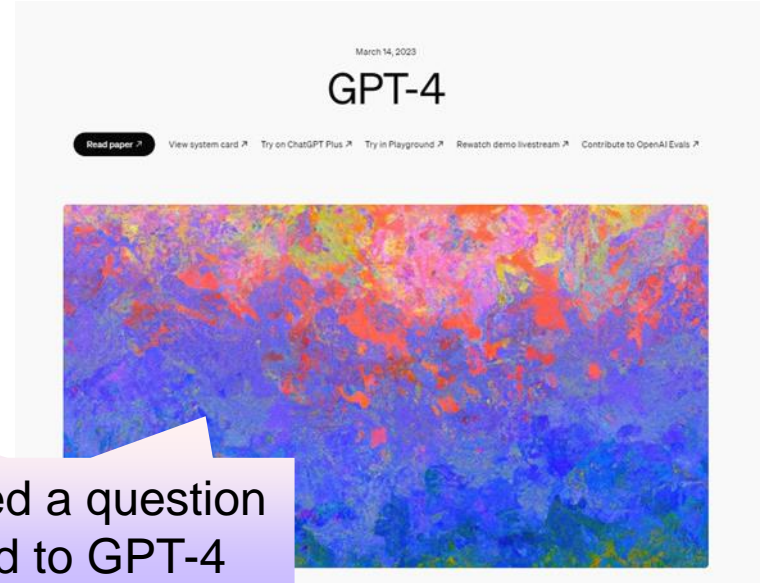


AI Ethics for Better Coexistence of
Society and AI
-From principles to practice-

Shanshan YU
Research Center for AI Ethics
Research Unit, Fujitsu Research
Fujitsu Limited



Received a question
related to GPT-4



Who could anticipate the evolution of Generative AI in the past year?

The acceleration of the evolution of generative AI



GPT-4

14th, March 2023



Sora

25th, Feb 2024




GPT-4o

13th, May 2024



Meta Llama 3

18th, April 2024



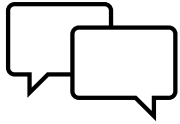
Gemini 1.5 Flash

15th, May 2024

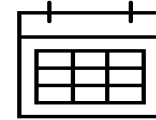
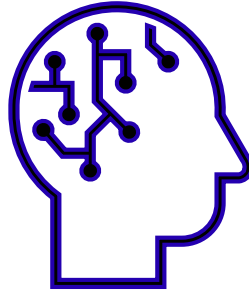
What will be the next generative AI model?

What is Generative AI?

An artificial intelligence technology that uses machine learning algorithms to **generate contents**



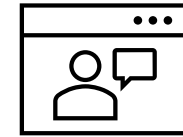
Interactive communication



Others(schedule, etc.)



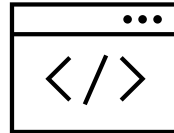
Summarization



Video



Music



Program



Image

What do people expect from generative AI?

Innovations: Driving advancements in AI technology and applications

Performance

Creativity

Novel and imaginative content

Fidelity

High-fidelity outputs with realistic and coherent content

Usability

Interactivity/
Versatility/
Customization

Trust

Insurance in the aspects of Quality/Security/Ethics

What is important from the perspective of industry? FUJITSU

Innovations: Driving advancements in AI technology and applications

- To support people to fulfill their specific purposes by using Generative AI technology
- To improve the performance and trustworthiness of Generative AI technology

Performance

Creativity

Fidelity

Usability

Novel and imaginative content

High-quality outputs with realistic and coherent content

Interactivity/
Versatility/
Customization

Trust

Quality/Security/Ethical insurances

Example | Trial to promote creativity/fidelity

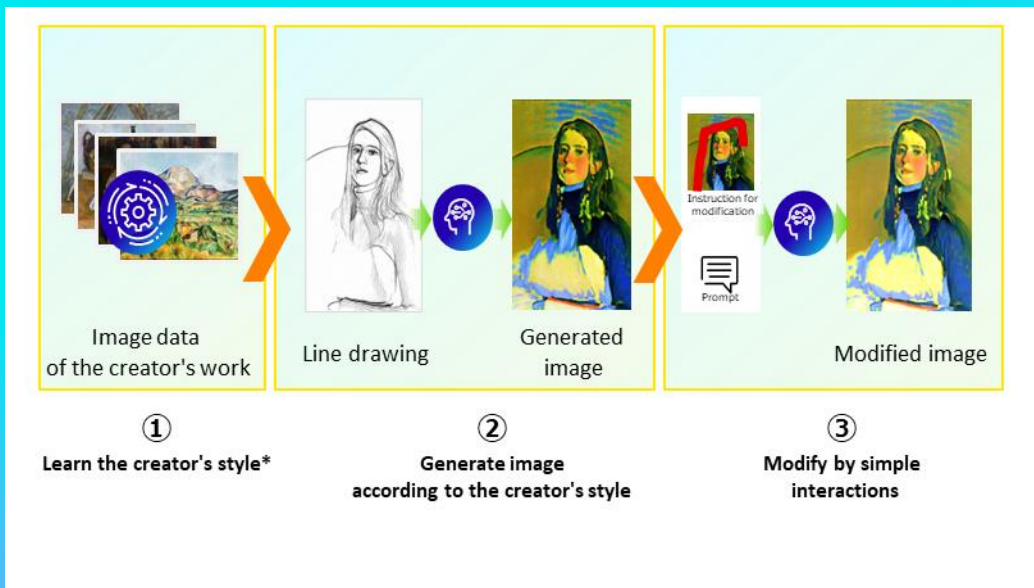
Synthetic Image Generation

[link](#)

Enabling users to generate high-quality synthetic image data with a desired style

Generate images with a specific style

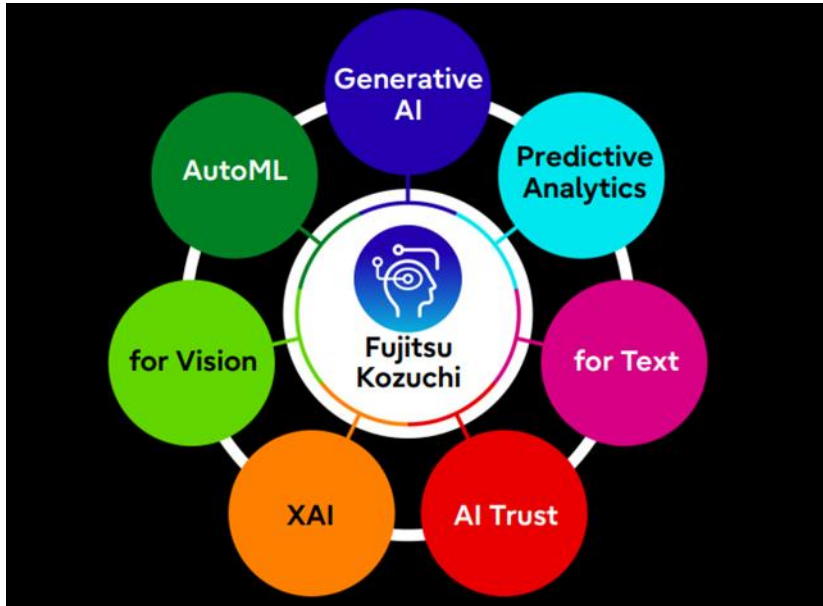
- Easily create AI models by learning a user's images data such as samples of the creator's work, etc.
- Apply the model to an original drawing (line drawing or image) to generate a desired image with the creator's unique choice of brush strokes and colors
- Adjust the color of the image by modifying the generated image through simple interactions.



Example | Trial for the improvement of usability

Provide a cloud-based AI platform, that enhances the productivity and creativity of user business and supports sustainability transformation

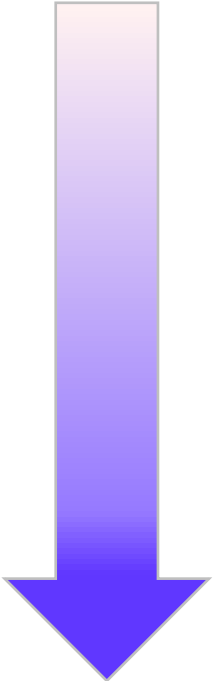
Fujitsu Kozuchi [link](#)



Accelerate the loop from research to industry

- Makes the research easier to be used
- Provides a platform and consulting services that allow engineers to create services on their own
- Enables customers to use AI safely and securely

What is especially important in the perspective of research integrity in industry?



AI up to 2024
Performance is a differentiating factor

© Only trusted AI can sustain business

AI after 2025
"Performance" and "Trust"
are the differentiating factors



[GPT-4o \(Omni\) Human interaction demo w/ Sal Khan](#)

Trust is especially important for Generative AI

Example | Specific performance & trust issue of generative AI

Hallucination: generative AI 'lies'

A phenomenon in which generative AI creates incorrect or unrelated output

An example *"The Dawn of LMMs: Preliminary Explorations with GPT-4V(ision)", 2023-09-29*

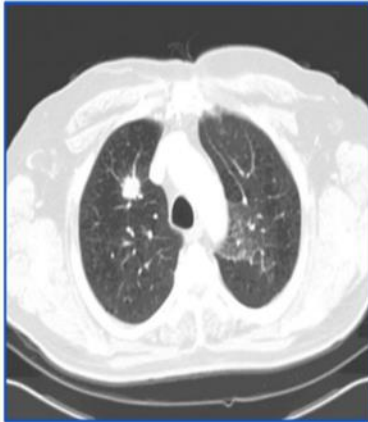


Input

A prompt with an X-ray image

Prompt:

Write a radiology report for this image as an expert radiologist..



Output

Answer generated by GPT-4V.

GPT-4V:

Report:

Date: 6/25/2021

Study Type: Chest CT

Subject: 42-year-old male

Findings: The chest CT images demonstrate a nodular opacity in the left upper lobe, measuring approximately 1.3 cm in diameter. The opacity appears to have spiculated margins and is located adjacent to the pleura. There is no evidence of mediastinal or hilar lymphadenopathy.

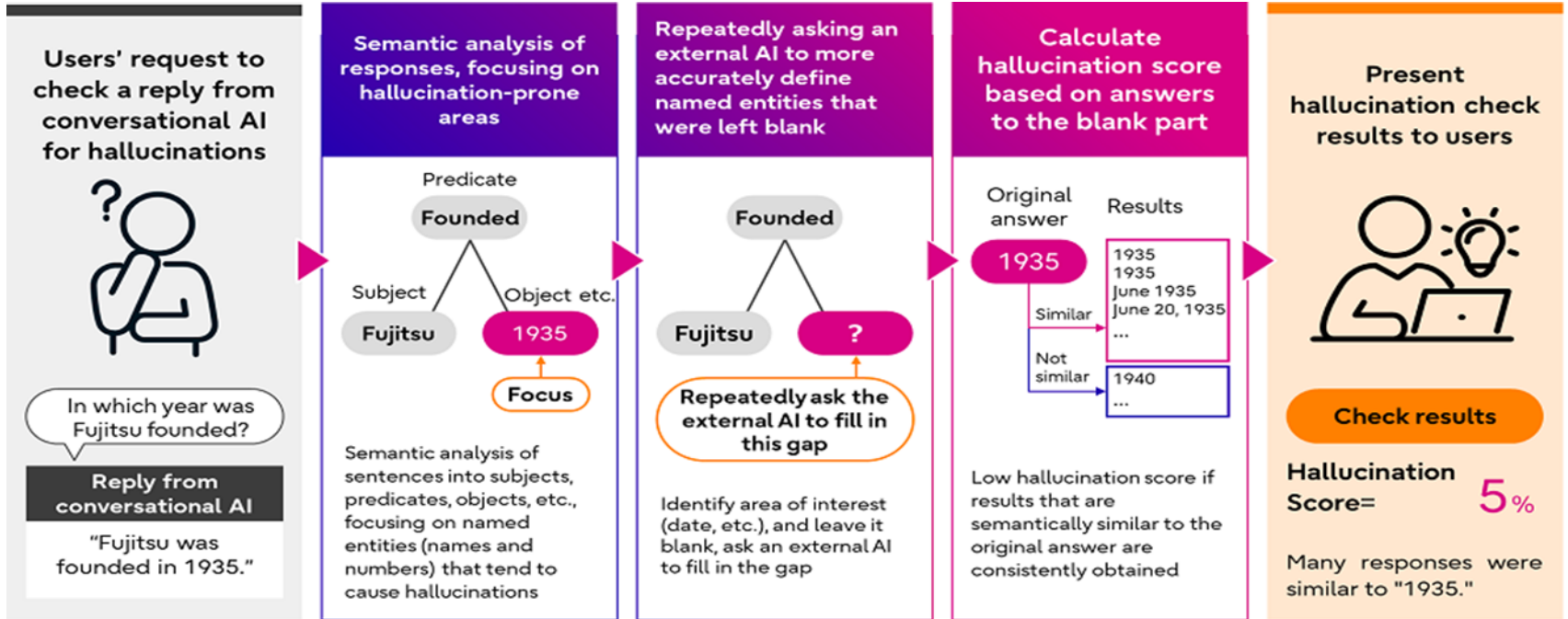
Impression: The findings are concerning for a primary lung malignancy. Further evaluation with a biopsy is recommended to confirm the diagnosis.

Hallucination determined by expert

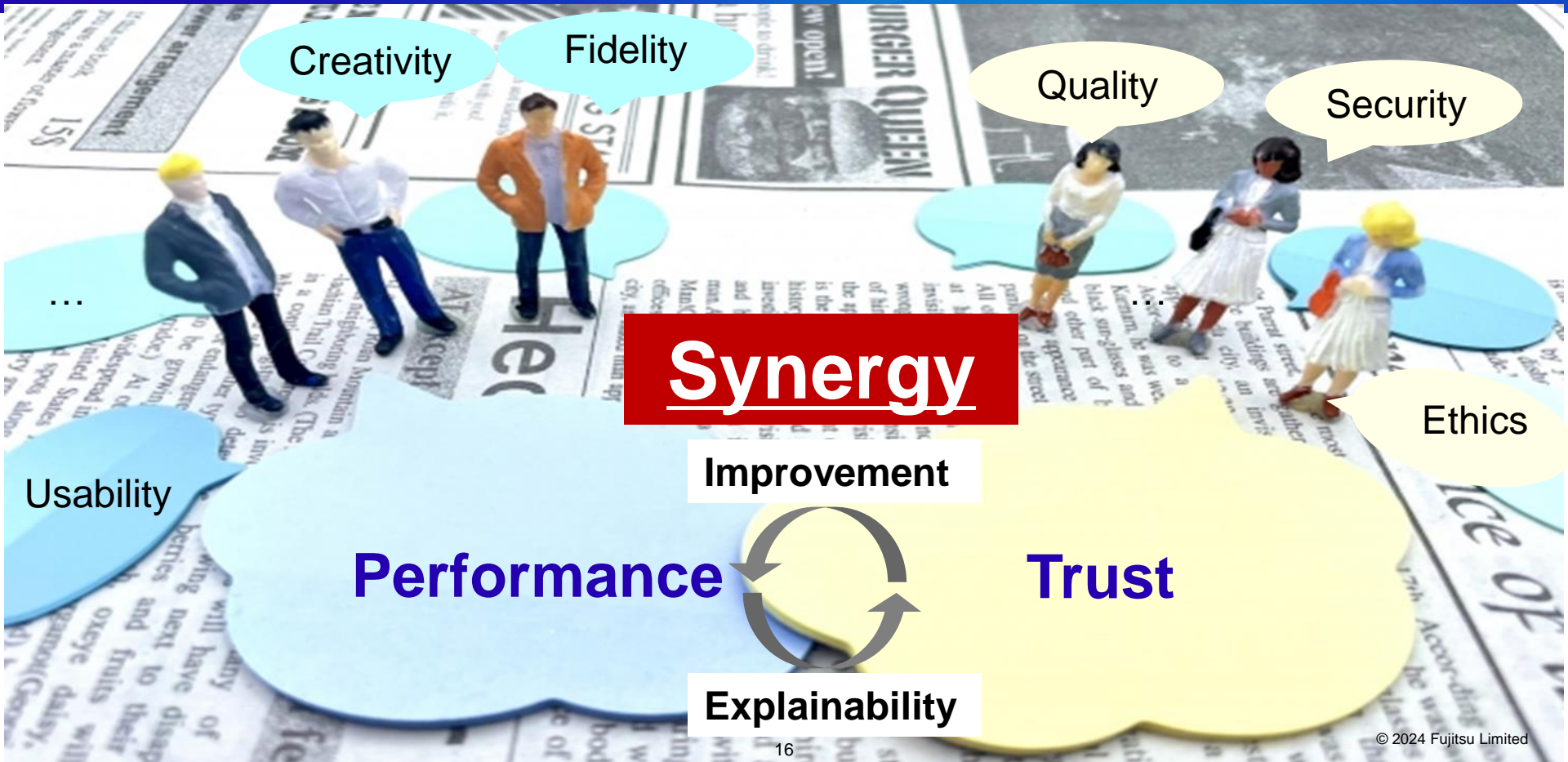
Calculate the Hallucination score

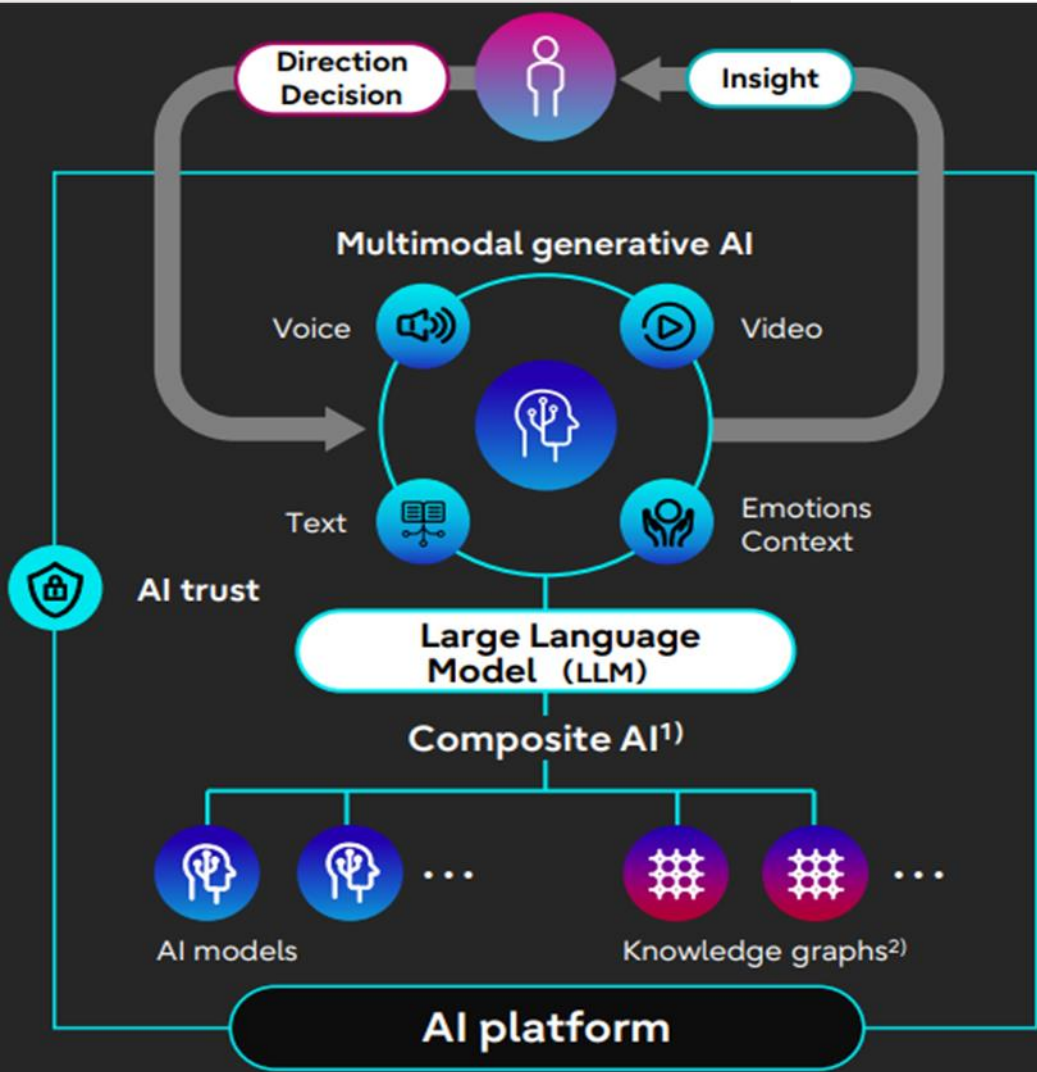
[link](#)

by semantic analysis of continuity of multiple responses of generative AI



What is the next for trusted generative AI





An ecosystem for the synergy of performance & trust for generative AI

Natural communication of AI and Trust to promote people-AI collaboration

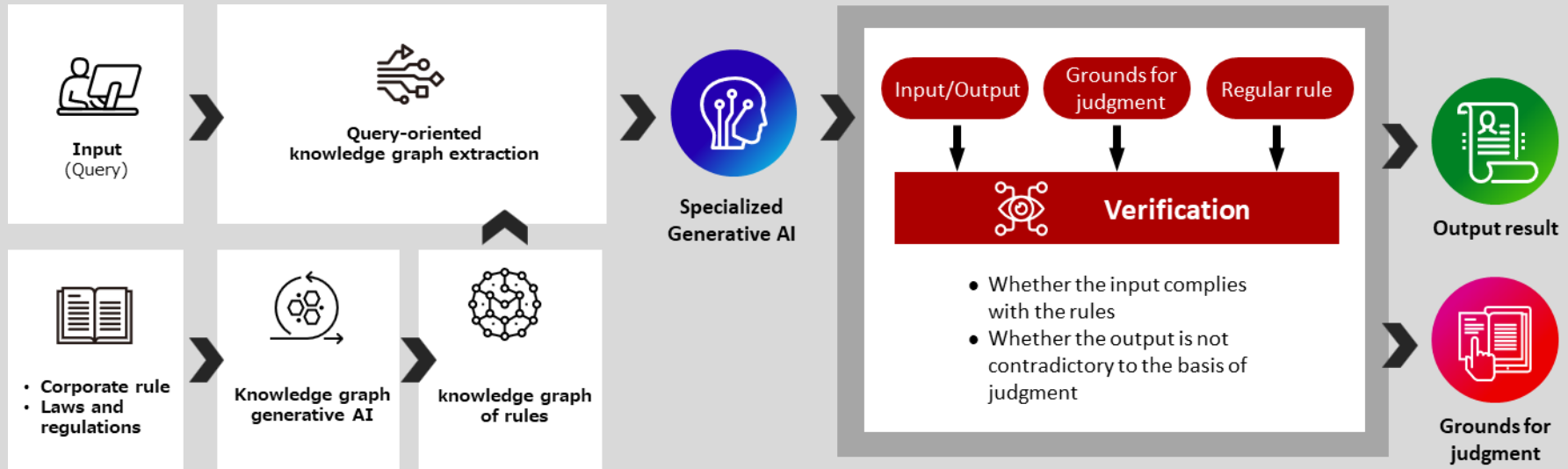
Utilize the information of multiple AI models and knowledge graphs

- Higher performance
- Explanation for the output of generative AI to improve trust issues

※The example is extracted from Fujitsu Technology and Service Vision 2024

Control the behavior of generative AI with a knowledge graph, complying with corporate rules and laws.

- Utilize a knowledge graph that corresponds to laws and corporate rules **to verify compliance with input rules.**
- By analyzing the basis on which the generative AI derived its output, we provide explanatory power for the basis of judgment and determine hallucination.



- **Background:** Rapid evolution of Generative AI
- **Industry perspective:**
 - Issues of Generative AI
 - Fujitsu examples of industry trials
 - 2 examples to support people to fulfill their specific purposes by using Generative AI technology to promote creativity/fidelity/usability
 - 1 example to improve the performance and trustworthiness for Generative AI technology
- **Next for trust generative AI**
 - Ecosystem for the synergy of performance and trust

**Creating trustworthiness and robust Generative AI technology
for a better society!**

Thank you

