



Visual General Intelligence: Vision Research Toward the AGI Era

Hirokatsu Kataoka

AIST / University of Oxford

VGI Organizing Team

<http://www.hirokatsukataoka.net/>

Definition of Intelligence & AGI (so far)

Intelligence:

“the ability to learn, understand and think in a logical way about things” – Oxford Dictionary

“intelligence measures an agent’s ability to achieve goals in a wide range of environments” – Legg & Hutter (2007)

AGI (Artificial General Intelligence):

“this framework introduces levels of AGI performance, generality, and autonomy” – Morris et al. (2023)

“can match or exceed the cognitive versatility and proficiency of a well-educated adult” – Hendrycks et al. (2025)

Universal Intelligence:
A Definition of Machine Intelligence

Shane Legg

IDSIA, Galleria 2, Manno-Lugano CH-6928, Switzerland
shane@vetta.org www.vetta.org/shane

Marcus Hutter

RSISE @ ANU and SML @ NICTA, Canberra, ACT, 0200, Australia
marcus@hutter1.net www.hutter1.net

December 2007

Abstract

A fundamental problem in artificial intelligence is that nobody really knows what intelligence is. The problem is especially acute when we need to consider artificial systems which are significantly different to humans. In this paper we approach this problem in the following way: We take a number of well known informal definitions of human intelligence that have been given by experts, and extract their essential features. These are then mathematically formalised to produce a general measure of intelligence for arbitrary machines. We believe that this equation formally captures the concept of machine intelligence in the broadest reasonable sense. We then show how this formal definition is related to the theory of universal optimal learning agents. Finally, we survey the many other tests and definitions of intelligence that have been proposed for machines.

Keywords

AIXI; Complexity theory; Intelligence; Theoretical foundations; Turing test; Intelligence definitions; Intelligence tests.

Position: Levels of AGI for Operationalizing Progress on the Path to AGI

Meredith Ringel Morris¹ Jochen Söhl Dickschels² Noah Finkel³ Tris Warkentin⁴ Allan Dufresne⁵
Aleksandra Faust⁶ Clement Furbert⁷ Shane Legg⁸

Abstract

We propose a framework for classifying the capabilities and behavior of Artificial General Intelligence (AGI) models and their precursors. This framework introduces levels of AGI performance, generality, and autonomy, providing a common language to compare models, assess risks, and measure progress along the path to AGI. To develop our framework, we analyze existing definitions of AGI and distill its principles that a useful ontology for AGI should satisfy. With these principles in mind, we propose “Levels of AGI” based on depth (performance) and breadth (generality) of capabilities, and reflect on how current systems fit into this ontology. We discuss the challenging requirements for future benchmarks that quantify the behavior and capabilities of AGI models against these levels. Finally, we discuss how these levels of AGI interact with deployment considerations such as autonomy and risk, and emphasize the importance of carefully selecting Human-AI interaction paradigms for responsible and safe deployment of highly capable AI systems.

1. Introduction

Artificial General Intelligence (AGI) is an important and sometimes controversial concept in computing research, used to describe an AI system that is at least as capable as a human at most tasks. Given the rapid advancement of Machine Learning (ML) models, the concept of AGI has grown from a subject of philosophical debate, to one which also has near-term practical relevance. Some experts believe that “peak” of AGI (Bostik et al., 2023) are already present in the latest generation of large language models (LLMs); some predict AI will broadly outperform humans within about a decade (Bergio et al., 2023), some even assert that current LLMs are AGI (Agrava & Arora & Norcia, 2023). The concept of AGI is important as it maps onto goals for, predictions about, and risks of AI.

Goals:

Achieving human-level “intelligence” is an implicit or explicit north-star goal for many in our field, from the 1956 Dartmouth AI Conference (McCarthy et al., 1955) that kick-started the modern field of AI, to today’s leading AI research firms, whose mission statements include goals such as “ensure transformative AI helps people and society” (Anthropic, 2023a) and “ensure that artificial general intelligence benefits all of humanity” (OpenAI, 2023).

Predictions:

The concept of AGI is related to a prediction about progress in AI, namely that it is toward greater generality, approaching and exceeding human generality. Additionally, AGI is typically intertwined with a notion of “emergent” properties (Wei et al., 2023), i.e. capabilities not explicitly anticipated by the developer. Such capabilities offer promise, perhaps including abilities that are complementary to typical human skills, enabling new types of interaction or novel industries. Such predictions about AGI’s capabilities in turn predict likely societal impacts; AGI may have significant economic implications, i.e., reaching the necessary criteria for widespread labor substitution (Ellingrud et al., 2023; Dill’Acqua et al., 2023; Eloundou et al., 2023), as well as geo-political implications relating not only to the economic advantage AGI may confer, but also to military considerations (Kosinger et al., 2022).

Risks:

Lastly, AGI is viewed by some as a concept for identifying the point when there are extreme risks (Ovchinnikov et al., 2023; Bengio et al., 2023), as some speculate that AGI systems might be able to deceive and manipulate, accumulate resources, advance goals, behave aggressively, overtake humans in broad domains, displace humans from key roles, and/or recursively self-improve.

In this position paper, we argue that it is critical for the AI research community to explicitly reflect on what we mean by “AGI,” and aspire to quantify attributes like the performance, generality, and autonomy of AI systems. Shared operationalizable definitions for these concepts will support comparisons between models; risk assessments

A Definition of AGI

Dan Hendrycks¹, Dawn Song^{2,3}, Christian Szegedy⁴, Honglak Lee^{5,6}, Yarin Gal⁷, Erik Brynjolfsson⁸, Sharon LF, Andy Zeng^{10,11}, Lionel Levine¹², Bo Han¹³, Ji Fu¹⁴, Zhiwei Liu¹⁵, Hanwei Shiu¹⁶, Kimin Lee¹⁶, Mantas Malchaire¹⁷, Long Phan¹, George Impeybertsen¹, Adam Khoja¹, Cihang Xie¹⁷, Olawale Salau¹⁸, Matthias Hein¹⁹, Kevin Zhao²⁰, Alexander Pan²¹, David Devenaud^{21,22}, Bo Li^{2,3}, Steve Omohundro²³, Gabriel Alfour²⁴, Max Tegmark¹⁸, Kevin McGrew²⁵, Gary Marcus²⁶, Jaan Tallinn²⁷, Eric Schmidt¹⁸, Yoshua Bengio^{28,29}

¹Center for AI Safety ²University of California, Berkeley ³Virtue AI
⁴Morph Labs ⁵University of Michigan ⁶LG AI Research
⁷University of Oxford ⁸Stanford University ⁹University of Wisconsin-Madison
¹⁰Gray Swan AI ¹¹Carnegie Mellon University ¹²Cornell University
¹³Hong Kong Baptist University ¹⁴HKUST ¹⁵Nanyang Technological University
¹⁶KASIT ¹⁷University of California, Santa Cruz ¹⁸Massachusetts Institute of Technology
¹⁹University of Tübingen ²⁰University of Washington ²¹University of Toronto
²²Vector Institute ²³University of Illinois Urbana-Champaign ²⁴Beneficial AI Research
²⁵Conjecture ²⁶Institute for Applied Psychometrics ²⁷New York University
²⁸CISER ²⁹Université de Montréal ³⁰LawZero

Abstract

The lack of a concrete definition for Artificial General Intelligence (AGI) obscures the gap between today’s specialized AI and human-level cognition. This paper introduces a quantifiable framework to address this, defining AGI as matching the cognitive versatility and proficiency of a well-educated adult. To operationalize this, we ground our methodology in Cartell-Horn-Cardell theory, the most empirically validated model of human cognition. The framework dissects general intelligence into ten core cognitive domains—including reasoning, memory, and perception—and adapts established human psychometric batteries to evaluate AI systems. Application of this framework reveals a highly “jagged” cognitive profile in contemporary models. While proficient in knowledge-intensive domains, current AI systems have critical deficits in foundational cognitive machinery, particularly long-term memory storage. The resulting AGI scores (e.g., GPT-4 at 27%, GPT-5 at 57%) concretely quantify both rapid progress and the substantial gap remaining before AGI.

[1] Legg & Hutter, Universal Intelligence: A Definition of Machine Intelligence, Minds and Machines, 2007.

[2] Morris et al., Levels of AGI for Operationalizing Progress on the Path to AGI, ICML, 2024.

[3] Hendrycks et al., A Definition of AGI, 2025.

CVPR 2026 VGI Workshop

2

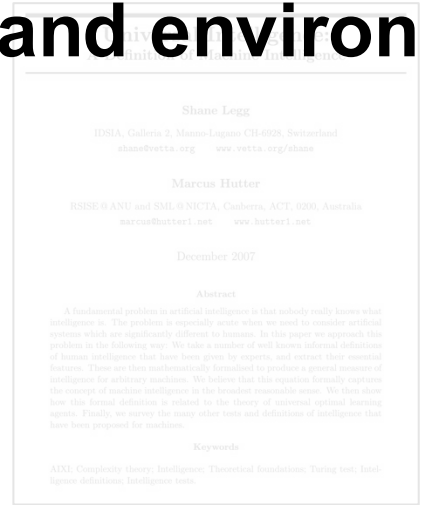
Definition of AGI (summary)

Intelligence:

“the ability to learn, understand and think in a logical way about things” – Oxford Dictionary

“intelligence measures an agent’s ability to achieve goals in a wide range of environments” – Legg & Hutter (2007)

AGI is human-level intelligence that can adapt and achieve goals across a wide range of tasks and environments



- [1] Legg & Hutter, Universal Intelligence: A Definition of Machine Intelligence, Minds and Machines, 2007.
- [2] Morris et al., Levels of AGI for Operationalizing Progress on the Path to AGI, ICML, 2024.
- [3] Hendrycks et al., A Definition of AGI, 2025.

One of the most successful moments in NLP



Next-token prediction (SSL) at scale

Model: 175B parameters & Data: 300B tokens

GPT-3 $\leftarrow$ $\times 100$ from GPT-2 $\leftarrow$ $\times 10$ from GPT-1

■ Emergent abilities

Human-level text generation

In-context learning

Emergent reasoning abilities

**If intelligence can emerge from language,
can intelligence also emerge from vision?**

Visual & Language Intelligence

Primitive visual function	700M years
Eyes	540M years
Mammalian intelligence	200M years
Humans	300k years
Advanced language	100k years
Writing systems	5k years
Old English	1.5k years
Modern English	Hundreds years

Think about the history of Earth

Perspective of vision (540M years ago)

Visual perception emerged far earlier than language

Perspective of language (100k years ago)

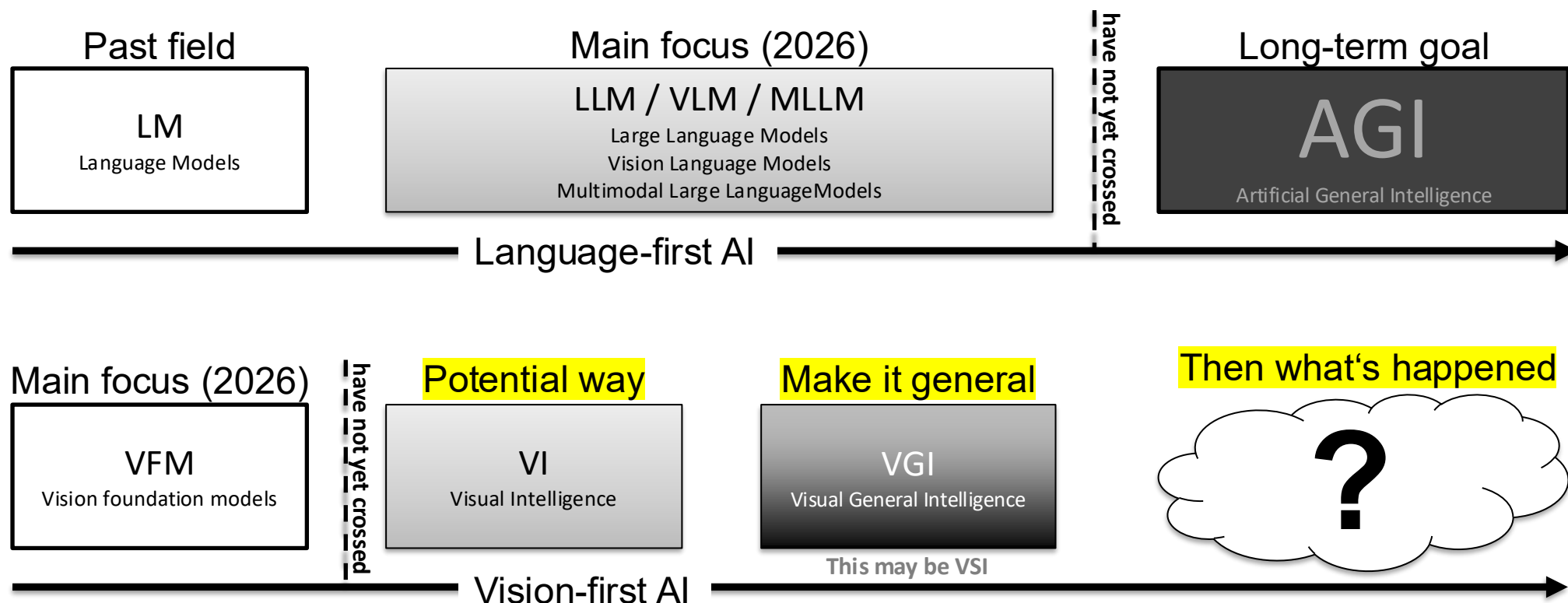
Advanced language emerged relatively recently

Vision came first

Intelligence evolved together with perception, long before symbolic language existed

Visual and/or Language Intelligence as a Pathway to AGI

Possible future from the vision community



We, as CVPR community, should push forward Visual Intelligence

Visual Intelligence (VI)?

Intelligence that emerges from visual experience

Visual Intelligence,

1) it's like a 'Visual GPT-3'

which has emergent abilities to solve general visual tasks

2) it's a 'multitask & multimodal VFM'

but it has capabilities that were not explicitly trained also emerge naturally

3) it's an 'ability to understand the diverse real world'

to understand, predict, compress, & generalize the dynamics of the world through visual experience

Hypotheses:

- **Visual intelligence may...**
 - **emerge before language**
 - **emerge from generative, autoregressive, and reconstruction tasks**
 - **require efficient learning geometry, dynamics, and hierarchy (e.g., synthetic data / high-efficient data collection)**

HELD AS PART OF

CVPR
JUNE 3-7, 2026



DENVER
COLORADO

Visual General Intelligence

-Vision Research Toward the AGI Era-

CVPR 2026 Workshop

📅 June 3, 2026, 8:30 AM - 4:50 PM 📍 Colorado Convention Center - Room 703
<https://cvpr2026-vgi-workshop.limitlab.xyz/>

That's why we organized this workshop to start a discussion with the CVPR community

Robert Geirhos

Google DeepMind



View Profile

Aditi Raghunathan

Carnegie Mellon University



View Profile

Matt Deitke

Meta Superintelligence Lab



View Profile

Kristen Grauman

University of Texas, Austin / Meta



View Profile

Yuki M. Asano

University of Technology Nuremberg



View Profile

Andrea Vedaldi

University of Oxford / Meta



View Profile

Alexei A. Efros

University of California, Berkeley



View Profile

Alex Kendall

Wayve



View Profile

Workshop Program

Time	Session	Presenter
08:30 - 08:40	Organizer Talk & Opening Remarks <i>Visual General Intelligence: Vision Research Toward the AGI Era</i>	Hirokatsu Kataoka
08:40 - 09:20	Invited Talk 1 <i>Are generative video models the path towards solving visual intelligence?</i>	Robert Geirhos
09:20 - 10:00	Invited Talk 2	Aditi Raghunathan
10:00 - 10:40	Invited Talk 3	Matt Deitke
10:40 - 11:00	Short Break	
11:00 - 11:40	Invited Talk 4	Kristen Grauman
11:40 - 12:20	Invited Talk 5 <i>Learning from moving, Generalising from speaking</i>	Yuki M. Asano
12:20 - 13:30	Lunch break / poster session (invite-only)	
13:30 - 14:10	Invited Talk 6 <i>Building a 3D Foundation for Spatial AI</i>	Andrea Vedaldi
14:10 - 14:50	Invited Talk 7	Alexei A. Efros
14:50 - 15:10	Coffee break	
15:10 - 15:50	Invited Talk 8 <i>Frontier challenges to bring general purpose driving AI to 10M cars</i>	Alex Kendall
15:50 - 16:00	Closing remarks	
16:00 - 16:50	Discussion time	

Invited posters:

Board 1-left

Theory of Space: Can Foundation Models Construct Spatial Beliefs through Active Exploration?

Pingyue Zhang, Zihan Huang, Yue Wang, Jieyu Zhang, Letian Xue, Zihan Wang, Qineng Wang, Keshigeyan Chandrasegaran, Ruohan Zhang, Yejin Choi, Ranjay Krishna, Jiajun Wu, Li Fei-Fei, Manling Li

 Paper

Board 1-right

Unique Lives, Shared World: Learning from Single-Life Videos

Tengda Han, Sayna Ebrahimi, Dilara Gokay, Li Yang Ku, Maks Ovsjanikov, Iva Babukova, Daniel Zoran, Viorica Patraucean, Joao Carreira, Andrew Zisserman, Dima Damen

 Paper

Board 2-left

Learning to See Through a Baby's Eyes: Early Visual Diets Enable Robust Visual Intelligence in Humans and Machines

Yusen Cai, Qing Lin, Bhargava Satya Nunna, Mengmi Zhang

 Paper

Board 2-right

From None to All: Self-Supervised 3D Reconstruction via Novel View Synthesis

Ranran Huang, Weixun Luo, Ye Mao, Krystian Mikolajczyk

 Paper

Board 3-left

Efficiently Reconstructing Dynamic Scenes One D4RT at a Time

Chuhan Zhang, Guillaume Le Moing, Skanda Koppula, Ignacio Rocco, Liliane Momeni, Junyu Xie, Shuyang Sun, Rahul Sukthankar, Joëlle K. Barral, Raia Hadsell, Zoubin Ghahramani, Andrew Zisserman, Junlin Zhang, Mehdi S. M. Sajjadi

 Paper

Board 3-right

Seeing without Pixels: Perception from Camera Trajectories

Zihui Xue, Kristen Grauman, Dima Damen, Andrew Zisserman, Tengda Han

 Paper

Invited posters:

Board 4-left

Next-Embedding Prediction Makes Strong Vision Learners

Sihan Xu, Ziqiao Ma, Wenhao Chai, Xuweiyi Chen, Weiyang Jin, Joyce Chai, Saining Xie, Stella X. Yu

[🔗](#) Paper

Board 4-right

4D-RGPT: Toward Region-level 4D Understanding via Perceptual Distillation

Chiao-An Yang, Ryo Hachiuma, Sifei Liu, Subhashree Radhakrishnan, Raymond A. Yeh, Yu-Chiang Frank Wang, Min-Hung Chen

[🔗](#) Paper

Board 5-left

Steerable Visual Representations

Jona Ruthardt, Manu Gaur, Deva Ramanan, Makarand Tapaswi, Yuki M. Asano

[🔗](#) Paper

Board 5-right

VGGT- Ω

Jianyuan Wang, Minghao Chen, Shangzhan Zhang, Nikita Karaev, Johannes Schönberger, Patrick Labatut, Piotr Bojanowski, David Novotny, Andrea Vedaldi, Christian Rupprecht

[🔗](#) Paper

Board 6-left

A Frame is Worth One Token: Efficient Generative World Modeling with Delta Tokens

Tommie Keressies, Gabriele Berton, Ju He, Qihang Yu, Wufei Ma, Daan de Geus, Gijs Dubbelman, Liang-Chieh Chen

[🔗](#) Paper

Board 6-right

Visual General Intelligence: Vision Research Toward the AGI Era

Hirokatsu Kataoka et al.

Invited posters:

Board 7-left

TBD

Kohsuke Ide et al.

Board 7-right

COG: Confidence-aware Optimal Geometric Correspondence for Unsupervised Single-reference Novel Object Pose Estimation

Yuchen Che, Jingtu Wu, Hao Zheng, Asako Kanezaki

[🔗](#) Paper

Board 8-left

Towards Structured Recognition and Generation of Human-Object Interactions

Takuma Yagi, Kohei Torimi, Jyun-Ting Song, Kris Kitani

Board 8-right

BabyVLM-V2: Toward Developmentally Grounded Pretraining and Benchmarking of Vision Foundation Models

Shengao Wang, Wenqi Wang, Zecheng Wang, Max Whitton, Michael Wakeham, Arjun Chandra, Joey Huang, Pengyue Zhu, Helen Chen, David Li, Jeffrey Li, Shawn Li, Andrew Zagula, Amy Zhao, Andrew Zhu, Sayaka Nakamura, Yuki Yamamoto, Jerry Jun Yokono, Aaron Mueller, Bryan A. Plummer, Kate Saenko, Venkatesh Saligrama, Boqing Gong

[🔗](#) Paper

Board 9

Planning in 8 Tokens: A Compact Discrete Tokenizer for Latent World Model

Dongwon Kim, Gawon Seo, Jinsung Lee, Minsu Cho, Suha Kwak

[🔗](#) Paper

Board 10

E-RayZer: Self-supervised 3D Reconstruction as Spatial Visual Pre-training

Qitao Zhao, Hao Tan, Qianqian Wang, Sai Bi, Kai Zhang, Kalyan Sunkavalli, Shubham Tulsiani, Hanwen Jiang

[🔗](#) Paper

Organizers

Hirokatsu Kataoka
AIST / Oxford VGG



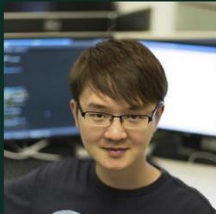
Website

Yoshihiro Fukuhara
AIST / CADDi



Website

Yonglong Tian
OpenAI



Website

Elliott / Shangzhe Wu
University of Cambridge



Website

Oishi Deb
Oxford VGG



Website

Ryosuke Yamada
AIST



Website

Christian Rupprecht
Oxford VGG



Website

Jianyuan Wang
Oxford VGG



Website

Kohsuke Ide
University of Tsukuba



Website

Xianzheng Ma
Oxford VGG



Website

Koichi Namekata
Oxford VGG



Website

Yiming Chen
Oxford VGG



Website

[Tentative title] Visual General Intelligence: A White Paper

Hirokatsu Kataoka^{1,2} Yoshihiro Fukuhara¹ Yonglong Tian³ Shangzhe Wu⁴ Oishi Deb²
Ryosuke Yamada^{1,5} Christian Rupprecht² Jianyuan Wang² Kohsuke Ide^{1,6} Xianzheng Ma²
Koichi Namekata² Yiming Chen² **+further inviting special authors**

¹ National Institute of Advanced Industrial Science and Technology (AIST)

² Visual Geometry Group (VGG), University of Oxford ³ OpenAI ⁴ University of Cambridge

⁵ University of Technology Nuremberg ⁶ University of Tsukuba

Organizing a white paper for visual intelligence & VGI

Let's get started the workshop & explore the future of VI

Sponsor & Supporter (many thanks)

 SB Intuitions



ASPIRE 先端国際共同研究推進事業
Adopting Sustainable Partnerships for Innovative Research Ecosystem