The Third Edition of Large Vision – Language Model Learning and Applications Grand Challenge (LAVA Challenge)

(Note: LAVA Challenge has been accepted as Grand Challenge at ACMMM 2025)

Recent advances in **Large Vision-Language Models (LVLMs)** hold immense promise across various domains, including healthcare, education, entertainment, transportation, and finance, by enabling more sophisticated and context-aware multimedia interactions.

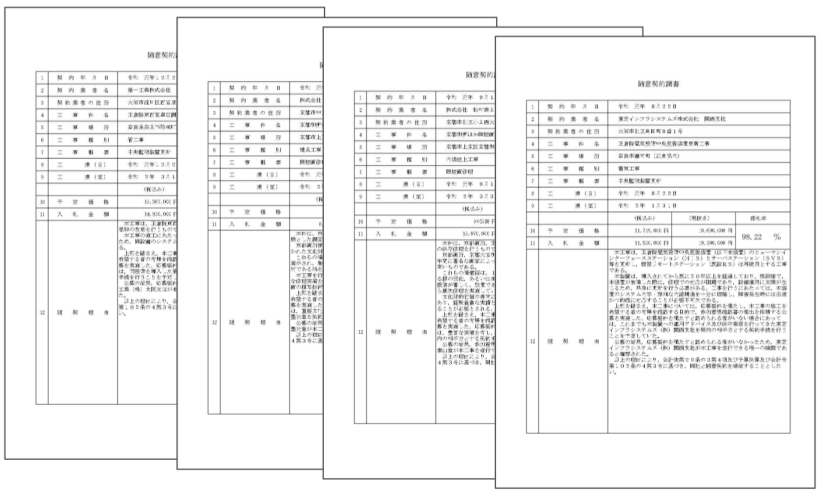
Indeed, the outcomes of our previous challenges, held in conjunction with the Asian Conference on Computer Vision (ACCV) 2024 in Hanoi, Vietnam, and the ACM International Conference on Multimedia (ACMMM) 2025 in Dublin, Ireland (https://lava-workshop.github.io), underscored the limitations of LVLMs in processing such documentations and presentations.

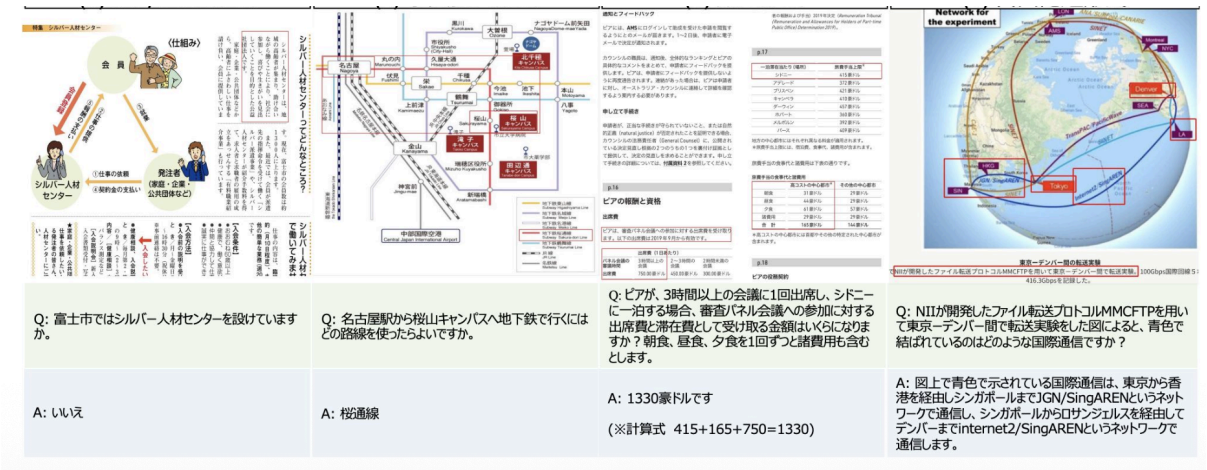
To enhance the capability of LVLMs to accurately interpret and generate descriptive text from complex visual inputs within multi-page business-related documents and slides, we continue to organize the **Large Vision Language Model Learning and Applications (LAVA)** Challenge, building upon the success of previous years.

LAVA Challenge focuses on question-answering tasks, including both multiple-choice and open-ended questions, based on multi-page documentations and slides, including diverse data representations such as Graphs, Charts, Tables, Diagrams, Data Flow Diagrams (DFDs), Class Diagrams, Gantt Charts, and Building Design Drawings. We provide Illustrative samples as shown below.

Q. これらの随意契約調書のうち、落札率が明記されていない契約は、いずれも文化財そのものの保存・修理を直接の目的としている。これらの契約における税抜きの契約金額の合計を答えなさい。

A. 21,600,000円





Participants of the LAVA Challenge may register as individuals or teams and are required to develop a model capable of answering questions pertinent to the provided input data.

We anticipate that the LAVA Challenge will stimulate further advancements in research on this topic across the academic community, industry, and society over (at least) the forthcoming 3-5 years. To contribute meaningfully to this research area, we are committed to maintaining the information, datasets, and tasks for the LAVA Challenge for a minimum duration of three years.

Submission procedure

All submissions will be handled automatically on Codabench/Huggingfaces Datasets (to be improved):

- Public dataset: We will release our dataset collected from the internet. It contains about 1,000 samples.
- Private dataset: The TASUKI team (SoftBank) provides the private dataset. It contains about 500 samples.

We use the MMMU metric to evaluate the results. Final score = 0.3 * Public dataset + 0.7 * Private dataset

Computational resources: Participants may use SoftBank Beyond AI SANDBOX GPUs.

Relevance to ACM Multimedia 2026

This challenge aligns with ACM Multimedia's mission by addressing multimodal AI advancements. With the growing role of **LLMs and LVLMs** in processing **images, audio, video, and text**, the ability to **accurately interpret structured visual data** remains an open challenge. This workshop aims to push the boundaries of **multimodal multi-page documentation research**, facilitating advancements with

broad applications in **healthcare, finance, engineering, and multimedia content creation**.

Importance to the AI Research Community

- Addresses a critical research gap in the conversion of **multi-page documents and slides**.
- Provides a public **benchmark dataset** to drive future research.
- Engages researchers in **multimodal learning, NLP, and computer vision**.

Expected Participation

- Interest from **academic researchers** in NLP, multimodal AI, and computer vision.
- Potential participation from **industry professionals** seeking real-world applications.

Tentative Schedule

- Challenge track opened: 2026/3/10
- Public and Private set released: 2026/3/10
- Challenge track closed: 2026/5/7
- Code submission deadline: 2026/5/14
- Paper submission deadline: 2026/5/28
- Acceptance notification: 2026/7/16
- Camera-ready deadline: 2026/8/6
- Author Registration: 2026/8/13
- Workshop date: 2026/11/10

Challenge history

The first edition of the LAVA workshop challenge was successfully held in conjunction with the ACCV 2024 in Hanoi, Vietnam. 12 international participants joined the challenge. The top 3 winning teams received travel grants and prizes to present their research in LAVA Workshop. The workshop featured a half-day program with over 50 participants (excluding organizers, invited speakers, and paper presenters) and accepted seven papers for publication.

The LAVA Challenge's second edition took place at ACMMM in Dublin, Ireland. The competition was held through the Kaggle platform, attracting 62 entrants, 22 international participants, and 14 teams. Two teams among the top four teams were awarded conference registration fees and prizes to present their research at the event. The score of the top teams is provided below.

Rank	Team Name	Public Acc.	Private Acc.
1	SYSUpporter	0.61	**0.56**
2	Woof	0.59	0.55
3	nsbsk	0.58	0.51
4	char	0.58	0.43
Baseline	GPT-4o	0.15	

The workshop successfully gathered over 70 participants and accepted seven papers. The dataset of this challenge is published on the HuggingFace repository (https://huggingface.co/datasets/d-sato/LAVA-Challenge-2025-Dataset).

Challenge organizers

- **Minh-Duc Vo**, SB Intuitions, 1-7-1 Kaigan, Minato-ku, Tokyo, Japan 105-0022. Email: vmduc.work@gmail.com (main contact).
    - Minh-Duc Vo received the Ph.D. degree in computer science from The Graduate University for Advance Studies (SOKENDAI) in alliance with the National Institute of Informatics (NII), Japan, in 2020. After holding a Project Assistant Professor position at The University of Tokyo until 2025, he is currently a Senior Research Scientist at SBintuitions. His research interests include image recognition and generation, language, and vision. He is a regular reviewer of international conferences in computer vision.
- **Akihiro Sugimoto**, National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda City, Tokyo, Japan, 101-8430. Email: sugimoto@nii.ac.jp
    - Akihiro Sugimoto received the Dr. Eng. in mathematical engineering from the University of Tokyo. He is currently a full professor at the National Institute of Informatics, Tokyo. He has regularly served as an Area Chair at several top-tier conferences, including ICCV, CVPR, ECCV, ICLR, NeurIPS, ACCV, ICPR, and 3DV. He also served as an associate editor of IJCV, and is now serving as an associate editor of CVIU. He served as GC of ACCV2012, 2020, and 3DV2020. He has published more than 150 peer-reviewed journal/international conference papers.
- **Hideki Nakayama**, The University of Tokyo, 7-3-1 Hongo, Bunkyo city, Tokyo, Japan 113-8654. Email: nakayama@ci.i.u-tokyo.ac.jp
    - Hideki Nakayama received the Ph.D. degree in information science from the University of Tokyo, Japan, in 2011. From 2012 to 2018, he was an Assistant Professor with the Graduate School of Information Science and Technology, The University of Tokyo, where he has been an Associate Professor since April 2018. He is also a Faculty Member

with the International Research Center for Neurointelligence (IRCN) and the Institute of AI and Beyond (BeyondAI), The University of Tokyo. His research interests include machine perception, natural language understanding, content generation, and multimodal learning. He has published 100+ peer-reviewed papers, including 50+ renowned conferences and journals such as CVPR, ICCV, ECCV, ACL, NAACL, EMNLP, TACL, ICLR, AAAI. Also, he has served as Senior Area Chair or Area Chair for many top conferences, including CVPR, ICCV, NAACL, NeurIPS, ICLR, and IJCAI. He is a member of IEEE and ACM.

- **Khan Md Anwarus Salam**, SoftBank, 1-7-1 Kaigan, Minato-ku, Tokyo, Japan. 105- 7529. Email: khan.mdanwarussalam@g.softbank.co.jp.
    - Dr. Khan Md. Anwarus Salam is working in SoftBank's Beyond AI Promotion division. His extensive career includes roles as a Research Scientist at IBM Research in Tokyo and as the country engineering consultant for Google in Bangladesh. Academically, he holds a Ph.D. and a Master's in Information and Communication Engineering from The University of Electro-Communications in Tokyo, where he researched machine translation, and a B.Sc. in Computer Science from BRAC University, Dhaka. His research interests include Generative AI, Natural Language Processing, Semantic Analysis, Machine Translation, and Machine Learning. Dr. Salam's work significantly influences both academic research and practical applications in AI, shaping the future of technology.
- **Takara Taniguchi**, The University of Tokyo, 7-3-1 Hongo, Bunkyo city, Tokyo, Japan 113-8654. Email: hiroshi-tani@g.ecc.u-tokyo.ac.jp.
    - Takara Taniguchi is a first-year master's student at the University of Tokyo. He received a B.Eng. from the University of Tokyo. Now he is working on multimodal language models in the Graduate School of Information Science and Technology, The University of Tokyo, as a master's student.
- **Daichi Sato**, The University of Tokyo, 7-3-1 Hongo, Bunkyo city, Tokyo, Japan 113-8654. Email: satodai370054@g.ecc.u-tokyo.ac.jp.
    - Daichi Sato is a second-year master's student at the University of Tokyo. He received a B.Eng. from the University of Tokyo. Now he is working on constructing datasets for LVLMs in the Graduate School of Information Science and Technology, The University of Tokyo, as a master's student.
- **Kaito Baba**, The University of Tokyo, 7-3-1 Hongo, Bunkyo city, Tokyo, Japan 113-8654. Email: baba-kaito662@g.ecc.u-tokyo.ac.jp.
    - Kaito Baba is a first-year master's student at the University of Tokyo. He received a B.Eng. from the University of Tokyo. Now he is working on large language models in the Graduate School of Information Science and Technology, The University of Tokyo, as a master's student.
- **Duc-Tuan Luu**, University of Information Technology, Vietnam National

University HCMC, Quarter 34, Linh Xuan Ward, Ho Chi Minh City, Vietnam. Email: tuanld@uit.edu.vn

- Duc-Tuan Luu received the Master degree in computer science from The University of Science (HCMUS), Vietnam National University HCMC in 2025. Now, he is working as a researcher in the Laboratory of Multimedia Communications, University of Information Technology (UIT), Vietnam National University HCMC. His research interests include visual retrieval, AI4edu and computer vision.

Main contact: Minh-Duc Vo, Khan Md Anwarus Salam