Hi, everyone! My name is Kai and I’m a PhD candidate in HKUST. I’m glad to have this opportunity to enjoy the second half of today’s workshop with you. I’m primarily working on omni-modal large language models and multi-view video generation for autonomous driving, which is also matching the topic of our next invited talk.

Our next talk will be given by Dr. Lorenzo Bertoni. Lorenzo is a Senior Applied Scientist at Wayve, where he works at the intersection of generative AI and autonomous driving. He earned his Ph.D. with distinction from EPFL in 2022, focusing on 3D perception for autonomous driving. His research emphasized improving safety for vulnerable road users and addressing corner cases. Let’s welcome Lorenzo!

The next talk will be industrial talk given by Dr. Chufeng Tang. Chufeng is now a researcher in IAS BU of Huawei. His research interests primarily lie in computer vision and autonomous driving. He received his Ph.D. degree from Tsinghua University in 2023, and B.E. degree from Huazhong University of Science and Technology in 2018.

Thanks, Chufeng and all the other speakers for their wonderful talks! Now let’s get into the challenge section. In this year’s W-CODA challenge, we hold two tracks on different perspectives of multi-modal foundational models for autonomous driving.

Track 1 is on corner case scene understanding. This track focuses on enhancing multimodal perception and comprehension capabilities of MLLMs for autonomous driving, emphasizing global scene understanding, local area reasoning, and actionable navigation. The goal is to promote the development of more reliable and interpretable autonomous driving. To do that, we construct the CODA-LM dataset, which includes around 10000 images and textual descriptions with general perception, regional perception and driving suggestions.

We are glad to have more than 50 teams from all over the world to participate this track, covering more than 100 individuals and 30 institutions. This tight competition results in significant performance improvements. The normal organizer baseline based on LLaVA-1.5 is surpassed by 156.02% by the winners, which also improve our advanced organizer baseline by 13.36%.

The 3rd place is achieved by Chongqing University of Posts and Telecommunications. Let’s now welcome the team winners on stage and Dr. Songcen Xu from Huawei Noah’s Ark Lab to present the awards! Now let’s give the speaker to the winners to present their solutions.

The 2nd place is achieved by Peking University. Unfortunately, the winners cannot make their trip here, so I can play the video for them.

The 1st place is achieved by The Chinese University of Hong Kong (Shenzhen) and HKUST. Unfortunately, the winners cannot make their trip here, so I can play the video for them.

Track 2 is about corner case scene generation. This track aims to enhance diffusion models for generating multi-view street scene videos that align with 3D geometric descriptors, such as Bird's Eye View (BEV) maps and 3D LiDAR bounding boxes. Building on the state-of-the-art MagicDrive framework, developed by the organizer team, this track strives for improved scene generation capabilities in autonomous driving, emphasizing consistency, higher resolution, and extended video duration.

We receive submissions from more than 10 teams from all over the world, covering more than 40 individuals and 20 institutions. The quality of the generated videos (estimated by the FVD) and the consistency with the provided descriptors (estimated by the mAP and mIoU values) are significantly enhanced, and we are glad to see impressive results from our winner submissions.

The 3rd place is achieved by Fudan University. Unfortunately, the winners cannot make their trip here, so I can play the video for them.

The 2nd place is achieved by Zhejing University, Shanghai AI Lab and the Technical University of Munich. Unfortunately, the winners cannot make their trip here, so I can play the video for them.

The 1st place is achieved by HIT SZ, Li Auto Inc, THU, Westlake University and NUS. Let’s now welcome the team winners on stage and from Amy Sun from Huawei to present the awards! Now let’s give the speaker to the winners to present their solutions.