

With the rapid evolution of wireless communication, the operational frequencies continue to increase, now extending into the millimeter-wave (mmWave) and sub-terahertz (sub-THz) bands. This progression unlocks significant potential, offering advantages such as increased bandwidth, substantially higher data rates, and remarkably reduced latency, thereby supporting emerging applications including augmented reality (AR), virtual reality (VR), ultra-high-definition streaming, and ultra-reliable low-latency communications [1]. These advancements promise significant improvements in user experience and enable new, bandwidth-intensive, and latency-sensitive services critical for future communication systems. Despite these benefits, higher frequency communications face considerable challenges due to signal sensitivity, particularly in line-of-sight (LoS) and nonline-of-sight (NLoS) conditions. At mmWave and sub-THz frequencies, signals experience severe attenuation and are highly susceptible to blockage from common environmental objects and human bodies, leading to frequent disruptions [2]. Moreover, beamforming techniques, essential for maintaining high antenna gains and increasing data rates, introduce additional complexities. The narrow beams required for high gains demand precise alignment between the transmitter and the receiver, which can be hard to maintain in dynamic environments, increasing the overhead associated with beam training and adaptation [3].

To overcome these limitations, integrating additional sensing modalities, such as visual sensors like cameras, has emerged as a promising approach [3][4]. By capturing detailed visual information, the system gains a richer environmental context, which significantly improves its capability to discern spatial relationships and anticipate obstructions in the signal path. Through advanced computer vision and machine learning algorithms, future blockage events and UE position can be effectively predicted, enabling proactive network responses such as beam switching, handovers, or adaptive resource allocation. The obtained environmental context is also valuable for channel prediction, reducing the overhead associated with channel estimation.

In this challenge, participants are invited to develop innovative machine learning solutions using visual and radio data to tackle the unique demands of fast-changing, high-frequency communication environments. Participants will work on one or more tasks using visual data captured from cameras and radio-frequency data collected by wireless nodes:

- **Blockage Prediction:** Accurately forecast future blockage conditions to enable proactive network adjustments.
- **User Equipment (UE) Localization and Position Prediction:** Determine and predict UE positions to support wireless communication.
- **Channel Prediction:** Predict the Sounding Reference Signal (SRS) to reduce the channel estimation overhead.
- **Beam Prediction:** Predict the optimal beam index to reduce the overhead associated with beam training.

This challenge is grounded in and supported by the experimental infrastructure developed in the CONVERGE project [5], which provides a unique platform for multimodal experimentation. The CONVERGE chamber integrates a mobile FR2-capable gNB and UE, a programmable obstacle, stereo RGB-D cameras, Reconfigurable Intelligent Surfaces (RIS), and a programmable control architecture. This environment enables the generation and collection of synchronized radio and visual datasets under realistic, controlled indoor conditions with configurable occlusions and mobility patterns, which reflects the complexity of near-field beam management in dynamic environments.

We warmly invite researchers and industry practitioners to participate in this CONVERGE Challenge: Multimodal Learning for 6G Wireless Communications. With such a diverse and realistic dataset, as well as a wide range of well-designed tasks to choose from, we believe this challenge will attract broad interest and participation from a wider community. We hope that through this challenge, participants will gain a deeper understanding of how visual data integration can enhance and revolutionize wireless communications, driving forward research and innovation in this emerging interdisciplinary field.

The top five teams will be invited to submit a 2-page paper describing their approaches, which will be presented at ICASSP 2026, and accepted papers will be published in the ICASSP proceedings. In addition, teams presenting in person at ICASSP will be encouraged to submit a full-length paper to the IEEE Open Journal of Signal Processing (OJ-SP).

NOTE: All intellectual property (IP) rights for the data and baseline code provided in this challenge remain with the organizers. Participants retain ownership of any methods or models they develop. By submitting results, participants grant the organizers to use the submitted solutions for evaluation and publication purposes related to the challenge. Participants are responsible for ensuring that their submissions do not infringe upon the rights of third parties.

These are the important dates for the Challenge:

- **Competition Launch and Data Release:** September 30 December 22, 2025
- **Registration Deadline:** October 31, 2025 December 31, 2025
- **Submission Deadline:** November 15, 2025 January 16, 2026
- **Results and Rankings Notification:** December 01, 2025 January 23, 2026
- **2-page Papers Due (by invitation):** December 07, 2025 January 31, 2026
- **2-page Paper Acceptance Notification:** January 11, 2026 February 7, 2026
- **Camera-ready Submission:** January 18, 2026 February 14, 2026 (FIRM)