

# 251<sup>th</sup> WPI-IIIS Seminar

## From Sound to Structure: Learning Representations in Multimodal and Audiovisual Systems

This talk presents our work on building multimodal AI systems that learn structured representations from audiovisual data. We develop methods that leverage cross-modal interactions between sound and vision to enable alignment, generalization, and learning from weak supervision. I will highlight approaches for audiovisual scene understanding, spatial and temporal modeling, and generative frameworks that capture the structure of multimodal signals. Finally, I will discuss how these systems connect to neuroscience and how they enable more unified, robust, and adaptive multimodal AI.



## Dr. Eli Shlizerman

Department of Applied Mathematics and  
Electrical and Computer Engineering,  
University of Washington

Date: **Monday, April 13, 2026**

Time: **11:00 – 12:00**

Venue: **1F Auditorium, IIIS Building**

**\* On-site participation only**



Contact: International Institute for Integrative Sleep Medicine, University of Tsukuba  
029-853-5859 (ext.5859) | [wpi-iiis-alliance@ml.cc.tsukuba.ac.jp](mailto:wpi-iiis-alliance@ml.cc.tsukuba.ac.jp)