Key Contributions:

- **VLA model evaluation on a real robot:**
  Trained and evaluated multiple models (iDP3, 3DDA, 3DFA, Pi0). Identified critical limitations—iDP3 struggled with precise 3D tasks, while Pi0 showed slower action execution and a strong dependence on wrist-camera input. Evaluate the model's performance of trajectory and keypoint outputs.

- **Perception and Data Collection Pipelines:**
  Implemented real-time multi-camera streaming and depth estimation for large-scale data collection. Benchmarked depth cameras (RealSense, Azure Kinect, ZED X), and optimized the ZED X pipeline for low-latency, high-frequency streaming.

- **Language-Guided Data Collection:**
  Built a voice-controlled assistant using OpenWakeWord, SileroVAD, Whisper, DeepSeek, and Kokoro-TTS. Enabled a single operator to collect large datasets with natural language commands, tightly integrating speech with robot data capture.

- **Robotic Control Systems:**
  Designed and delivered a real-time end-effector controller for the Aloha robot, achieving 50 Hz control using a custom numerical inverse-kinematics solver. Outperformed the official Interbotix controller, which was significantly slower.

- **Benchmarking and Calibration:**
  Conducted large-scale comparisons across major datasets (RH20T, Open X-Embodiment, BridgeData). Performed system-wide calibrations (cameras, hand-eye alignment, forward kinematics), reducing overall error margins to within 1 cm.