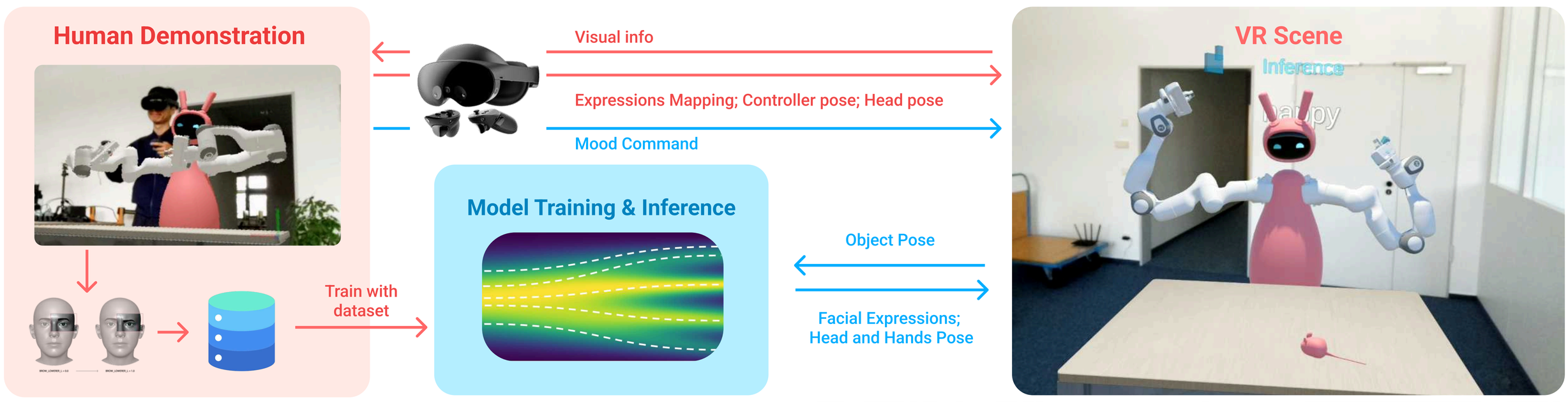
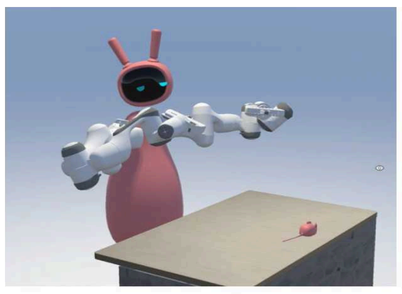
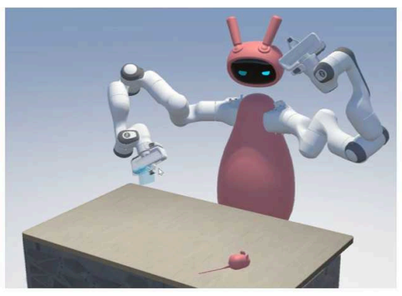
Real-Time Robotic Emotional Expression from Mixed-Reality Demonstrations via Flow Matching

Expressive behaviors in robots are critical for effectively conveying their emotional states during interactions with humans. In this work, we present a framework that autonomously generates realistic and diverse robotic emotional expressions based on expert human demonstrations captured in Mixed Reality (MR). Our system enables experts to tele-operate a virtual robot from a first-person perspective, capturing their facial expressions, head movements, and upper-body gestures, and mapping these behaviors onto corresponding robotic components including eyes, ears, neck, and arms. Leveraging a flow-matching-based generative process, our model learns to produce coherent and varied behaviors in real-time in response to moving objects, conditioned explicitly on given emotional states.

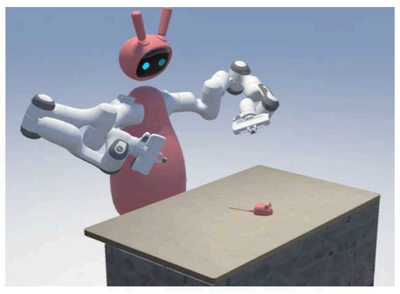




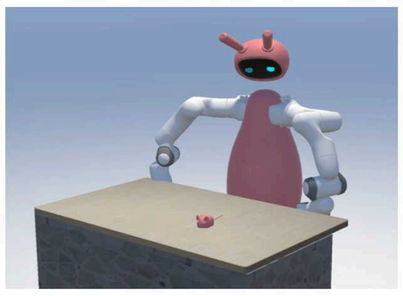
Fear: Hands flutter nervously while the head turns away; the eyes steal sideways glances at the mouse.



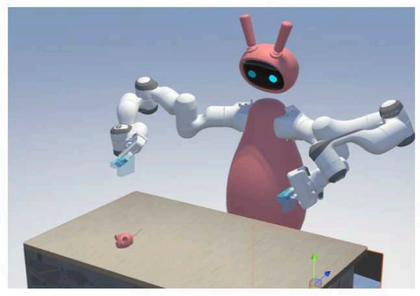
Angry: Fixes the mouse with a furious stare and, as it approaches, slams downward with both arms.



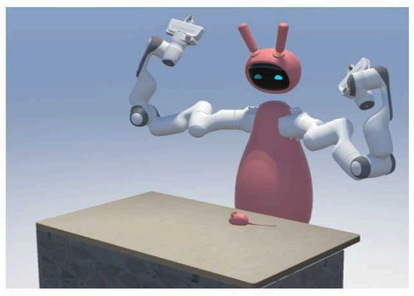
Curious: Tilts and slowly rolls the head, blinking inquisitively as it studies the mouse.



Sad: Arms sag, gaze drops to the mouse with down-cast eyes, and the head gives a slow shake.

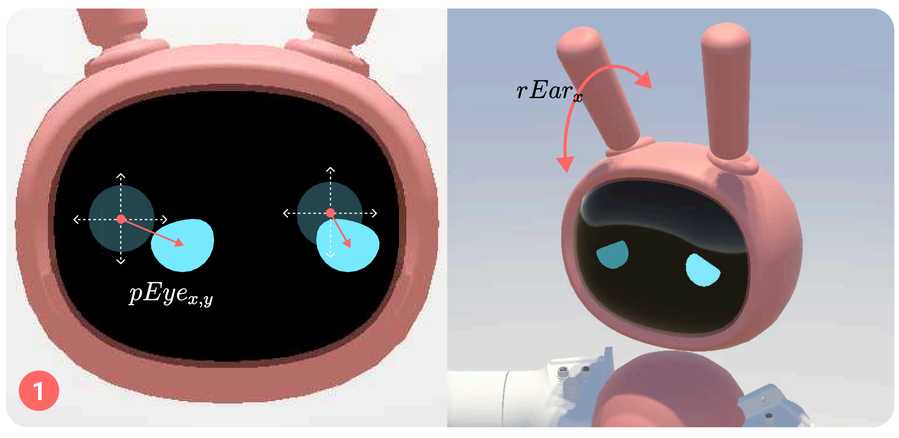


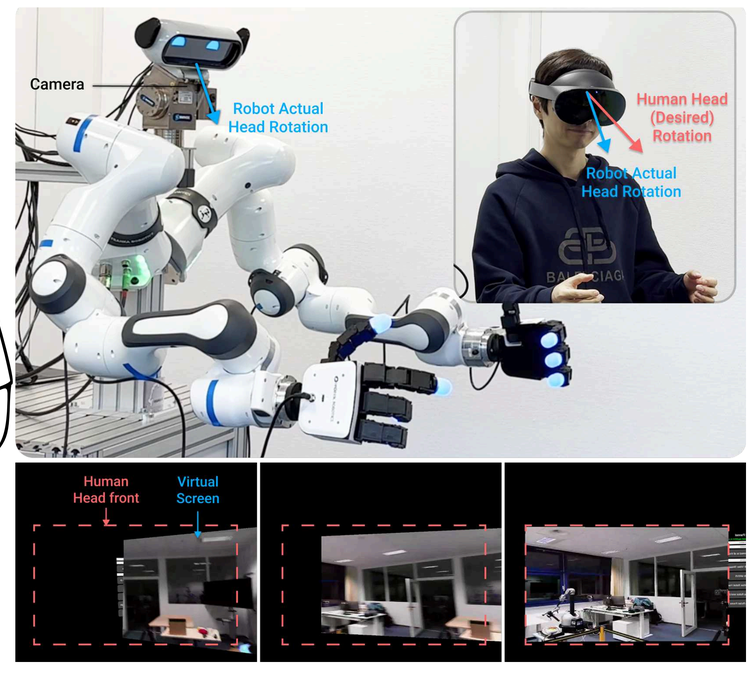
Calm: Maintains a relaxed posture and peaceful gaze on the mouse, punctuated by gentle blinks.

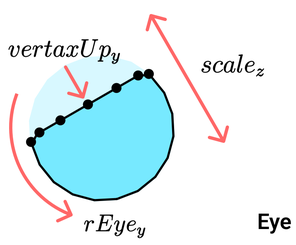


Happy: Smiling eyes, nodding head, and rhythmic arm waves convey cheerful excitement toward the mouse.

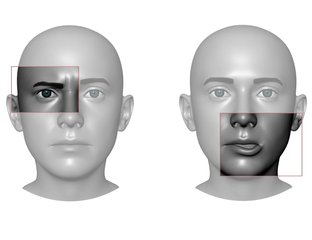
Mixed-reality pipeline for robotic emotions. Tele-operated demos (face/head/hand signals) train a flow-matching policy that maps mood + percepts to joint targets; at 120 Hz it controls eyes, ears, neck, and arms to express emotion. Red/pink =training-only; blue = runtime. Bottom: the generated emotional expressions directed toward the moving target object.

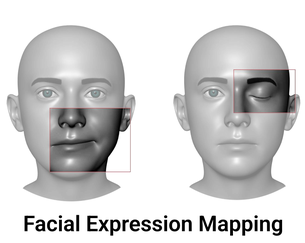




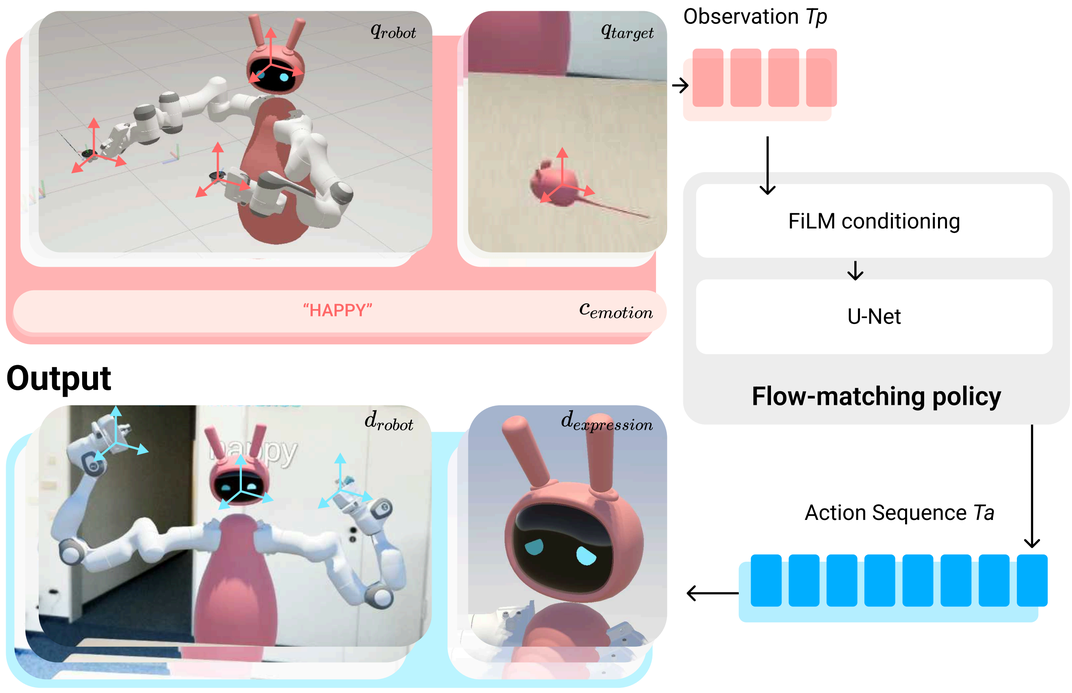


Eye Model





Facial Expression Mapping



UP: The XR platform: 1. 7 facial-expression values detected by the XR-headset map the robot's ears angle and shape of the eyes, the gaze direction also maps the position of the eyes on the plane of robot face screen. Some value of the facial expression also maps the movement of the robot's ear. 2. Human's head position and orientation maps the robot's end effector, relative to the operator's head pose as the origin. The positional value is sceled by 1.5 for enhancing operator's reachablitiy. 3. There is an virtual screen floating in front of the operator, which allows the operator to observe the environment from the first person perpective.

Left: Overview of flow matching for expression generation. A history window of robot and target poses plus an emotion label (pink) is fed through FiLM-conditioned U-Net to predict the blue action sequence executed on the robot.

Chao Wang,
Michael Gienger
Fan Zhang

Visit the paper webpage for more details