2026 START Program
CFP Brief

THEME: **01.** **Robotics/Physical AI**

SUB-THEME: **1.2.** **Foundation models for close human-robot interaction**

Context/ Overview

Future applications of robots in the home, offices, and hospitals require navigating amongst humans, communicating with humans, touching humans, observing humans from close and far distances, and collaborative manipulation. Learning-based approaches have been successful at modeling many aspects of human behavior, such as speech, social navigation, facial expressions, and bodily motion. But existing models are largely limited to individual parts of the body and/or disparate tasks. This project envisions that a general-purpose **HRI foundation model** that can predict diverse aspects of human-robot interaction will be a crucial step toward building general-purpose robots that are accepted and safe for consumers.

Problem Statement

Although agent identification and behavior prediction in autonomous driving is a well-studied field, the interactions between humans and robots in close proximity are far less understood. The proposed problem is to develop practical ML models that perform identification and prediction functions during close-range interactions involving bodily movement, language, and facial expressions. Cooperative, competitive, and neutral interactions may be present during navigation, manipulation, and communication. An envisioned HRI foundation model is expected to implement multiple functions, such as interpreting the intent of speech and gesture, recognizing anatomy from near and far ranges, predicting full-body motions, and interpreting human emotion state through appearance, language, and behavior. To enable robot planning and policy training, the model should also predict human responses to future robot behavior, not simply to predict based on history or how a human would act in the absence of stimuli.

Objectives & Scope

Applicants should consider incorporating multi-modal data sources, including video, audio, and body pose into their model. A successful HRI foundation model may require new methods for capturing and learning from data at scale containing human reactions to robots. Such a model may also include cross-embodiment learning techniques that enable predicting differences between robot morphology, size, and appearance. Other potentially valuable capabilities of a model include variation due to personal and multi-cultural differences.

A successful proposal will describe a high-impact, technically sound, multi-year plan that builds toward the desired functionality. A critical component of a winning proposal is an evaluation plan that considers not only accuracy of methods on datasets, but end-to-end impact of the proposed research on human interaction. A single human participant is an acceptable scope during early stages of the project, while longer-term plans may consider the interaction of multiple humans and multiple robots.

Specific Topics & focus areas*

Addressing the proposed problem may include research in the following areas:

1. Multi-modal foundation models
2. Generative human models
3. Human motion tracking
4. Human motion prediction
5. Social navigation and human-robot safety

※ The topics are not limited to the above examples and the participants are encouraged to propose other original ideas.

END OF DOCUMENT