



(a) Explicit human intervention

This section shows four input methods for explicit human intervention, each in a blue box with an icon:

- Lang.** (Language): Represented by a speech bubble icon.
- Keys** (Keyboard): Represented by a keyboard icon.
- GUIs** (Graphical User Interfaces): Represented by a monitor with a cursor icon.
- 3DSG** (3D Scene Graph): Represented by a hand pointing at a screen icon.

A large blue curved arrow points from this section towards the Task Planner in the main architecture.

(b) Implicit behavior cues

This section shows four input methods for implicit behavior cues, each in a green box with an icon:

- Pose**: Represented by a hand icon.
- Expr** (Expression): Represented by a smiley face icon.
- Gaze**: Represented by an eye icon.
- EEG**: Represented by a head with a brain wave icon.

A large green curved arrow points from this section towards the Intervention point in the main architecture.

(c) Proactive clarification

This section shows a proactive clarification interface within a dashed orange box, containing three messages:

- Human icon: **Put the fruit on the plate.**
- AI icon: **Which one, pear, orange?**
- Human icon: **Put the pear on the plate.**

A large orange curved arrow points from this section towards the Control Policy in the main architecture.