

The diagram illustrates the UI-TARS environment architecture. At the top, a computer icon represents the environment. Below it, a horizontal timeline of colored squares (yellow, red, orange) represents the sequence of actions. The environment is divided into two main sections: **Initialize** (left) and **Observation** (right). The **Initialize** section contains a **User Query** (yellow box) and an **Action Space** (yellow box). The **Observation** section contains a **PyAutoGUI** (red box) and a **Thought** (red box). The **PyAutoGUI** box is connected to the **Initialize** section and the **Observation** section. The **Thought** box is connected to the **PyAutoGUI** box and the **Observation** section. The **PyAutoGUI** box is connected to the **Initialize** section and the **Observation** section. The **Thought** box is connected to the **PyAutoGUI** box and the **Observation** section. The **PyAutoGUI** box is connected to the **Initialize** section and the **Observation** section. The **Thought** box is connected to the **PyAutoGUI** box and the **Observation** section.

Initialize

PyAutoGUI

Observation

User Query

Action Space

Thought

UI-TARS

The diagram illustrates the XUI-TARS architecture, which is designed for GUI-based tasks. It is organized into four main functional areas:

- Perception:** This component processes visual information from the GUI. It includes:
 - Element Description
 - Dense Captioning
 - Transition Captioning
 - Question Answering
 - Set-of-Mark
- Action:** This component generates actions based on the perceived state. It includes:
 - Unified Action Space (containing Click, Type, ...)
 - Our Annotated Dataset
 - Open-Source Data (including AITZ ... and AITW ...)
 - Multi-Step Trajectory Data (represented as a database cylinder)
- System-2 Reasoning:** This component enhances the system's reasoning capabilities through:
 - Reasoning Enrichment with GUI Tutorials
 - Reasoning Stimulation with Thought Augmentation
- Learning from Prior Experience:** This component leverages past experiences for learning, including:
 - Online Trace Bootstrapping & Reflection Tuning
 - Agent DPO (Direct Preference Optimization)

The central part of the diagram shows the **XUI-TARS** system, which integrates these components to perform GUI tasks. It is depicted as a computer monitor displaying a GUI, with the XUI-TARS logo and a user icon next to it.