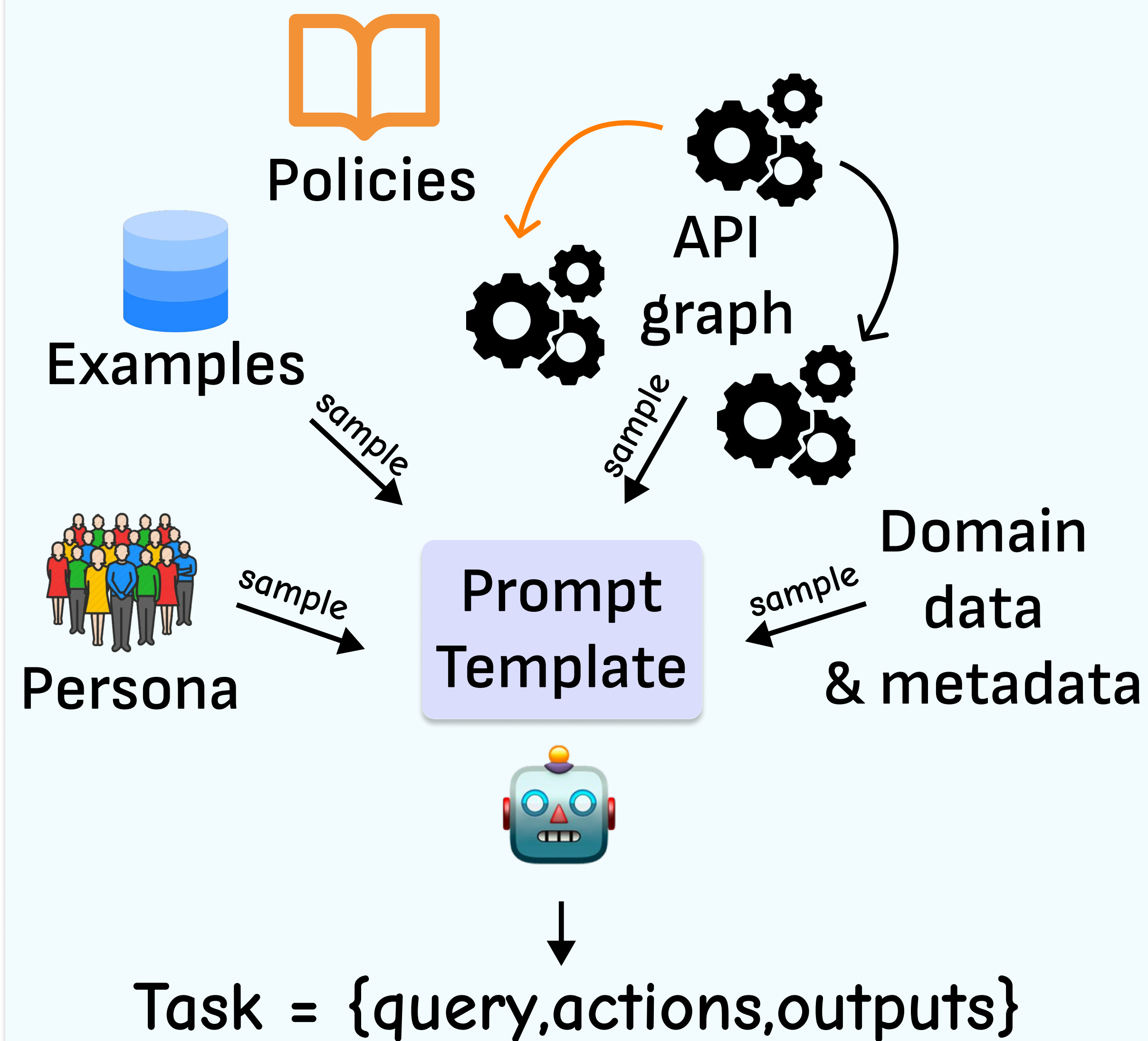


1 Task Generation



2 Task Validation

1. Action Validation

- Format Check
- Execution Check
- Policy Check

```
@@ -60631,11 +60631,11 @@
{
  "name": "Smart Watch",
  "product_id": "6945232052",
-  "item_id": "1706622510",
-  "price": 328.67,
+  "item_id": "2860956907",
+  "price": 315.61,
```

Diff Patch

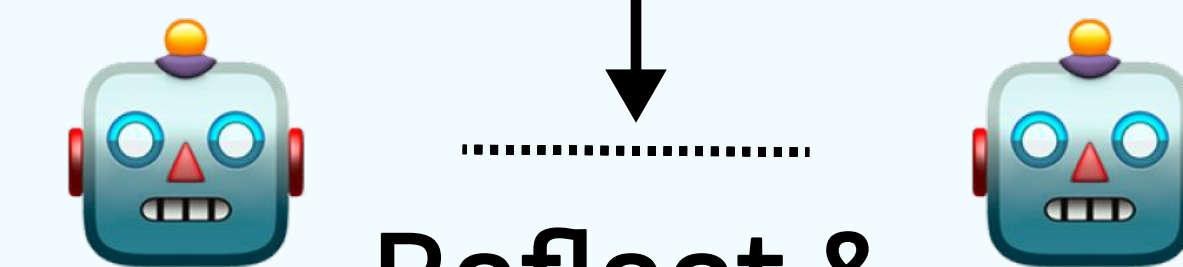
3. Final Review

Aggregate Reflections & Summarize

Realign task using feedback

2. Alignment LM Committee

{Task data, Query, Diff patch}

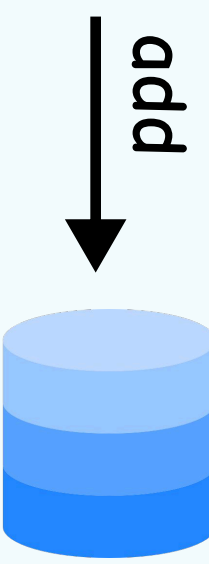


Reflect & Score

avg_score > threshold

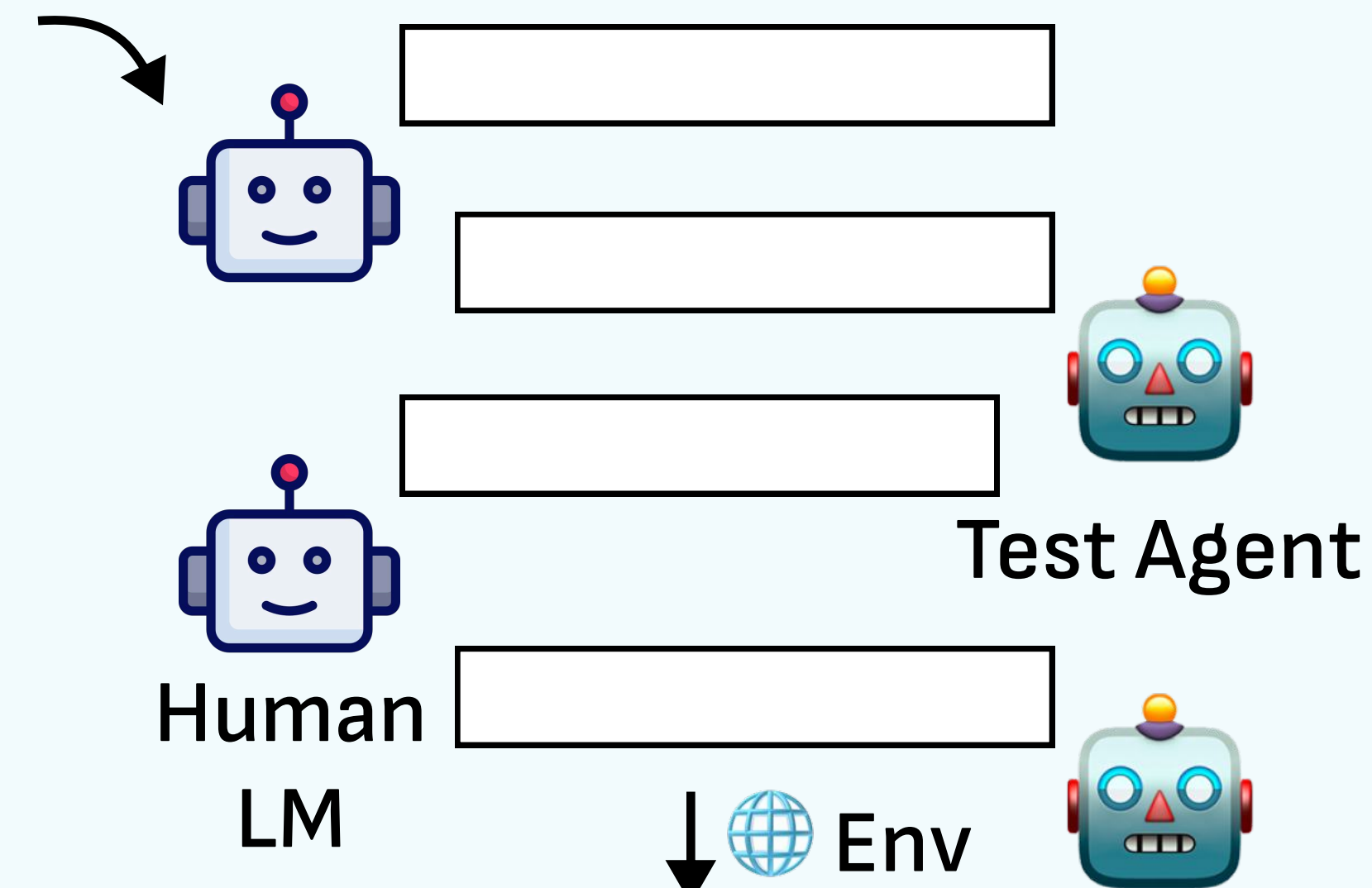


Validated Task



3 Simulated Agent-Human Interplay

Prompt to reveal query details in turns



- State based reward
- Output based reward

Successful Trajectory