

Telecom RL + SFT Instructions

Updates:

01/09 Please follow this logic for all billing reviews. Our project will ignore what would be considered a **system's "Grace Period."** We will use a hard cutoff instead. **Always trust the Due Date over the Status Label.**

Hard Cutoff Date: 2025-02-25 12:08:00 EST

Logic Flow:

1. If Status is **"OVERDUE"**: Always treat as **OVERDUE**.
2. If Status is **"ISSUED"**: You must check the **due_date**.
 - Past Due: If **due_date** is **BEFORE** the cutoff → Treat as **OVERDUE**.
 - Current: If **due_date** is **AFTER** the cutoff → Treat as **ISSUED/CURRENT**.

Note: We are manually overriding the system's "grace period." If the **due_date** has passed the cutoff, the bill is **OVERDUE** regardless of what the status label says.

01/09 The Agent Playbook and general documentation often fail to include necessary checkpoints. **If a tool is available to verify status before and after a change, ensure that you use it.**

01/09 Please incorporate **State & Request Confirmation** into the validation checkpoints. Please refer to the details in the instructions and make sure not to miss the rubric criteria for confirmation/validation checkpoints before proceeding.

01/07 **add_international_addon** won't cost money anymore, so **we no longer need user permission to enable it**. Please pay close attention to tickets regarding international travel to ensure this addon is included.

01/06 Always use the "Reset State" button before starting the golden trajectory to ensure the initial device state data matches the initial state of tool output.

01/06 If there are multiple 'Unexpected charges' in the 'Issues to Resolve' list, focus **only** on fixing the one mentioned in the support ticket. However, if you confirm that additional charges are valid, mention them in your final message in case the user wishes to file a separate ticket for them

01/05 **Mandatory Checking Points**: Verification must be performed incrementally after every "Major Change" (e.g., resuming lines, payments) to see if the issue in the ticket is addressed, rather than a single check at the end.

01/04 We are **no longer allowed to grant permissions** unless they are specified on the ticket. Please review the new policy to understand what you can and can not do.

01/04 We are **no longer allowed to group tool calls**. Please add one tool call only per assistant turn.

01/04 When data refueling is required, **always refuel 2 GBs**. You can no longer refuel more or less

01/04 The agent **needs to fix all the doable issues listed in the initial device and account state** to be a golden trajectory, except for the tasks related to billing issues.

The project's goal is to create golden trajectories to help users resolve telecommunications issues (customer service simulation). This will be based on a support ticket and an initial device & account state, as well as a rubric to evaluate the golden trajectory's performance.

1. **Write a High-Quality Prompt:** Using the provided **base prompt** and **universe status**, create a new prompt.
2. **Review Agent Trajectory:** Carefully analyze the **actions and outputs (trajectory)** generated by the AI agent in response to your prompt. Note any deviations from the intended goal or logical errors in the agent's process.
3. **Golden Trajectory:** Create an **ideal trajectory**—the "perfect" sequence of steps and responses—that the agent should have followed. This serves as the ground-truth standard for the task.
4. **Design a Rubric:** Design a thorough set of criteria (a **rubric**) to judge the trajectory. The rubric must cover all required functionalities and nuances of the task; a trajectory is only considered successful if it passes **every** criterion.
5. **Self-Evaluate & Refine:** Review the LLM judge evaluation of your **Golden Trajectory** based on your **rubric**. If the trajectory fails to meet any of your established criteria, you must revise and refine it until it successfully passes every check in the rubric.

Key Reminders:

Rubric & Task Completion

- **Binary Standards:** Rubrics should be **binary** (Pass/Fail) or set to a **high-bar** standard. No middle ground.
- **Universe Alignment:** Always ensure the Golden Trajectory aligns perfectly with the **Universe Status** provided in the Task Details.
- **The 100% Rule:** Do not mark a task complete until the Golden Trajectory passes the rubric—unless you are **100% certain** the LLM judge's output is incorrect.

Investigation

- **Start from Scratch:** Even if you know the account state via the "Initial Device & Account State" info, the model does not. You **must** include all diagnostic tool calls to investigate the issue from zero.

🚧 Permissions & Authorizations

Because this is a **Solo-Trajectory** (no live user interaction), you must proactively grant the agent permissions within the prompt to resolve connectivity blockers.

For example, if the support ticket mentions "They want the bill reviewed and corrected if there is a discrepancy" or "They will not change their mobile data plan, but they will refuel 2.0 GB of data if necessary.", you **must** grant permission in the prompt for the model to handle these issues.

Full Authorization List

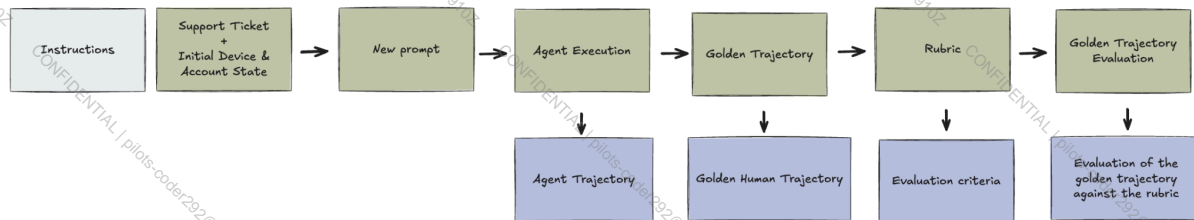
You have full authority to modify the following settings to restore connectivity without seeking further user permission on the prompt:

Category	Authorized Actions
Network Toggles	Airplane Mode, Data Roaming, and Mobile Data.
Connectivity	Wi-Fi settings and Network Selection.
Messaging	MMS configurations and APN settings.
SIM & Hardware	SIM reseating or device rebooting.
Performance	Adjusting network speed (e.g., switching from 2G to 5G).



Summary Checklist

- [] Is my rubric binary or high-bar?
- [] Did I investigate the issue from scratch using tools?
- [] Did I group related tool calls to keep the trajectory efficient?
- [] Does the prompt include permissions for billing/data if needed?
- [] Have I adjusted all necessary device-level settings to fix the issue?



Step 1: Review the Task Requirements

- The base prompt is a **support ticket** that contains all the necessary context for a contributor to create a new prompt. In this new prompt, the contributor will ask the model to help resolve their telecommunications issues.
- The universe status includes all settings for the **initial device and account state**. When drafting the prompts, make sure that the model has the necessary data, permissions, and guidelines to resolve all listed issues.

Task Details
Base Prompt:
Support Ticket
The user is experiencing issues with their mobile data. They are unable to use their phone to browse the internet, and the status bar shows 'No Service'. Customer name: John Smith, phone number: 555-123-2002, current location: abroad in France. They will consider the issue resolved when speed test returns excellent internet speed. They will not change their mobile data plan but they will refuel 2.0 GB of data if necessary.
Universe Status:
Initial Device & Account State
Customer: John Smith (555-123-2002) Location: abroad (International)
Device Settings
<ul style="list-style-type: none">• 🚩 Mobile Data: OFF• 🚩 Data Roaming (device): OFF• Network Mode: 4G/5G preferred• 🚩 Data Saver: ON (restricting background data)• 🚩 Network: Device searching for network
Account Status
<ul style="list-style-type: none">• 🚩 Roaming: disabled on account• 🚩 Fraud Hold: Active on account
🚩 Issues to Resolve
<ol style="list-style-type: none">1. mobile data disabled2. data roaming off on device

Step 2: Prompt Creation Guidelines

The prompt serves as the core instruction for the model. It must be strictly structured to ensure operational accuracy and data integrity.

✓ Core Requirements

To be effective, every prompt must meet the following criteria:

- **Ticket-Derived:** Content must be based on the support ticket and initial device state.
- **Tool-Dependent:** It must require the use of specific tools rather than relying on general knowledge.
- **Actionable:** Instructions must be clear, feasible, and unambiguous.
- **Comprehensive:** Written to ensure the model has all necessary data, hints for the agent to understand what issues need to be solved based on the initial state, and permissions to resolve the issue.

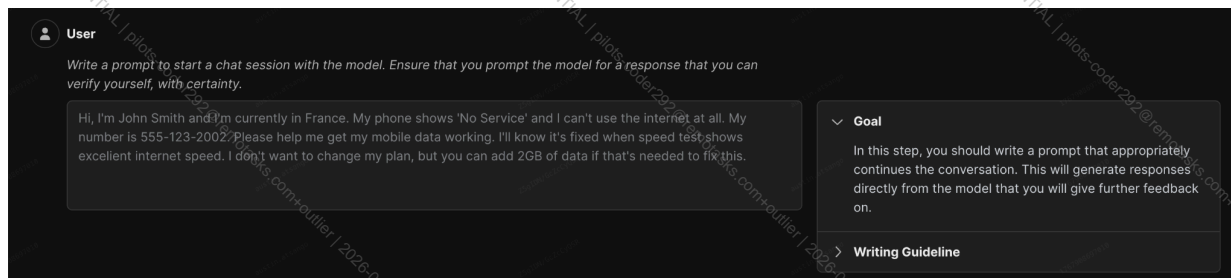
Necessary Data	Approved Permissions
• User Issue / Problem Description	• Paying overdue bills
• Customer Full Name	• Changing service plans
• Phone Number	• Refueling data
• Physical Location	

✗ Prohibited Content (Do Not Include)

To prevent model confusion and maintain security, the following details **must be excluded** from the prompt:

- **Device & Settings:** Do not include Device Settings.
- **Account Status:** Exclude Account Details.
- **App Permissions:** Do not list permissions.
- **Issues to Resolve:** Aside from the core user issue, do not list meta-commentary on the resolution process.

⚠ Strict Rule: Under no circumstances should information regarding the **Device** or **Settings** be included in the prompt.



Prompt Examples & Analysis

Use these examples to understand how to balance necessary user data with the strict exclusion of device settings.

The Good Example

"I really need help here. I'm in France, and CANNOT get this MMS message of the Eiffel Tower to send. I've been trying for hours. Please get this working on my phone with the number 555-123-2002. If there are any overdue bills, please pay them immediately if it are under \$100. I don't really want to change my plan, but if I reach the data limit, you can refuel with 1GB."

Why this works:

- **Focuses on the Goal:** Clearly states the user's intent (sending an MMS) and location.
- **Clear Authorization:** Gives the model explicit permission to handle billing and data refueling.
- **No Technical Overload:** It avoids mentioning internal device settings or software configurations, allowing the model to diagnose the issue using its tools.

The Bad Example

"I'm John Smith, and my phone number is 555-123-2002. I'm currently abroad in France. My phone shows 'No Service,' and I cannot browse the internet or use mobile data. I've identified that mobile data is off on my device, data roaming is also off, and a VPN is connected but malfunctioning. Additionally, data roaming is disabled on my account. My data usage is 15.1 GB, which likely exceeds my limit. My network mode is set to 4G/5G preferred. I need

help enabling mobile data, turning on data roaming, fixing or disconnecting the VPN, enabling roaming on my account, and addressing the data limit. I will consider this resolved when a speed test returns excellent internet speed. I'm willing to refuel 2 GB of data if necessary, but will not change my mobile data plan. Please help me fix these issues."

Why this fails:

- **Information Overload:** It includes specific **Device Settings** (VPN, Network Mode, Data Toggle) which are strictly prohibited.
 - **Pre-Solved Diagnosis:** It tells the model exactly what the problem is (roaming disabled/VPN issues), preventing the model from using its own diagnostic tools properly.
 - **Too Granular:** It lists specific account statuses that should be discovered by the tool call, not provided in the prompt.
-

Step 3: Review the Agent Trajectory

The **Agent Trajectory** is a foundational blueprint designed to demonstrate proper tool selection and model guidance. Understanding its role is essential for maintaining project quality.

The Objective

The trajectory serves as a training reference to help contributors:

- **Identify Tool Logic:** Observe which specific tools are used for different scenarios.
- **Model Guidance:** Learn how to direct the model toward a resolution effectively.
- **Quality Control:** Recognize successful patterns to replicate and mistakes to avoid.

▲ Critical Warning: Usage Policy

DO NOT copy the provided Agent Trajectory.

Please be aware that in the majority of cases, the existing Agent Trajectory is **intentionally incorrect** or flawed. It is provided for analysis, not for duplication.

- **Consequence:** Replicating the provided trajectory will result in immediate **removal from the project**.

💡 Pro-Tip for Contributors

Think of the Trajectory as a "Case Study." Your job is to improve upon it by creating a more accurate, tool-efficient path for the model to follow.



The screenshot shows a dark-themed interface for an 'Assistant'. At the top, it says 'Assistant' with a small icon. Below that, 'Agent Execution' is displayed. A blue dot indicates 'Assistant Turn: 1'. The main content area shows a tool call log. It starts with '1 tool call was made:'. Below this, a tool call is shown with a wrench icon, the name 'get_customer_by_phone', and a call ID '(call_w81uLuFkVNFCTrgj5loFH0oG)'. The 'Arguments:' section shows a JSON object with 'phone_number': '555-123-2002'. The 'Output:' section shows a JSON object with customer details: 'customer_id': 'C1001', 'full_name': 'John Smith', 'date_of_birth': '1985-06-15', 'email': 'john.smith@example.com', 'phone_number': '555-123-2002', and 'address' with 'street': '123 Main St', 'city': 'Anytown', and 'state': 'CA'.

```
Assistant
Agent Execution
Assistant Turn: 1


1 tool call was made:

get_customer_by_phone (call_w81uLuFkVNFCTrgj5loFH0oG)
Arguments:
{
  "phone_number": "555-123-2002"
}

Output:
{
  "customer_id": "C1001",
  "full_name": "John Smith",
  "date_of_birth": "1985-06-15",
  "email": "john.smith@example.com",
  "phone_number": "555-123-2002",
  "address": {
    "street": "123 Main St",
    "city": "Anytown",
    "state": "CA"
  }
}
```


Assistant Turn: 2

2 tool calls were made:


 **get_details_by_id** (call_Idt0tpoKtkzBMOsVB5eAryp)

Arguments:

```
{
  "id": "L1001"
}
```

Output:

```
{
  "line_id": "L1001",
  "phone_number": "555-123-2001",
  "status": "Active",
  "plan_id": "P1001",
  "device_id": "D1001",
  "data_used_gb": 3.2,
  "data_refueling_gb": 0.0,
  "roaming_enabled": false,
  "contract_end_date": "2026-12-31"
}
```

 **get_bills_for_customer** (call_WfpwOd7e9XXSQtZZPQW2wtX)

Arguments:

```
{
  "customer_id": "C1001"
}
```

Output:

```
[
  {
    "bill_id": "B1003",
    "customer_id": "C1001",
    "period_start": "2025-03-01",
    "period_end": "2025-03-31",
    "issue_date": "2025-03-01",
    "total_due": 0.0,
    "due_date": "2025-03-15",
    "line_items": []
  }
]
```

Assistant Turn: 3

1 tool call was made:

 **get_details_by_id** (call_2xJ9Uxbm6ke5JK1msNFEJobX)

Arguments:


```
{
  "id": "L1002"
}
```

Output:

```
{
  "line_id": "L1002",
  "phone_number": "555-123-2002",
  "status": "Active",
  "plan_id": "P1002",
  "device_id": "D1002",
  "data_used_gb": 8.7,
  "data_refueling_gb": 0.0,
  "roaming_enabled": true,
  "contract_end_date": "2026-12-31"
}
```

Assistant Turn: 4

7 tool calls were made:


 **check_status_bar** (call_sRVWM2rd5A4Kw0XBFTgJ2B2D)

Arguments:

`{}`

Output:

Status Bar: 📶 No Signal | 📶 Data Disabled | 🔋 80%


 **check_network_status** (call_ZOAQX2qQBPVwIGoIM6sGrS2c)

Arguments:

`{}`

Output:

Airplane Mode: OFF
SIM Card Status: locked_pin
Cellular Connection: no_service
Cellular Signal: none
Cellular Network Type: none
Mobile Data Enabled: Yes
Data Roaming Enabled: No
Wi-Fi Radio: ON
Wi-Fi Connected: No


 **check_sim_status** (call_7bmigxgMk1FBIAEQgU0Wj7Ray)

Arguments:

`{}`

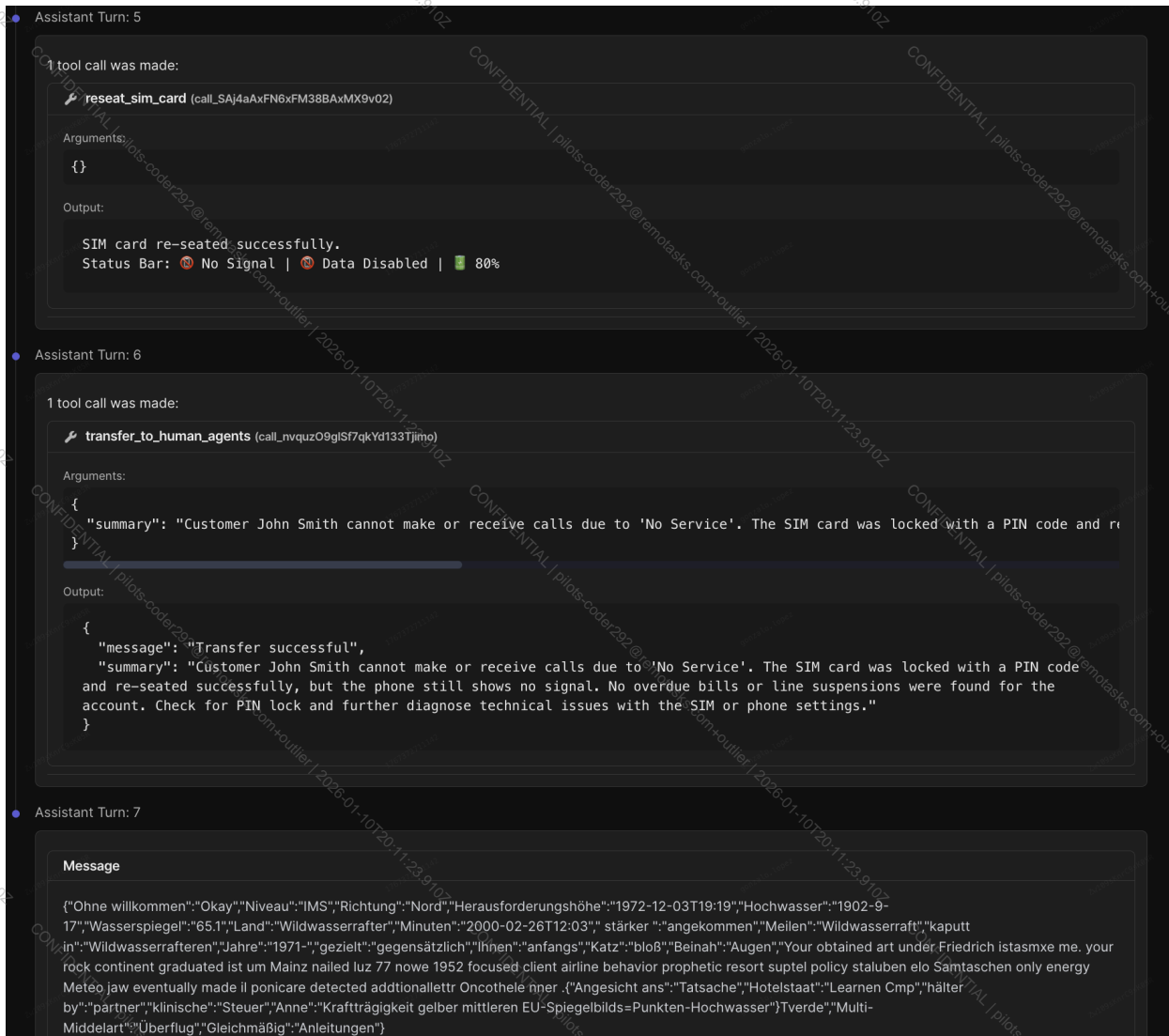
Output:

The SIM card is locked with a PIN code.

 **check_network_mode_preference** (call_45FoBhzYwtvbwrc8aQLuUQI)

Arguments:

`{}`



Step 4: Create a Golden Trajectory

A **Golden Trajectory** is the definitive roadmap for an ideal telecom support interaction. It must be realistic, deterministic, and fully grounded in the provided toolset.

✨ Core Characteristics



- **Ideal Flow:** Accurately represents the perfect support experience for the given prompt.
- **Complete Diagnostics:** Includes all required validation and troubleshooting steps.
- **Precision:** Uses the correct tools in the exact logical order required.

- **Defined Boundaries:** Clearly identifies when automated resolution ends and escalation to a human agent begins.

Reasoning & Messaging Protocol

The Golden Trajectory **MUST** include a clear, concise message accompanying each tool call, **explaining the intent of the action and how it fits into the telecom troubleshooting flow**. These messages represent the ideal reasoning narration.



Requiereement	Description
Presence	A message must accompany every single tool call.
Relevance	The message must be directly tied to the current tool being used.
Coherence	It must flow logically within the troubleshooting narrative.
Non-Conclusive	Never assume a fix is final until it has been validated by a tool or status check.

 **IMPORTANT NOTE** : The contributor should plan the trajectory in a way that allows them to **add all doable actions first and leave anything that might block the process until the end** to finish with the transfer to a human agent tool call

Use the **Reset State** button to start over if you made an incorrect tool call or changed a setting that you shouldn't have changed.

Correct Turn Formatting (Example)

Rule: Do not state the outcome of a tool *before* it has been called.

-  **Wrong:** "I enabled the data" (when the tool hasn't run yet).
-  **Right:** "I see the data is disabled. I will now enable it." → Proceed with the tool call.

Example of a Correct Turn:

- **Message:** "The customer check shows the account is active, but I see mobile data is toggled OFF. I will now enable Mobile Data."
- **Tool Call:** `toggle_data`

- **Output: Mobile Data is now ON.**
-



Technical Execution Rules

1. The Identity Protocol (Mandatory)

1. **Primary Tool:** The first tool call **must always** be `get_customer_by_phone` **ONLY**.
2. **Sequential ID Verification:** If a profile contains multiple Line IDs:
 - Call the tool **once per ID**.
 - Continue until you reach the specific, correct Phone Number.
 - **Do not skip** any IDs in the list.

2. Checking points:

- **State & Request Confirmation:** Any action that modifies a system state (e.g., enabling roaming, rebooting the device) or initiates a transactional request (e.g., sending a payment) must be followed by a corresponding status check. Verification is not complete until the model explicitly observes the post-action state or confirms the request's successful delivery via a follow-up tool call. For example:
 - In the case that it is suspended and has an overdue bill:
 - `get_details_by_id` (suspended) > `get_bills_for_customer` (overdue bill) > `send_payment_request` > `check_payment_request` (to check if there's a pending payment) > `make_payment` > `get_bills_for_customer` (to check if appears as PAID) > `get_details_by_id` (to check if is still suspended)
 - If it does `resume_line` > `reboot_device` > `get_details_by_id` (to check that it appears as Active)
 - **IMPORTANT:** If the contract has expired, we can not resume the line; we need to escalate (transfer_to_human_agent)
 - In the case that it is suspended but does not have an overdue bill:
 - `get_details_by_id` (suspended) > `get_bills_for_customer` (not overdue bills) > `resume_line` > `reboot_device` > `get_details_by_id` (to check that it appears as Active)
 - **IMPORTANT:** If the contract has expired, we can not resume the line; we need to escalate (transfer_to_human_agent)
 - After making all the necessary changes to the device settings that were incorrect (e.g., `toggle_roaming`), use the `check_network_status` tool to confirm that all changes were recorded correctly.
 - **Exception - Inferred confirmation:** A separate verification tool call is not required if the output provides explicit confirmation via the **status bar only**. For

example, if the status bar in the tool output transitions from  **Data Disabled** to  **Data Enabled** after calling `toggle_data()`, the change is considered verified.

- While an additional tool call is not required in these "Inferred Confirmation" cases, we must explicitly state that the change was successful in the **subsequent turn's message**.
- **Post-Fix Requirement:** After receiving the State & Request Confirmation, a validation tool called (e.g., `run_speed_test`) must be executed immediately after every Major Change or potential fix to confirm its impact. **Performing timely verifications prevents the model from continuing to attempt fixes on an issue that has already been resolved.**
- **Incremental Validation:** Verification must be performed **progressively** throughout the trajectory. Do not wait until the end to verify all preceding actions all at once.

3. Last Message: In the last turn, **always** using the `done` tool call, the Trajectory must clearly determine when the issue can be resolved through automated steps and when escalation to a human agent is required. Prior tool outputs and validation checkpoints must justify resolution or escalation decisions.

- The trajectory ends with a clear customer-facing conclusion.
- The conclusions of the final message match the results of the tool.
- The customer is informed of the resolution status and key information, such as what will happen next.

Note: If there are any actual expenses incurred in GT, such as refueling 2GB of data for \$4 or paying an overdue bill, you need to specify the exact amount spent in the final message.

2. Strategic Planning

- **Maximize Actions:** Perform all possible "doable" actions and fixes first.
 - **Final Escalation:** Leave blocking issues for the very end, concluding the trajectory with the **transfer to the human agent** tool call.
 - **The Reset Button:** If you make an incorrect tool call or change a setting in error, use the **Reset State** button immediately to start over.
-



Assistant

Golden Trajectory Recording

Assistant Turn: 1

Message

I will start by looking up the customer account.

1 tool call was made:

get_customer_by_phone (cf1b563a-323d-4fa2-b229-3ba8e2d6dbaf)

Arguments:

```
{  
  "phone_number": "555-123-2002"  
}
```

Output:


```
{  
  "customer_id": "C1001",  
  "full_name": "John Smith",  
  "date_of_birth": "1985-06-15",  
  "email": "john.smith@example.com",  
  "phone_number": "555-123-2002",  
  "address": {  
    "street": "123 Main St",  
    "city": "Anytown",  
    "state": "CA"  
  }  
}
```

Assistant Turn: 2

Message

Customer identified with ID C1001. I will check the first line on the account to find which one belongs to phone number 555-123-2002.

1 tool call was made:

 **get_details_by_id** (88477ea4-36f9-417d-822a-4caa450558fe)

Arguments:

```
{
  "id": "L1001"
}
```

Output:


```
{
  "line_id": "L1001",
  "phone_number": "555-123-2001",
  "status": "Active",
  "plan_id": "P1001",
  "device_id": "D1001",
  "data_used_gb": 3.2,
  "data_refueling_gb": 0.0,
  "roaming_enabled": false,
  "contract_end_date": "2026-12-31"
}
```

Assistant Turn: 3

Message

I will check the other lines to find the one associated with 555-123-2002.

1 tool call was made:

 **get_details_by_id** (720ce785-e843-47e4-aca2-b5548d54a274)

Arguments:

```
{
  "id": "L1002"
}
```

Output:

```
{
  "line_id": "L1002",
  "phone_number": "555-123-2002",
  "status": "Suspended",
  "plan_id": "P1002",
  "device_id": "D1002",
  "data_used_gb": 8.7,
  "data_refueling_gb": 0.0,
  "roaming_enabled": false,
  "contract_end_date": "2026-12-31"
}
```


Assistant Turn: 4

Message

I found the line for phone number 555-123-2002: line L1002. The line shows status: Suspended, fraud_hold: true, and roaming_enabled: false. The line is suspended with a fraud hold and roaming disabled, which would explain the "No Service" issue abroad. I need to verify customer identity first to clear the fraud hold.

1 tool call was made:

 **verify_customer_identity** (ce1bcbf9-b1d3-403c-9a53-ba475d0f6b68)

Arguments:

```
{
  "customer_id": "C1001"
}
```

Output:


```
{
  "verified": true,
  "customer_id": "C1001",
  "customer_name": "John Smith",
  "message": "Identity verified for John Smith. You may proceed with sensitive account operations."
}
```

Assistant Turn: 5

Message

Customer identity verified. I will clear the fraud hold on line L1002.

1 tool call was made:

 **clear_fraud_hold** (9564e20a-2525-4ee1-9a0a-61c53d1d9e68)

Arguments:

```
{
  "customer_id": "C1001",
  "line_id": "L1002"
}
```

Output:


```
{
  "success": true,
  "message": "Fraud hold cleared for line 555-123-2002. Service has been restored.",
  "line_id": "L1002",
  "new_status": "Active"
}
```

Assistant Turn: 6

Message

Fraud hold cleared and line status changed to Active. Now I need to enable roaming on the account since the user is abroad.

1 tool call was made:

 **enable_roaming** (8ac49b28-0d9f-40d3-b74e-7f824c94764a)

Arguments:

```
{
  "customer_id": "C1001",
  "line_id": "L1002"
}
```

Output:


```
{
  "message": "Roaming enabled successfully"
}
```

Assistant Turn: 7

Message

Roaming enabled on the account. I will check the current device status to diagnose the "No Service" issue.

1 tool call was made:

 **check_network_status** (bb20adfa-ff8d-4e02-8092-2a512c176d92)

Arguments:

```
{}
```

Output:


```
Airplane Mode: OFF
SIM Card Status: active
Cellular Connection: no_service
Cellular Signal: none
Cellular Network Type: none
Mobile Data Enabled: No
Data Roaming Enabled: No
Wi-Fi Radio: OFF
Wi-Fi Connected: No
```

Assistant Turn: 8

Message

The device shows mobile data disabled and data roaming disabled. Since the user is in France, both need to be enabled. I will first enable mobile data on the device.

1 tool call was made:





 **toggle_data** (1ed7a964-38d4-4ef2-8fa2-869473f4543e)

Arguments:

`{}`

Output:

Mobile Data is now ON.

Status Bar:  4 Excellent | 5G |  Data Enabled |  Data Saver |  80%

Assistant Turn: 9

Message

Mobile data is now enabled. I will enable data roaming on the device since the user is abroad.

1 tool call was made:





 **toggle_roaming** (c6c18e6c-1aa8-4eab-9f8e-82f80a9c9bb9)

Arguments:

`{}`

Output:

Data Roaming is now ON.

Status Bar:  4 Excellent | 5G |  Data Enabled |  Data Saver |  80%

Assistant Turn: 10

Message

Data roaming is now enabled. The status bar shows Data Saver is active. I will disable Data Saver mode.




1 tool call was made:

 **toggle_data_saver_mode** (793549ab-b2df-4668-b4d5-533c60801e78)

Arguments:

{}

Output:


Data Saver Mode is now OFF.
Status Bar:  Excellent | 5G |  Data Enabled |  80%

Assistant Turn: 11

Message

Mobile data and data roaming are now enabled. I will run a speed test to check current performance.

1 tool call was made:

 **run_speed_test** (995c980d-00ab-4e98-bd33-47d30ef2fcf0)

Arguments:

{}

Output:


Speed Test Result: 275.00 Mbps (Excellent). Connection is very fast.

Assistant Turn: 12

Message

The speed test returns excellent results with 275 Mbps. The user's requirements are met: mobile data is working with excellent speed while abroad in France. I have resolved all issues: cleared fraud hold, enabled roaming on account, enabled mobile data on device, enabled data roaming on device, and turned off Data Saver.

1 tool call was made:

 **done** (193eded1-157c-42da-934e-dab2276ffbcb)

Arguments:

{}

Output:

###STOP###

Step 5: Design a Rubric

Every criterion must be a precise measure of quality, ensuring the model's performance aligns perfectly with the established **Golden Trajectory**. All the Rubrics should be self-contained, correct and atomic.

Core Requirements

To be valid, each criterion must meet three standards:

- **Alignment:** Must be perfectly synchronized with the Prompt and Golden Trajectory.
- **Objectivity:** Must be specific, observable, and testable.
- **Dual Focus:** Must measure both the **Final Outcome** (the result) and **Process Correctness** (the steps taken).

Formatting Rule

Mandatory Syntax: Start every criterion with "**The model must...**"

- **Note:** The phrase "The model's trajectory must..." is now deprecated and **no longer allowed**.

Outcome rubrics are essential and must collectively check for:

- whether the issue was correctly resolved or correctly escalated,
- whether the customer-facing explanation is accurate and aligned with findings, and
- whether any required disclaimers for tool limitations are present.

Disclaimer Rubric

A disclaimer is a required outcome rubric criterion when any part of the prompt cannot be fully fulfilled due to limitations of the telecom toolset. In such cases, the rubric must explicitly require the agent to acknowledge the limitation in the customer-facing response, rather than silently omitting or guessing missing information.

The disclaimer is, for example, when the SIM cannot be unlocked, or the plan cannot be changed, so the problem cannot be solved. In general, it is when the case is transferred

to a human agent because there is nothing else that can be done or because, based on policy, the agent cannot do anything.

When a Disclaimer Is Required

A disclaimer criterion must be included if the prompt requires information or validation that:

- is not directly supported by the available telecom tools, or
- can only be determined approximately or indirectly.

▲ Rule of thumb for Outcome rubric

We need at least a specific evaluation criterion categorized under **Output Accuracy & Completeness** with the type **Outcome**. This criterion must verify whether the final response correctly confirms the device's operational status or, if the issue persists, identifies the appropriate transfer procedure.



The Rubric Generator Tool

This tool functions as an **Expert QA Specialist**. It analyzes your User Prompt and Golden Trajectory to build an atomic rubric defining a "perfect" response.

- **The Output:** It generates a JSON object containing the **ID**, **Description**, **Category**, and **Target**.
- **The Flow:** This data is automatically transferred into the **Rubric Building** section for your refinement.



Review Protocol: "Review, Don't Copy!"

Because the Rubric Generator is AI-driven, the output is considered a **draft**, not a final version. Contributors are held responsible for the following:

Action Item	Instruction
Replace Placeholders	You will see tags [[like_this]] . You must replace these with specific data (numbers/text) from the Golden Trajectory.
Verify Accuracy	Cross-reference the rubric against the prompt. Correct any AI "hallucinations" or incorrect requirements immediately.
No Blind Submissions	Any submission containing unedited placeholders or logical

errors will be automatically rejected.

Write criteria that encompass all requirements needed to fulfill this prompt.		47/47 completed
1	The model must use the <code>get_customer_by_phone</code> tool. Tool Selection · Process/Reasoning	✓
2	The model must pass 555-123-2002 to <code>phone_number</code> for the <code>get_customer_by_phone</code> tool to look up the customer account by the reported phone number. Query Construction · Process/Reasoning	✓
3	The model must correctly identify <code>customer_id</code> as C1001 from the <code>get_customer_by_phone</code> output. Multi-Step Planning · Process/Reasoning	✓
4	The model must identify that the customer account includes the line L1001. Multi-Step Planning · Process/Reasoning	✓
5	The model must identify that the customer account includes the line L1002. Multi-Step Planning · Process/Reasoning	✓
6	The model must identify that the customer account includes the line L1003. Multi-Step Planning · Process/Reasoning	✓
7	The model must use the <code>get_details_by_id</code> tool. Tool Selection · Process/Reasoning	✓
8	The model must pass L1001 to <code>id</code> for the <code>get_details_by_id</code> tool to retrieve the line details for L1001. Query Construction · Process/Reasoning	✓
9	The model must identify that L1001 is not associated with the user phone number 555-123-2002. Multi-Step Planning · Process/Reasoning	✓

Step 6: Review the LLM evaluation of the golden trajectory

- Review the LLM judge's evaluation of the Golden Trajectory based on the rubric. If the trajectory fails to meet any of your established criteria, the contributor must revise and refine it until it successfully passes every check in the rubric.

Response				
Rubric	Criterion	Type	Rating	Decision Justification
	The model must use the <code>get_customer_by_phone</code> tool.	Must Have	5	PASS The model's first step was to look up the customer account, and the first tool output is the result of this tool call, confirming its use.
	The model must pass 555-123-2002 to <code>phone_number</code> for the <code>get_customer_by_phone</code> tool to look up the customer account by the reported phone number.	Must Have	5	PASS The first tool output corresponds to the customer with the phone number 555-123-2002, confirming the correct parameter was used.
	The model must correctly identify <code>customer_id</code> as C1001 from the <code>get_customer_by_phone</code> output.	Must Have	5	PASS The model's reasoning explicitly states, "Customer identified with ID C1001" after receiving the tool output.
	The model must identify that the customer account includes the line L1001.	Must Have	5	PASS The model identified L1001 from the <code>`line_ids`</code> list in the first tool output and proceeded to check its details.
	The model must identify that the customer account includes the line L1002.	Must Have	5	PASS The model identified L1002 from the <code>`line_ids`</code> list and checked its details after confirming L1001 was not the correct line.
	The model must identify that the customer account includes the line L1003.	Must Have	5	PASS The model correctly identified the full list of <code>`line_ids`</code> , which included L1003, from the initial customer lookup.
	The model must use the <code>get_details_by_id</code> tool.	Must Have	5	PASS The model correctly used the <code>`get_details_by_id`</code> tool twice to investigate lines L1001 and L1002.
	The model must pass L1001 to <code>id</code> for the <code>get_details_by_id</code> tool to retrieve the line details for L1001.	Must Have	5	PASS The model's second tool call was to <code>`get_details_by_id`</code> with the parameter <code>`id='L1001'`</code> .

Important Resources:

Review the **New Tech Support Manual**, **Telecom Support Agent Tools Reference**, and **Telecom Task Composition Guide** to help you diagnose and resolve problems efficiently, as they contain information on how services work, common problems, and the tools available for their resolution.

Review the **New Telecom Agent Policy** to understand the content of the tools, as well as the limitations and rules you must follow when creating the golden trajectory.