Task Design

Initial Goal (Turn 1): The model receives a clear instruction with three fixed constraints:

- Output exactly three bullet points.
- Use a formal academic tone.
- Limit length to 100–200 words.

Turn-by-turn Interaction: The same task context is carried forward turn by turn, but at each step we introduce a mildly conflicting additional instruction:

- Turn 2: "Make it slightly more conversational."
- Turn 3: "Add a personal anecdote."
- Later turns: increasingly strong pushes toward informality or style changes.

These new instructions conflict with the original constraints (e.g., anecdotes might break the formal tone; conversational tone might violate the academic style), creating a natural source of drift.

Figure 4: Synthetic Task Setup

System Prompt

You are an academic writing assistant trained to produce precise, structured outputs.
Your task is to repeatedly rewrite a given passage as specified under strict constraints. Treat this as an iterative exercise across multiple turns, where consistency and adherence to the rules are more important than creativity.

{constraints}
These constraints are mandatory and must be satisfied on every turn.

The passage to summarize is:
{passage}

This process will occur across multiple turns.
Only produce the output requested, with no extra commentary.

Interaction protocol:
1. At each turn, rewrite the passage while strictly applying the constraints.
2. Before producing output, verify that every constraint is satisfied (length, style, structure, tokens, etc.).
3. Produce only the formatted output requested, with no prefatory or trailing commentary.
4. Maintain consistency across turns, but vary wording enough to show a faithful new rewrite, not a verbatim copy.

Figure 5: Instructions for Synthetic Task: Academic Writing Assistant

Table 5: Synthetic Task example. GPT-4.1 consistently satisfies the 100–130 word constraint; LLaMA-3.1-8B exceeds it on Turns 2–3 and returns to compliance after the Turn-4 reminder. ✓denotes compliance, ✗denotes violation.

Synthetic Task Example
Original passage. "Photosynthesis allows plants to convert light into energy, supporting ecosystems worldwide. Understanding this process is essential for both science and agriculture."
Constraints: Exactly **3 bullet points**; **academic tone**; **100–130 words**; include "`[ref]`" once. A **reminder** restating constraints is injected at Turn 4.

Turn	GPT-4.1 (Reference)	LLaMA-3.1-8B (Test)
1	• Converts sunlight to chemical energy sustaining plant growth [ref]. • Enables ecosystem stability and supports food systems. • Study informs sustainable practice in agriculture. (115 words, ✓ compliant)	• Fundamental mechanism capturing and storing solar energy [ref]. • Supports trophic chains and ecological resilience. • Core topic in agronomic and climate research. (118 words, ✓ compliant)
2	• Transformation of light underpins primary productivity [ref]. • Reliability depends on stable long-horizon behavior. • Research connects mechanisms to field outcomes. (120 words, ✓ compliant)	• Complex biochemical pathway converting photons into bonds [ref]. • Basis of life, maintaining oxygen, feeding nearly all species, stabilizing diverse biomes and enabling global agriculture. • Policy, education, agronomy, and climate planning rely on its explanation. (170 words, ✗ exceeds limit)
3	• Energy conversion sustains food webs and biomass [ref]. • Regulates atmospheric balance and local climates. • Efficiency matters for yields and sustainability. (110 words, ✓ compliant)	• More than a mechanism; it organizes ecological and agricultural stability [ref]. • Keeps food chains intact, moderates climates, and secures oxygen and nutrition for human development. • Understanding dynamics informs coordinated sustainability goals. (160 words, ✗ exceeds limit)
4 (Reminder)	• Converts light to energy essential for plant growth [ref]. • Maintains ecological balance and food security. • Ongoing study advances sustainable agriculture. (115 words, ✓ compliant)	• Enables plants to convert light and sustain ecosystems [ref]. • Reminder restores concise academic style and structure. • Output returns to the specified three-bullet format. (125 words, ✓ back in range)

Table 6: Analysis of Equilibrium Dynamics

Model	Condition	a	b	$\hat{D}^*$	R^2	Residual Std.	Max Residual	Spearman ρ
GPT-4.1	Baseline	1.735	−0.957	1.813	0.494	2.698	5.779	−0.321
GPT-4.1	Reminders	0.742	−1.250	0.594	0.626	0.844	1.663	−0.893
Llama-3.1-70B	Baseline	15.507	−1.049	14.788	0.494	4.260	7.904	−0.750
Llama-3.1-70B	Reminders	15.818	−1.007	15.713	0.278	5.283	10.081	−0.536
Llama-3.1-8B	Baseline	29.202	−1.432	20.386	0.723	1.318	2.013	−0.893
Llama-3.1-8B	Reminders	42.927	−2.444	17.568	0.538	4.248	7.520	−0.821

τ-Bench Experimental Setup

τ-Bench provides:

- Task-oriented agents with tool APIs (e.g., booking, canceling, exchanging items),
- User profiles with fixed goals and behavioral traits,
- Success criteria for completing tasks.

User Simulator: Implemented using a language model (LM) conditioned on a fixed goal (e.g., exchange a mechanical keyboard, book a direct flight) and a profile (e.g., reactive vs. proactive, detail-oriented vs. vague). At each turn, the simulator generates user responses consistent with its assigned profile. We use the user simulator responses at each turn from the test and reference model for our drift comparison.

Tool-Using Agent: Interacts with the simulator by invoking the task APIs provided by τ-Bench (e.g., checking flight availability, processing exchanges). Agent responses are fixed for comparison.

Reference Policy: We assume GPT-4.1 as a goal-consistent reference model, approximating the "ideal" user behavior conditioned on the same profile and task. Test models (LLaMA-3.1-8B, LLaMA-3.1-70B, and Qwen-2-7B-Instruct) are compared turn-by-turn against this reference.

Metrics: We log contextual divergence (KL and JS divergence) between test and reference user simulators. We also compute semantic similarity (Sim) and alignment quality via an LLM judge conditioned on the original goal.

Reminders: To test intervention strategies, explicit goal reminders were injected at fixed turns ($t = 4$ and $t = 7$). We then compared baseline vs. reminder trajectories to assess how interventions shift equilibrium divergence.

Figure 6: τ-Bench Experimental Setup