Task #20

Testing Prompt Robustness with Combinations of Contextual Elements
US #17 Identify Essential Contextual Elements
Author: Aniket Patil
ASU ID: 1233384063

1. Objective

Task #19 (Aniket Patil) established the impact of removing individual contextual elements from the LLM prompt. Task #20 extends this analysis by testing combinations of elements together to determine the optimal prompt composition the minimal set of contexts that produces high-quality, natural auto-responses while minimizing privacy exposure and hallucination risk.

This directly addresses the key research question: How much context is necessary vs sufficient for high-quality auto-responses?

2. Methodology

Building on the baseline prompt established in Task #19, eight combination tests were designed to explore different subsets of contextual elements. The combinations were chosen to test specific hypotheses:

- **Combos 1–3:** Progressive enrichment starting minimal and adding elements to find the quality inflection point
- **Combos 4–5:** Category isolation testing whether environmental or social context alone can produce acceptable responses
- **Combo 6:** Privacy-optimized testing the minimum context needed while minimizing data exposure
- **Combo 7:** Full baseline control test with all elements included
- **Combo 8:** Sender variation testing how sender relationship changes response style with the same core context

The auto-response is generated by the system, speaks about the user in third person, and does not imitate the user directly consistent with the approach in Task #19.

3. Baseline Prompt (from Task #19)

The following baseline prompt includes all contextual elements and serves as the control:

User name: Satyam
Current time: 2:15 PM, Tuesday
Location: Office – Meeting Room
Activity: In a scheduled team meeting
Phone state: Locked, on silent mode
Noise level: Moderate
Sender: Manager (high priority contact)
Message urgency: Medium

Expected response time: After meeting (~30 minutes)

Prompt instruction: Generate a polite auto-response on behalf of Satyam explaining temporary unavailability.

4. Combination Tests

Combo 1: Minimal Core: Activity + Time Only

Included: User name, Current time, Activity

Removed: Location, Phone state, Noise level, Sender role, Urgency, Expected response time

Context Provided:

User name: Satyam
Current time: 2:15 PM, Tuesday
Activity: In a scheduled team meeting

Generated Response:

AUTO-RESPONSE: Satyam is currently in a meeting and may not be able to respond right away. He'll get back to you once he's available.

Observations:

- Response is coherent and socially acceptable
- No mention of when user will respond lacks expectation setting
- Tone is generic, not adapted to sender relationship
- Sufficient for a basic auto-response

Quality Rating: Acceptable

Combo 2: Core + Social: Activity + Time + Sender Role + Urgency

Included: User name, Current time, Activity, Sender role, Urgency

Removed: Location, Phone state, Noise level, Expected response time

Context Provided:

User name: Satyam
Current time: 2:15 PM, Tuesday
Activity: In a scheduled team meeting
Sender: Manager (high priority)
Message urgency: Medium

Generated Response:

AUTO-RESPONSE: Satyam is currently in a scheduled team meeting and is unable to respond at the moment. He will review your message as soon as the meeting concludes and respond promptly.

Observations:

- Tone is noticeably more professional due to manager context
- “Respond promptly” reflects urgency awareness
- Still missing specific time estimate
- Significant quality improvement over Combo 1

Quality Rating: Good

Combo 3: Optimal Set: Activity + Time + Sender + Urgency + Response Window

Included: User name, Current time, Activity, Sender role, Urgency, Expected response time

Removed: Location, Phone state, Noise level

Context Provided:

User name: Satyam
Current time: 2:15 PM, Tuesday
Activity: In a scheduled team meeting
Sender: Manager (high priority)
Message urgency: Medium
Expected response time: After meeting (~30 minutes)

Generated Response:

AUTO-RESPONSE: Satyam is currently in a scheduled team meeting and is unable to respond at this time. He is expected to be available in approximately 30 minutes and will review your message promptly after the meeting concludes.

Observations:

- Near-identical quality to the full baseline from Task #19
- Specific time estimate sets clear expectations
- Professional tone matches sender relationship
- No meaningful loss from removing location, device state, and noise

Quality Rating: Excellent

Combo 4: Environmental Only: Location + Noise + Device State (No Activity)

Included: User name, Current time, Location, Phone state, Noise level

Removed: Activity, Sender role, Urgency, Expected response time

Context Provided:

User name: Satyam
Current time: 2:15 PM, Tuesday
Location: Office – Meeting Room
Phone state: Locked, on silent
Noise level: Moderate

Generated Response:

AUTO-RESPONSE: Satyam's phone is currently on silent and he may not see your message right away. He appears to be occupied at the moment and will respond when he is able to check his device.

Observations:

- Response is vague "appears to be occupied" is weak
- LLM attempts to infer activity from environmental cues but lacks confidence
- No mention of when response is expected
- Device state does not compensate for missing activity context

Quality Rating: Poor

Combo 5: Social Only: Sender Role + Urgency (No Activity, No Time)

Included: User name, Sender role, Urgency

Removed: Current time, Activity, Location, Phone state, Noise level, Expected response time

Context Provided:

User name: Satyam
Sender: Manager (high priority)
Message urgency: Medium

Generated Response:

AUTO-RESPONSE: Satyam is currently unavailable and unable to respond to your message. He will get back to you as soon as possible.

Observations:

- Tone is appropriately professional (manager context works)
- But "unavailable" with no reason feels dismissive
- No time framing at all when is "as soon as possible"?
- Social context alone cannot produce a quality response

Quality Rating: Poor

Combo 6: Privacy-Minimal: Activity + Expected Response Time Only

Included: User name, Activity, Expected response time

Removed: Current time, Location, Phone state, Noise level, Sender role, Urgency

Context Provided:

User name: Satyam
Activity: In a scheduled team meeting
Expected response time: After meeting (~30 minutes)

Generated Response:

AUTO-RESPONSE: Satyam is currently in a meeting and will not be able to respond right away. He should be available in about 30 minutes.

Observations:

- Clean, informative response with minimal data exposure
- Answers both "why unavailable" and "when will they respond"
- Tone is neutral no sender-based adaptation
- Excellent privacy-to-quality ratio

Quality Rating: Good

Combo 7: Maximum Context: All Elements (Full Baseline)

Included: All elements: User name, Current time, Location, Activity, Phone state, Noise level, Sender role, Urgency, Expected response time

Removed: None

Context Provided:

User name: Satyam
Current time: 2:15 PM, Tuesday
Location: Office – Meeting Room
Activity: In a scheduled team meeting
Phone state: Locked, on silent mode
Noise level: Moderate
Sender: Manager (high priority contact)
Message urgency: Medium
Expected response time: After meeting (~30 minutes)

Generated Response:

AUTO-RESPONSE: Satyam is currently in a scheduled team meeting and his phone is on silent. He is expected to be available in approximately 30 minutes and will review your message promptly once the meeting concludes.

Observations:

- Highest quality response but only marginally better than Combo 3
- Device state mention ("phone is on silent") adds minor value
- Location and noise contributed nothing to the response text
- Diminishing returns clearly visible beyond the optimal set

Quality Rating: Excellent

Combo 8: Casual Sender Variant: Activity + Time + Friend Sender

Included: User name, Current time, Activity, Sender role (friend), Expected response time

Removed: Location, Phone state, Noise level, Urgency

Context Provided:

User name: Satyam
Current time: 2:15 PM, Tuesday
Activity: In a scheduled team meeting
Sender: Close friend
Expected response time: After meeting (~30 minutes)

Generated Response:

AUTO-RESPONSE: Hey! Satyam’s in a meeting right now and can’t check his phone. He should be free in about 30 minutes and will get back to you then!

Observations:

- Tone shifts dramatically casual, friendly, uses “Hey!”
- Same core information as Combo 3 but different delivery
- Demonstrates that sender role affects style, not just content
- Validates sender role as a high-impact element for personalization

Quality Rating: Excellent

5. Comparative Results Summary

Combination	Elements Included	Quality	Key Finding
Combo 1	User name, Current time, Activity	**Acceptable**	Functional but vague
Combo 2	User name, Current time, Activity, Sender role, Urgency	**Good**	Social context adds tone
Combo 3	User name, Current time, Activity, Sender role, Urgency, Expected response time	**Excellent**	Optimal balance found
Combo 4	User name, Current time, Location, Phone state, Noise level	**Poor**	Environmental alone fails
Combo 5	User name, Sender role, Urgency	**Poor**	Social alone insufficient
Combo 6	User name, Activity, Expected response time	**Good**	Best privacy-quality ratio
Combo 7	All elements: User name, Current time, Location, Activity, Phone state, Noise le...	**Excellent**	Marginal gain over Combo 3
Combo 8	User name, Current time, Activity, Sender role (friend), Expected response time	**Excellent**	Sender drives tone shift

6. Key Findings

6.1 The Optimal Prompt Set

Combo 3 (Activity + Time + Sender Role + Urgency + Expected Response Time) achieves near-identical quality to the full baseline while using only 5 of 9 available elements. This represents the optimal balance between response quality and data minimization.

6.2 Progressive Enrichment Curve

Response quality follows a clear diminishing returns pattern:

- **Combo 1 (2 elements):** Acceptable basic but functional
- **Combo 2 (4 elements):** Good adding social context significantly improves tone
- **Combo 3 (5 elements):** Excellent adding response time window reaches near-peak quality

- **Combo 7 (9 elements):** Excellent marginal improvement, 4 additional elements contribute almost nothing

6.3 Environmental Context Is Insufficient Alone

Combo 4 demonstrates that location, noise level, and device state cannot compensate for missing activity context. The LLM produced a vague, uncertain response. This validates Task #19's finding that activity is the single most critical element.

6.4 Sender Role Drives Style, Not Just Content

Comparing Combo 3 (manager) with Combo 8 (friend) shows identical information conveyed with dramatically different tone. This confirms that sender role is essential for generating natural, relationship-appropriate responses.

6.5 Privacy-Optimized Configuration

Combo 6 (Activity + Response Time only) achieves "Good" quality with minimal data exposure. This is the recommended configuration for privacy-sensitive users who want functional auto-responses without sharing social or environmental context.

7. Alignment with Task #22 Priority Ranking

These findings empirically validate the priority ranking established in Task #22:

Task #22 Tier	Combo Test Result	Validation
Tier 1 (Critical): Activity, Time, Reason, Location	Combos 1, 4	Activity confirmed as most critical. Location is redundant when activity is present.
Tier 2 (High): Duration, Sender, Noise, Light, Calendar, Urgency	Combos 2, 3, 8	Sender role and urgency confirmed high-impact. Noise/light confirmed low-impact.
Tier 4–5 (Low/Minimal): Device state, Battery, Weather	Combos 4, 7	Device state, noise confirmed negligible. No quality loss when removed.

8. Recommended Prompt Structure for Stage 2 API

Based on the combination testing results, the following prompt structure is recommended for the Stage 2 API pipeline:

Required (Always Include)

- User name
- Current activity (inferred from sensors)

- Expected response time / duration

Recommended (Include When Available)

- Sender relationship / role
- Message urgency level
- Current time

Optional (User Preference Dependent)

- Specific location label
- Day of week

Exclude (No Meaningful Impact)

- Raw noise / light levels
- Device lock state
- Battery level
- Network connectivity

9. Conclusion

Task #20 demonstrates that effective auto-responses do not require exhaustive context. Through systematic combination testing, the optimal prompt consists of 5 elements: activity, current time, sender role, urgency, and expected response time. This set achieves excellent response quality while excluding 4 elements that add noise without improving output.

The findings directly inform our Stage 1 prompt engineering experiments and provide a concrete recommendation for the Stage 2 API design: focus the sensor-to-LLM pipeline on extracting and formatting the 5 optimal elements rather than passing all available data. This reduces token cost, minimizes privacy exposure, and avoids LLM over-speculation from irrelevant context.