Priority Ranking of Contextual Elements

US #17 — Identify Essential Contextual Elements

Task #22

Author: Satyam Shekhar

1. Ranking Methodology

This document prioritizes the contextual elements identified in Task #17 (Prerana Kumsi) based on their impact on generating high-quality, natural LLM auto-responses. The ranking is informed by the original "Laila is in a Meeting" research paper, our team's meeting discussions, and the key research questions guiding our Sprint 1 work.

Ranking Criteria

1. **Response Impact:** How much does including this element improve the naturalness and usefulness of the auto-response?
2. **Data Availability:** Is this element available from the sensors and data sources in the Laila system or readily obtainable on Android?
3. **Privacy Risk:** Does increasing specificity introduce unacceptable privacy exposure?
4. **Hallucination Risk:** Does this element help ground the LLM or could it lead to over-speculation?
5. **Abstraction Level:** Is this raw sensor data, mid-level inference, or high-level activity? Higher abstraction is generally more useful for the LLM.

2. Priority Tier Overview

Elements are organized into five tiers reflecting diminishing returns for auto-response quality. Tier 1 elements are essential for any meaningful response, while Tier 5 elements can be safely excluded in most scenarios.

Tier	Count	Description
Tier 1 — Critical	4	Must-have for any response. Without these, the response is meaningless or misleading.
Tier 2 — High	6	Significantly improves naturalness, specificity, and sender satisfaction.
Tier 3 — Medium	7	Adds nuance and personalization. Useful but not essential.
Tier 4 — Low	7	Nice-to-have. Situational value; may add noise if always included.
Tier 5 — Minimal	6	Deprioritize or exclude. Marginal impact, high noise or privacy risk.

3. Detailed Priority Ranking

The table below ranks all 30 contextual elements across the five tiers, with rationale for each placement grounded in the project's research questions.

#	Element	Category	Rationale
TIER 1 — CRITICAL (Must-Have for Any Response)			
1	**Current Activity**	Activity	Core determinant of availability. Drives the primary content of the response (e.g., sleeping, in a meeting, driving).
2	**Current Time**	Temporal	Essential for temporal grounding. Enables responses like "I'll reply in the morning" vs "I'll reply after the meeting."
3	**Why Unavailable (Reason)**	Reason/Intent	Directly answers the sender's implicit question. Without a reason, responses feel evasive or robotic.
4	**Semantic Location Label**	Location	Provides human-readable context (home, work, library) that the LLM can use to infer appropriateness of the response.
TIER 2 — HIGH (Significantly Improves Response Quality)			
5	**Duration / End Time**	Temporal	Enables the LLM to set expectations: "I'll be free in 30 minutes" vs leaving the sender uncertain.
6	**Who Is Contacting**	Social	Relationship context (boss vs friend) allows tone calibration: formal for manager, casual for close friend.
7	**Noise Level**	Environmental	Key sensor from the Laila paper. Helps disambiguate activity (e.g., high noise + calendar event = meeting).
8	**Light Level**	Environmental	Key sensor from the Laila paper. Combined with time and noise, helps infer sleeping, theater, outdoor activity.
9	**Calendar Event Data**	Activity	Structured source of activity, duration, and participants. Directly available from the original Laila system.
10	**Message Urgency**	Reason/Intent	High-urgency messages may warrant a different response (e.g., "I'm unavailable but will check soon").
TIER 3 — MEDIUM (Adds Nuance and Personalization)			
11	**Time of Day (morning/night)**	Temporal	Allows more natural phrasing: "good morning" greetings, "I'll get back to you tomorrow."
12	**Day of Week**	Temporal	Weekend vs weekday affects tone and expected response time.
13	**Sender Priority**	Social	Enables differentiated urgency handling for VIP contacts.
14	**Moving vs Stationary**	Location	Helps distinguish commuting from being at a fixed location.
15	**Screen**	Device	Proxy for phone attention. Locked = likely away from

	Locked/Unlocked		device.
16	**User Tone Preferences**	Behavioral	Personalizes LLM output to match user's communication style (casual, formal, humorous).
17	**Expected Response Time**	Reason/Intent	Helps the LLM set accurate expectations with the sender.
TIER 4 — LOW (Nice-to-Have, Situational Value)			
18	**Phone Silent/Vibrate**	Device	Confirms user likely does not want interruptions, but adds little to response content.
19	**Time Since Last Unlock**	Device	Useful proxy for extended absence, but redundant with activity context.
20	**GPS Location**	Location	Raw coordinates are rarely useful for response text; semantic labels (Tier 1) are preferred.
21	**Indoor vs Outdoor**	Location	Marginal improvement over semantic location; occasionally useful for exercise or commute.
22	**Typical Response Time**	Behavioral	Useful for personalized ETAs but requires historical data collection.
23	**Number of Participants**	Social	Adds specificity ("large meeting") but rarely changes the core response.
24	**Message Topic**	Reason/Intent	Could enable topic-aware responses but risks over-speculation by the LLM.
TIER 5 — MINIMAL (Deprioritize or Exclude)			
25	**Battery Level**	Device	Rarely relevant to auto-response content. Edge case: critically low battery.
26	**Weather Conditions**	Environmental	Almost never affects response content meaningfully.
27	**Headphones Connected**	Device	Weak signal; could mean music, call, or nothing.
28	**Network Connectivity**	Device	System-level concern, not response content.
29	**Crowd Density**	Environmental	Difficult to measure reliably; marginal value over other context.
30	**App Currently Active**	Device	Privacy concern outweighs value; what app is open rarely belongs in a response.

4. Privacy & Sensitivity: Cross-Cutting Filter

Privacy elements from Task #17 are not ranked alongside contextual elements because they function as filters applied on top of the ranking. Regardless of an element's priority tier, privacy rules can suppress or redact it before it reaches the LLM. These elements act as gatekeepers rather than content contributors.

Privacy Element	Category	Role in Pipeline
Confidential Meeting Status	Privacy	Must be checked before any meeting details are disclosed.
Sensitivity of Location	Privacy	Some locations (e.g., hospital, therapist) must never appear in responses.
User-Defined Sharing Rules	Privacy	User overrides take absolute precedence over system defaults.
Data Minimization Settings	Privacy	Controls how much context the LLM is permitted to reveal.

5. Mapping to Key Research Questions

How much context is necessary vs sufficient?

Tier 1 alone (activity, time, reason, location label) provides a minimum viable context for generating a coherent auto-response. Adding Tier 2 elements (duration, relationship, sensor data) moves the response from adequate to natural. Beyond Tier 2, returns diminish and risk introducing noise or hallucination.

What level of abstraction yields the best responses?

High-level inferred activities (Tier 1) are far more useful to the LLM than raw sensor values. For example, "user is in a meeting" is more actionable than "noise: 72dB, light: 340 lux." However, raw sensor values (Tier 2) remain important as inputs to the inference pipeline and for disambiguation.

Does increasing specificity improve clarity or increase privacy risk?

There is a clear inflection point. Specificity up through Tier 2 improves response quality with manageable privacy implications. Tier 3 and below introduce granular details (GPS, app usage, participant count) where the privacy cost may exceed the quality benefit. The privacy filter layer (Section 4) should gate these elements.

How does context selection affect hallucination?

Providing too little context (only Tier 1) may cause the LLM to fill gaps with plausible but incorrect details. Providing too much low-tier context can trigger over-specification, where the LLM mentions details the user would not want shared. The optimal zone is Tier 1 + Tier 2, supplemented selectively by Tier 3 based on user preferences.

6. Recommendation for Stage 1 Prompt Engineering

For our Stage 1 prompt experimentation using the professor's sensor data, the recommended approach is to start with Tier 1 elements as the baseline prompt context and progressively add Tier 2 elements to measure the improvement in response naturalness. This aligns with the meeting discussion to "first try abstract and then make it more specific." The experimental levels would be:

- **Level A (Minimal):** Activity + Time only ("Pranut is not available right now")
- **Level B (Standard):** Tier 1 full ("Pranut is in a meeting and unavailable")
- **Level C (Rich):** Tier 1 + Tier 2 ("Pranut is in a meeting until 6:30 PM and can't respond right now")
- **Level D (Full):** Tier 1 + 2 + 3 ("Hey! Pranut is in a meeting until 6:30 PM. It's a work meeting so he might be a while — he'll get back to you after.")

Comparing response quality across these levels will directly answer the research question of how much context is necessary vs sufficient for high-quality auto-responses.