Project Based Learning Project

Uncertainty-First Agent Council

An Agentic AI System that Explicitly Models Unknowns

Team Member 1
Amitkumar Racha
24070126501

Team Member 2
Bontha Mallikarjun Reddy
23070126026

Team Member 3
Neil Cardoz
23070126079

To
Mentor: Dr. Aniket Shahade
Panel member: Dr. Sumit Kumar

Problem Statement

⚠️The Core Problem

Modern AI chatbots generate confident-looking answers even when information is incomplete, ambiguous, or outdated.

❌ Key Limitations of Existing Systems

- Optimize for fluent answers
- Hide uncertainty from users
- Do not state what they don't know

🇮🇳 Real-World Indian Contexts

- Government scheme eligibility
- Legal pre-screening
- Financial compliance

🔧 Engineering Gap

No systematic mechanism to identify, represent, and communicate uncertainty in AI-generated responses.

Problem Objectives

Primary Objective: Design an AI system that prioritizes **uncertainty awareness** over answer confidence.

1 **Identify Epistemic States**
Explicitly identify **known facts, assumptions, and unknowns** in AI responses

2 **Calibrate Confidence**
Calibrate confidence levels based on available and missing information

3 **Reduce Hallucinations**
Reduce **false confidence and hallucinations** in decision-critical queries

4 **Guide Users Safely**
Guide users safely when information is insufficient

🎯 **Outcome:** Safer, more transparent AI-assisted decision-making

Engineering Context

🔬 Domain Intersection


Artificial Intelligence


Multi-Agent Systems


Decision Support Systems


AI Safety & Reliability Engineering

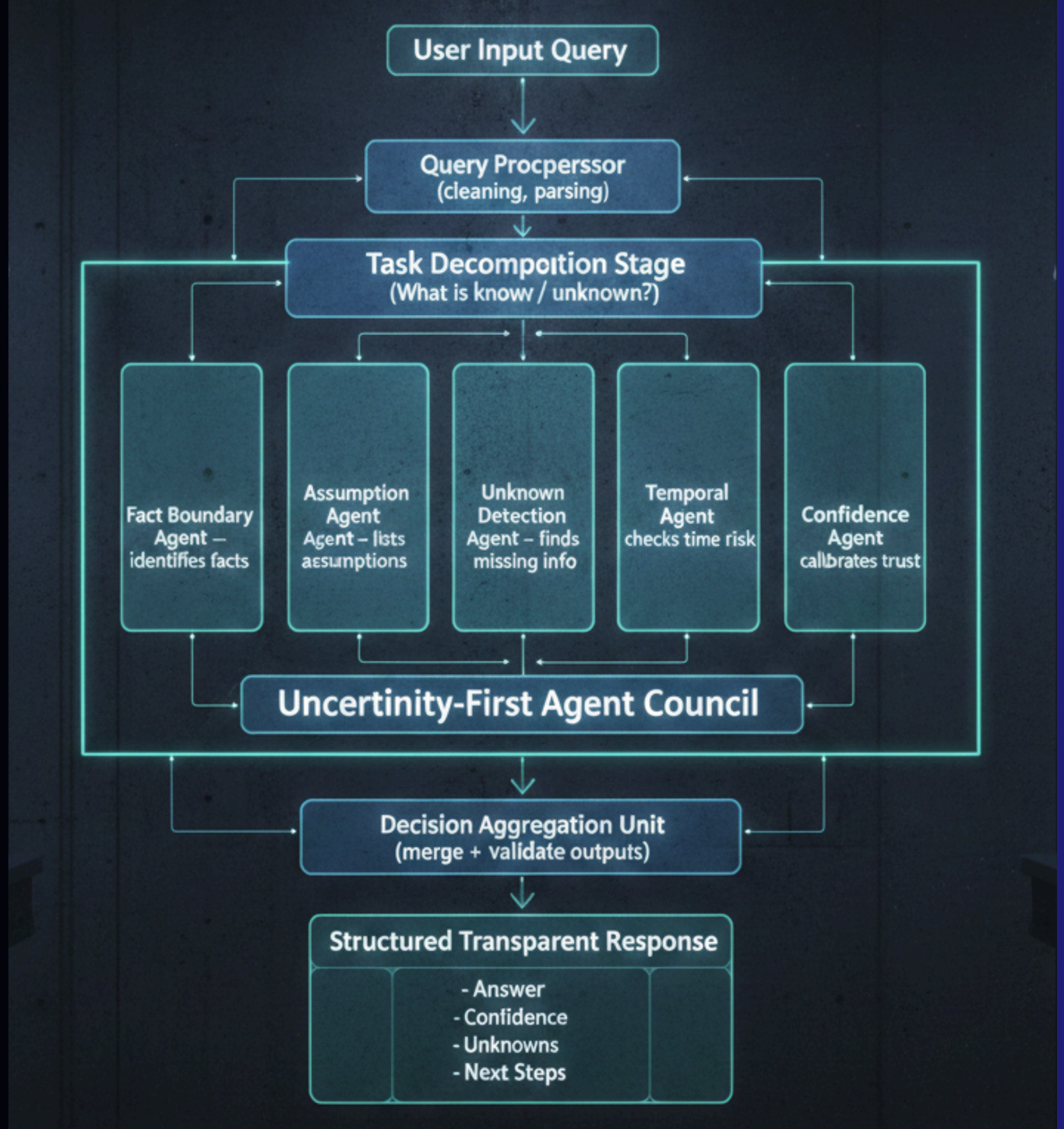
🛠️ Why This Is an Engineering Problem


Black Box Behavior
AI systems behave as opaque black boxes


Systemic Failure
Overconfidence is a system-level failure, not user error

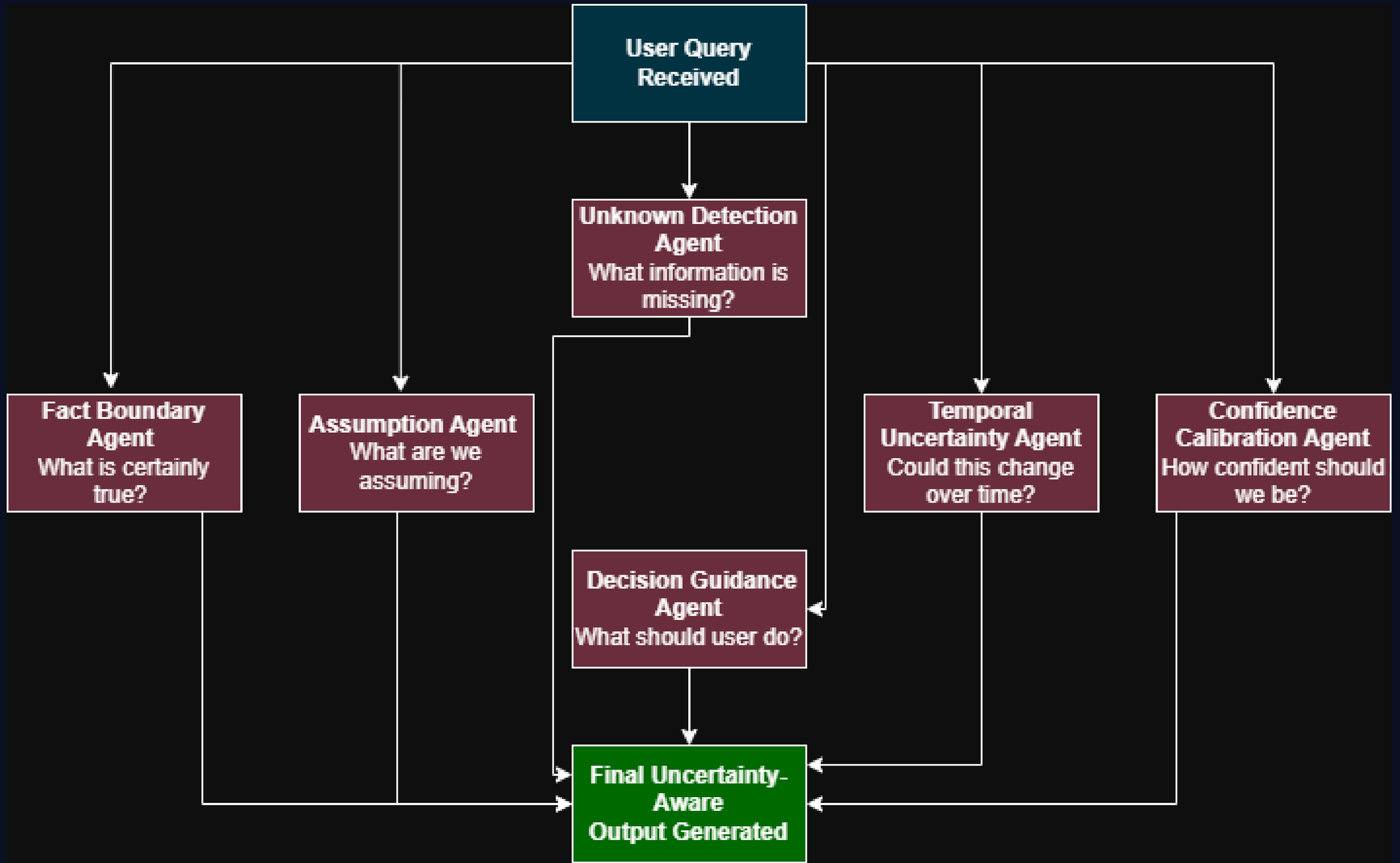

System-Level Design
Requires architectural design, not prompt tweaking

Key Insight: We treat uncertainty as a **first-class engineering variable**



Uncertainty-First Agent Council End-to-End Process

HOW UNCERTAINTY IS EXPLICITLY MODELED



Engineering Variables & Parameters

Category	Variables / Parameters
Input Variables	• Query complexity score • Domain classification • Information completeness
Agent Parameters	• Confidence threshold (0-1) • Agreement weight matrix • Uncertainty propagation factor
System Metrics	• Epistemic uncertainty score • Agent consensus level • Information gap ratio
Output Variables	• Calibrated confidence (0-100%) • Known/Unknown classification • Safety recommendation flag

Why a New Approach Is Needed

- Eligibility rules already exist, but users fail due to missing information and unclear processes
- Government portals assume users are digitally literate and document-ready

- Existing AI/chatbots:

1. Give confident yes/no answers
2. Do not highlight missing or risky information
3. Push users to apply even when failure is likely

The real problem is not eligibility — it is lack of clarity about readiness, risks, and unknowns.

- Identified Need
- An AI system that:

1. Explicitly shows what is known and unknown
2. Warns users when not to proceed
3. Guides users to reduce uncertainty before applying

➡️ This need motivates the Uncertainty-First Agent Council approach.

Summary & Key Takeaways

🎯 What We're Building

A Multi-Agent AI Council that prioritizes uncertainty awareness over confident-sounding answers

🛡️ Why It Matters

Prevents misleading AI responses in critical domains like government schemes, legal, and finance

⚙️ Engineering Approach

Treating uncertainty as a first-class variable with measurable parameters and calibration

📊 Expected Outcome

Safer, transparent AI-assisted decision-making with explicit uncertainty communication