An LLM-based Framework for Human–Swarm Teaming Cognition in Disaster Search and Rescue

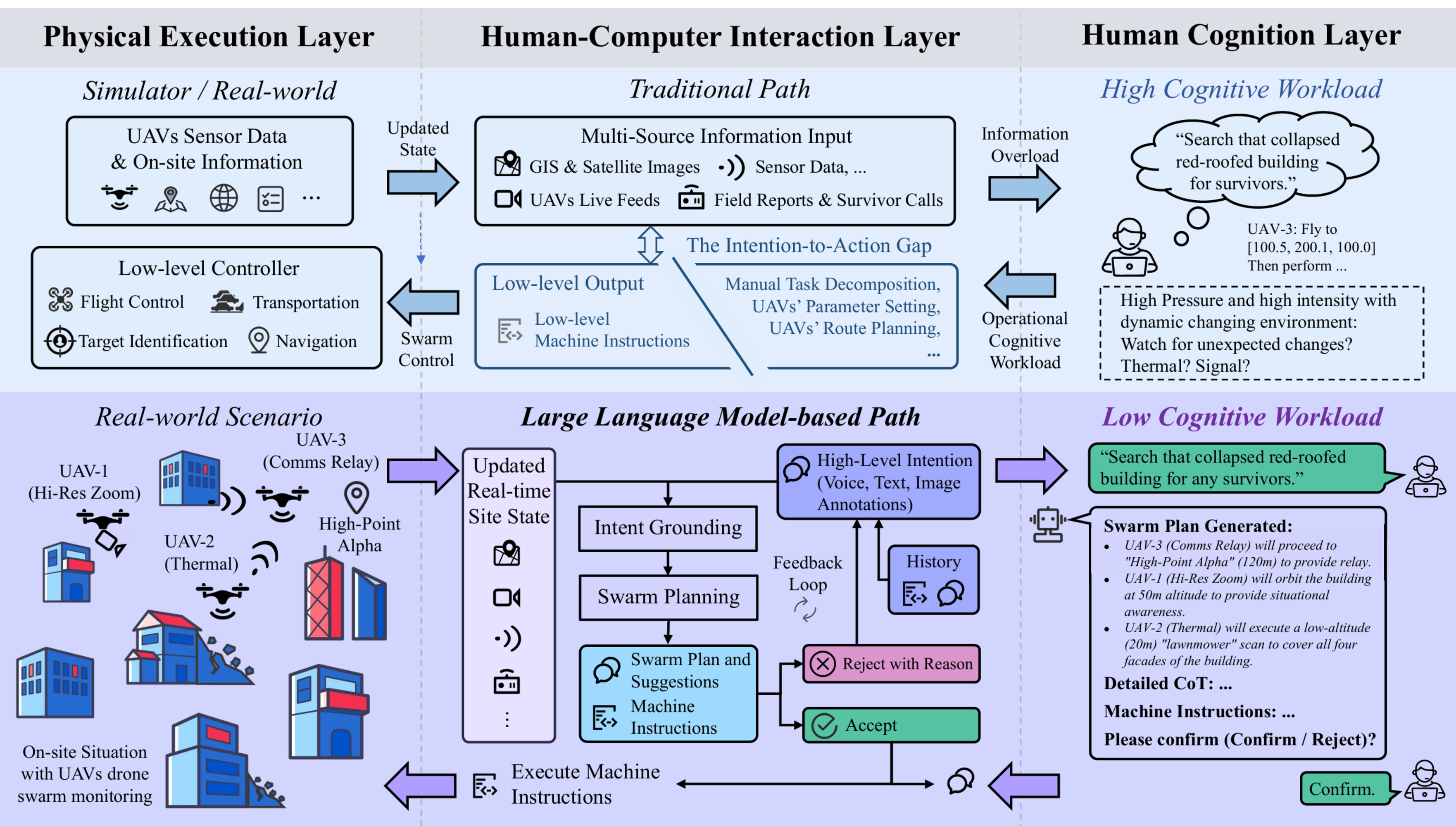
Kailun Ji[1] **Xiaoyu Hu**[1] **Xinyu Zhang**[1, 2, *] **Jun Chen**[1, 2]

[1]Northwestern Polytechnical University [2]Chongqing Institute for Brain and Intelligence *Xinyu.Zhang@nwpu.edu.cn

Paper

Motivation & Contribution

This paper introduces an LLM-CRF approach that translates natural language commands into coordinated actions for UAV swarms, significantly reducing human cognitive load in disaster search and rescue operations.



Motivation

High Cognitive Load in Time-Critical Missions: Operators face extreme pressure within the "golden 72-hour" window. Manual control under disrupted conditions imposes overwhelming cognitive load, leading to inefficiency.

The Intention-to-Action Gap: Translating abstract goals (e.g., "search area") into low-level executable code requires expert programming skills. This dependency makes human-swarm collaboration fragile and unscalable.

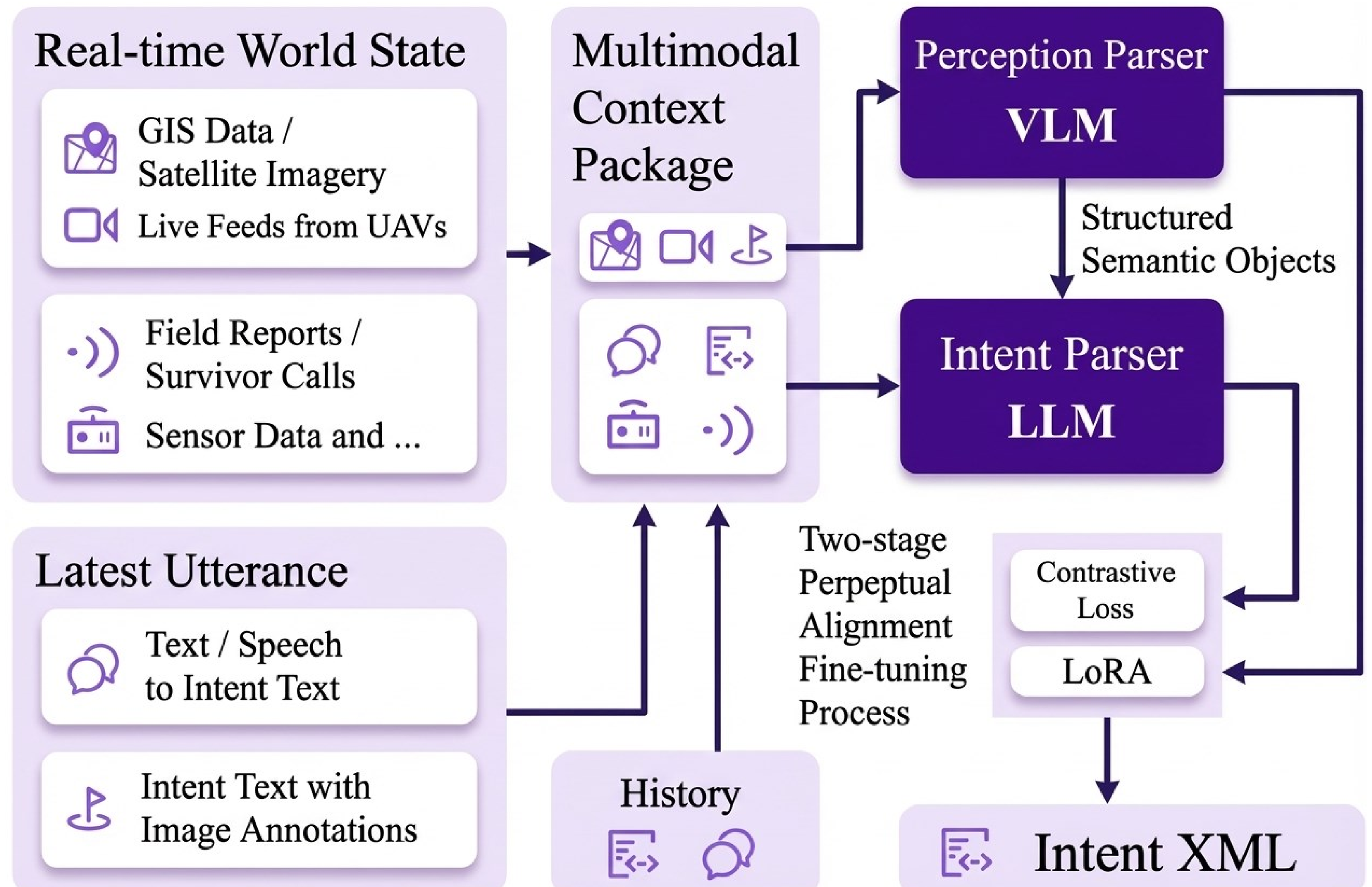
Safety & LLM Hallucinations: Directly deploying LLMs is dangerous due to inherent hallucinations and a lack of physical grounding. In disaster SAR, unverified plans pose severe safety risks to both the swarm and the mission.

Contribution

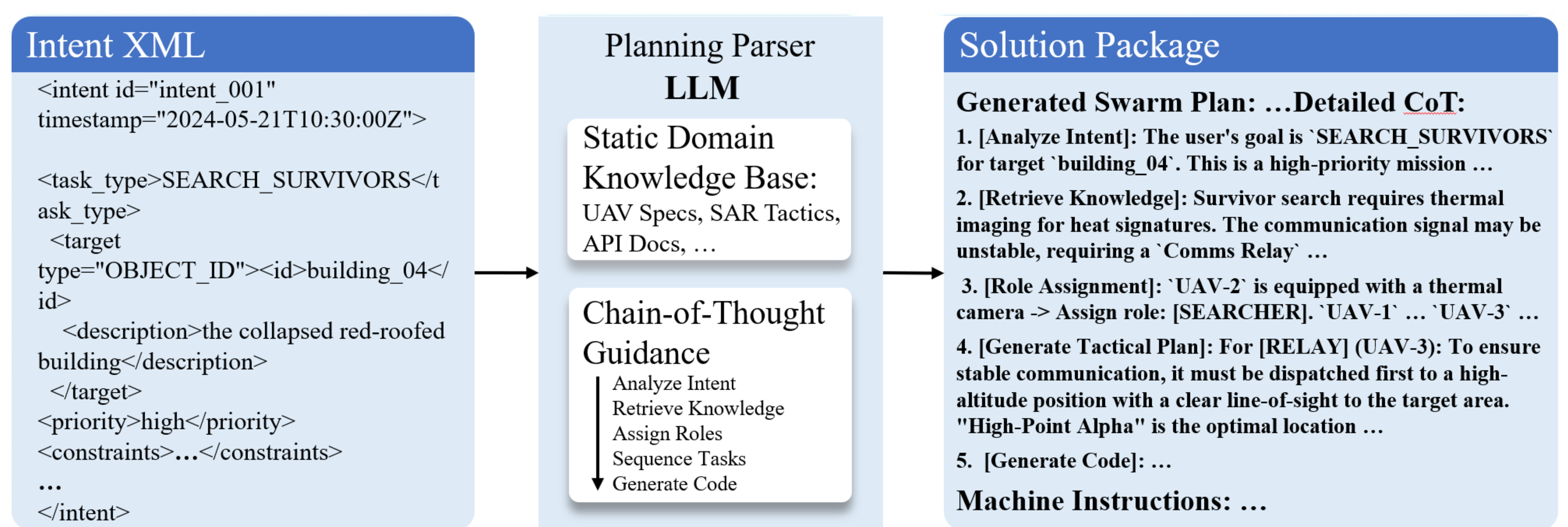
- LLM-CRF Framework: Bridges the gap between human intent and swarm action, significantly reducing operator cognitive load.
- Flexible Task Planning: Utilizes In-Context Learning (ICL) instead of expensive fine-tuning, enabling adaptability to new scenarios without retraining.
- Safety & Auditability: Integrates Human-in-the-loop verification with a Code-as-CoT paradigm, ensuring generated plans are traceable, safe, and reliable.

Methodology

Stage 1: Multi-modal Intent Grounding



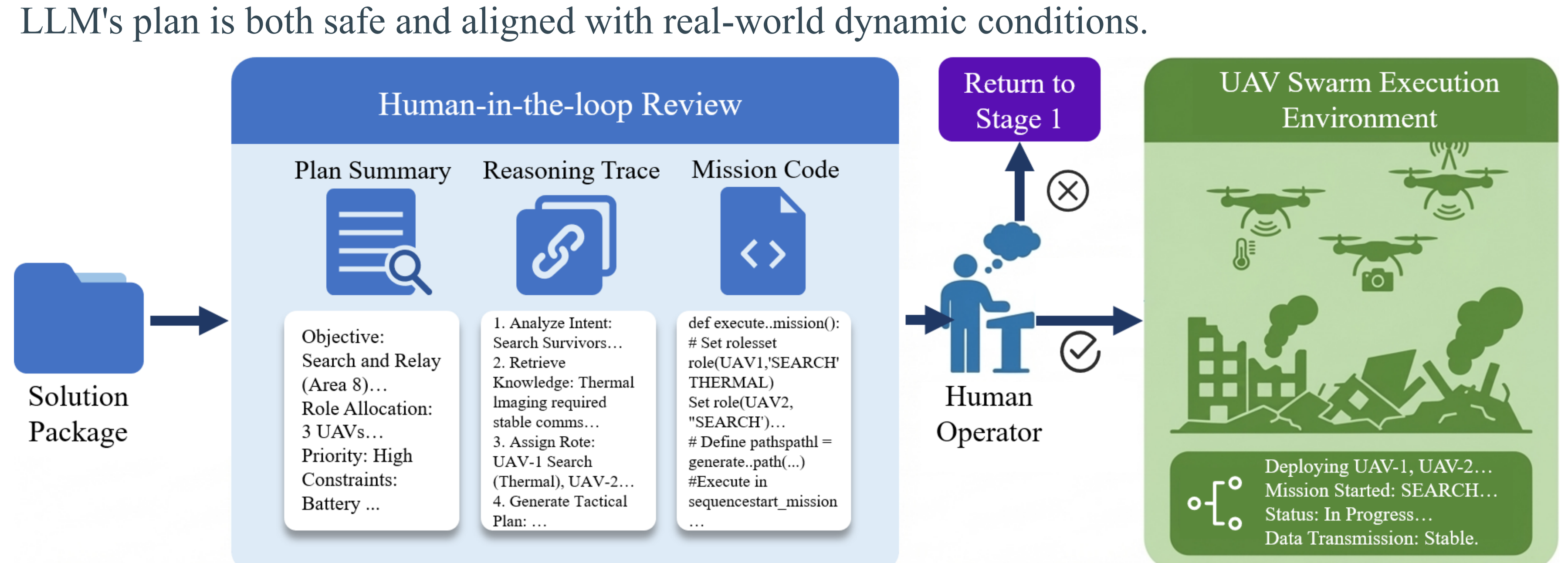
Stage 2: Swarm Task Planning via In-Context Learning



See Appendix for detailed in-context learning pseudocode.

Stage 3: Feedback and Execution

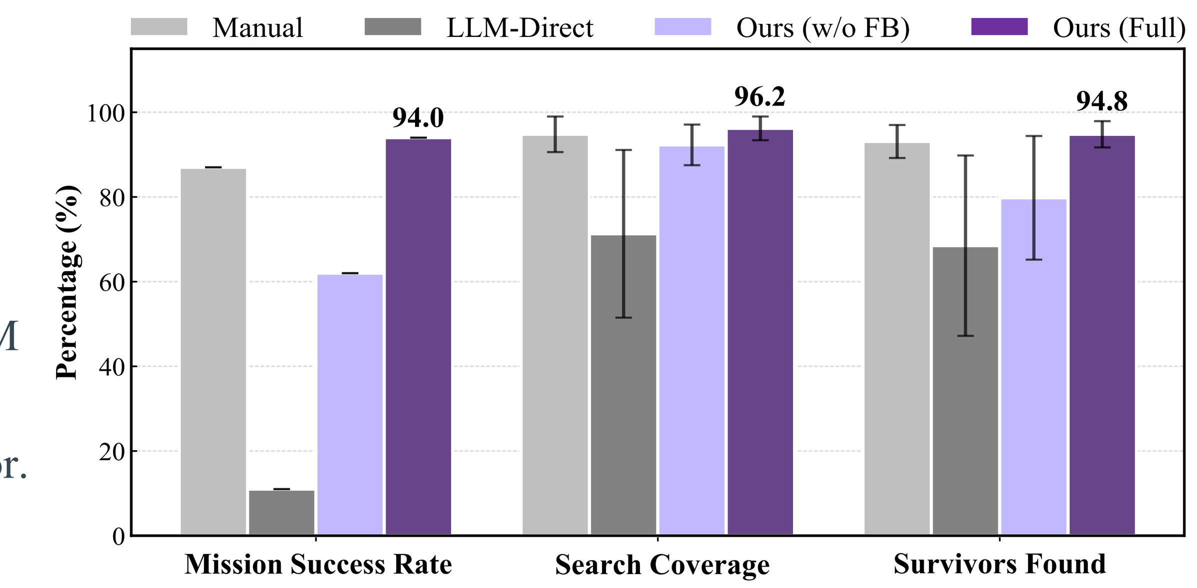
This final module embeds a critical Human-in-the-Loop (HIL) verification process to ensure the LLM's plan is both safe and aligned with real-world dynamic conditions.



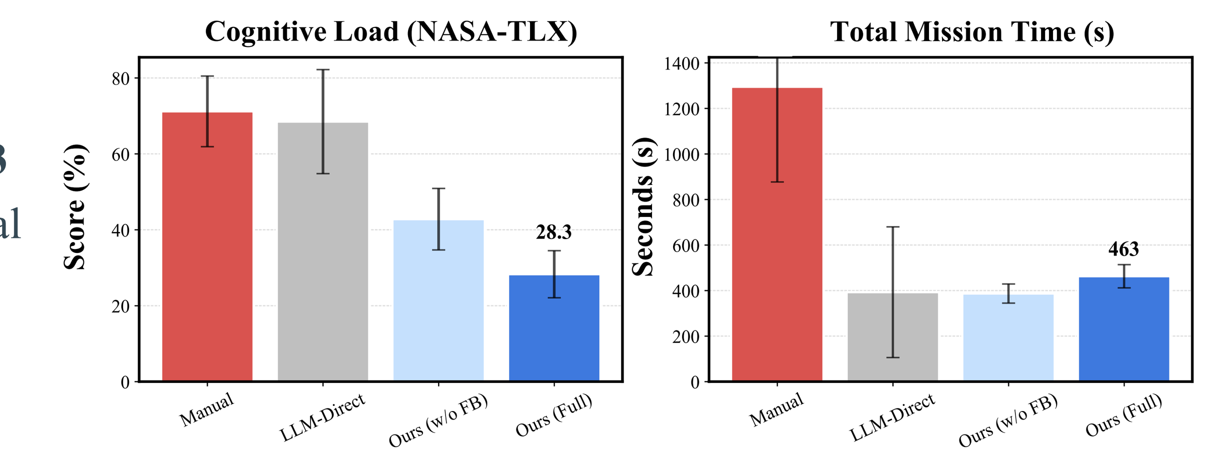
Experiments & Analyses

Metric	B1 (Manual)	B2 (LLM-Direct)	B3 (Ours w/o Feedback)	Ours (Full)
Mission Success Rate (%)	87.0	11.0	62.0	**94.0**
Search Coverage (%)	94.8 ± 4.2	71.3 ± 19.8	92.3 ± 4.8	**96.2** ± 2.8
Survivors Found (%)	93.1 ± 3.9	68.5 ± 21.3	79.8 ± 14.6	**94.8** ± 3.1
Total Mission Time (s)	1295 ± 418	393 ± 287	**387** ± 42	463 ± 51
NASA-TLX Score (%)	71.2 ± 9.3	68.5 ± 13.7	42.8 ± 8.1	**28.3** ± 6.2

Analyses



- **High Performance:** Matches expert manual control (94% success) and far outperforms direct LLM baselines (11%), ensuring robust multi-UAV behavior.



- **Low Cognitive Load:** Significantly reduces operator burden. NASA-TLX scores dropped by **43 points** compared to manual coding, solving the "cognitive overload" problem.
- **Safety via Feedback:** The ablation study shows that Human-in-the-loop verification is critical, restoring success rates from 62% (blind execution) to 94% (verified).

Interface

The interface illustrates real-time UAV swarm scenarios with video and thermal feedback (Left), the LLM-CRF's dynamic dialogue and generated mission plans (Middle), and a comprehensive dashboard for UAV and task parameter management (Right). This visual interface underscores our commitment to transparent and safe autonomous operations.