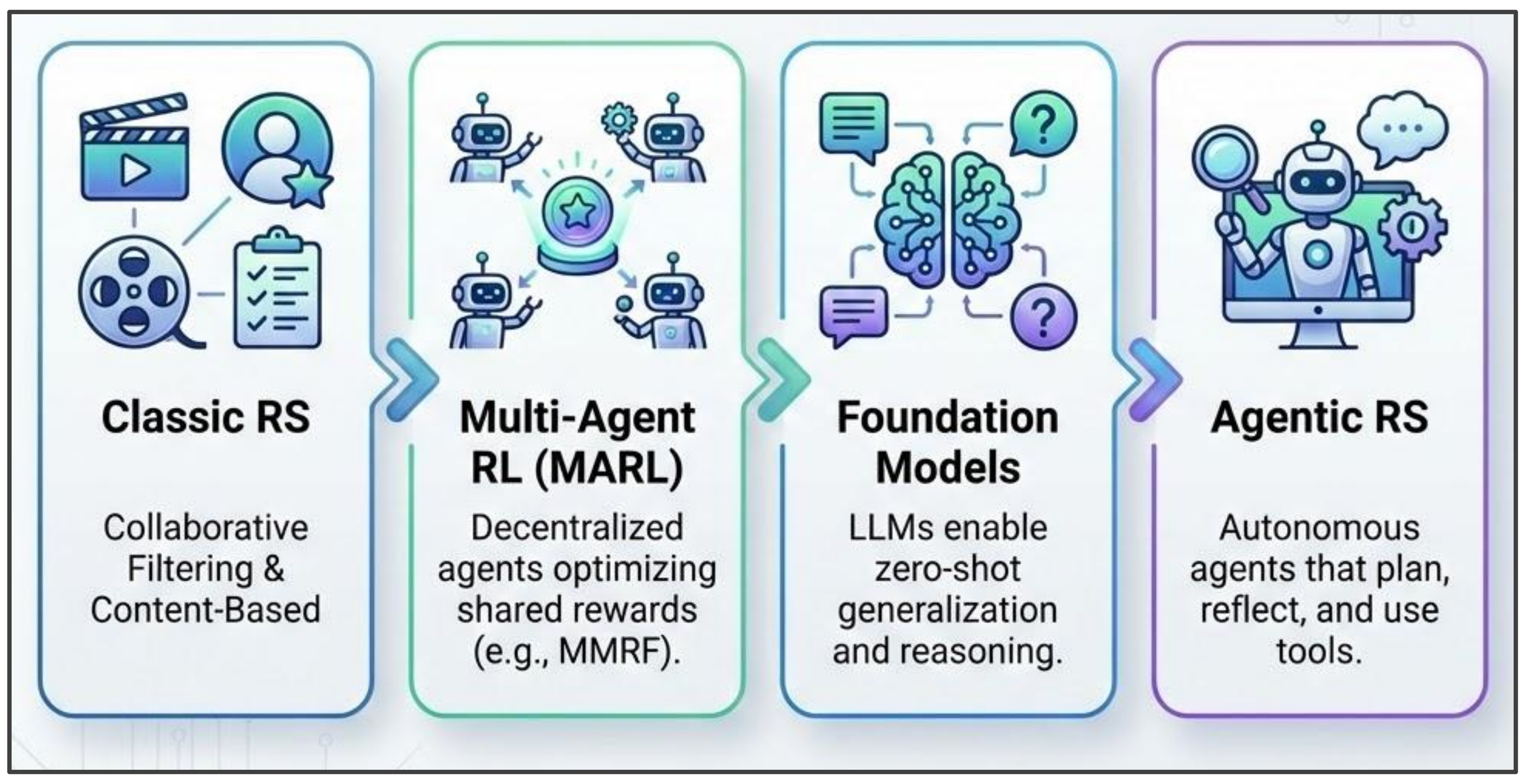
Multi-Agent Video Recommenders: Evolution, Patterns, and Open Challenges

Authors: Srivaths Ranganathan, Abhishek Dharmaratnakar, Anushree Sinha, Debanshu Das
Google

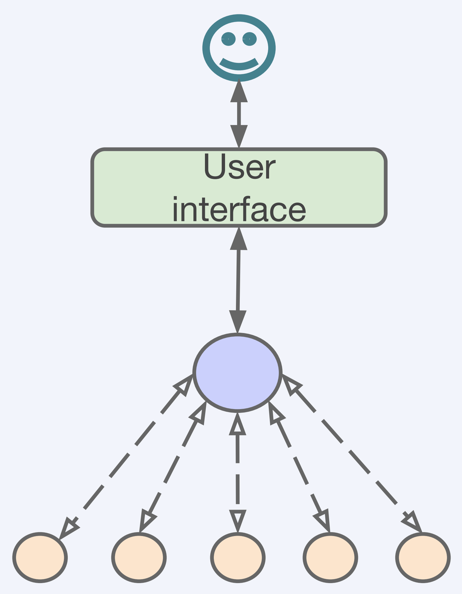
Why Multi-Agents for Video RecSys

- Traditional single-model recommenders (Collaborative Filtering, Deep Sequential) optimize static global objectives like CTR or watch time.
- This neglects competing goals (diversity, fairness, explainability) and struggles to adapt to complex, dynamic user feedback loops.
- Decomposing recommendation into specialized, interacting agents (perception, reasoning, feedback) allows for precise and explainable systems.

Evolution of Video Recommenders

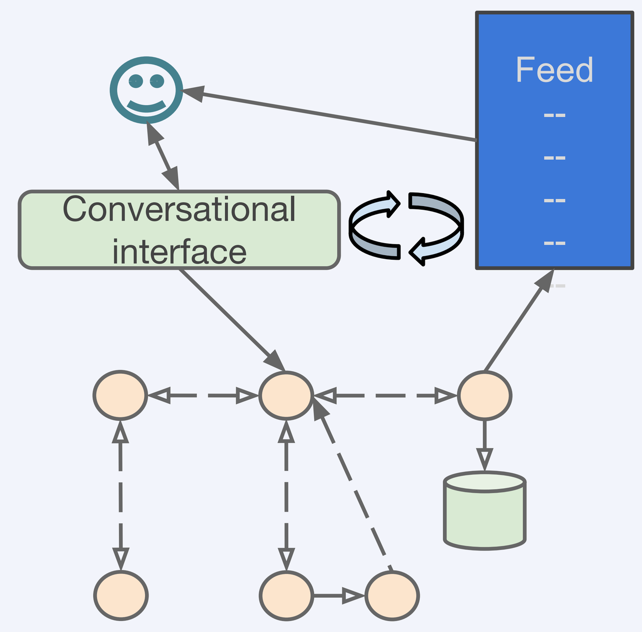


Patterns



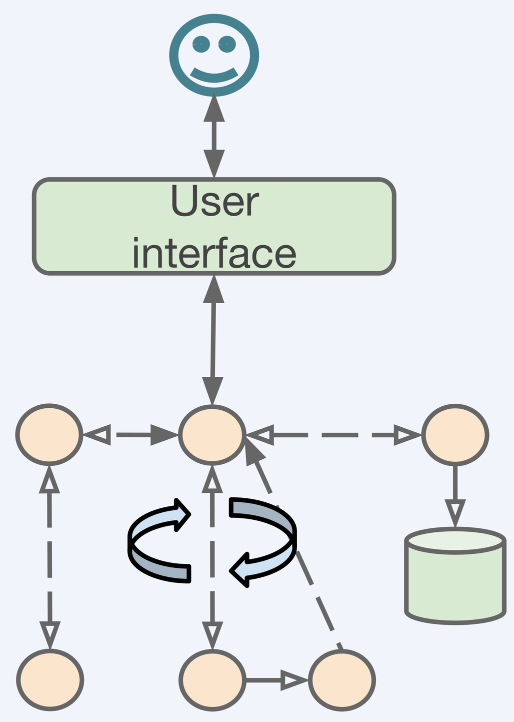
Hierarchical orchestration

- Central coordinator optimizes for primary engagement metric (e.g., WatchTime)
- Dynamically weighs inputs from sub agents tracking secondary signals like Likes, Shares, or Comments.



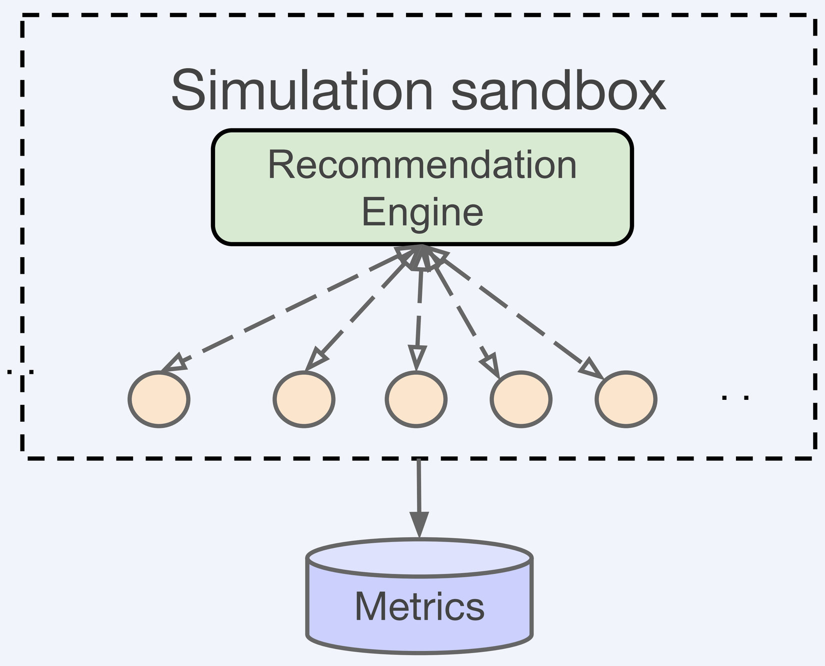
User-agent collaboration

- Internal agents collaborate to translate user conversational commands into algorithmic adjustments
- Users iteratively interact with conversational agent to customize their feed.



Pipeline based collaborative agents

- Sequential workflow: specialized agents first extract deep semantic meaning from raw video content (Perception)
- Refined semantic summaries passed to downstream agents that predict user interest (Reasoning)



User simulation ensembles

- A sandbox of thousands of LLM-driven agents simulates diverse viewer behaviors and traits
- Allows engineers to test video ranking policies and social phenomena (like filter bubbles) offline

Challenges and Open Research

Challenge: Cost & Scalability
High computational/financial overhead of LLM agents; impractical for real-time inference.

Challenge: Multimodal Grounding
Agents struggle to reason deeply on dense video (visual, audio, temporal), relying on lossy text summaries.

Challenge: Incentive Alignment
Ensuring agents with conflicting goals (e.g., WatchTime vs. Likes) cooperate truthfully toward a global objective.

Challenge: Controllability & Trust
Agents may diverge from goals, misinterpret user intent, or fail opaquely; ensuring value alignment is difficult.

Challenge: Evaluation
Offline metrics (nDCG) are insufficient; difficult to validate if simulated users (alignment) reflect real behavior.

Research: Hybrid RL-LLM
Use LLMs as high-level planners to set goals/rewards for fine-grained RL (policy) agents.

Research: Lifelong Personalization
Develop agents with long-term memory that learn *with* the user; explore federated, on-device agents for privacy.

Research: Human-in-the-Loop
Use direct user feedback (critiques, rankings) as continuous supervision; build transparent dashboards.

Research: Self-Improving Systems
Design meta-agents that self-evaluate, detect data shifts, and autonomously evolve policies to stay value-aligned.