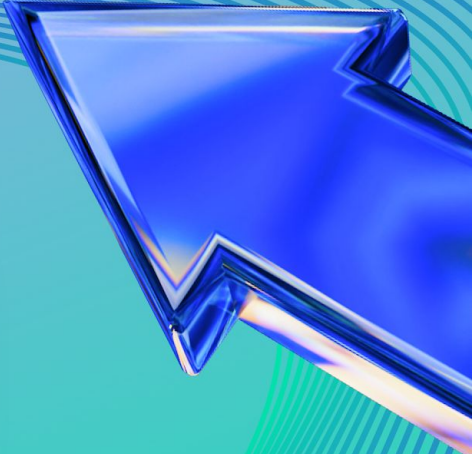




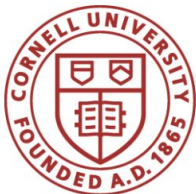
CALM SUMMIT



20
24

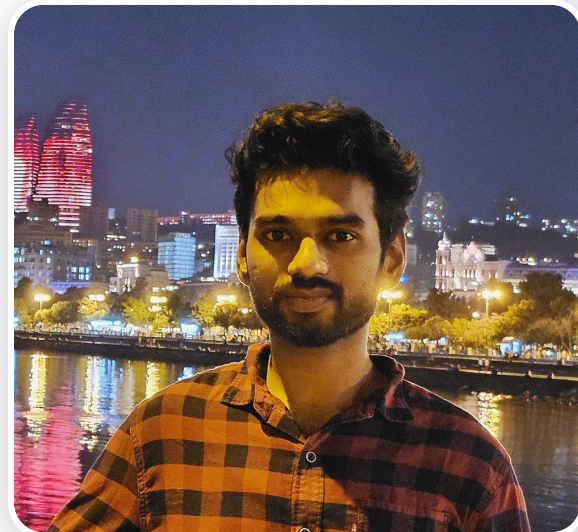
October 30, 2024 | New York, NY

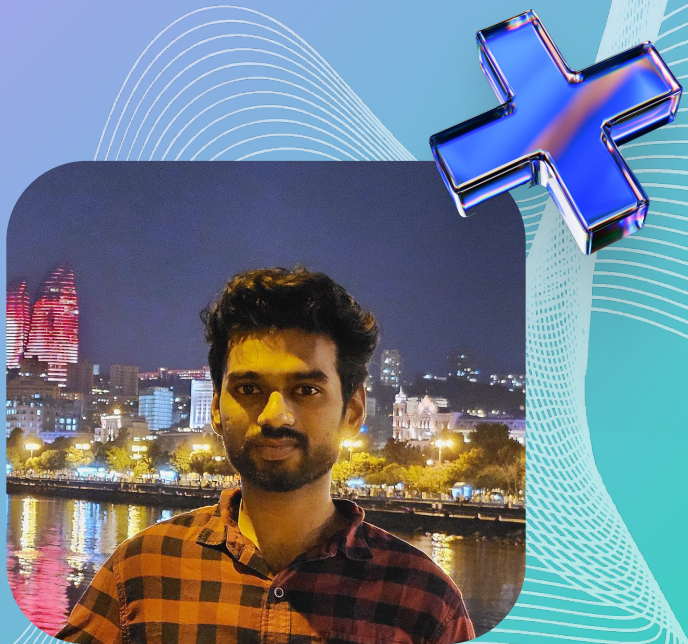
Praneet Pabolu Researcher



01:21pm - 01:41pm

MACAP: Designing
Conversational AI Platform via
Multi-Agent Systems and
Reinforcement Learning

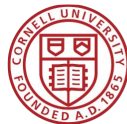




📅 October 30

Designing Conversational AI via Multi-Agent Systems & Reinforcement Learning

Researcher – Generative AI



Cornell University

01

Art of Designing Conversational AI Workflow

A quick look at the current state of Conversational AI

02

Multi-Agent Systems

Solving the challenge of Multiple Domain Collaboration

03

Reinforcement Learning to the rescue

Aligning LLMs with Human Preferences

04

Hierarchical Task Decomposition

Breaking larger task to subtasks for enhanced focus

05

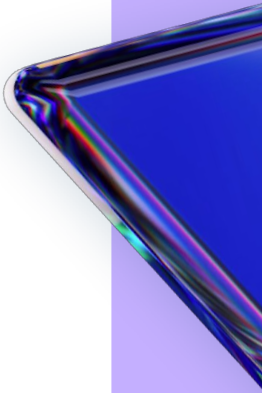
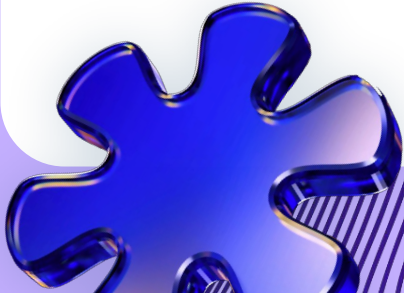
Limitations

Dealing Sycophancy Bias and Long-term consequences

Agenda

01

Art of Designing Conversational AI Workflow



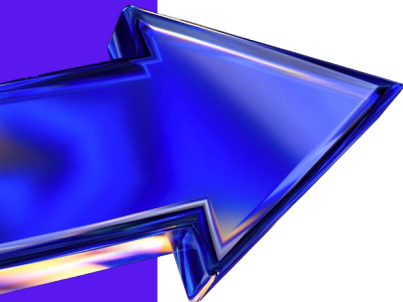
Conversational AI Design

Considerations for Interactive Dialogue Workflow

- ▢ Responses tailored to user's sentiment and context
- ▢ Mimicking non-robotic real-life conversations
- ▢ Efficient navigation through user's request
- ▢ Handle human variance
- ▢ Active dialogues in contrast to generic
- ▢ Graceful Fallbacks
- ▢ Consistency in tone while aligning to brand/domain



Generated using [Microsoft Designer](#)



02

Multi-Agent Systems



RAG, Agents & Multi-Agents

Retrieval Augmented Generation (RAG):

- LLM's haven't seen your data
- Dealing with constraints focused on proprietary, licensed and recent data

LLM Agents:

- Solves Action Planning, Domain Analysis, Active Reasoning, Context Window Limitations, etc.
- Connect with external data sources

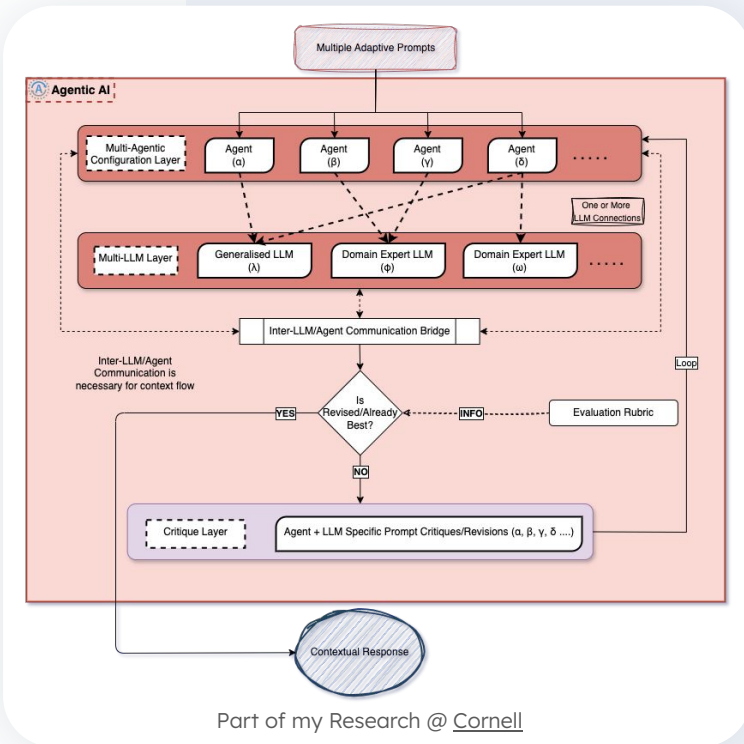
Multi-Agents:

- Domain Specific Expert LLM Agents and Critic Evaluators
- Contextual Semantic Communication for solving user queries

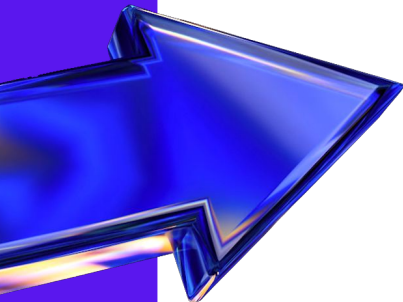
Agentic AI

Multi-Agent Design

- Context Aware Adaptive Prompt Formulation
- Inter-LLM Collaboration via Communication Bridge
- Action-Planning to tackle domain-specific tasks
- Automated Query Translation and Query Rewriting
- Response Generation based on Conversational Rubric



Part of my Research @ [Cornell](#)



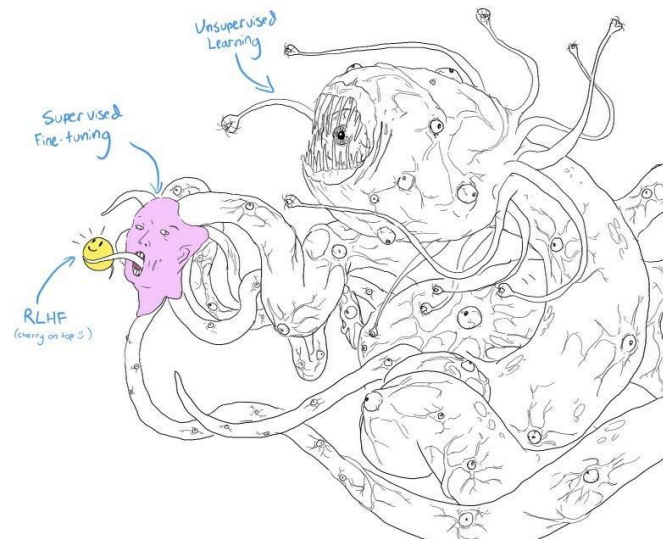
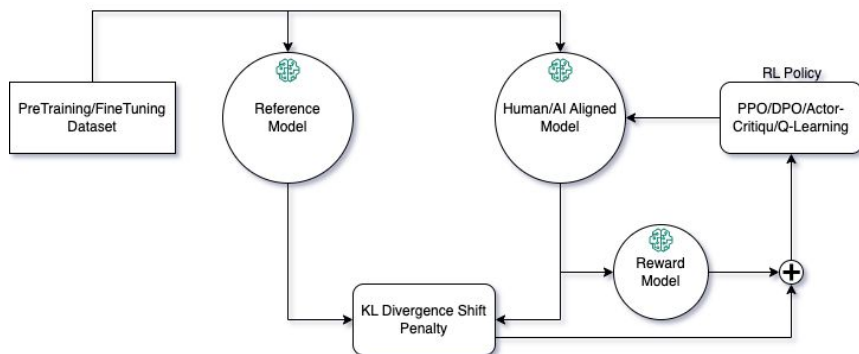
03

Reinforcement Learning to the rescue



Aligning LLMs with RL

- ML Models are great but suffer from aligning with Human values
- Need for Helpful, Harmless and Honest responses
- Preference Learning with Human and AI Demonstrations



Courtesy by Anthrupad

Under the hood

- ChatGPT used Proximal Policy Optimization (PPO) for RLHF
- PPO works in 2 phases (Create Completions + Model Update)
- We adapt a version of PPO to support Multi-Agentic RL formulation

$$L^{POLICY} = \min \left(\underbrace{\frac{\pi_{\theta}(a_t | s_t)}{\pi_{\theta_{old}}(a_t | s_t)}}_{\text{Guardrails: Keeping the policy in the "trust region"}}, \underbrace{\hat{A}_t, \text{clip} \left(\frac{\pi_{\theta}(a_t | s_t)}{\pi_{\theta_{old}}(a_t | s_t)}, 1 - \epsilon, 1 + \epsilon \right) \cdot \hat{A}_t}_{\text{Defines "trust region"}} \right)$$

Policy Loss in PPO

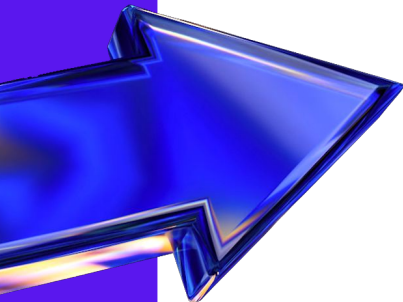
Ref. from [GenAI Course on Coursera](#)

Bellman Equation

$$V_{\theta}^{\pi}(s, i) = V_{\theta}^{\pi}(s, \pi_i(s)) + \sum_{s' \in S_{i,N}} Pr_i^{\pi}(s', N | s, \pi_i(s)) \gamma^N V^{\pi}(s', i)$$

Learning Update

$$V_{t+1}(s, i) = (1 - \alpha_t(i))V_t(s, i) + \alpha_t(i)r$$



04

Hierarchical Task Decomposition



Automated Hierarchical Reward Signals

Decomposing tasks

- **Target task is divided into hierarchical subtasks**
- **Helps LLM Agents to focus on high level goals instead of primitive actions**

Cooperation through Hierarchy

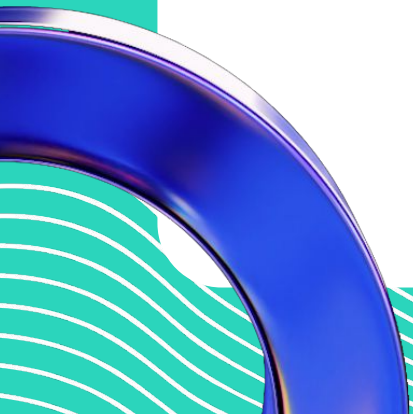
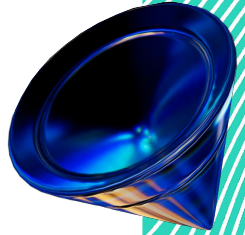
- **For Joint Value Function Learning**
- **Facilitates better coordination for solving subtasks**

$$V^\pi(i, o) = Q^\pi(i, o, \pi_i(o)) = \max_{a \in A_i} [V^\pi(a, o) + C^\pi(i, o, a)]$$

Multi-Agent Subtask Cooperative Learning in Conversational AI

05

Limitations



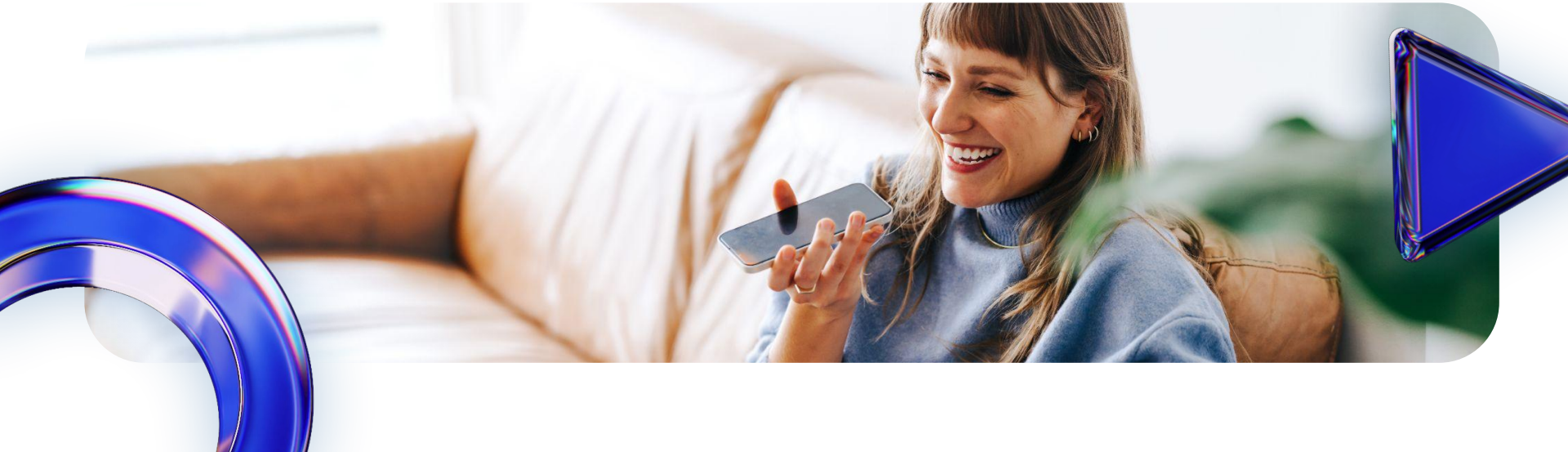
Challenges of Multi-Agent Systems in Conversational AI

Hierarchical formulation works great, but...

- Biggest risk – Hallucination across Agents
- Challenge of optimally scaling Multiple Sub-Agents
- Conversational AI receives rewards at the end of session leading to delayed rewards for RL
- Latency and Response Times
- Maintenance Costs

Got A Question?

Thank you for your Attention.
Please don't hesitate to ask
questions.



JOIN

US



ON THIS JOURNEY

Get in *Touch* *with* me

**CALM
SUMMIT** 



Address

Ithaca, NY

Phone Number

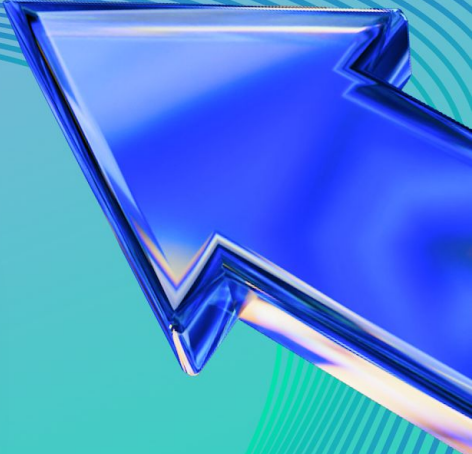
+1-607-262-7836 ("Personal")

Social Media

linkedin.com/in/deathreaper0965

deathreaper0965.github.io

praneet.pabolu@gmail.com



CALM SUMMIT



20
24

October 30, 2024 | New York, NY

