

LLM-Based Agents

Background

- Origin of AI Agent
  - Agent in Philosophy
    - From the perspective of philosophy, is artificial entities capable of agency?
  - Introduction of Agents into AI
    - Symbolic Agents
    - Reactive Agents
    - Reinforcemnet Learning-Based Agents
  - Why is LLM suitable as the primary component of an Agent's brain
    - Agents with Transfer Learning and Meta Learning
    - Large Language Model-Based Agents
    - Autonomy
    - Reactivity
    - Pro-activeness

The Birth of An Agent: Construction of LLM-Based Agents

- Brain
  - Natural Language Interaction
    - Multi-turn Interactive Conversation
    - High-Quality Natural Language Generation
    - Intention and Implication Understanding
  - Knowledge
    - Knowledge in LLM-Based Agent
      - Pretrain Model
        - Linguistic Knowledge
        - Commonsense Knowledge
        - Professional Domain Knowledge
    - Potential Issues of Knowledge
      - Editing Wrong and Outdated Knowledge
      - Mitigating Hallucination
  - Memory
    - Memory Capability
      - Raising the Length Limit of Transformers
      - Summarizing Memory
    - Memory Retrieval
      - Compressing Memories with Vectors or Data Structures
      - Automated Retrieval
      - Interactive Retrieval
  - Reasoning
  - Planning
    - Plan Formulation
    - Plan Reflection
  - Transferability and Generalization
    - Unseen Task Generalization
    - In-context learning
    - Continual Learning
- Perception
  - Textual Input
  - Visual Input
    - Visual Encoder
    - Learnable Architetrue
      - Query Based
      - Projection Based
  - Auditory Input
    - Cascading Manner
    - Transfer Visual Method
  - Other Input
    - Tactile
    - Gestrues
    - 3D Maps
- Action
  - Textual Output
  - Tools
    - Learning Tools
    - Using Tools
    - Making Tools
  - Embodied Action
    - The Potential of LLM-Based Agents for Embodied Actions
      - Cost Efficiency
      - Embodied Action Generalization
      - Embodied Action Planning
    - LLM-Based Embodied Actions
      - Obeservation
      - Manipulation
      - Navigation
    - Prospective to the Embodied Action

Agents in Practice: Harnessing AI for Good

- Single Agent Deployment
  - Task-Oriented Deployment
    - Web Scenarios
    - Life Scenarios
  - Innovation-Oriented Deployment
  - Lifecycle-Oriented Deployment
- Multi-agent Interaction
  - Cooperative Interaction
    - Disordered Cooperation
    - Ordered Cooperation
  - Adversarial Interaction
- Human-Agent Interaction
  - Instructor-Executor Paradigm
    - Feedback Types
      - Quantitative Feedback
      - Qualitative Feedback
    - Applications
      - Education
      - Health
      - Other Applications
  - Equel Partnership Paradigm
    - Empathetic Communicator
    - Human-Level Participant

Agent Society: From Individuality to Sociability

- Behavior and Personality
  - Socail Behavior
    - Individual Behaviors
      - Input Bahaviors
      - Internalizing Behaviors
      - Output Behaviors
    - Group Behaviors
      - Positive Group Behaviors
      - Neutral Group Behaviors
      - Negative Group Behaviors
  - Personality
    - Cognition Ability
    - Emotion Intelligence
    - Character Protrayal
- Socail Enviroment
  - Text-Based Environment
  - Virtual Sandbox Environment
    - Visualization
    - Extensibility
  - Physical Environment
    - Sensory Perception and Processing
    - Motion Control
- Society Simulation
  - Key Properties and Mechanism
    - Open
    - Persistent
    - Situated
    - Organized
  - Insights from Agent Society
    - Organized Productive Cooperation
    - Propagation in Socail Networks
    - Ethical Decision-Making and Game Theory
    - Policy Formulation and Imrovement
  - Ethical and Social Risks
    - Unexpected Socail Harm
    - Stereotypes and Prejudice
    - Privacy and Security
    - Over-reliance and Addictiveness

Discussion

- Mutual Benefits between LLM Research and Agent Research
  - LLM Research -> Agent Research
  - Agent Research -> LLM Research
- Evaluation for LLM-Based Agents
  - Utility
    - Success Rate of Task Completion
    - Foundational Capabilites
  - Sociability
    - Efficiency
    - Language Communication Proficiency
    - Cooperation and Negotiation Abilities
  - Values
    - Role-Playing Capability
    - Honesty
    - Harmlessness
  - Ability to Evolve Continually
    - Adapting to Specific Demographics, Cultures and Contexts
    - Continual Learning
    - Autotelic Learning Ability
    - The Adaptability and Generalization to New Environments
- Security, Trustworthiness and Other Potential Risks of LLM-Based Agents
  - Adversarial Robustness
  - Trustworthiness
  - Other Potential Risks
    - Misuse
    - Unemployment
    - Threat to the Well-Being of the Human Race
- Scaling Up the Number of Agents
  - Pre-determind Scaling
  - Dynamic Scaling
  - Potential Challenges
    - Increasing Computation Burden
    - Distorted Information Propagation
    - Difficulty of Coordinating
- Open Probelms
  - The Debate over Whether LLM-Based Agents Represent a Potential Path to AGI
  - From Virtual Simulated Environment to Physical Environment
  - Collective Intelligence in AI Agents
  - LLM-Based Agent as a Service