Conversational AI. Dialogsysteme, Chatbots, Assistenten

Veranstalter: Christoph Ringlstetter

Sitzung 8: Agents

Was machen wir denn heute.

- Besprechung Restprogramm.
- News of the Week
- Nachreichung: Besprechung Paper Tula Masterman et al. The Landscape of Emerging AI Agent Architectures
- Besprechung Paper Wang Lee et al. A Survey on LLM based Agents

Wang Lee et al. A Survey on LLM based Agents. Überblick.

Vor LLM Zeit. Limitiertes Wissen für isolierte Umgebungen. Divergenz zu menschlichem Lernen. Entscheidungen anders.

Jetzt: Welle von Agent Research.

Agent Frameworks.

A Survey on Large Language Model based Autonomous Agents

Lei Wang[1], Chen Ma*[1], Xueyang Feng*[1], Zeyu Zhang[1], Hao Yang[1], Jingsen Zhang[1], Zhi-Yuan Chen[1], Jiakai Tang[1], Xu Chen(✉)[1], Yankai Lin(✉)[1], Wayne Xin Zhao[1], Zhewei Wei[1], Ji-Rong Wen[1]

1 Gaoling School of Artificial Intelligence, Renmin University of China, Beijing, 100872, China

Abstract Autonomous agents have long been a research focus in academic and industry communities. Previous research often focuses on training

LLM-based autonomous agents. Based on the previous studies, we also present several challenges and future directions in this field.

Wang Lee et al. A Survey on LLM based Agents. Überblick

Definition: „An autonomous agent is a system situated within and a part of the environment that senses that environment and acts on it. Over time, in pursuit of ist own agenda and so as to effect what it senses in the future.“ Franklin Graesser (1997)

Kürzer: Ein Agent ist ein gestaltenderTeil seiner Umwelt.

Vielversprechender Ansatz für AGI.

Wang Lee et al. A Survey on LLM based Agents -- Introduction

Agents Paradigma vor LLM: Einfache heuristische Policy Funktionen. Gelernt und isoliert für sehr begrenzte Umwelten.

Human Mind: komplex, kann sich in einer weiten Bandbreite verschiedener Umwelten lernen zurecht zu finden.

=> frühere Agenten hatten keine Performanz, Anwendung in Open Domain, unconstrained Settings nicht möglich.

=> LLM als „Intelligenzbaustein“: jetzt zentraler Controller für Agenten mit „menschenähnlichen“ Entscheidungsfähigkeiten [11 -17]

Vergleich RL: umfassendes Weltwissen. Natural Language Interface für Interaktion mit Menschen und auch zwischen Agenten – nochmal überlegen ob das ein Zwischenschritt ist.

Wang Lee et al. A Survey on LLM based Agents -- Introduction

Haben jetzt LLM basierte Agent Modelle mit Memory und Planning Modul.
Werden unabhängig oder integriert angesteuert.
Construction, Application und Evaluation können über Prompting integriert werden.
Implementation:

(1) Wie den Agent designen um LLMs besser auszunutzen im Rahmen eines Agent Frameworks. Hardware Analogie.

(2) Agenten für verschiedene Aufgaben ertüchtigen. Software Analogie.

Wang Lee et al. A Survey on LLM based Agents – Agentenkonstruktion.

LLM-basierte autonome Agenten erstellen:
Zwei Aspekte: welche Architektur soll angewandt werden, um LLMs besser auszunutzen andererseits, gegeben die Architektur, wie es ermöglichen dass der Agent spezifische Aufgaben erledigt.

Analogie zum NN: Netzwerk Struktur <-> Parameter Lernen

Wang Lee et al. A Survey on LLM based Agents – Agentenkonstruktion.

(1) Agentenarchitektur: Unterschied zu QA Systemen: spezifische Rollen des Agenten: Umwelt wahrnehmen, nachdenken, Umwelt verändern, wieder wahrnehmen, über Iteration verbessern.
Design rationaler Agentenarchitekturen um LLMs zum maxiamalen Nutzen zu bringen => Vorschlag hier; unified Framework:
profiling, memory, planning, action.

Wang Lee et al. A Survey on LLM based Agents – Agentenkonstruktion.

Profiling Modul: Agenten führen Aufgaben aus indem sie Rollen annehmen die diese Aufgaben in Ihrer Rollenbeschreibung haben: Coder, Lehrer, Domänenexperten [18,19]

Das Profiling Modul hält diese Rollen vor: werden dann in den Prompt geschrieben.

Basisinformation: Geschlecht, Alter, Karriere, soziale, psychologische Information, anwendungsabhängige Information.

=> nach der Informationssammlung wird ein Profil für den Agenten erstellt.

Wang Lee et al. A Survey on LLM based Agents– Agentenkonstruktion

Wie funktioniert die Erstellung des Agentenprofils:

1) Manuell über einen Prompt erstellen

2) Durch LLM generiert. Regeln und Seed Agentenprofile als Few Shot.

z.B. Recommendation Agent: Alter, Geschlecht, Persönlichkeitsmerkmale, Movie Präferenzen,

=> Seed Profile

=> dann GPT

3) Dataset Alignment Methode: BasisInfo manuell in Prompts umwandeln.

Profile erstellen und dann mit Daten alignieren.

Wang Lee et al. A Survey on LLM based Agents– Agentenkonstruktion

Memorymodul: wichtiges Element beim Design der Agentenarchitektur.

Experience, Self-Evolve umsetzen. Agent muss besser, persönlicher werden.

Strukturen:

- Menschen: sensorisch, shortterm, longterm
- Unified: eigentlich nur shortterm
- Longterm: Hybrid. Auch erfolgreiche Pläne Abläufe strukturiert speichern.

Memory Sandbox: Oberfläche mit 2D Canvas auf dem Memory Objekte in den Fokus gezogen werden können.

Wang Lee et al. A Survey on LLM based Agents – Agentenkonstruktion

Memory Formate: Natural Language Memory, Embedding Memory.

Memory Bank [39], ChatDev[8]

Datenbanken: ChatDB [40] => Symbolisch SQL

Strukturierte Listen. GITM Action Lists.

Memory Operationen: acquire, accumulate, utilize: signifikantes Wissen.

read, write, reflect

Lesen im Memory nach Kriterien: recency, relevance, importance [20]

wann, hat damit zu tun, Gewicht

Formel: arg min $_{m \in M}$ .... Implementiert mit FAISS, etc

Wang Lee et al. A Survey on LLM based Agents – Memory

Writing: Information speichern mit Informationen in der Zukunft abgleichen

2 Probleme:

(1) wie Information speichern die ähnlich zu bereits existierender im Memory ist

(2) Information entfernen wenn das Speicherlimit erreicht ist

ad (1): z.B. erfolgreiche Aktionen die ähnlich sind in eine Liste speichern. Wenn das Limit erreicht wird – Anzahl Einträge – kondensieren in eine vereinte Lösung

Wang Lee et al. A Survey on LLM based Agents – Memory

ad 2) Overflow: älteste gesteuert löschen. Kondensieren über Summaries

Memory Reflexion: „witness and evaluate their own cognitive, emotional, behavioral processes“
Agenten: zusammenfassen, abstraktere, komplexere high level Information ableiten: „denke darüber nach“
Generative Agent [20]: Memory in Eisichten zusammenfassen. Details im Paper.
Agent arbeitet mit 3 Schlüsselfragen am Memory.

Wang Lee et al. A Survey on LLM based Agents – Memory

GIT [16]. Sammelt erfolgreiches Vorgehen in Listen. Mehr als 5 Elemente: Abstraktion über Prompting.

EXPEL: Erfolgreiche Trajectories vergleichen und „Erfahrung“ bilden.

Wang Lee et al. A Survey on LLM based Agents– Planung

Planungsmodul: Aufgaben in kleine Schritte zerlegen. Feedback in den laufenden Planungsprozess einarbeiten. Ohne Feedback: nach Aktion kein Abgleich zwischen Planung und Resultat.

Single Path Reasoning: Aufgabe in mehre Zwischenschritte zerlegen.

Kaskadierende Verbindung: ein Schritt hat nur einen Nachfolgeschritt.

Chain of Thought als Grundlage des Vorgehens.

Reprompting [47] checkt die Voraussetzungen des Nachfolgeschritts – Plan

Hugging GPT [13]: Dekomposition der Task und Zwischenlösungen bewerten.

Wang Lee et al. A Survey on LLM based Agents – Planung

Multi-Path Reasoning: Baumartige Struktur. Mehrere Unterschritte.

Selbst Consistent: [49] CoT-SC

Für Komplexere Probleme, die verschiedene Wege zulassen um eine finale Antwort zu erzeugen. Wähle die Antwort mit höchster Konfidenz, höchstem Vote der Agenten.

Tree of Thoughts [50] jeder Knoten ist ein Gedanke. Auswahl der Zwischenschritte: Evaluation durch das LLM. Suche Breadth oder Depth.

Wang Lee et al. A Survey on LLM based Agents – Planung ohne Feedback

Rec Mind: vorhergehnde Schritte self inspiring mit einbezogen

Graph of Thought. RAP: Weltmodel mit Monte Carlo Suche.

Externer Planner: Effektives planen für Spezialdomänen noch immer schwierig für viele Modelle. Daher: Externe Module mit einbringen. LLM + P [57] Transformation de Task Beschreibungen in formale Planning Domain Definition Language PDDL

Externer Planer verwendet dann die PDDL Statements

Wang Lee et al. A Survey on LLM based Agents – Planung mit Feedback

Real World Szenarios: erfordern oft Langzeitplanung. LLMs und auch Menschen haben damit Schwierigkeiten: (1) Übersicht zu Preconditions behalten. (2) Unvorhersehbare Transitionsdynamik, die den ursprünglichen Plan non-executable setzt.

Menschen lösen solche Herausforderungen iterativ z.B. mit externem Feedback.

Feedback Mechanismen implementieren über: environment, humans, models

Wang Lee et al. A Survey on LLM based Agents– Planung mit Feedback

Feedback aus der Umwelt: Objektive, sachliche oder virtuelle Umwelt.

React [59]: konstruieren von Prompts mit thought-act-observation Triplets.

Plane etwas, Agentenaktion, Observation des Outcomes: Bloßes Resultat oder mit Feedback.

Nächster Thought <-- Observation => adapt

DEPS [33]: Argumentation, dass die bloße Outcome Info nicht ausreicht zu Plananpassung → Info detailliert zu Gründen der Task Failure. Teilweise mehrere Informationsebenen, Szenenbeschreibungen. Erfolgsinformationen.

Wang Lee et al. A Survey on LLM based Agents – Planung mit Feedback

Human Feedback: z.B. Inner Monologue [61] implementiert Feedback in Bezug auf Szenenbeschreibungen.

=> Feedback dann in den Prompt übernehmen

=> Informiertes Planning, Reasoning

→ noch mehr Literatur dazu suchen

Wang Lee et al. A Survey on LLM based Agents – Planung mit Feedback

Model Feedback: Nutzen von internem Feedback des Agenten selbst durch den Agenten. [62] Selfrefine Mechanismus

- LLM produziert Output
- LLM produziert Feedback
- LLM führt eigenständig Promptrefining durch

STOP KRITERIUM nach Anwendung definieren

Mehrere finegetunte oder promptoptimierte Implementierungsvorschläge um das Selfrefinement zu gewährleisten.

Reflexion [12] benutzt die Agent Trajectory als Input. Detailliertes verbales Feedback statt bloß numerischem Wert zur Erfolgsmessung.

Wang Lee et al. A Survey on LLM based Agents – Action Modul

Action Modul: agiert direkt mit der Umwelt.

Profile+Memory+Planing => Action

PRE 1) Action Ziel → was soll rauskommen
2) Action Production → wie machen
IN 3) Action Space → welche Aktionen gibt's denn
POST 4) Action Impact → Konsequenzen

Wang Lee et al. A Survey on LLM based Agents – Action Modul

Action Goals:

1) Task Completion – was soll eigentlich gemacht werden
2) Kommunikation – wie wird interagiert
3) Exploration -- welche Wege werden zur Lösungsfindung beschritten

Wang Lee et al. A Survey on LLM based Agents – Action Modul

Action Production: Standard LLM Verwendung – vor Agent Paradigma: Input und Output sind direkt verbunden.
Der Agent dagegen kann unterschiedliche Strategien und Quellen benutzen.

(1): Action via Memory Recollection. Agenten Memory anzapfen durch Retrieval gemäß der aktuellen Aufgabe: dann Aufgabe und extrahierte Memory Episoden als Prompt in das LLM einbringen um Agenten Aktionen zu triggern: Generative Agents [20]
GITM: Fragt nach einem Subgoal und dessen Completion über das Memory um historische Lösungen zu finden.

Wang Lee et al. A Survey on LLM based Agents – Action Modul

Action vial Plan Following: generiert Pläne und erzeugt dann über Dekomposition den Action Space: Menge der möglichen Aktionen.

(1) Externe Tools: Domänenwissen als Schwachpunkt der LLMs. Verstärktes halluzinieren für spezifische Domänenanfragen.

1.1 APIs zu Search, Wikipedia etcd

1.2 Datenbanken, KB: ChatDB, MRKL, OPENAGI (OpenAGI: When LLM Meets Domain Experts)

1.3 Externe Modelle: spezielle Modelle für spezielle Aufgaben. E.g. Memory Bank. LLM Encoeder , LLM Query Matcher. Chem-Crow, MM React

Wang Lee et al. A Survey on LLM based Agents – Action Modul

Internes Wissen: viele Agenten benutzen ausschließlich das interne Wissen der LLMs um ihre Aktionen auszulösen.

1) Planning Kapazität: LLMs haben Task Dekompositionsfähigkeit:sogar 0-Shot Beispiele sind schon aus 2022/2023 Minecraft DEPS, Voyager

2) Konversationsfähigkeiten: LLMs können wie Menschen kommunizieren, Aufgaben diskutieren. Neue große Modelle haben Reflexionsfähigkeit: können über sich selbst nachdenken.

3) Common Sense: simulieren tägliches Leben von Menschen, emulieren Entscheidungen. Papers Generative Agent, Kec Agent [21] ,S3 [77]

Wang Lee et al. A Survey on LLM based Agents – Action Modul

Action Impact: Konsequenzen der Aktionen.

1) Veränderung der Umwelt. GITM [16], Voyager [38]
Position verändert sich, Items werden eingsammelt
2) Internen Zustand des Systems verändert: Memory update, neue Pläne
3) Aktionen neu triggern: eine Aktion wird durch eine andere Aktion ausgelöst.
Material bereit → Hausbau beginnt: Minecraft Agent

Wang Lee et al. A Survey on LLM based Agents – Capability Akquisiton

Architektur: Hardware

Skills, Experience: Software

- **mit Finetuning**
- **ohne Finetuning**

Wang Lee et al. A Survey on LLM based Agents – Capability Akquisiton

Mit Finetuning: Erweitern der Agent Fähigkeiten für Task-Completion durch Finetuning auf Task-abhängigen Datensets: der Agent lernt z.B. die Task besser zu machen indem das LLM auf die Task optimiert wird.

- menschliche Annotation
- LLM Generation für Annotationen
- Realworld Anwendungen als Annotation z.B. industrielle Datensets

Wang Lee et al. A Survey on LLM based Agents – Capability Akquisiton

<u>Finetuning mit menschenannotierten Datensets:</u>

Designe eine Annotationsaufgabe und dann rekrutiere ArbeiterInnen um sie ausführen zu lassen – also die Daten explizit zu produzieren.

Cott [84] Menschliche Werte als favorisiertes Verhalten

RET LLM: Memory Tuning: ***„in order to better convert natural languages into structured memory information, the authors fine-tune LLMs based on a human constructed dataset, where each sample is a "triplet-natural language" pair."***

WEB Shopping Datenset – Webshop [85] 1 Mio Produkte im Annotationsset

Wang Lee et al. A Survey on LLM based Agents – Capability Akquisiton

Erziehungsbereich: Swiftsage [87] Agentendesign beeinflusst durch die dual-process Theory aus dem Bereich der Kognitionswissenschaft => komplexes interaktives Reasoning.

Paper lesen!

Wang Lee et al. A Survey on LLM based Agents – Capability Akquisiton

Finetuning mit LLM generierten Datensets.

Toolbench: → Fähigkeit Tools zu benutzen verstärken. Im Llama Originalpaper beschrieben.

[82] Sandbox für LLM soziale Interaktion. Simulierte menschliche Gesellschaft.

Wang Lee et al. A Survey on LLM based Agents – Capability Akquisiton

Fine Tuning mit Real World Datensets: Mind2WEB tatsächliche Aufgaben im Webbereich. SQLPalm für SQL Statements. Schwierig zu beurteilen ob die in den Papers beschriebenen Vorteile gegeben die neuen fortgeschrittenen LLMs Bestand haben: für Standardaufgaben eher nicht.

Wang Lee et al. A Survey on LLM based Agents – Capability Akquisiton

Ohne Finetuning: Model Fähigkeiten sollen durch spezifische Prompts erschlossen werden: Prompt als Intellectual Property – Firmenkapital.

???LLMs sind dann wie Belegschaften? – Workforce – wir müssen eine Art AI HR bilden → hence Promptengineer???

Wertvolle Informationen die über die Prompts die Model Kapazitäten erweitern. Ist eine von drei Strategien zur Erweiterung der Modelfähigkeiten: Model Finetuning, Prompt Engineer, Agent Evolution

Wang Lee et al. A Survey on LLM based Agents – Capability Akquisiton

Prompting Engineering: Menschen können jetzt direkt und natürlich mit dem System interagieren. Man kann die gewünschte Fähigkeit formulieren und mitteilen.

COT: Complex Task Reasoning Few Shot [45]

Social AGI [30] „Selbstbewußtsein“ als Basis der Verbesserung

Retroformer [91] Reflexion von Misserfolgen durch ein retrospektives Modell + Reinforcement Learning um das Modell zu verbessern.

Wang Lee et al. A Survey on LLM based Agents – Capability Akquisiton

Mechanism Engineering: System vertrial and error. Aktion-Kritik-Promptanpassung

RAH: Recomender System – werde einem Menschen ähnlich in den Empfehlungen

Plan + Explainer: erklärt die Gründe für Fail. DEPS. Kann auch auf einem Daten Preset so durchgeführt werden.

Crowd Sourcing: [45] Debatiermechanismus der die Weisheit der Gruppe benutzt: Iteration bis ein Konsens zwischen Vorschlägen und Outcome erreicht wird.

Wang Lee et al. A Survey on LLM based Agents – Capability Akquisiton

Erfahrung Akkumulieren: GITM. Ausgangspunkt: Der Agent weiss nich wie er das Problem lösen kann: Lösungsweg finden.

Explore (Umwelt + Memory) → Evaluate: Solved → write to Memory mit Actions: VOYAGER

APPAGENT lernt auch von Menschlichen Demonstrationen [96].

Wang Lee et al. A Survey on LLM based Agents – Capability Akquisiton

Self Driven Evolution

LMA3[98]: „Agent can autonomously set goals for itself and gradually improve capability by exploring the environment and receiving feedback from a reward function“ Eigene Ziele setzen und über reward function verbessern.
SALLM-MS[99]: mit fortgeschrittenen LLMs → z.B. GPT4 wird ein Multiagentensystem mit selbstgetriebener Evolution implementiert: Lösung komplexer Aufgaben, Fortgeschrittene Kommunikation.
Weiterhin: Teacher – Student Setups, Verbesserung struktureller Komponenten: Reasoning Skills Theory of Mind getriggert [100].

Wang Lee et al. A Survey on LLM based Agents – Capability Akquisiton

NLSOM: Eingreifen nur wenn der Student nicht selbst weiter kommt:

Selfdriven learning: erlaubt dynamische Updates: Rollen, Aufgaben, Beziehungen.

Bemerkung: Finetuning mit Anpassung der Modellparameter:

- brauchen sehr viel taskspezifische Information
- Nur open Source LLMs
- aber Prompting Kontext Windows sind begrenzt oder teuer

und: Design Spaces der Prompt Mechanik ist sehr groß, das Optimum kann nicht gefunden werden.

Wang Lee et al. A Survey on LLM based Agents – Anwendungen

Social Science, Natural Science, Engineering:

Social Science: Psychologie, Simulationsexperimente – vorher nur mit Menschen möglich.

GPT4: Aber: hyperaccuracy distortion – zu schalu

Politische Wissenschaft. VWL soziale Simulation [29,105,106]

Agent Simulations [34] Stadt mit Agents tägliches Leben simulieren.

Rechtswesen: Simulation von Richtern, Entscheidungen, Beratung.

Research: Abstracts erstellen, Streamline Research, Schreibassistenten.

Wang Lee et al. A Survey on LLM based Agents – Anwendungen

Naturwissenschaften: Beschreibung, Verständnis und Vorhersage natürlicher Phänomene.

Dokumentation, Datenmanagement, Literaturorganisation.

Planen von Experimenten. Eigenschaften und Strukturen von Stoffen.

Frameworks: ChatMOF, Chemcrow

Experimental Agents[115]

Wang Lee et al. A Survey on LLM based Agents – Anwendungen

Erziehung: Math Agents [117]. Explore, Dis over, Solve, Proof mathematical Problems.

CodeX für mathematische Probleme auf Uni Level.

Wang Lee et al. A Survey on LLM based Agents – Anwendungen

Engineering: komplexe Strukturen im Civil Engineering. Interaktiver Agent insbesondere für Engineering Design. GPT Engineer [128]

CS Software Engineering: Coding, testing, debugging, documenting

Pentest GPT [125]

Industrielle Automatisierung: LLM based agents für intelligentes Planen und die Kontrolle von Produktionsprozessen.

[129] Digitale Zwillinge

IELLM [130] Öl, Gas Tubing

Wang Lee et al. A Survey on LLM based Agents – Anwendungen

Robotics and Embodied AI: Reinforcement Learning spiet eine große Rolle [140 - 143]. Fokus die planung, reasoning, und collab/communication module für den embodied Fall zu verbessern.

[140]: Unified Agent System für embodied Reasoning und Task Planning.

High Level Commands ermöglichen verbessertes Planning.

Low Level Controllers: Kommandos in Aktionen.

Zusätzliche Beispiele:

SayCan [78] multiple Skills Robot in einer Küchenumgebung. 551 Skills, 17 Objectives

OS libraries [144 - 157] Basierend auf Langchain: Xlang.

Militärische Bots/Agents und Strategie

Wang Lee et al. A Survey on LLM based Agents – Evaluation

1) Subjective: basierend auf der menschlichen Beurteilung.

- keine Datensets
- quantitative Metrik schwer zu designen

e.g. Intelligenz, Benutzerfreundlichkeit

11) Menschliche Annotation: output des Agenten bezüglich Schlüsselfragen annotieren. Harmlos, ehrlich, hilfreich, engagiert, unbiased.

Turing Test: outputs: agents, humans = humanlike performance

zt LLMs selbst benutzen, z.B. GPT für Bias etc

Wang Lee et al. A Survey on LLM based Agents – Evaluation

Objective Evaluation:

Successrate

Reward Score

Coverage

Accuracy

Similarity

Trajectory, Location ACC

Length of Planning

Speed, Cost

Wang Lee et al. A Survey on LLM based Agents – Evaluation

Soziale Evaluation: Kooperation, Empathie, Generalisierung, MultiTask

Benchmarks in VE Welten z.B. Minecraft [16,33,38]

Socket [165] Benchmark für soziales Handeln und soziale Sprachfähigkeiten

BEBE [125] Penetration Scenarios für Agents

Wang Lee et al. A Survey on LLM based Agents – Evaluation

Andere Surveys:

[178] human alignment

[179] Reasoning

[178] ALLMs Augmented LLMs

Wang Lee et al. A Survey on LLM based Agents – Challenges

Roleplaying: selten im Web diskutierte Rollen sind nicht gut simulierbar

ähnlich Domänenproblem

Mangel an “Selbstbewußtsein“ der Modelle – fängt an mit o1?

Finetuning nicht einfach großer Parameterraum

Wang Lee et al. A Survey on LLM based Agents – Challenges

Generalisierung:

- Simulation sollte auch negative Werte umfassen → nicht weichgespühlt
- Für Problemlösungsszenarien ist natürlich die Simulation negativer Effekte sehr wichtig
- Agent build a bomb: GPTs durch „universale humane Werte“ gehindert: wie durch Prompts egalisieren.

Wang Lee et al. A Survey on LLM based Agents – Challenges

<u>Prompt Robustness:</u>

Zur Erweiterung der Modelle. Erhöhung der Rationalität.

- Zusatzmodule: Memory, Planning
- Prompts dürfen nicht bei kleinen Änderungen massive Ergebnisauswirkungen haben. „unified and resilient prompt framework accross divers LLMs“.
- manual
- durch LLMs selbst: GPT general prompt besser als Azure

Wang Lee et al. A Survey on LLM based Agents – Challenges

<u>Halluzination:</u>

falsche Information wird mit hoher Konfidenz ausgegeben [186]

Korrektur, Feedback, Grounding [23, 176]

Wang Lee et al. A Survey on LLM based Agents – Challenges

<u>Knowledge Boundary:</u>

human real-world Verhalten [20]

Vorsicht für der Überkapazität von LLMs: AI kann menschliche Simulation durch Überperforming fehlerhaft machen: too good to be true

Constraining von Wissen auf ein natürliches Maß

Efficiency: aufgrund der autoregressiven Architektur haben LLMs „langsame Inferenz“ Lösungen in der Literatur verfolgen: next big thing?