Making Sense of Large Language Model-Based AI Agents

Completed Research Paper

Andreas Göldi
University of St.Gallen
andreas.goeldi@unisg.ch

Roman Rietsche
Bern University of Applied Science
roman.rietsche@bfh.ch

Abstract

Large Language Models (LLMs) have had major impact in society even though most LLM applications use single model calls to generate output. Recent innovations have uncovered that multiple chained calls tend to produce better results. Even more impactful is the discovery that these chains do not need to be predefined. LLM-based AI agents use frameworks to generate written intermediate reasoning that decides which steps to take next and when to return with a final output. LLM-based AI agents can use external tools like search engines, calculators, code engines, etc. to gather information and act on the world. Developments in this area are rapid and potentially consequential. However, it is difficult to keep apace with the developments. To address this, we introduce a typology grounded in recent research that provides a structured framework for understanding LLM-based agents, facilitating proactive engagement with future developments.

Keywords: LLM, generative AI, agents, typology

Introduction

The rapid development of Large Language Models (LLMs) has marked a significant milestone in the realm of artificial intelligence (AI), particularly in the domain of natural language processing. Many posit the next milestone will be the advancement of agent capabilities (Wu et al. 2023; Zhiheng et al. 2023). These capabilities are based on reasoning in multiple steps, and using external tools or sources where necessary to reach a final response[1].

These LLM-based AI agents use multiple calls to the LLM because, firstly, this enhances output quality (Wu et al. 2022). Secondly, it allows using tools, such as calculators, web search, or document retrieval, in the reasoning loop (Parisi et al. 2022). While LLMs on their own already present functional language competence (Mahowald et al. 2023), more complex competencies are enabled by making multiple calls to the model. The novelty in LLM-based AI agents is that they determine on their own when to call the LLM to generate additional reasoning steps, generate instructions for their tools, e.g., a web search, or return with the final response (Göldi and Rietsche 2024). These agents, therefore, can allow LLM-based agents, e.g., to think before responding (Chen et al. 2023), retrieve facts from corporate data (Lewis et al. 2020), or even simulate the tasks of software developers (Zhou et al. 2023). Since these agents consist of code with prompts that enable chaining calls to LLMs, creating novel agents is far less expensive than iterating on LLMs. Therefore, not only can more people take part in advancing this sub-field of generative AI, its cycles of improvement are likely to be even shorter than those of more capital-intensive pursuits, such as model training. Further amplifying this trend, LLM providers strategically simplify agent creation to create

[1]A popular framework for agents is langchain: https://python.langchain.com/docs/modules/agents/; other efforts focus on multi-agent teams, such as Microsoft's AutoGen (Wu et al. 2023).

dependency. A prime example is OpenAI's development of a natural language interface for customizing GPT-based agents (OpenAI 2023a). Since even LLM advancements have been rapid, it is not a courageous bet to predict that the field of LLM-based AI agents will evolve even more quickly.

Understanding the distinctions and assessing the impacts of LLM-based AI agents is inherently complex due to their integration with disparate LLMs that alter their decision-making processes, the diversity in their applications and outputs, their dynamic interactions, and the crucial need to evaluate them on both performance and ethical grounds. One common approach to addressing such complexities is creating classifications and structuring the field. Characterizing LLM-based AI agents through a standardized framework ensures consistency in evaluation and comparison, enabling objective assessments that facilitate informed decision-making and technological improvement.

However, despite the valuable existing classifications, a gap remains in the literature: there is no typology specifically focusing on single-agent artifacts that is extensible to new developments in this fast-advancing field. Current classifications do not allow for anticipating the possibilities and consequences of LLM-based agents before their arrival. Our typology aims to bridge this gap. Therefore, we are aiming to address this gap by developing a typology and answering the research question:

RQ: How can we make sense of Large Language Model based AI agents and promote, prevent, or prepare for specific types?

This paper contributes significantly to understanding LLM-based AI agents. It proposes a new, precise definition for these agents, introduces a detailed typology for categorizing them, and explores how to use it to consider existing and potential future agents and their impact. Our typology offers at least three essential contributions: *First*, it provides a means for comparing agents, facilitating evaluating their progress or differences. This comparison aids researchers in making more informed assessments of the agents' potential capabilities. *Second*, our typology helps by providing a framework for thinking about agents, helping with decisions among offerings, and allocating attention to agent projects that drive innovation. *Third*, the typology aids in envisioning future scenarios, acting as a foundational tool for creating prototypes of future agents to evaluate their impact on user perceptions and for speculative analysis. This allows researchers to explore potential future developments and their wider consequences.

The subsequent sections will briefly touch on the history of agent artifacts, provide a primer on LLMs, and then go into definitional issues of LLM-based AI agents. This definition provides the scope for the typology, which will then be developed using the emerging literature. Finally, we will provide examples of how to use our typology to think about the different potential futures of and with agents.

Background

Agents

The concept of agents has a long history in both the information systems and artificial intelligence literatures. In the information systems literature, the term 'agent' has mainly been used metaphorically. A common expression, emerging already in the 1990s (Bavaresco et al. 2020; Cassell and Thorisson 1999) in this sense is that of the 'conversational agent', mostly used as a synonym to chatbot (Brendel et al. 2023; Göldi and Rietsche 2023). The use of 'agent' here is merely to allow transfer of knowledge we have of beings with agency, namely humans, to artifacts that seem and are designed to be alike, such as chatbots (Göldi 2024). For example, the term 'agent' refers to various software systems, including smart personal assistants, chatbots, and similar platforms; in short, systems designed to interact with users using natural language (Elshan et al. 2023). Additionally, in the realm of conversational agents, these agents are further distinguished by their communication mode, embodiment, and context. This includes distinctions such as voice or text communication, disembodied or physically embodied agents, and general-purpose versus domain-specific applications (Diederich et al. 2022). In a similar vein, 'agent' is also used to describe human-like conversational agents, emphasizing the design and evaluation of agents with features that enhance their human-like appearance (Brendel et al. 2023). Moreover, the term extends to intelligent software capable of interacting with humans through an embodied avatar, thus broadening the scope of agent design and functionality (Nunamaker et al. 2011).

The concept of agency in these systems, therefore, is not primarily about what they are but about what they seem. Such a perspective allows for a deeper understanding of the role and impact of these agents in various service environments, contrasting them with human service providers and highlighting potential advantages like consistency, scalability, and personalization (Luo et al. 2023). Furthermore, the comparison between AI and human agents is sometimes drawn explicitly, offering insights into the unique characteristics and benefits of each (Han et al. 2023). While this definition of 'agent' is clearly useful as an umbrella term for human and AI agents as well as to ease the understanding of chatbots as social actors, these systems do not actually possess agency in any meaningful sense. They merely resemble humans, and are often used in the place of humans, and therefore referred to as agents metaphorically. We will later restrict the definition of 'agent' used in this paper to only incorporate artifacts capable of more than the mere display of agency. In the field of artificial intelligence, the term 'agent' encompasses a broad and nuanced spectrum of concepts. Starting from a fundamental perspective, AI agents are viewed as artificial entities capable of perceiving their surroundings using sensors, making decisions based on these perceptions, and taking actions (Zhiheng et al. 2023). This basic definition underpins the functionality of AI agents, emphasizing their capability to interact with and respond to their environment.

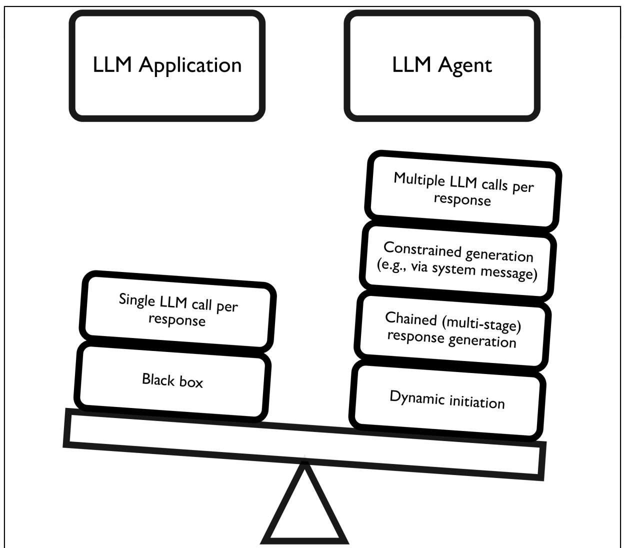

Figure 1. LLM agents versus non-agent LLM apps.

Expanding on this, an agent in AI is often characterized by certain properties such as autonomy, social ability, reactivity, and pro-activeness. This definition presents a weak and general notion of agency, common in computer science and software engineering. Conversely, a more specific and stronger notion of agency in AI attributes human-like qualities to agents, such as beliefs, desires, intentions, and emotions. This approach aligns more closely with the field's traditional objectives, viewing agents as systems conceptualized or implemented with characteristics typically ascribed to humans (Wooldridge and Jennings 1995). Further delving into the philosophical aspects, an agent in AI can also be described through the intentional stance (Dennett 1988). This stance involves attributing mental states to complex systems to explain and predict their behavior. This philosophical notion of agency has motivated the use of formal logic and languages for representing and reasoning about agents, providing a framework for understanding their actions and interactions. It does not necessarily imply that artifacts have intentions, merely that it is useful to think of them as if they had.

Large Language Models (LLMs) in the Context of Agents

Advances in language modeling have revolutionized the field of natural language processing. These advances have enabled language models to produce coherent and contextually relevant responses to natural language queries. The upscaling of these models has been a pivotal factor in enhancing the quality of their outputs, leading to significant improvements in text generation capabilities (Brown et al. 2020). The performance of modern Large Language Models (LLMs) in generating responses is now comparable to human-level performance (Bubeck et al. 2023) and can be used in a variety of human-level tasks (e.g. Göldi et al. 2024; Meier et al. 2024).

Language, which is intrinsically used by humans for both communication and reasoning (Kompa 2023), is being effectively modeled by LLMs. These models exhibit the ability to form reasoning chains when prompted effectively, suggesting an emerging capacity for complex thought processes or at least their simulation (Wei et al. 2022). To further enhance the quality of outputs, multiple calls to LLMs can be chained, i.e., outputs of one call can be used as input to subsequent calls until an acceptable final answer has been generated (Wu et al. 2022). This chaining of reasoning steps is essential in the development of AI agents (Park et al. 2023), which represent a significant evolution beyond the capabilities of standalone LLMs.

LLM based AI agents, leveraging the power of LLMs, can dynamically utilize external tools such as retrieval or search engines during their reasoning loops (Parisi et al. 2022). This dynamic initiation of steps, including the selection of tools, is a hallmark of agency (Schlosser 2019). Agents, therefore, provide additional functional capabilities over and above the formal competencies of LLMs (Mahowald et al. 2023), similar to how television relies on but is distinct from display technology. These agents are not only closely linked to LLMs but also represent an emerging technology that is distinct and increasingly sophisticated.

In the context of LLMs, the notion of agency may be able to take on the specific and strong claim of agents having intentions. LLMs are capable of inferring and representing properties of an agent likely to have produced a given textual context. These representations can influence subsequent text generation, mirroring the way human agents' communicative intentions influence their language (Andreas 2022). However, going so far as to ascribe intention, as the cited author does, and not just to take an intentional stance, is an extreme position. We will take note that the possibility is seriously discussed in the artificial intelligence community but keep our definition more restricted. We are of the opinion that the incorporation of generating explicit reasoning does imbue agency, even if the reasoning does not follow intentions per se but is only simulated.

LLM-based AI Agents

To define LLM-based AI agents, it is essential to consider several foundational concepts, beginning with the most general and progressing to the specific.

Agency: An agent is fundamentally an artifact possessing agency (Händler 2023; Zhiheng et al. 2023). Agency, the capacity to act, is often decomposed into the capacity for intentional behavior. Ascribing intention to artifacts is controversial, raising several implications. For instance, if some artifacts do have intentions, how do we distinguish them from others that do not? This problem of what is called vagueness is not the only one. More importantly, with intentions comes the possibility of frustrated intentions. Humans suffer from frustrated intentions, i.e., whenever they do not achieve them. Does that mean we cause suffering if we stop an artificial agent before it has finished its reasoning loop? Yet a third argument from apagoge, i.e., the demonstration of inconsistent, unlikely, or seemingly absurd implications, is that artificial intentions are simply not expected. Again, this does not preclude them, as for most of recorded history, animal sentience was put into question as well. However, it would require a complete rethinking of our ethical frameworks, which is not a task lightly undertaken. Luckily, for the purpose of defining agency in this context, we can sidestep these complexities by adopting an alternative conception of agency focused on the *capacity to initiate action* (Schlosser 2019). This approach, which emphasizes initiation capacity over intention, allows a clear classification of artifacts capable of determining their next steps as possessing agency (or initiation-agency). LLM-based AI agents exemplify this, capable of initiating actions unanticipated by users or developers.

Minimal Agency: To define LLM-based AI agents, we can take this conception of agency and use the minimal criteria for agency to derive the definition's components. Barandiaran et al. (2009) define such minimal agency: It necessitates adaptive regulation of the agent's interaction with its environment and self-maintenance. In the context of LLM-based AI agents, the environment is the data they process, and self-maintenance refers to maintaining operational integrity during request handling. This concept aligns with the essential functions of LLM-based AI agents, demonstrating their capacity for minimal agency. Barandiaran et al. (2009) outline the criteria for agency as *individuality, asymmetry, and normativity*. Individuality means that a system is able to define its own identity as an individual and distinguish itself from its environment. The system has a self-generated and self-maintained organization that specifies which environmental features are relevant to it and demarcates its boundary from the rest of the world. Asymmetry means that a system is able to modulate its coupling with the environment in a way that the environment does not. The system is the active source of its interactions and can change some of the constraints and conditions that affect the coupling according to its own organization. Normativity means that a system is able to regulate its interactions according to some norms or goals that are generated or sustained by its own organization. The system can evaluate its interactions as successful or failed in relation to these norms and act accordingly.

If we now apply these three criteria, individuality, asymmetry, and normativity, we find them present in LLM-based AI agents. Individuality, in the specific and constrained sense of retaining characteristics over time, is derived in LLM-based AI agents from instructions defining their operational scope, currently most often with a system prompt that instructs on the purpose or role of the agent in natural language. A specifically fine-tuned model may serve the same purpose. Asymmetry is evident in the dynamic action determination observed in LLM agents, meaning that they create and initiate all steps required to fulfil their purpose. This allows positive identification of both the asymmetry criterion of Barandiaran et al. (2009) and the initiation criterion of Schlosser (2019). Normativity is manifested by the non-random nature of these actions, achieved through chaining (Wu et al. 2022). Namely, each intermediate step depends on the last generation step but still follows the purpose set out, e.g., in the system prompt. This method not only differentiates agents from mere models with system prompt but also ensures that the actions taken are regulated by self-sustained goals, meeting the normativity criterion of agency Barandiaran et al. (2009).

AI Agent: Building upon the concept of an agent, an AI agent is an artifact capable of initiating actions based on its code, distinguishing it from natural agents. The 'intelligent' part of artificial intelligence is already implied by the definition and operationalization of agency we use here (similar to another definition found in Cheng et al. 2024, who define LLM agents as a subclass of intelligent agents but do not explicitly consider the different agency criteria). This may be different from other definitions, where non-intelligent agency is possible. For our definition, since the agency criteria are fulfilled via natural language understanding, we assume at least verbal intelligence performance, as indeed LLMs themselves already exhibit, even before considering them as engines for agents (Bubeck et al. 2023).

LLM-based AI Agent: This leads to the last criterion. An LLM-based AI agent specifically employs a Large Language Model (LLM) for their core processing and decision-making capabilities (Zhiheng et al. 2023). The LLM functions as the central processing unit, akin to (part of) a "brain", guiding reasoning and response generation in these agents. In fact, humans can use similar predictive techniques for language comprehension as LLMs (Pickering and Gambi 2018).

Definition

Basing the elements for our definition on this last point of demarcation, the use of an LLM, and on the initiation-agency criteria, we arrive at our definition of an LLM based AI agent, which will guide the scope of our typology.

LLM-based AI agents:

(1) rely on a Large Language Model (LLM)
(2) the LLM is pre-constrained to produce text with consistent characteristics ("system prompt")
(3) create output by making multiple calls to the LLM ("chaining")
(4) determine and initiate the steps they take dynamically ("initiation-agency")

Previous Classifications

Previous efforts to classify LLM based AI agents exist (see Table 1 for the search strings used to identify them). A search on SCOPUS and ACM revealed 21 and 13 hits, with some of the SCOPUS hits being about LLM-based AI agents. Notably, this had not been the case when we first searched this just a few months ago.

ACM did not include any relevant reviews while in arXiv, of 22 papers, 20 had content about LLM-based AI agents. Of the reviewed papers, six were especially relevant: One paper on LLM agent definitions (Cheng et al. 2024), three surveys (two with some classificatory work) and two dedicated classification papers. These papers tackle various aspects such as the core idea of LLMs in agents, application areas, balancing autonomy and alignment, easy definition for LLM agent production, and a mere classification of applications. Cheng et al. (2024) define LLM-based agents in terms of a subclass of intelligent agents, which have autonomy, perception, decision-making and can perform actions. This aligns well with our definition. However, there are some critical differences. Firstly, Cheng et al. (2024) treat LLM-based agents as yet another instance of intelligent agents. While clearly interesting and capable, they do not explicitly draw the connection between the novel capability to dynamically initiate the next step and the definition of LLM-based AI agents. More broadly, their definition, while useful and well-founded, is not grounded in a theoretical principle as ours is with agency. By grounding our definition in this way, we connect the definition explicitly to the novel feature of actual (initiation-) agency with a better understanding of the emerging phenomenon of LLM-based AI agents in general. Five recent studies have offered surveys and classifications on LLM-based AI agents. Zhiheng et al. (2023) delve into the problems and opportunities presented by LLMs as the core of agents' brains, providing a structured approach to understand their components and application scenarios, thus signaling the emerging potential of these agents. Wang et al. (2023b) further accentuate the "era of agents", classifying these systems into various application areas and underlining the autonomous capabilities they bring to diverse domains. Händler (2023) contributes a significant understanding of autonomous multi-agent architectures, focusing on the crucial balance between autonomy and alignment, an essential factor in the design and functionality of these systems. Crouse et al. (2023) introduce a novel approach using temporal logic for the easy definition and production of LLM agents, streamlining the process of agent creation. Finally, Zhao et al. (2023) provide a comprehensive classification of 15 applications of LLM-based agents, showcasing the wide array of practical implementations. Each of these studies, in its unique way, underscores the complex and dynamic nature of LLM-based AI agents, revealing both the current state and the expansive potential of this rapidly advancing field. Despite the valuable existing classifications, a gap remains in the literature: there is no typology specifically focusing on single-agent artifacts that emphasize understanding potential future agents. This means that the existing classifications are not geared towards enabling contemplation of the possibilities and consequences of LLM based agents before they arrive. This is the gap we aim to fill with the typology in this paper.

Database	**Search String**
SCOPUS	((ABS (llm) OR ABS ("large language model")) AND (ABS (agent) OR ABS (agents)) AND (ABS (review) OR ABS (survey) OR ABS (classification) OR ABS (taxonomy) OR ABS (typology)) AND (ALL (reasoning) OR ALL (chaining) OR ALL ("tool-use")) AND (PUBYEAR > 2021))
ACM	[[Abstract: llm] OR [Abstract: "large language model"]] AND [[Abstract: agent] OR [Abstract: agents]] AND [[Abstract: review] OR [Abstract: survey] OR [Abstract: classification] OR [Abstract: taxonomy] OR [Abstract: typology]] AND [[All: reasoning] OR [All: chaining] OR [All: tool-use]] AND [E-Publication Date: Past 2 years]
arXiv	order: -announced_date_first; size: 50; include_cross_list: True; terms: AND title=LLM OR "Large Language Model"; AND title=agent OR agents; AND abstract=LLM OR "Large Language Model"; AND abstract=agent OR agents; AND all=review OR survey OR classification OR taxonomy OR typology

Table 1. Search Strings Used in Different Databases

LLM based AI Agent Characteristics in the Literature

In a first step, we extracted relevant concepts from all the surveyed literature that did describe LLM-based AI agents. After removing duplicates, this were 23 papers. Concepts mostly focused on either definitional aspects and capabilities, such as task decomposition capabilities (Huang et al. 2024) or the ability to use local resources (Li et al. 2024).

Characteristic	1	2	3	4	5
Multimodal perception capabilities	×			×	
Environmental observation and evaluation		×			×
Memory storage, information processing, decision-making	×	×			
Reflection and evaluation cycles			×	×	
Continuous interaction loops for conversational agents				×	×
Embodied actions and tool use	×		×		
Action-reflection cycles			×	×	
Individual behaviors: planning, reasoning, reflection	×	×			×
Personality traits: cognition, emotion, character		×	×		
Memory operation strategies			×	×	×
Memory structure: unified vs. hybrid		×		×	
Feedback mechanisms in planning			×	×	×
Use of tools in action models	×		×		×
Fine-tuning methods for skill development		×		×	
Tool-building			×		×
Educational function enhancement	×			×	×
Fine-tuning with annotated datasets			×	×	
Collaboration, unity	×		×		
Mimicry, spectating		×		×	
Conflict, confrontation			×		×
Human-agent collaboration in decision-making	×			×	×
Multiple GPT agents with collaborative roles		×			×
Chemistry agent (ChemCrow)		×			×
Autonomous scientific experiment design	×		×		
Task-related process management			×	×	
Broad application LLM agents	×	×			×
Domain-specific expertise			×	×	
Technical platform interfacing		×			×
Synthesis of human-like behavior and memory	×			×	
Task decomposition and self-reflection capabilities			×		×
Autonomous planning and execution	×	×			
Adaptability and feedback incorporation			×	×	×
Interaction-based learning and behavioral adaptation	×			×	×
Emotional companionship chatbots		×			×
Minecraft gaming agent			×		×
Code generation and style learning	×		×		
AI-powered design platform			×	×	

Table 2. Mapping of agent characteristics to literature, numbers defined in-text.

In the literature, a number of characteristics that can differentiate LLM based AI agents are discussed (see Table 2: 1=Zhiheng et al. 2023; 2=Wang et al. 2023b; 3=Zhao et al. 2023; 4=Händler 2023; 5=Crouse et al. 2023). Individual-level characteristics refer to features and abilities at the single-agent level. Perception characteristics involve multimodal perception and environmental observation. Function characteristics include memory storage, information processing, decision-making, reflection, evaluation cycles, and ongoing interaction loops. Motor activity characteristics encompass embodied actions, tool use, and action-reflection cycles. Behavioral characteristics cover behaviors like planning, reasoning, reflection, and personality traits related to cognition and emotion. Capability acquisition and enhancement characteristics pertain to the development and improvement of agents' skills and knowledge. This includes memory module strategies and structures, planning module feedback mechanisms, and action module incorporation in action models. Interaction-level characteristics describe agent interactions among themselves and with humans. This includes group behavior characteristics such as collaboration, mimicry, spectating, conflict, and confrontation, as well as collaborative agents involved in human-agent decision-making and multiple agents in collaborative roles. Other agent characteristics describe the purposes of the agents. This encompasses specialized task agents for tasks like chemistry or scientific experiment design, general purpose agents for broad applications, technical role agents for domain-specific expertise, vision-language model agents for synthesizing human-like behavior and memory, and specialized application agents for unique tasks like emotional companionship or gaming.

Methodology

To understand emerging technologies such as LLM-based AI agents, we must balance existing knowledge with the potential for unforeseen developments. Therefore, we adopt a typological rather than a taxonomical approach. A taxonomical approach classifies existing work while a typology merely uses empirical data to ground a conceptual classification that can inform on what does not yet exist (Bailey 1994). As an illustration, by combining characteristics from our typology, we can conceptually build agent types that do not yet, and may never, exist. This allows it to think about specific agent types before they make an impact; with this, we can avoid overgeneralizations and promote or prevent specific outcomes. To reach a useful typology, we will use an iterative approach. In a first step, we substruct an initial classification by varying characteristics of an LLM-based AI agent that has the existing agent features extrapolated to their extreme. Then, we refine the classification with reference to its purpose and internal logic. Finally, we test and improve upon it with reference to empirical data, namely the agents that already exist.

Iteration 1: Ideal Type and Substruction

An established approach to developing typologies, based on Max Weber's approach is to construct ideal types, which are methodological constructs representing the purest embodiment of the observed characteristics (Bailey 1994, p. 18). An ideal type is not a utopia but an extreme type, embodying the maximum values on all dimensions, serving as a tool for analysis and comparison (Bailey 1994, p. 22). Translating this to the realm of LLM based agents, and taking the characteristics outlined in the existing literature, the ideal type would be characterized by the highest level in all surveyed characteristics. It would have multimodal capabilities based on needs; it would demonstrate proactive interaction with the environment, actively sourcing information and contacting users; its memory capability would be able to support complex decision-making processes; it would have emotional intelligence and could adeptly modify its role in group dynamics. Enhanced problem decomposition would be a key trait, allowing the agent to simplify complex problems into manageable components. Lastly, domain-specific knowledge application would be evident in its continuous knowledge accrual and specialization. This ideal type serves as a benchmark for evaluating the evolution of agents towards it (see the ideal type in this online appendix: https://osf.io/3fzr5/?view_only=903e02a095cd4a368541c9a4a703daba). The ideal type delineates the upper extreme, but we are interested in the spectrum of agent possibilities. Substruction helps here; it involves a systematic variation of the presence, absence, or degree of various characteristics in different instances (Bailey 1994), thereby enabling the categorization of LLM agents. We thus use the ideal type's characteristics, Modalities, Interaction, Memory, Emotional Intelligence, Personality, Problem Decomposition, Knowledge, Planning, Task Management, Solution Engineering, and Tool Proficiency and vary their expression to form different types. These characteristics are descriptive and useful, and make for a good initial typology. However, we identified two challenges: unclear demarcation between its characteristics and mixed levels of measurement, e.g., many different memory types may exist concurrently while interaction levels are ordinal. These fundamental issues lead to divergence in terms of established quality criteria, such as mutual exclusivity (Gregor 2006; Hunt 1991; Meyer 2007; Nickerson et al. 2013) and present difficulties for clear operationalization.

Iteration 2: Characteristics as Ordinal Levels of Meta-Characteristics

In our revised approach, we adopt a refined meta-characteristic method (Nickerson et al. 2013), treating each typology level as a unique characteristic within a meta-characteristic, organized ordinally. We further exclude constitutive characteristics, which are already implied by our defined scope, and irrelevant characteristics that don't aid in differentiating between LLM agent scenarios or their potential impacts. We also omit non-exclusive (nominal) characteristics to ensure efficient categorizations, crucial for evaluating the technology's future. This helps focus on characteristics that reveal LLM agents' progression over time. Crucially, this would be an unwise omission if the purpose of the typology was a technical overview of agents. However, our purpose with this typology is to understand agents in their capabilities and impact. We detail our new classification in Table 3, excluding memory and emotional intelligence characteristics due to their nominal nature, such as different potentially co-occurring memory systems (Tulving 1985) or emotional perception abilities (Mayer et al. 2012). The typology begins with levels reflecting minimal requirements. This restructured typology (see Table 4 for post-empirical classification) provides a comprehensive framework for understanding LLM based AI agents.

Iteration 3: Conceptual-to-empirical

Guided by the principles of Nickerson et al. (2013) and Bailey (1994), our approach in this final iteration was grounded in empirical evidence, aiming to ensure comprehensive coverage of LLM-based AI agents within our framework. This was facilitated by employing a curated list of LLM agent literature from Zhiheng et al. (2023)[2], which provided a foundational dataset for informing our typological categories, reflecting the current state of research and practice. The rationale behind eschewing a traditional literature review stems from several factors. The term "agent" in this domain is homonymous, creating challenges in accurately capturing all relevant literature. Moreover, the rapid evolution of this field and the scarcity of peer-reviewed publications make a traditional literature review less suitable. Additionally, given the anticipated growth in the number and variety of agents, obtaining an exhaustive list of current agents is less critical. Our focus was not on an exhaustive enumeration but on sourcing a representative set of artifacts and capabilities for the further development of the typology. This approach aligns with our goal of conceptualizing the future trajectories of LLM-based AI agents rather than cataloging their current state. To assist in the process of instantiation, we used LLM-based workflows, once using the LLM gpt-4-1106-preview with custom prompts (OpenAI 2023b), first as a retrieval-agent, scanning the papers for LLM agents fitting the definition outlined in our Definition. We validated our approach using a second approach, namely an elicit.org[3] notebook, which yielded comparable results in both quality and further information for our typology. The second use we made of LLMs in this process was to employing the LLM as a reclassifying advisor, taking the output of the previous step, namely the extracted agents with their descriptions, to comment on whether each type can be placed in the typology abbreviated in Table 4, or whether demarcation lines would need to be redrawn or new characteristics added. We manually followed up on this generated advice to complete the process. While current agents do not yet seem capable of fully automated literature review or classificatory work, they are already capable screening and advisor tools, potentially replacing some traditional work of student research interns.

[2](https://github.com/WooooDyy/LLM-Agent-Paper-List, date of access: 01.11.2023; k=222)
[3]elicit.org

(Meta)-Characteristic	Description
Knowledge Scope	Distinguishes the range and adaptability of the underlying LLM's knowledge.
Narrow	Focused on domain-specific functionalities.
General	Versatile functions across multiple domains.
Dynamically Narrowing	Ability to specialize dynamically, e.g., via mixture-of-experts (Masoudnia and Ebrahimpour 2014).
Multimodality	Indicates how LLMs process multimodal inputs.
Unimodal	No multimodal capability
Transcriptive	Language-based understanding of other modalities.
Direct Understanding	Beyond-language, e.g., spatial, understanding.
Built-in Augmentation	Indicates on what external context final answers are generated.
Only Chaining	Language generation using only the underlying LLM.
Retrieval-augmentation	Retrieval-augmented generation, using external documents, databases, or web searches.
External Processing	Augmenting generation with context produced during the reasoning loop with tools, such as calculators, programming shells, or other language models.
Active Probing	Augmenting generation with context requiring intervening in the world, such as attending a product demo or suing to judge legal realities.
Interactivity Level	Reflects the degree of interaction capability of an LLM.
Only Input/Output	Basic interaction capabilities.
During Processing	Interaction capability during processing.
Proactive Engagement	Initiation of communication without external prompts.
Operational Autonomy	Highlights the operational independence of LLMs.
Short-Term Operation	Designed for brief tasks.
Checkpointing	Requires intermittent inputs.
Continuous Operation	Capable of extended autonomous operation.
Personalization	Indicates the extent of adaptation to the user.
None	All users are treated the same.
Memory-based	The agent takes into account prior experience with the user.
Profile-based	The agent infers and stores information about the user and takes it into account.
Self-Improvement Capability	Reflects the LLM's ability to evolve and adapt over time.
None	No self-improvement capabilities.
Tool Creation	Ability to create new tools.
Orchestration Modification	Capability to modify operational structures.
LLM Self-Update	Potential for updates by the LLM itself.

Table 3. Meta-characteristics and characteristics of LLM based AI agents

Meta-Characteristic	Basic	Intermediate	Advanced	Futuristic
Knowledge Scope	Narrow	General	Dynamically Narrowing	-
Multimodality	Unimodal	Transcriptive	Direct Understanding	-
Built-in Augmentation	Only Chaining	RA*	External Processing	Active Probing
Interactivity Level	Only Input/Output	During Processing	Proactive Engagement	-
Operational Autonomy	Short-Term Operation	Checkpointing	Continuous Operation	-
Personalization	None	Memory-based	Profile-based	-
Self-Improvement	None	Tool Creation	Orchestration Modification	LLM Self-Update

Table 4. Classification of LLM based AI agents. *RA = Retrieval-augmentation

After thus having analyzed the fit of our typology to the agents in the literature in the curated list by Zhiheng et al. (2023), we made several changes. Firstly, the omission in iteration 2 of a tool dimension, which was present in iteration 1, did result in placing agents with very different capabilities under the same type. To avoid reintroducing a nominal meta-characteristic, we used a principle allowing for ordinal classification, namely the nature of augmentation. Tools are used by agents to retrieve or create context useful for prompting for the next step in its reasoning chain (Parisi et al. 2022). Besides the observed characteristics of only chaining, simple retrieval, and external processing of given information, we also added active probing, which encapsulates interaction with the world to generate context for answering a question. While unobserved in our dataset, this is distinct from only processing existing context, and already reality in smart personal assistants (Chandwani et al. 2023). We specified this augmentation to be built-in, since self-improvement capabilities of future agents may lead to them acquiring additional augmentation capabilities, which could lead to a historically contiguous agent being of two types, once at launch and later after improving itself. This is prevented by specifying that the augmentation capability is built-in, and thus only considers what is present at launch. Furthermore, the inclusion of this new meta-characteristic allowed us to more clearly distinguish adaptation to knowledge requirements between LLM-side and tool-side, the former of which is classified using the knowledge scope meta-characteristic, which was renamed from functional scope. Similarly, the omission of memory capability and social intelligence led to unsatisfactory instantiation results. If agents use memory of previous conversations to shape their answers, this results in very different, more tailored user experiences. We therefore added a new meta-characteristic Personalization to incorporate memory-enhanced interactions. While it is possible to categorize various forms of memory, the main distinction we found is whether the basis of personalization are the chat logs themselves, or whether knowledge about the user was inferred. Lastly, many current agents are not multimodal; therefore, we added the basic level as unimodal and moved the transcriptive characteristic, i.e., the capability to generate language describing other modalities, such as images or audio, to intermediate.

Using the Typology

After these iterations, we present a typology of LLM-based AI agents, scoped according to our definition. The typology is designed to represent the most relevant agent dimensions for thinking about their future and the future they bring. To allow for an easier estimation of progression in the field, we furthermore designed the base characteristics of agents as ordinal levels of the relevant meta-characteristics. Both the meta-characteristics and their ordinal levels are designed to be extensible. We found it necessary to include more meta-characteristics after our empirical iteration. As new artifacts are published, we expect some meta-characteristics will become less important, and may be replaced for ease of use by others. However, the design of ordinal levels allows us to leave open space for higher levels. Currently, we added two futuristic capabilities, which are not presently common. The first involves active probing of the environment, which is as of now a core human pursuit. The second may be even more futuristic: If LLM itself could be updated continuously, that would make an agent much more human-like in terms of neural condition, capable of real-time learning. The typology can be used by indicating the characteristic as a level per meta-characteristic using digits 1-4. For example, we can classify Voyager by Wang et al. (2023a) as 2131312 (See Figure 2). It has general knowledge scope since GPT-4, a generalist model, is used; unimodal multimodality, since it uses a text-based representation of the game; its built-in augmentation uses external processing, given that Voyager uses skill libraries as tools; its interactivity level is only input/output; it seems to be in continuous operation as long as it is not stopped; there is only one run, and therefore no user interaction memory; Voyager does display self-improvement capability with tool creation level, since it has a mechanism for iterative prompting and self-verification for program improvement, indicating the ability to create tools. Measures such as the mean would only be useful for comparison between agents if the importance of each meta-characteristic is similar for its purpose. The typology we propose offers multiple applications for different stakeholders. One application is comparison of agents, either existing offerings or proposed agent projects that require funding. The typology can help developers or strategy heads define aims and target gaps in technology. It can further be applied to contemplate the future and potential ethical implications of advances in agent technology. To illustrate the usefulness of the typology for comparing existing agents, consider a scenario where the management of a large institution is interested in participating in the the technological race and pilot the use of LLM-based AI agents to enhance the work of or replace their software developers. There are some recent agents that offer such capabilities, but which project is the most promising? Our typology facilitates a systematic analysis[4] (See Table 5 for agents and details): Of the assessed agents, two stood out, MetaGPT[5] and Devin from Cognition Labs[6]. Devin, in particular, has indications of superior capabilities by not only meeting higher relative levels in all categories. Thus, this structured classification empowers institutions to make informed decisions by clearly aligning agent capabilities with organizational needs and objectives, hopefully shortening costly trial and error processes.

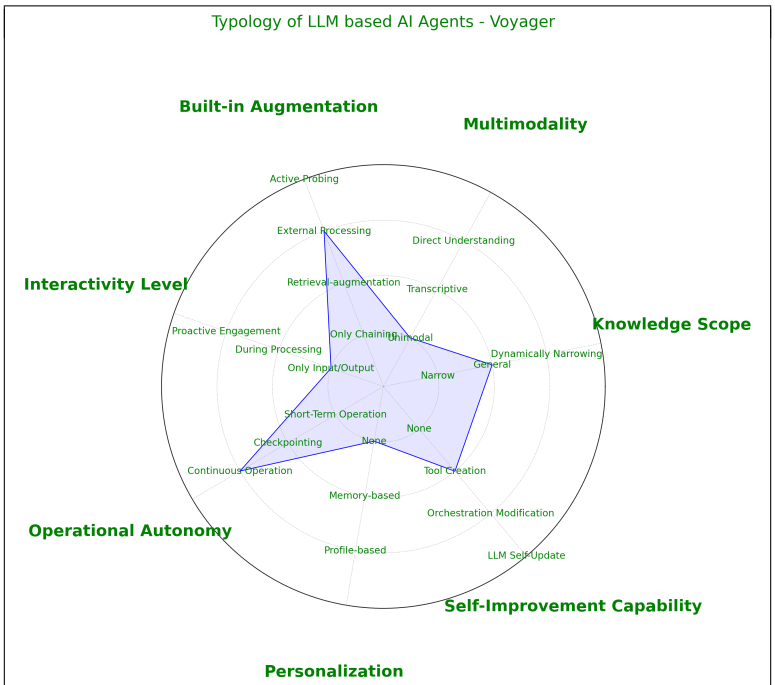

Figure 2. Spiderplot representation of the typology.

Another application of the typology lies in aiding the timing of investment decisions. By situating a prospective agent within the typology and juxtaposing its capabilities with those of existing or competitive agents, stakeholders can discern whether the current level of advancement in the field aligns with their requirements. This strategic placement allows for the identification of key developmental indicators, guiding investors to wait for an agent that meets the necessary criteria, thus optimizing the timing of their investments. On the other hand, if many agents already exist that meet the criteria, this information can help to shift resources onto as of yet unsolved-for aims. When ideating on how to enhance existing agents, the typology helps to identify areas where improvement is needed, especially where certain functionalities lag behind others. For entirely new functionalities, it offers a perspective on what to build-upon, namely if looking

[4]We omit knowledge scope and personalization for brevity, as the descriptions of the agents we found did not allow for a clear classification on these dimensions.

[5]https://github.com/geekan/MetaGPT

[6]https://www.cognition-labs.com/introducing-devin

at the mostly empty cells in the futuristic level. For example, the futuristic cell could involve anticipating changes in user preferences to preempt them becoming bored with the product. The typology can also help in backcasting. This method involves envisioning a future goal and then working backwards to identify the necessary steps to achieve it (Markus and Mentzer 2014).

Our typology, with its classification of meta-characteristics reflecting underlying technologies, simplifies this process. It provides a structured approach to understanding the advancements needed in areas like LLM knowledge scope or the integration of external tools for augmentation. Our typology further facilitates the envisioning of future scenarios. It serves as a foundational tool not just for creating tangible mock-ups of future agents to study their impact on user perceptions, but also for speculative analysis, aiding in the contemplation of future possibilities and their broader implications. This aspect of the typology is particularly vital for forward-looking research in the field of LLM agents.

Addressing ethical implications is another significant application of the typology. Instead of general discussions about the impact of agents on our lives, the typology allows for a more detailed and specific exploration of ethical consequences associated with various forms and functionalities of future agents. This focus enables a more nuanced and actionable discourse on the ethical dimensions of LLM technologies.

Agent	Funct. Scope	Interactivity Level	Oper. Autonomy	Multimodality	Self-Impr. Cap.
MetaGPT	3 (Dyn. Narrow.)	3 (Proact. Eng.)	3 (Cont. Oper.)	1 (Transcriptive)	3 (Orchestr. Mod.)
AzureOpenAI	2 (General)	2 (During Processing)	2 (Checkpointing)	1 (Transcriptive)	2 (Tool Creation)
Mistral AI	2 (General)	2 (During Processing)	2 (Checkpointing)	1 (Transcriptive)	2 (Tool Creation)
IBMWatson	2 (General)	2 (During Processing)	2 (Checkpointing)	3 (Dir. Underst.)	1 (None)
Devin	3 (Dyn. Narrow.)	3 (Proact. Eng.)	3 (Cont. Oper.)	1 (Transcriptive)	3 (Proact. Eng.)
Intervenor	1 (Narrow)	2 (During Processing)	2 (Checkpointing)	1 (Transcriptive)	2 (Tool Creation)

Table 5. Classification of LLM Agents

Practitioner Typology

Figure 3. Simplified practitioner typology

Our simplified practitioner typology for LLM based AI agents is designed to provide a clear and concise overview of the developmental stages of agents, using the heuristic that similar capabilities will likely evolve at similar speeds. This typology is structured into three main categories: Firstly, *Core Functional Capabilities*, encompassing Knowledge Scope, Multimodality, and Built-in Augmentation. It traces the evolution from basic, narrowly focused, and unimodal systems to advanced, dynamically narrowing, and, for example, spatially aware multimodal systems with sophisticated augmentation capabilities. Secondly, *Operational and Interactive Dynamics*, covering Interactivity Level and Operational Autonomy. This category highlights the progression from basic input/output interactivity and short-term operations to advanced levels of proactive engagement and continuous operation. Thirdly, *Adaptive and Self-Evolving Features*, including Personalization and Self-Improvement Capability. This category traces the development from systems with no personalization or self-improvement abilities to those with advanced profile-based personalization and the capability for self-updating.

Discussion

This paper introduces a typology for Large Language Model (LLM) based AI agents, offering a framework to understand and project their development. It outlines key characteristics and provides a comprehensive view of LLM-based AI agents, shedding light on their abilities and future directions.

The paper's key contribution is the precise definition of LLM-based AI agents. It provides a clear scope for LLM agent research while avoiding mere metaphorical use of the concept previously prevalent in information systems. The proposed typology classifies LLM agents across development stages and capabilities. It includes meta-characteristics like Knowledge Scope, Multimodality, and Operational Autonomy, spanning basic to advanced levels. This classification assists in comparing agents, guiding investments, developing new functionalities, and envisaging future scenarios. It serves as an academic and practical tool in the field. The typology aids future research, especially in forecasting new capabilities from current ones. It could help to identify correlations in LLM agent capabilities, offering insights into their developmental paths. This is vital in a rapidly advancing field, as it helps anticipate future trends and innovations. The typology also acts as a heuristic for understanding practical applications of different LLM agents. It demystifies the technology, providing a straightforward approach to interacting with LLM agents. Moreover, the paper's typology encourages exploring beyond current technologies, promoting imaginative and innovative thinking about future agents. It also has implications for understanding societal and user-centric aspects of LLM agents, offering a framework to consider their practical and ethical impacts.

The typology can help IS researchers understand LLM agent types that are not yet available to study. This is important considering the speed of progress and magnitude of impact in LLM-based AI agent development. By using the typology, researchers can promote the development of valuable agent types while preventing the development of harmful ones. It helps practitioners prepare their organizations and workflows for the future, easing the transition into an agent-filled environment. To illustrate this, let us consider a type each that urges promotion, prevention, or simply preparation[7]. *A type of LLM-based AI agent that many may want to promote* can be derived by considering an agent with the following characteristics. Let us assume a typology user is an Information Systems professor. The chosen agent characteristics would include: general knowledge scope, transcriptive multimodality, retrieval-augmentation built-in augmentation, interactivity during processing, checkpointing operational autonomy, memory-based personalization, and tool creation capability. These characteristics enhance the teaching and administrative tasks of IS professors by providing versatile, up-to-date, and personalized support. A general knowledge scope ensures the agent can handle a broad range of topics, while transcriptive multimodality aids in understanding various content formats. Retrieval-augmentation keeps the agent's information current, and interactivity during processing allows for refined queries. Checkpointing ensures the agent can perform extended tasks with periodic input, fitting academic workflows. Memory-based personalization tailors the agent's assistance to the professor's style, and tool creation adapts to evolving educational needs. For example, an IS professor could use the agent to create personalized quizzes based on student performance, ensuring targeted practice. The typology aids in identifying these beneficial features, allowing for proactive promotion and integration of such advanced agents into the educational environment. *A type that many may want to prevent the development of* can be derived by considering an agent with the following characteristics. Let us assume a typology user is an ethi-

[7]We have created a custom GPT to help apply the typology. The 3 use cases discussed below are the sample prompts provided if you follow the link. Disclaimer: The GPT may be pay-walled by the provider, its functioning depends on the operation of OpenAI and its proprietary models, which can change behavior or stop being accessible at any time. We are not associated with OpenAI and cannot guarantee for the outputs of the custom GPT. This is only for illustrative purposes, we did not validate the outputs of the GPT systematically: https://chatgpt.com/g/g-YreiOnxKo-llm-agent-typology-expert

cist working in a healthcare organization. The chosen agent characteristics would include: narrow knowledge scope, unimodal processing, built-in augmentation relying solely on chaining, proactive engagement, continuous operation, profile-based personalization, and orchestration modification. These characteristics can cause harm due to limited versatility and understanding, leading to frequent errors, misinterpretations, and outdated information. In healthcare, a narrow knowledge scope might lead to misdiagnosis, while unimodal processing could miss critical non-textual data. Built-in augmentation relying on chaining can result in unvalidated outputs, and proactive engagement may overwhelm patients with unsolicited advice. Continuous operation without oversight could perpetuate errors, while profile-based personalization raises privacy concerns. Orchestration modification allows unpredictable changes, deviating from medical protocols. To prevent these harms, the ethicist should advocate for strict oversight, limit agent autonomy, and ensure comprehensive testing. For instance, any proactive engagement by the agent must be reviewed by a human provider. Regular audits and updates of the knowledge base ensure current medical guidelines. Using the typology to anticipate these harmful features helps prevent the deployment of high-risk agents, ensuring patient safety and trust in healthcare services. *A type that will likely cause less friction if practitioners can prepare for it* can be derived by considering an agent with the following characteristics. Let us assume a typology user is working in a customer relationship department. The chosen agent characteristics would include: dynamically narrowing knowledge scope, direct multimodal understanding, built-in external processing, proactive engagement, continuous operation, profile-based personalization, and orchestration modification. These characteristics can initially cause friction due to integration challenges, such as ensuring access to relevant knowledge bases, handling input formats, integrating tools, and maintaining robust infrastructure. To prepare, practitioners should organize their knowledge bases, develop multimodal interfaces, and secure APIs for external processing tools. Updating engagement protocols to leverage proactive capabilities around the clock and ensuring reliable infrastructure for continuous operation is essential. Engagement protocols can be expanded to include automated follow-ups and reminders tailored to the customer's interaction history, setting clear guidelines for proactive outreach. For instance, implementing a system where the agent sends follow-up messages after a support interaction can check if issues were resolved satisfactorily and offer further assistance. The typology aids in anticipating these features, allowing proactive preparation and smoother implementation.

Our methodological approach has some limitations. It fundamentally carries subjectivity, a common trait in conceptual classifications, especially when exploring future scenarios with limited empirical data. The typology's complexity might challenge those less versed in its concepts (outlined in Table 3). The simplified version in Figure 3 does address this, however in exchange for comprehensiveness, which is why we only recommend it for a quick overview. We further have not operationalized the levels beyond face validity, which is an important step to take for future research.

Conclusion

This paper contributes to understanding and projecting the trajectory of LLM based AI agents. The detailed classification and practical implications of the proposed typology make it a valuable resource for academic and practical purposes. It facilitates informed decision-making and fosters innovative thoughts about these technologies' future. The typology is a significant addition to the discussion on LLM-based AI agents and their societal impact, helping to prepare our thinking about futures before they arrive.

References

Andreas, J. 2022. "Language Models as Agent Models," in *Findings of the Association for Computational Linguistics: EMNLP 2022,* Y. Goldberg, Z. Kozareva, and Y. Zhang (eds.). Abu Dhabi, United Arab Emirates: Association for Computational Linguistics, pp. 5769–5779.

Bailey, K. D. 1994. *Typologies and Taxonomies: An Introduction to Classification Techniques,* Thousand Oaks: Sage Publications, Inc.

Barandiaran, X. E., Di Paolo, E., and Rohde, M. 2009. "Defining Agency: Individuality, Normativity, Asymmetry, and Spatio-temporality in Action," *Adaptive Behavior* (17:5), pp. 367–386.

Bavaresco, R., Silveira, D., Reis, E., Barbosa, J., Righi, R., Costa, C., Antunes, R., Gomes, M., Gatti, C., Vanzin, M., Junior, S. C., Silva, E., and Moreira, C. 2020. "Conversational agents in business: A systematic literature review and future research directions," *Computer Science Review* (36), p. 100239.

Brendel, A. B., Hildebrandt, F., Dennis, A. R., and Riquel, J. 2023. "The Paradoxical Role of Humanness in Aggression Toward Conversational Agents," *Journal of Management Information Systems* (40:3), pp. 883–913.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., and Amodei, D. 2020. *Language Models are Few-Shot Learners*.

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M. T., and Zhang, Y. 2023. *Sparks of Artificial General Intelligence: Early experiments with GPT-4*.

Cassell, J. and Thorisson, K. R. 1999. "The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents," *Applied Artificial Intelligence* (13:4-5), pp. 519–538.

Chandwani, A., Shrivastava, K., Sadashiv, S., Saha, I., and Mishra, P. 2023. "Virtual Assistant for Appointment Booking," in *2023 IEEE 8th International Conference for Convergence in Technology (I2CT),* IEEE, pp. 1–5.

Chen, Y., Ding, N., Zheng, H.-T., Liu, Z., Sun, M., and Zhou, B. 2023. *Empowering Private Tutoring by Chaining Large Language Models*.

Cheng, Y., Zhang, C., Zhang, Z., Meng, X., Hong, S., Li, W., Wang, Z., Wang, Z., Yin, F., Zhao, J., and He, X. 2024. *Exploring Large Language Model based Intelligent Agents: Definitions, Methods, and Prospects*.

Crouse, M., Abdelaziz, I., Basu, K., Dan, S., Kumaravel, S., Fokoue, A., Kapanipathi, P., and Lastras, L. 2023. *Formally Specifying the High-Level Behavior of LLM-Based Agents*.

Dennett, D. C. 1988. "Précis of The Intentional Stance," *Behavioral and Brain Sciences* (11:03), p. 495.

Diederich, S., Brendel, A. B., Morana, S., and Kolbe, L. 2022. "On the Design of and Interaction with Conversational Agents: An Organizing and Assessing Review of Human-Computer Interaction Research," *Journal of the Association for Information Systems* (23:1), pp. 96–138.

Elshan, E., Ebel, P., Söllner, M., and Leimeister, J. M. 2023. "Leveraging Low Code Development of Smart Personal Assistants: An Integrated Design Approach with the SPADE Method," *Journal of Management Information Systems* (40:1), pp. 96–129.

Göldi, A. 2024. "Flexibility in Chatbot Identity Perception," in *AMCIS 2024 Proceedings,* 2024.

Göldi, A. and Rietsche, R. 2023. "Whereto for Automated Coaching Conversation: Structured Intervention or Adaptive Generation?," in *ECIS 2023 Research-in-Progress Papers,* 2023.

Göldi, A. and Rietsche, R. 2024. "Insert-expansions for Large Language Model Agents," in *AMCIS 2024 Proceedings,* 2024.

Göldi, A., Wambsganss, T., Neshaei, S. P., and Rietsche, R. 2024. "Intelligent Support Engages Writers Through Relevant Cognitive Processes," in *Proceedings of the CHI Conference on Human Factors in Computing Systems,* ACM, 2024, pp. 1–12.

Gregor 2006. "The Nature of Theory in Information Systems," *MIS Quarterly* (30:3), p. 611.

Han, E., Yin, D., and Zhang, H. 2023. "Bots with Feelings: Should AI Agents Express Positive Emotion in Customer Service?," *Information Systems Research* (34:3), pp. 1296–1311.

Händler, T. 2023. *Balancing Autonomy and Alignment: A Multi-Dimensional Taxonomy for Autonomous LLM-powered Multi-Agent Architectures*.

Huang, X., Liu, W., Chen, X., Wang, X., Wang, H., Lian, D., Wang, Y., Tang, R., and Chen, E. 2024. *Understanding the planning of LLM agents: A survey*.

Hunt, S. 1991. "Modern marketing theory: Critical issues in the philosophy of marketing science," ().

Kompa, N. A. 2023. "Inner Speech and 'Pure' Thought – Do we Think in Language?," *Review of Philosophy and Psychology* ().

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., Rocktäschel, T., Riedel, S., and Kiela, D. 2020. "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," in *Advances in Neural Information Processing Systems,* H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.). Vol. 33. Curran Associates, Inc, pp. 9459–9474.

Li, Y., Wen, H., Wang, W., Li, X., Yuan, Y., Liu, G., Liu, J., Xu, W., Wang, X., Sun, Y., Kong, R., Wang, Y., Geng, H., Luan, J., Jin, X., Ye, Z., Xiong, G., Zhang, F., Li, X., Xu, M., Li, Z., Li, P., Liu, Y., Zhang, Y.-Q., and Liu, Y. 2024. *Personal LLM Agents: Insights and Survey about the Capability, Efficiency and Security*.

Luo, B., Lau, R. Y., and Li, C. 2023. "Emotion-regulatory chatbots for enhancing consumer servicing: An interpersonal emotion management approach," *Information & Management* (60:5), p. 103794.

Mahowald, K., Ivanova, A. A., Blank, I. A., Kanwisher, N., Tenenbaum, J. B., and Fedorenko, E. 2023. *Dissociating language and thought in large language models: a cognitive perspective*.

Markus, M. L. and Mentzer, K. 2014. "Foresight for a responsible future with ICT," *Information Systems Frontiers* (16:3), pp. 353–368.

Masoudnia, S. and Ebrahimpour, R. 2014. "Mixture of experts: a literature survey," *Artificial Intelligence Review* (42:2), pp. 275–293.

Mayer, J. D., Salovey, P., and Caruso, D. R. 2012. *PsycTESTS Dataset,*

Meier, A., Rietsche, R., and Blohm, I. 2024. "An AI Approach for Predicting Audience Reach of Presentation Slides," in *ECIS 2024 Proceedings,* 2024.

Meyer, R. 2007. "Mapping the mind of the strategist," Diss. Univ. Rotterdam, 2007. Rotterdam.

Nickerson, R. C., Varshney, U., and Muntermann, J. 2013. "A method for taxonomy development and its application in information systems," *European Journal of Information Systems* (22:3), pp. 336–359.

Nunamaker, J. F., Derrick, D. C., Elkins, A. C., Burgoon, J. K., and Patton, M. W. 2011. "Embodied Conversational Agent-Based Kiosk for Automated Interviewing," *Journal of Management Information Systems* (28:1), pp. 17–48.

OpenAI 2023a. *OpenAI DevDay*.

OpenAI 2023b. *OpenAI Platform Documentation*.

Parisi, A., Zhao, Y., and Fiedel, N. 2022. *TALM: Tool Augmented Language Models*.

Park, J. S., O'Brien, J. C., Cai, C. J., Morris, M. R., Liang, P., and Bernstein, M. S. 2023. *Generative Agents: Interactive Simulacra of Human Behavior*.

Pickering, M. J. and Gambi, C. 2018. "Predicting while comprehending language: A theory and review," *Psychological bulletin* (144:10), pp. 1002–1044.

Schlosser, M. 2019. "Agency," in *The Stanford Encyclopedia of Philosophy,* Edward N. Zalta (ed.). Metaphysics Research Lab, Stanford University.

Tulving, E. 1985. "How many memory systems are there?," *American Psychologist* (40:4), pp. 385–398.

Wang, G., Xie, Y., Jiang, Y., Mandlekar, A., Xiao, C., Zhu, Y., Fan, L., and Anandkumar, A. 2023a. *Voyager: An Open-Ended Embodied Agent with Large Language Models*.

Wang, L., Ma, C., Feng, X., Zhang, Z., Yang, H., Zhang, J., Chen, Z., Tang, J., Chen, X., Lin, Y., Zhao, W. X., Wei, Z., and Wen, J.-R. 2023b. *A Survey on Large Language Model based Autonomous Agents*.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., and Zhou, D. 2022. *Chain-of-Thought Prompting Elicits Reasoning in Large Language Models*.

Wooldridge, M. and Jennings, N. R. 1995. "Intelligent agents: theory and practice," *The Knowledge Engineering Review* (10:2), pp. 115–152.

Wu, Q., Bansal, G., Zhang, J., Wu, Y., Zhang, S., Zhu, E., Li, B., Jiang, L., Zhang, X., and Wang, C. 2023. *AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation*.

Wu, T., Terry, M., and Cai, C. J. 2022. "AI Chains: Transparent and Controllable Human-AI Interaction by Chaining Large Language Model Prompts," in *CHI Conference on Human Factors in Computing Systems,* S. Barbosa, C. Lampe, C. Appert, D. A. Shamma, S. Drucker, J. Williamson, and K. Yatani (eds.). New York, NY, USA: ACM, pp. 1–22.

Zhao, P., Jin, Z., and Cheng, N. 2023. *An In-depth Survey of Large Language Model-based Artificial Intelligence Agents*.

Zhiheng, X., Wenxiang, C., Xin, G., Wei, H., Yiwen, D., Boyang, H., Ming, Z., Junzhe, W., Senjie, J., Enyu, Z., Rui, Z., Xiaoran, F., Xiao, W., Limao, X., Yuhao, Z., Weiran, W., Changhao, J., Yicheng, Z., Xiangyang, L., Zhangyue, Y., Shihan, D., Rongxiang, W., Wensen, C., Qi, Z., Wenjuan, Q., et al. 2023. *The Rise and Potential of Large Language Model Based Agents: A Survey*.

Zhou, W., Jiang, Y. E., Li, L., Wu, J., Wang, T., Qiu, S., Zhang, J., Chen, J., Wu, R., Wang, S., Zhu, S., Chen, J., Zhang, W., Zhang, N., Chen, H., Cui, P., and Sachan, M. 2023. *Agents: An Open-source Framework for Autonomous Language Agents*.