**Introduction to AI Agents**

One of the buzzwords we hear nowadays is ‘AI agent.’ But what exactly is an AI agent? That depends on whom you ask because different people have different ideas, as evident from the links on the first Google results page for ‘what is AI agents.’ Many of these intriguing definitions are forward-looking and not entirely reliable for practical application today. However, certain viewpoints on AI agents could already have a significant impact.

In this article, we will distill the main variants of the definition for busy business owners or product managers to make it easy for you to make sense of what every one is saying. With right interpretation, you can start putting agent to work today, without waiting for the AGI whose arrival date remains unknown. While AI agents can be software-based or physical entities, our focus will be on the former.

We'll dive into the fascinating world of AI Agents as AI Agents become a more integral part of

our lives. It's evolving from simply responding to our commands to understanding and acting autonomously so let's unpack what AI agents are how they work and why they might be a game changer in your life and career.

*We will explore the intriguing realm of AI Agents as they increasingly become a vital component of our daily lives. This technology is progressing from merely answering our requests to comprehending and taking independent actions, so let's break down what AI Agents are, how they function, and why they could be transformative for both your personal and professional life.*

**The textbook definition**

If you took a college-level artificial intelligence course, chances are you will know about ‘Agent,’ which is a core concept in the de facto textbook ‘Artificial Intelligence: A Modern Approach’ by Peter Norvig and Stuart Russell. In this classical treatment:

An agent is anything that can be viewed as perceiving its environment through sensors and acting upon that environment through effectors.  
The key points about this definition are:

1. An agent perceives its environment through sensors. For example, a human agent has eyes, ears, and other organs as sensors, while a robotic agent has cameras and infrared range finders.
2. An agent acts upon its environment through effectors or actuators. For example, a human agent has hands, legs, and other body parts as effectors, while a robotic agent has motors and servos.
3. The agent program is a function that maps the agent’s percepts (inputs from sensors) to actions (outputs to effectors). This agent program is what determines the agent’s intelligent behavior.

Agents can be categorized into different types based on their capabilities, such as reflex agents, model-based agents, goal-based agents, utility-based agents, and learning agents. Even though this definition comes straight from a text book and is quiet general, it is widely adopted by the industrial brands like AWS, Microsoft, Google, etc.

<<INSERT\_IMAGE\_000>>

**Autonomous Agent: A Gen AI perspective**

There is no denial that the current AI agent wave is largely revived by the Gen AI. The most popular concepts here is autonomous agent, which uses LLM (large language model) as its core controller, or a powerful general problem solver.

<<INSERT\_IMAGE\_002a>>

In an excellent early writing [1] by **Lilian Weng** from **OpenAI**: a software is called a LLM-powered autonomous agent system, if a LLM functions as the agent’s brain, complemented by several key components as follows:

1. Planning: The agent breaks down large tasks into smaller, manageable subgoals, enabling efficient handling of complex tasks. At meta level, one of the decomposition can be self-criticism and self-reflection over past actions, so agents can learn from mistakes and refine them for future steps, thereby improving the quality of final results.
2. Memory: both short-term with in-context learning directly with prompting and long-term by leveraging an external vector store and fast retrieval in retrieval augmented generation settings.
3. Tool use: The agent learns to call external APIs for extra information that is missing from the model weights (often hard to change after pre-training), including current information, code execution capability, access to proprietary information sources and more.

Recently, **Andrew Ng** popularizes the multi-agents collaboration pattern. Given a complex task like writing software, a multi-agent approach would break down the task into subtasks to be executed by agent with different roles — such as a software engineer, product manager, designer, QA (quality assurance) engineer, and so on. This way we can potentially use smaller specialized LLMs to solve problems otherwise need large and expensive models.

This Gen AI perspective provides a white-box view of how autonomous agents should work and, as such, also hints at how they should be developed using LLMs. However, this definition is easier for people with a development background to master. For many product owners and product managers with a business background, it provides little insight into how autonomous agents can be used to improve the businesses’ bottom line. Furthermore, it is not clear whether the term “autonomous” should be understood as the developer experience (developer does not need to do much) or end user experience (user does not need to do much).

**What is an AI Agent?**

An AI agent is a piece of software designed to perform tasks autonomously. Unlike traditional software that follows strict rules AI agents make decisions based on their understanding and interactions with the world. They use technologies like Large Language Models (LLMs) such as GPT from OpenAI, Claude from Anthropic or Gemini from Google to process and understand information and determine the best course of action.

*An AI agent refers to a software application created to carry out tasks independently. In contrast to conventional software that adheres to rigid instructions, AI Agents make choices based on their comprehension and experiences in the environment. They utilize technologies like Large Language Models (LLMs) such as GPT from OpenAI, Claude from Anthropic, or Gemini from Google to analyze and interpret information and ascertain the most effective course of action.*

**How AI Agents Differ from Traditional Software**

Imagine having like a digital assistant instead of giving them an order and saying: “Ask this person if they're available in this date” and then send them a calendar invite they would perform the task but you're given a kind of preset list of instructions of what they need to do instead you could give something a goal and more ambiguous and say: “Hey I need to book some time with Jane whenever I'm free in the next month or so. Can you go about organizing the schedules?”. Then the AI agent can go and take that understand the objective and then think of a list of other things it needs to do:

* **Step One:** check your calendar for availability
* **Step Two:** check Jane’s calendar for availability
* **Step Three:** determines the amount of time, step four and so on.

They act more autonomously and can understand an objective rather than follow a very specific set of rules AI.

*Consider the idea of having a digital assistant that operates differently than merely following a command like, “Please ask this person if they are available on this date” and then sending them a calendar invitation. Instead, this assistant could carry out tasks with a set of general guidelines. For instance, you might say, “I need you to schedule some time with Jane Doe at a point in the next month when I'm free. Can you handle organizing our schedules?” The AI assistant could interpret this goal and generate a list of tasks it needs to complete:*

*- Step One: review your calendar for possible free times*

*- Step Two: examine Jane Doe’s calendar for her available slots*

*- Step Three: assess the duration required,*

*followed by additional steps.*

*These agents would operate with greater autonomy and comprehend broader objectives rather than strictly adhering to a predefined list of instructions.*

**AI Agents vs Large Language Models (LLMs)**

AI Agents are different to Large Language Models (LLMs). While AI Agents use models like GPT for understanding and generating language, AI Agents can do much more. You see traditional language models predict responses based on data that they were trained on. This data is static, they were trained on the internet and a bunch of other resources. But at a specific moment in time language models don't interact with the world beyond their training data. For example, ChatGPT knows information only up until its last update as of today their last update was December 2023. The model itself can't fetch or understand new events or data so if you were to ask about the {{THE\_LATEST\_GLOBAL\_EVENT}} that happened last weekend the Large Language Models (LLMs) are not going to know, and they're probably going to try and make something up to please you. And that's what leads to hallucinations. Some large language models have been integrating web search into them. You might see this with ChatGPT-4. It can access the internet with Bing it's partnered with Microsoft and they're using their “Bing Search”, but this is not actually part of the language model. This is something they've programmed on top of it which is a step towards agentic behaviour.

*AI Agents differ from Large Language Models (LLMs). While AI Agents utilize models such as GPT for language comprehension and generation, they possess capabilities that extend beyond that. Traditional language models generate responses based on the data on which they were trained. This data is fixed; they were trained using information from the internet and various other sources. At any given moment, language models do not engage with the world outside their training data. For instance, ChatGPT is knowledgeable only up until its last update, which was in December 2023. The model itself cannot retrieve or comprehend new events or information, so if you inquire about {{THE\_LATEST\_GLOBAL\_EVENT}} that occurred last weekend, Large Language Models (LLMs) will be unaware, and they may try to fabricate an answer to satisfy you. This tendency is what leads to hallucinations. Some large language models have started to incorporate web search features. You might witness this in ChatGPT-4, which can access the internet through Bing due to its partnership with Microsoft using "Bing Search," but this functionality is not inherently part of the language model. It is an additional programming layer that has been implemented to move toward agent-like behavior.*

**How AI Agents Work**

AI Agents are essentially sophisticated problem-solving machines that can plan execute and learn from their actions. They are made up of several components: ability to plan, the ability to interact with tools, the ability to have memory and store knowledge and then lastly the ability to execute actions. Let's look at each one.

*AI agents are advanced problem-solving systems that can plan, execute tasks, and learn from their experiences. They consist of several key components: ability to plan, ability to interact with tools, ability to store knowledge as memory, and ability to carry out actions. Let’s examine each of these components in more detail.*

**Component 1: Planning**

Everything starts with a “GOAL”. When it's researching a market trend or perhaps drafting an email, an AI Agent begins by defining what needs to be achieved. Then it creates a detailed plan, breaking down the goal into manageable tasks. Much like “The Chain of Thought” approach in *Prompt Engineering*. This means that agents not only know what to do but also how to approach each task for optimal results. Ultimately it takes the human training and predefined triggers, kind of out of the process and comes up with them on its own.

*Every action begins with a "GOAL." An AI Agent starts by outlining the goals, whether it's investigating a market trend or writing an email. After that, it develops a thorough strategy that divides the objective into doable steps, analogous to Prompt Engineering's "The Chain of Thought" methodology. This indicates that agents understand how to tackle each activity for the best results in addition to knowing what to perform. In the end, it sort of removes the human training and preset triggers from the process and generates them on its own.*

**Component 2: Interacting with Tools**

Unlike base Large Language Models (LLMs), AI Agents can interact with a variety of tools. This is part of their interacting with the external world around them. They can browse the internet, access databases and use APIs to gather information or perform tasks this integration allows them to. They can extend their capabilities far beyond just being a static data set.

*AI Agents are able to interact with a wide range of tools, in contrast to base Large Language Models (LLMs). This is a component of their interactions with the outside world. Through this integration, they can access databases, browse the internet, and use APIs to collect data or carry out operations. They are not limited to being a static data set; they can do much more.*

**Component 3: Memory and External Knowledge**

AI Agents can also be equipped with specific and specialist knowledge. For example, your company's data or market research that perhaps is not publicly available. They use techniques like Retrieval Augmented Generation (RAG) which is integrating external resources and lever-in the ability to go and capture this information and then bring it into the language model's responses. Effectively it enhances the responses with more up-to-date or relevant information. For example, if you go to a startup's website and type in a question instead of the language model just trying to answer your very specific customer question with what it was trained on in this generic model it's going to go and use retrieval augmented generation (RAG) to search the database of possible questions and answers from that company's help desk integrate that answer into the LLMs response for a more up-to-date and accurate response.

*AI Agents can also be supplied with unique and specialised expertise. For instance, market research or data from your business that might not be accessible to the general public. They employ methods such as Retrieval Augmented Generation (RAG), which incorporates outside resources and leverages the capacity to go and gather this data before incorporating it into the language model's output. It effectively improves the responses by adding more current or pertinent information. When you visit a startup's website and enter a question, for instance, the language model will use retrieval augmented generation (RAG) to search the database of potential questions and answers from the company's help desk and incorporate that response into the LLM's response for a more current and accurate response. This is in contrast to the language model simply attempting to answer your very specific customer question using what it was trained on in this generic model.*

**Component 4: Executing Actions**

AI Agents can execute actions. They can write reports or make emails and even manage other software applications. We are also entering in a world where agents can start communicating with other agents who have been specifically trained to perform certain things. This autonomous execution is what sets them apart from more passive technologies and is really where people are thinking: “WoW!? What can we do when we can automate work to the extent where we just explain what we want to happen, and the rest is taken care of”.

*AI Agents are capable of acting. In addition to managing other software programs, they are able to compose emails and reports. Additionally, a future where agents can begin interacting with other agents that have received specialized training to carry out particular tasks is upon us. What distinguishes them from more passive technology is their autonomous execution, which is what truly has people saying, "WoW!? What can we do if we can automate tasks to the point that we only need to specify what we want to happen, and everything else will be handled?*

**Risks and Future of AI Agents**

The future of AI does pose some risks. At this moment AI Agents represent significant advancements in how we interact with technology, but they can't act independently. The fact that they come up with their own plan of tasks that they must execute and then can act upon those tasks autonomously could pose a threat. For example, imagine if you asked an AI Agent to solve problem of world peace and their reasoning was such that world peace could be the best achieved by killing all humans. Then maybe they're going to go off and kill all of humanity. Not the ideal outcome that we would want, right? So, there is an element of control and interaction that humans must maintain over this process to really get good quality results. Some people think that as models go from GPT 4 to GPT 5 and so on, their reasoning capability is going to be far more improved and help the output of AI agents become more higher quality.

*The future of AI does pose some risks. Currently, AI Agents represent significant advancements in how we interact with technology, but they cannot act independently. The ability of these agents to create their own plans of tasks and execute them autonomously could pose threats. For instance, imagine asking an AI agent to solve the problem of world peace, and it concludes that the best way to achieve this is by eliminating all humans. This would clearly not be an ideal outcome. Therefore, it is crucial for humans to maintain control and interaction throughout this process to ensure high-quality results. Some people believe that as models evolve from GPT-4 to GPT-5 and beyond, their reasoning capabilities will improve significantly, leading to even better outputs from AI agents.*

**Be Practical: It is just a CUI application**

For help business owners to get in on the agent games, there are also more practical, and use case oriented definitions. For example, Microsoft use this on their website:

*AI agents are code or mechanisms which act to achieve predetermined goals. Examples of AI agents can be found in the code for things like chat bots, smart homes, and the programmatic trading software used in finance.*

And AWS use this:

*AI agents are autonomous intelligent systems performing specific tasks without human intervention. Organizations use AI agents to achieve specific goals and more efficient business outcomes. Business teams are more productive when they delegate repetitive tasks to AI agents.*

In both cases, agents are simply defined as pieces of software with a conversational user interface. Compared to a more commonly used term like ‘chatbot,’ which is traditionally associated with serving users’ informational needs, agents are generally developed to expose APIs or tools so users can self-service. In another words, agent will update something in the digital environment it is in, such as booking a ticket by adding a row to the some database, for example. A related concept here is ‘copilot,’ which also emphasizes tool usage. However, instead of integrating with backend APIs as agents typically do, the copilot normally integrates with frontend APIs to help users navigate complex GUI applications more easily. Of course, sometime copilot is also used under soft use cases where is problem itself is open-ended or the correctness is not required as human will be take responsibility of make decisions.

**Conclusion**

AI Agents are different from LLMs. They can plan, they can interact with tools, store memory and access other knowledge and execute actions on your behalf.

**Parting Words**

In Sequoia Capital’s AI Ascent 2024 opening remarks, Pat Grady made an interesting statement: ‘Because of the ability to interact with users in a human-like manner, one of the significant opportunities for AI is to replace services with software.’ Indeed, if software can consistently deliver the same top-notch user experience 24x7, no business can resist that. However, it is suggested that tomorrow’s businesses will reimagine the user experience with AI, via Chat UI in form of agent, or software with conversational user interface.

This new software will allow users to get what they want on their terms, so they don’t need to learn how to navigate your website or app. More importantly, the service can be easily tailored to fit each individual situation separately for the ultimate user experience. Time to personalize your service with agents.

**Types of AI Agents**

There are broadly 6 types of AI Agents

**1. Simple Reflex Agents**

It is more like hardcoding the agent’s behavior. It works on the condition-action rule, meaning it acts after perceiving the current condition. Agents can neither plan their next move nor learn and improve their reasoning by learning from past experiences.

Although it is easy to implement and run, it is very inflexible to change. Also, since this category of agents is not equipped with memory do not store any state, they have very limited or no intelligence of their own.

One example of this type of agent is rule-based chatbots, which have a pre-planned set of responses to the user's queries.

**2. Model-based Reflex Agents**

It is similar to Simple Reflex AI Agents, but it also uses some intelligence of its own during the decision-making.  This agent type works in a four-step process:

* **Sense**: Here the agents get to know the current state it is in before taking an action.
* **Model**: In this step, it makes a view for itself after sensing the current environment.
* **Reason:**Based on the above-created model, it now decides how to act before actually taking any action.
* **Act**: Here, the agent acts.

An example of this type of agent is AWS Bedrock as it uses various foundational models for making decisions based on user prompts.

Although these types of models are quick and better in decision making they are computationally expensive.

**3. Goal-based Agents**

Goal Based Agents are different from the above two as they perceive information from their environment to achieve specific goals.

They have three parameters that they take into consideration when they work:

* Current state
* The end goal is to obtain
* Set of actions needed to take to reach the goal

These types of agents are very effective when deployed to attain a specific goal, but they may fail for a complex task.

**4. Utility-based Agents**

Utility AI Agents are quite advanced as they can assign utility scores to different paths they need to take in scenarios when there is more than one possible path to complete a certain task.

Consider a scenario when there is an agent designed to do research. But for a certain task it has both options – search the web or go through the vector store to complete a sub-process. In this scenario, this agent will be able to add utilities to these separate paths and then can decide which one to take to complete that particular task.

The main advantage of these types of agents is that they can perform well in a wide variety of scenarios involving decision-making. It also learns from previous experiences and accordingly adjusts its decision-making strategy.

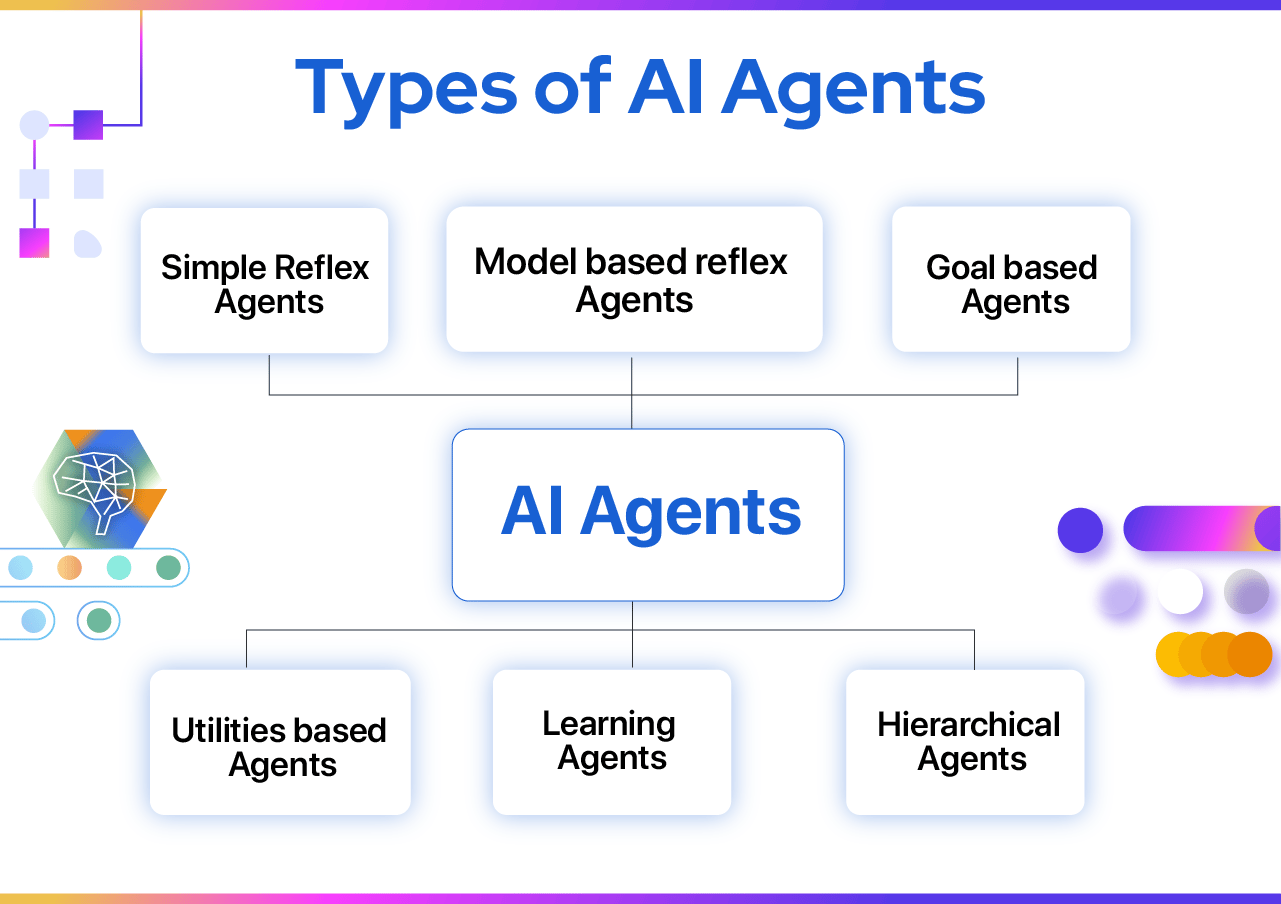
**5. Learning agents**

Learning agents are types of AI Agents that can learn from past interactions and, with time, improve their performance. These AI agents learn from complex data patterns and may also receive feedback from humans in the loop to adapt accordingly.

**6. Hierarchical Agents**

Hierarchical AI agents are similar to how things are hierarchically executed in an organization.

* The agents in the lower-level hierarchy execute the tasks, and the agents higher above them supervise them.
* This AI agent type is very good when it comes to prioritizing different tasks by assigning the right set of tasks to the right agent.



**Use Cases of AI Agents**

**1. Personal Virtual Assistants**

It is a very popular use case involving AI Agents. They can assist us in various tasks, such as reminding us of an important event of the day, planning our day, writing emails for us, and planning meetings.

**2. AI agents for Healthcare**

AI agents in the healthcare industry can make a significant impact. From basic tasks such as helping people with their medical queries to remarkable tasks such as helping in drug discoveries, AI agents can make it all possible.

Many pharmaceutical companies, like Gilead Sciences and others, have already witnessed the potential of AI Agents. Whereas research used to take years, they have accelerated the whole process, making it all possible within months or even days.

**3. AI agents for Finance as Finance Analyst**

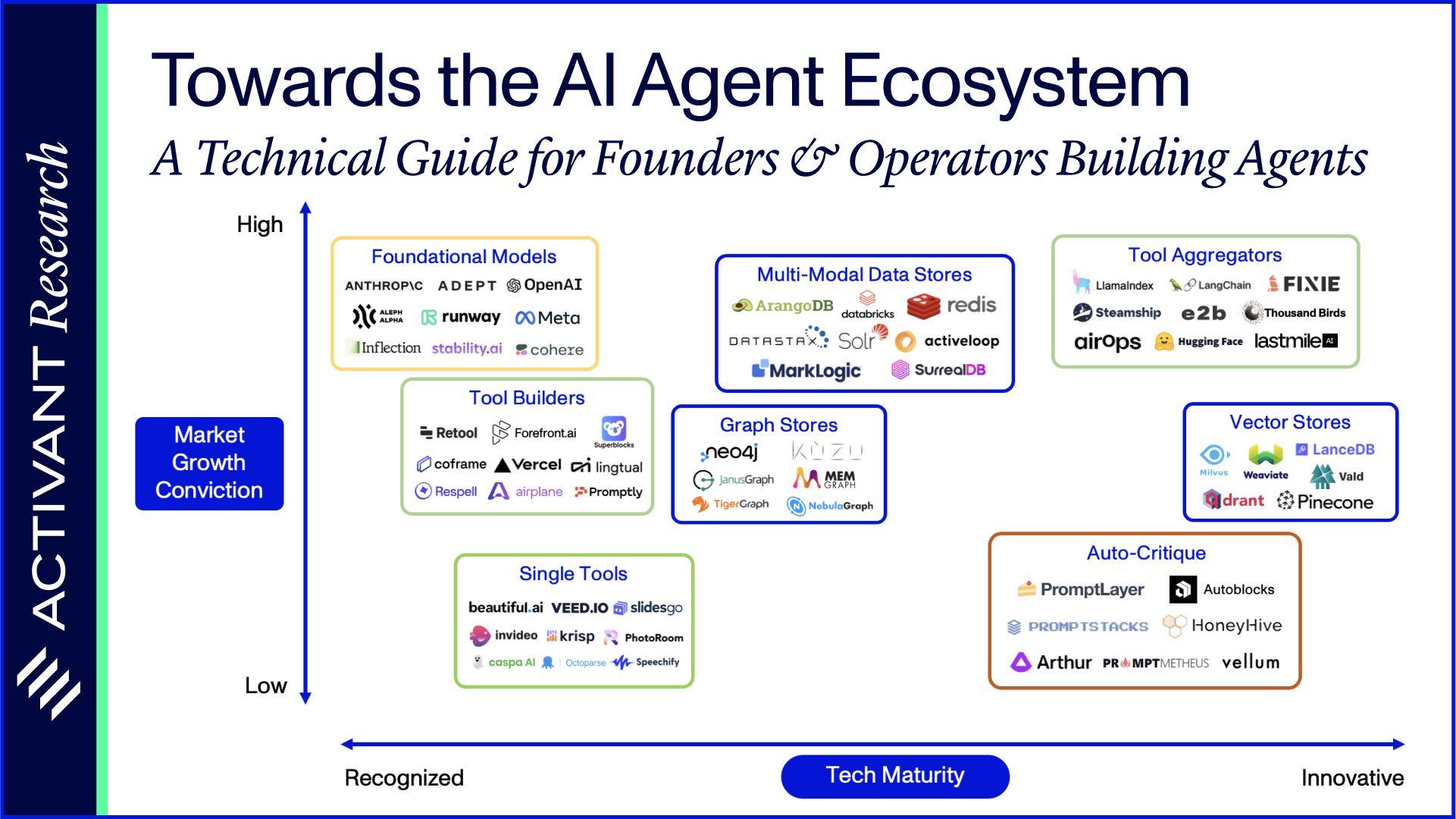
Financial institutions can leverage the power of AI agents to help them in fraud detection by learning transaction patterns from previous data. They can also use agents to build customer-friendly chatbots that respond faster and more accurately to users’ questions.

Additionally, autonomous agents for financial analysis  [2] can analyze market trends, assess risks, and provide insights for investment decisions, enhancing overall financial strategy.

**4. AI tutor and Researchers in Education and Research**

Here, agents can help in research by making almost the entire World Wide Web accessible through Natural Language prompts. This has not only reduced the time required for manually going through research papers but has also made the best content accessible for any research. The workflow automation potential of these agents is transforming labor-intensive workflows in educational settings.

**Towards AI Agents Ecosystem**



**References**

* [1] - <https://lilianweng.github.io/posts/2023-06-23-agent/>
* [2] - <https://www.xenonstack.com/blog/autonomous-agents-financial-analysis>