Simbian

AI Agents in Cybersecurity

AN OPPORTUNITY TO SOLVE SECURITY WITH AI

Table of Contents

- What are AI Agents? 2
- Difference between AI Agents, Chatbots, and Co-pilots 5
- AI Agents in Cybersecurity 7
- How to adopt AI Agents in cybersecurity 8
- Expected gains from AI Agents 10
- Work suitable and unsuitable for AI Agents 11
- Pitfalls to watch while using AI Agents in security 12
- Conclusion 13

1. What are AI Agents?

Imagine that you have a very tedious process that you need to keep performing all the time - for example, gathering information, analyzing information, summarizing information, creating reports and so on. Traditionally, humans need time, expertise and energy to perform this process repeatedly.

Automation can significantly help to ease the load on humans and lead to time savings, cost savings, and better reliability. In the modern age, automation is traditionally performed using code, for instance Python, to run a process reliably. This automation works well for fixed and repeatable processes.

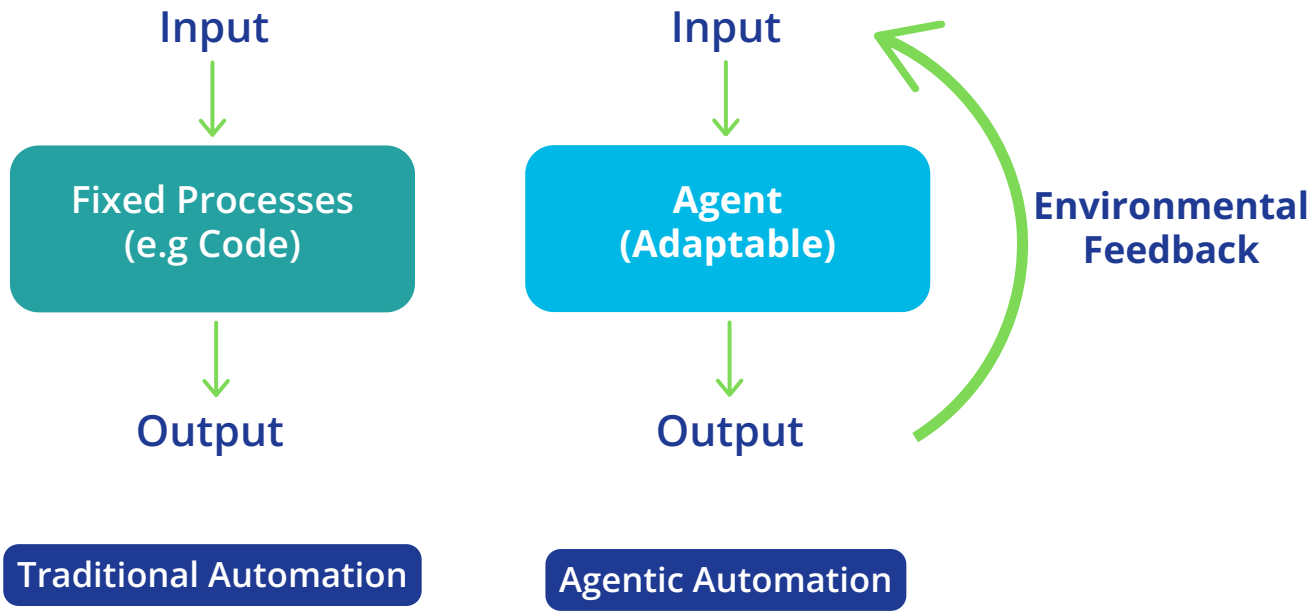

Figure 1 An Agentic-based automation is more dynamic and adaptable to real-world deviations compared to fixed processes like code

Increasingly, it turns out that automation based on fixed code cannot adapt to dynamic environments easily, as the code is static once created and cannot be changed easily, as illustrated in **Fig. 1.** To bridge the gap between a dynamic world and standard fixed automation pipelines, a new form of technology called AI Agents is born.

AI Agents are advanced software programs designed to autonomously interact with their environment, process information, and make decisions without human intervention. Unlike simple automation tools, AI Agents can dynamically adapt to

added information, and modify their responses based on user preferences or environmental feedback. Rather than having fixed process pipelines and executing them the same way each time, AI Agents can receive feedback from the environment and adapt its processes accordingly based on whether the previous steps were successful.

This is much more powerful than a traditional if-else loop in programming to select paths in a flowchart – it enables a much richer understanding of the environment feedback at run-time, and can even allow for understanding of scenarios which may not be planned for, so long as there is semantic meaning in the environment response.

The core of the Agentic AI loop is enabled by increased cognitive understanding of multimodal information via Large Language Models (LLM) or Large Multimodal Models (LMM). Simbian uses AI Agents at the core of its product to enable fast adaptation to a changing security landscape (**see Fig. 2**)

Figure 2 Simbian uses AI Agents at its core to make it easy to switch between vendors and thrive in a changing security landscape

An LLM or LMM alone is not sufficient to reason and understand about the world, as it is typically only a good pattern matcher for inputs to outputs. We will need a lot more modules interfacing with the LLM or LMM in order to have increased robustness in the pipeline. Such modules can be memory modules to store past knowledge and reasoning traces, planning modules to simulate and select an action plan, using reliable and robust tools for deterministic actions (e.g. code generator, logical modules, calculator) and extending the agent's capabilities (e.g. web browsing image processing), and learning agent's capabilities (e.g. web browsing, image, processing), and learning from experience via memory (see Fig. 3).

Once the LLM and LMM are interfaced with modules in a larger ecosystem, it is typically termed as an AI Agent and has numerous benefits to boosting performance of the baseline LLM or LMM. In fact, when we think about Agents, we do not just limit ourselves to just one Agent, but a multiple of them for the same pipeline or different pipelines. AI Agents are general, versatile and are able to be deployed across various environments.

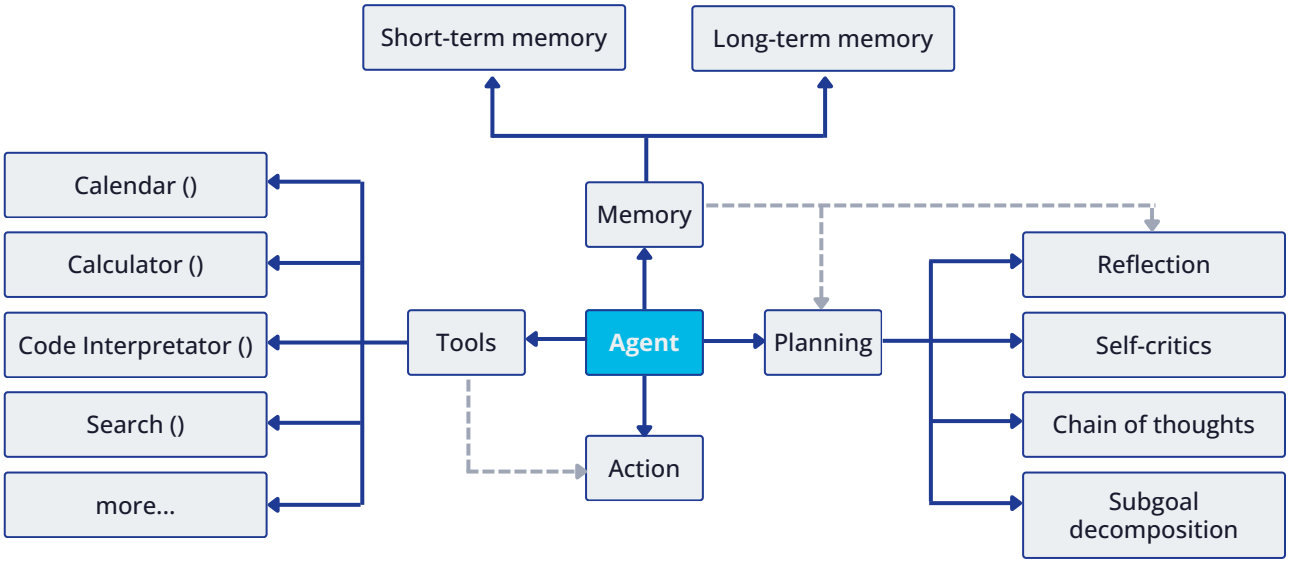

Figure 3 Components of an AI Agent. Extracted from Lilian Weng's blog post. https://lilianweng.github.io/posts/2023-06-23-agent/

To use an AI Agent in production, we can look at one of the many open-sourced Agentic frameworks, such as TaskGen (our very own in-house Agentic Framework), LangChain / LangGraph, AutoGen, Crew.ai, MetaGPT and many more. AI Agents have already replaced traditional pipelines like customer service (e.g. IBM - https://www.ibm.com/ai-customer-service), and are set to be more reliable and performant as technology advances:

- **Increase in Base Capabilities of LLM**: When the base capabilities of LLMs improve, such as latest models in 2024 like Meta's Llama 3.1, Claude 3.5 Sonnet, or OpenAI o1, the Agent's processing capabilities will increase, which can lead to better and more reliable agentic processes.

- **Better Agentic Frameworks**: With better structure guiding information flow, and better information representation in memory, better tools and planning capabilities, the AI Agent will be better poised to deal with the complex information of the modern world.

2. Difference between AI Agents, Chatbots, and Co-pilots

While AI Agents, chatbots, and co-pilots all function using AI, they differ in terms of complexity, autonomy, and purpose.

- **AI Agents**: These are autonomous systems capable of performing a wide range of tasks, making independent decisions, and interacting with complex systems. AI agents are designed to operate across multiple tasks and domains, orchestrating activities without ongoing human guidance. Example: Simbian SOC Agent, GRC Agent, etc.

- **Chatbots**: These are primarily designed for limited, often predefined, conversational interactions. They run within fixed parameters and handle specific user queries or requests, such as customer support or basic Q&A sessions. Traditional chatbots lack the ability to autonomously manage intricate workflows like AI agents. Example: Telegram Bots

- **Co-pilots:** Co-pilots are somewhere between AI Agents and Chatbots. They are collaborative tools that aid humans by augmenting their abilities. They do not operate autonomously like AI Agents, and can do more operations than a chatbot, and enhance human decision-making by offering suggestions, managing workflows, and automating simpler tasks. They collaborate with users to increase efficiency but are not designed to act independently over prolonged periods. Example: GitHub Copilot

A key difference between all three is in **autonomy and complexity** – AI Agents are highly autonomous problem-solvers, whereas chatbots and co-pilots provide limited assistance or responses to specific tasks. This is illustrated in **Fig. 4.**

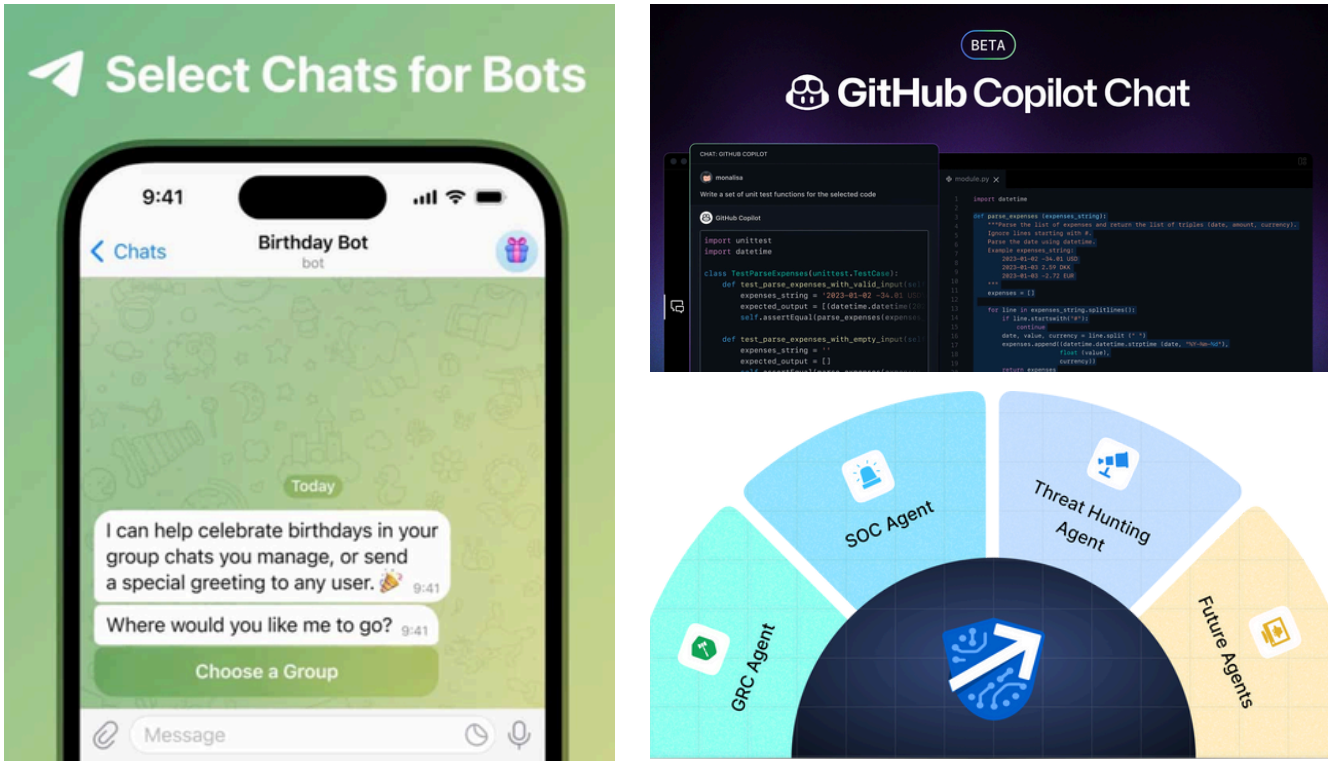

Figure 4 Examples of Chatbot (left: Telegram Bot), Copilot (right top: GitHub Copilot), AI Agents (right bottom: Simbian AI Agents)

That said, the difference between the three is no longer as big as before. Modern day chatbots may have AI Agents doing tasks at the backend in accordance with the chat information, which would make the chatbot more like an interface to the AI Agent. Such a chatbot might even be the interface for a co-pilot framework. Future automation systems may very well have AI Agents, Chatbots and Co-pilots at various stages of the pipeline.

3. AI Agents in Cybersecurity

AI Agents can take on one or more processes in cybersecurity, such as analyzing data, predicting cyber threats, or performing endpoint detection and response. These Agents are more versatile than traditional fixed processing systems and can learn from environmental experience, and correct and improve their processes accordingly.

The most impactful way to use AI Agents in cybersecurity is by taking those time-consuming tasks that are currently performed mostly by humans and augmenting them with Agents. Some examples include:

🔄 Automating Routine Security Tasks

A Generative AI Agent can take over routine security tasks such as patch management, vulnerability prioritization, and compliance checks. By managing these repetitive tasks, the Agent frees up valuable time for cybersecurity professionals to focus on more strategic initiatives.

🕵️ Conducting Advanced Threat Hunting

Advanced threat hunting requires continuous monitoring and analysis of network traffic and system logs. A Generative AI Agent can autonomously conduct threat hunting activities, finding anomalies and potential threats in real-time. This proactive approach enhances the ability to detect and respond to sophisticated attacks.

📊 Offering Real-time Security Insights and Recommendations

Generative AI Agents can offer real-time security insights and recommendations based on continuous data analysis. By using machine learning algorithms, the Agent can find patterns and trends, offering actionable insights to improve an organization's security posture.

4. How to adopt AI Agents in cybersecurity

To successfully integrate AI Agents into a cybersecurity ecosystem, organizations must follow these key steps:

1. Define Clear Objectives: Before implementing AI Agents, organizations must establish clear goals and understand how AI fits into their existing security strategy. AI Agents should not be used for everything, but only for the areas whereby reliability can be guaranteed, and the action space is benign enough for the AI Agent to make autonomous decisions.

2. Have Robust and Reliable Tools: As the saying goes, “a craftsman is only as good as his tools". Before moving to a full AI Agent pipeline, effort needs to be made to ensure that the tools to process information, such as querying databases or searching the web, are robust and dependable enough to aid the Agent perform its tasks.

3. Start with Low-risk Areas: Initially deploy AI Agents in less critical areas such as log analysis, data aggregation, or phishing detection. As technology becomes more reliable, AI Agents can be deployed in more parts of the cybersecurity pipeline.

4. Ensure Human-AI Collaboration: AI Agents should work in tandem with human analysts, and not replace them, to retain accountability in cybersecurity processes. AI Agents can be used to automate routine tasks, allowing security experts to focus on complex, high-priority issues.

5. Continuous Learning and Model Updates: AI Agents require continuous training and updates to be effective. As cyber threats evolve, the data that powers the AI Agent needs to evolve too via memory updating or fine-tuning the LLM or LMM.

6. Security of AI Models: There needs to be robust measures to protect AI models from tampering, adversarial attacks, and other vulnerabilities that could undermine their functionality. At Simbian, we use TrustedLLM™ to additionally improve the robustness of the pipeline so that the output is safe for your organization.

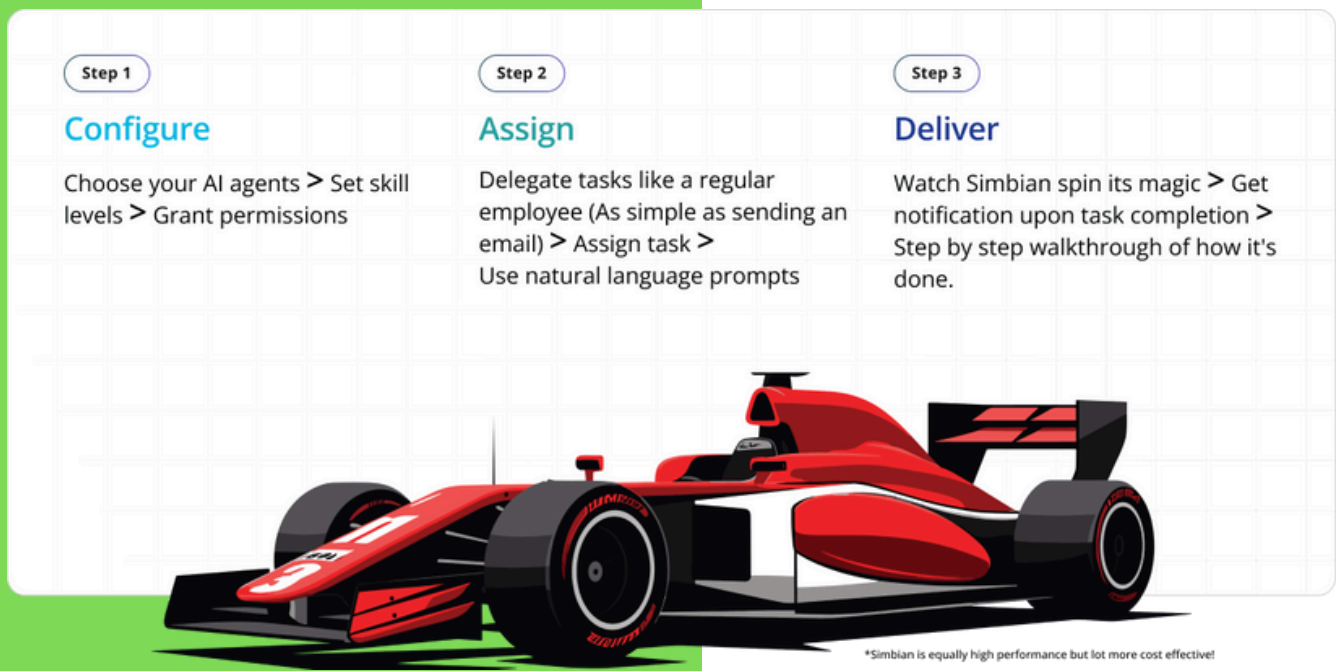

Figure 4 Three easy steps to configure an Agent with Simbian

At Simbian, our AI Agents are amazingly easy to configure and can be done in the following three steps **(see Fig. 5)**:

1. **Configure**: Choose your AI agents > Set skill levels > Grant permissions

2. **Assign**: Delegate tasks like a regular employee (As simple as sending an email) > Assign task > Use natural language prompts

3. **Deliver**: Watch Simbian spin its magic > Get notification upon task completion > Step by step walkthrough of how it's done

5. Expected gains from AI Agents

Adopting AI Agents, particularly in cybersecurity, offers significant benefits:

- **Reduced Human Cognitive Workload:** AI Agents automate repetitive tasks, reducing the workload on human teams and allowing them to focus on strategic decision-making.
- **Extensive Monitoring and Analyzing Capabilities:** AI Agents continuously monitor and analyze security data, detecting threats faster and with greater accuracy than manual processes.
- **Reduced Response Time:** AI Agents can autonomously trigger incident response mechanisms in real time, minimizing the damage caused by cyberattacks.
- **Cost Savings:** By automating routine tasks and improving overall security posture, AI Agents reduce the costs associated with manual operations and security incidents.
- **Scalability:** AI Agents can easily scale across large, complex environments, making them ideal for organizations of all sizes.
- **Adaptive Learning:** An AI Agent can learn from its past interactions via memory updating and can lead to dynamic updating of action plans to improve task success rates.

6. Work that is suitable and unsuitable for AI Agents

Suitable Work:

- **Repetitive Tasks:** AI Agents excel at performing repetitive and rule-based tasks with high efficiency. These include processing data, automating workflows, and performing routine cybersecurity monitoring.

- **Cognitive Layer Wrapping Tool Use:** An AI Agent with relevant tools is ideal for tasks involving use of the tool such as anomaly detection, behavioral analysis and many more. Such an Agent can do more than what the tool can do, as it can also decide how to use the tool based on experience and give the information of the tool in a humanly understandable manner.

- **24/7 Monitoring:** AI Agents are perfect for tasks requiring constant surveillance, such as analyzing security logs and detecting security threats continuously.

Unsuitable Work:

- **Logical Reasoning:** Despite recent hypes about LLMs being able to reason, it is still fundamentally a next token predictor and is subject to hallucinations. As such, logical reasoning tasks may not be suitable for AI Agents as the reasoning generated may not be correct.

- **Tasks Requiring Human Judgment:** Tasks with ethical considerations involved are better suited for an actual human. For example, HR decisions and whether to stop sensitive workflow processes.

- **Different Environment from Training Data:** Like most machine learning approaches, AI Agents are only good for environments that are like historical data and predefined patterns. When using AI Agents in an environment that is different from training data, they may not work well if they are not able to generalize using the right abstraction spaces.

7. Pitfalls to watch while using AI Agents in security

Although AI Agents offer powerful capabilities in security, there are several pitfalls to be aware of:

- **Over-Reliance on AI:** Completely relying on AI Agents without human oversight can lead to missed threats or misinterpretations of data, especially in scenarios where AI fails to detect sophisticated attacks.
- **Bias in Training Data:** AI models trained on biased or incomplete data can produce flawed outcomes. For example, if an AI Agent is trained on a dataset that excludes certain types of attacks, it may fail to detect new or unconventional attack methods.
- **False Positives and Negatives:** AI Agents might generate false positives (flagging benign activity as malicious) or false negatives (failing to detect real threats), both of which can undermine trust in the system and lead to inefficiencies.
- **Security Vulnerabilities in AI Models:** Like any software, adversaries can target AI Agents. Attackers can exploit vulnerabilities in the AI model itself through techniques like adversarial attacks or model poisoning, undermining the system's effectiveness.

8. Conclusion

AI Agents powered by LLMs or LMMs represent transformative technology for a wide range of industries, including cybersecurity. Their ability to autonomously learn, plan and act makes them invaluable for tasks requiring repeated interaction with an environment. However, organizations must adopt these AI Agents with caution, ensuring that human oversight and robust security practices accompany their deployment. That said, the technology is too good to be missed out on – integration of AI Agents promises substantial gains in efficiency, cost savings and reduced response time.

10x greater coverage - same team

- 10x your team's productivity
- Automate 90% of the common tasks
- Turn security analysts into programmers
- Zero coding needed
- Leverage their deep, historical expertise

Scale your team with AI Agents

- Slash costs
- Preserve best practices
- 24x7 availability
- Eliminate knowledge loss
- Augment your team with AI expertise

Meets you where you are

- No disruption to existing workflows
- No new tools to learn
- Integrates effortlessly into your browser, terminal, and scripts
- Communicate in natural language