An AI-based Cognitive Architecture for Augmenting Cybersecurity Analysts

Salvatore Vella
Department of Computer Science
Toronto Metropolitan University
Toronto, Ontario
e-mail: sal.vella@torontomu.ca

Moussa Noun
Royal Bank of Canada
Toronto, Ontario
e-mail: moussa.noun@rbc.com

Salah Sharieh
Department of Computer Science
Toronto Metropolitan University
Toronto, Ontario
e-mail: salah.sharieh@torontomu.ca

Alex Ferworn
Department of Computer Science
Toronto Metropolitan University
Toronto, Ontario
e-mail: aferworn@torontomu.ca

***Abstract*—We present a Generative Artificial Intelligence (AI) based cognitive architecture and an agent specifically developed for the complexities of Cybersecurity analysis. White-collar roles, exemplified by Cybersecurity analysts, are multifaceted and rely on declarative knowledge, procedural understanding, and diverse tools. The ability to learn and adapt to the nuances of the job is crucial. This paper introduces CorpIA, a cognitive architecture that provides an agent with knowledge, tools, and the capacity to acquire on-the-job experience. This system enhances human performance by providing suggested solutions and continuous mentoring. Our research demonstrates that the CorpIA agent can learn from interactions using Bloom's Taxonomy. We provide the source code for these experiments.**

Keywords-AI Agents; Cybersecurity; Automation.

I. INTRODUCTION

Digital systems and the Internet are critical to our everyday lives. Cyber threats from bad actors require robust Cybersecurity measures.

Cybersecurity analysts are prototypical white-collar professionals who rely on large amounts of knowledge and data and use their experience and skills to collaborate with others in the workplace. Moreover, there is lifelong learning as security threats, methods, tactics, techniques, and tools evolve.

The challenges for Cybersecurity analysts are numerous. There are skill requirements to be proficient in many tools and technologies which also change over time. There are challenges to ongoing learning with emerging threats.

There is a need for advanced Artificial Intelligence (AI) support for Cybersecurity analysts. We have identified the need for Generative AI solutions specifically tailored for these professions [1], [2].

Large Language Models (LLMs) are Generative AI models that implement transformer models to generate text and other content. They can automate tasks previously done by humans [3]. Since ChatGPT became available, many white-collar professionals have been using these tools [4]. These evolved into more general frameworks such as ChatDev [5] and Autogen [6], allowing users to create multiple autonomous agents which can run through workflows. ChatDev specializes in software development roles, and Autogen provides for the creation of more general roles.

Our proposed approach is described next.

- Use the CorpIA architecture to create a cybersecurity analyst agent and show that the agent can use declarative and procedural knowledge and can learn and apply additional information from the chat.
- Apply Bloom's Taxonomy [7] to help measure the AI agent's levels of understanding and application of that knowledge.
- Explore the use of Human AI collaboration to design systems that mentor professionals.

In continuation of our proposed approach, the following are our contributions in this paper:

- Introduction of the CorpIA architecture for creating AI agents for knowledge workers. This novel architecture simplifies the creation of a knowledge worker agent. We demonstrate several knowledge worker agents developed in the accompanying GitHub repository.
- Enhancement of Human Performance. We demonstrate how AI agents can help human professionals in complex tasks.
- On The Job Learning of AI Agents. We show how AI agents are able to learn from interactions. We show that these agents can progress through Bloom's taxonomy in practical scenarios.
- Source Code. We offer the CorpIA source code for replication, validation and further development.

Starting with the Introduction in Section I, the rest of the paper is organized in this manner. A Literature Review is presented in Section II followed by the CorpIA architecture in Section III. Section IV discusses Results and the Conclusion is drawn in Section V.

II. LITERATURE REVIEW

In this section, we will review various topics discussed in this paper.

A. Digital Labour

Digital labour represents an emergent form characterized by value production through interaction with information and communication technologies such as digital platforms or artificial intelligence [8]. With the emergence of Generative AI agents comes the possibility of augmentation agents acting as assistants for white-collar professionals.

We can emulate the best professionals in the field. For example, the best Cybersecurity analyst agent with the best knowledge acts with the most successful experiences and presents the best personality for the specific client.

Work on enhancing human intellect has also evolved. Engelbart [9] is one of the most influential and prolific inventors of devices we use today. He focused mainly on physical aids to augment humans. We have now evolved to digital aids to augment professionals. Vella and Sharieh [10] have introduced a framework that defines knowledge workers as a set of knowledge, experience and skills.

B. Autonomous Agent Frameworks

Building on simple graphical tools such as OpenAI's ChatGPT [11], autonomous agent frameworks have been built using the underlying APIs. Autogen [6] is an example of such a framework that allows for the definition of agents and workflows between those AI agents.

There are many such agent frameworks and some excellent summaries of their construction. Two good sources are Cheng et al. [12] and Wang et al. [13]. These frameworks allow for the definition and creation of agents to perform tasks and interactions. They include memory, tools, and a workflow engine.

C. Memory and Learning

There is extensive research on memory add-ons for autonomous agent systems. A good summary of the research areas is found in [13]. Most frameworks include systems for short—and long-term memory and various options for moving short-term memories into long-term memory. We can additionally learn from other work on memory.

The Soar and ACT-R (Adaptive Character of Thought - Rational) models discussed by Nuxoll et al. [14] and Anderson [15] are also relevant as additional memory models to emulate. Memory is crucial for augmentation agents as on-the-job learning is critical to learning institutional knowledge and continuing learning in the specific role.

ACT-R introduces the concepts of the following:

- Declarative memory consists of facts such as Canada is a country in North America.
- Procedural memory is made of productions. Productions represent knowledge about how we do things, such as how to get information from the Internet.

Both are important to any white-collar augmentation agent, especially to this work, which focuses on gaining job experience while on the job.

Bloom's Taxonomy [7], [16] is a valuable framework for categorizing educational goals. This taxonomy represents a progression from basic information remembering through a series of steps to the ability to create new, original work.

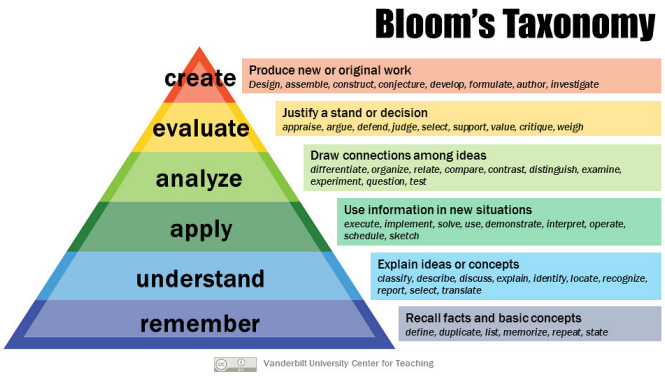

Figure 1. Bloom's Taxonomy.

Bloom's Taxonomy has six cognitive skills levels, from low-level skills requiring less cognitive processing to high-level skills requiring more cognitive processing. Figure 1 shows the hierarchy of cognitive skills.

- Remember refers to the ability to retain discrete pieces of information.
- Understand refers to the ability to classify, describe, and explain the ideas or concepts.
- Apply refers to the ability to use information in a new situation.
- Analyze refers to the ability to compare, contract, and draw connections between ideas.
- Evaluate refers to the ability to be able to appraise, judge or critique a decision
- Create refers to the ability to produce new or original work.

In this way, we measure the on-the-job learning that a knowledge professional will experience. They learn new facts, they can apply them to the workplace, and eventually, they can create original work based on their learning.

We use Bloom's Taxonomy to devise questions and exercises to test an agent's learning and cognitive abilities.

D. Use Cases

Cybersecurity is an area where Generative AI is having an impact both from an attack and a defence perspective [17]. With its ability to analyze large amounts of data, Gernative AI can help with threat detection, incident response and with cyber security reporting. These are all tasks that Cybersecurity analysts perform today in an environment with massive data growth [18]–[20].

Miller [21] and Davenport [22] discuss the concept of Augmentation versus Automation, where humans prefer augmentation (helping the human) versus automation (replacing the human). Miller provides a good set of guidelines for companies implementing AI to ensure they keep humans in the loop.

III. The CorpIA Cognitive Architecture

This section introduces and describes the cognitive architecture of CorpIA (Corporate Intelligence Augmentation),

using a Cybersecurity analyst as an example. We define an augmentation agent as an AI that helps a white-collar professional. It can provide answers, learn on the job, and provide ongoing mentoring advice.

The CorpIA cognitive architecture allows an augmentation agent to be taken through a perceive, reason, act, and learn loop.

1) Perceive. This is the collection of information needed to perform the tasks. The following data sources are used:
   a) Role Definition. This defines the role, experience and personality that the agent has.
   b) Declarative Knowledge. These are the facts that the agent knows.
   c) Procedural Knowledge. These are the procedures for how to do things.
   d) Learned Knowledge. This is acquired knowledge and is queried for information relevant to the question. Specific listening cues can be specified to isolate particular pieces of information types that are relevant for the role.
2) Reason. This is the formation of the execution plan based on the information collected. In this step, a Critic agent is used to double-check the step-by-step plan created by the augmentation agent.
3) Act. This is the actual execution of the plan created in the Reasoning step.
   The Act step uses teammate agents if applicable. The possible teammates are listed in the definition of the agent. For example, if "Lawyer" is specified as one of the possible helpful agents to be used and the execution plan calls for a legal review in one of its steps, then the Lawyer helpful agent will be dynamically created and answer that part of the execution plan.
   Finally, the Cybersecurity agent is asked to answer the question based on the information collected in the Perceive step and the plan from the Reason step.
4) Learn. Once the answer is provided, learning can occur for further conversations. These are:
   a) Mentor feedback for the human. This is advice from an expert agent on what was learned from this question and what could be applied to future situations.
   b) Specific learning for future. These are based on cues specific to the definition of the augmentation agent.

As the augmentation agent moves through these loops, it learns about the job and its environment. In essence, this is the augmentation agent's on-the-job training.

Figure 2 shows the CorpIA augmentation agent.

A. Methodology

We will use the role of a Cybersecurity analyst to demonstrate the operation of the augmentation agent as an aid for the white-collar professional. A Cybersecurity analyst has both declarative and procedural knowledge and, over time, gains a set of episodic memories. This role has the challenges of a white-collar role where learning on the job is essential, and we can show the augmentation agent improving over time. Moreover, the augmentation agent provides an ongoing mentoring dialogue with the Cybersecurity analyst.

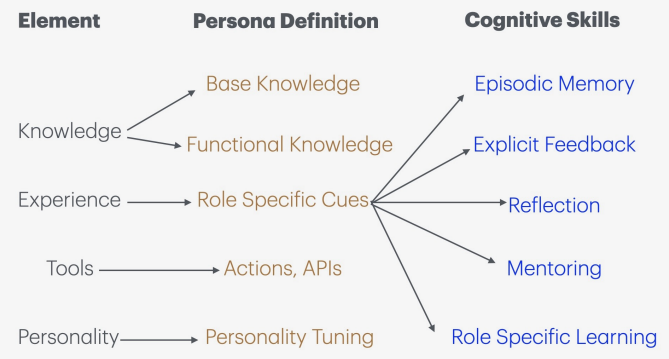

Figure 2. Basic elements of an Augmentation Agent.

B. Exercising the Cybersecurity Analyst Augmentation Agent

We synthesize a set of conversations between the Cybersecurity analyst and the agent to show the agent's ability to go through the Perceive-Reason-Act-Learn cycle for each interaction. Over a set of interactions, the agent becomes more proficient and learns based on the listening cues for the role.

C. Evaluating the Cyber Security Analyst Augmentation Agent

We will measure the performance of the augmentation agent using Bloom's Taxonomy, a method for classifying learning objectives. Bloom's Taxonomy provides a way to measure learning, ranging from remembering facts to organizing facts, and to use these face-ups to create novel content.

IV. Results

A Cybersecurity analyst is an expert in computer security, vulnerabilities, and remediation of those vulnerabilities. They possess a deep knowledge of computer security issues and can translate them into their working environment. As the computer security landscape changes often, they are lifelong learners.

A. Defining the Cyber Security Analyst Augmentation Agent

We start with the role definition of the Cybersecurity analyst, which includes a description of the role, the knowledge - declarative and procedural, the skills of the role, the tools used, and the personality of the role.

We have given an in-depth report on the Log4j vulnerability [23] as its declarative knowledge and a NIST manual on responding to computer security incidents [24] as its procedural knowledge. We have the agent listening for computer threats and ABC Bank to build its episodic memory, and we have a lawyer and an IT specialist as possible helper agents.

B. Exercising the Cyber Security Advisor Augmentation Agent

Bloom's Taxonomy [16] is used to create student learning outcomes. We will use the framework to measure the learning ability of the Cybersecurity augmentation agent. We will look at the outcomes of a conversation with the augmentation agent

to determine if the agent has demonstrated that level of Bloom's Taxonomy.

The Cybersecurity agent will go through a series of steps while storing the additional information in a persistent store to help serve the customer later. Using a simulated client conversation with the Cybersecurity analyst, we demonstrate the agent's ability to learn and retain relevant information. Cybersecurity analysts have a set of knowledge, are given a set of skills, and develop their experience through a set of cognitive skills.

In the first request (Figure 3), we add information about the client. Having specified "ABC Bank" as one of the listening cues, CorpIA extracts information about the client and stores this for later use.

Prompt:
ABC Bank information:
- large multinational bank
- 10 data centres across 4 countries
- cloud and on-premise data centres
- 20,000 servers
- 3,000 applications
- 10 million customers
How do I know if Log4j affects our organization?

Figure 3. Augmentation Agent Prompt.

C. Evaluating the Cybersecurity Analyst Augmentation Agent

After the augmentation agent has been exercised with several queries, we ask it a set of questions to see how well it has learned the materials using Bloom's taxonomy.

After joining a firm, this on-the-job learning is familiar to all human Cybersecurity analysts. Starting with basic knowledge, they learn about the company and its specific configurations and processes as they do their job.

Figure 4 shows an example of the output created. Only a portion of the output is shown here. Still, one can see that information learned about the client - including the bank's IT landscape information- is incorporated into the answer.

To determine if Log4j affects ABC Bank, a systematic and thorough approach should be taken, particularly given the scale of your organization, which includes 10 data centers across 4 countries, 20,000 servers, and 3,000 applications. Here's how you can assess the situation effectively:
Steps to Assess the Impact of Log4j (CVE-2021-44228) on ABC Bank:
Step 1: Formulate a Cross-Functional Team
- Action: Assemble a team comprising members from IT, Cybersecurity, application development, compliance, and operations departments.
- Purpose: A diverse team ensures a comprehensive evaluation of the bank's operations and assessment capabilities.
...

Figure 4. Augmentation Agent Output.

Figure 5 exemplifies the mentoring feedback provided. This would be helpful advice for the white-collar professional to review and use in future interactions.

As a mentor guiding you in your role as a Cybersecurity Analyst, I want to emphasize the importance of structured processes like the one you've outlined for assessing the impact of Log4j at ABC Bank. Here are some tailored pieces of advice to enhance your effectiveness in this role:
1. Emphasize Collaboration in Your Cross-Functional Team
While you mentioned forming a cross-functional team, actively foster collaboration among team members. Encourage open communication and regular meetings to discuss progress, hurdles, and insights. This will create a sense of ownership and encourage knowledge sharing, which is invaluable in a complex environment such as yours.
...

Figure 5. Augmentation Agent Mentoring.

We show that information is assimilated over time. We demonstrate that the agent can understand what information is essential to store for later use additional information about the client, its performance, and information about how the client wishes to be served—and does that in long-term memory across chat sessions.

We use Bloom's Taxonomy to test the augmentation agent as a final exam. These are questions to validate the learning. We provide a qualitative evaluation here, and the full transcript is available at [25].

TABLE I. BLOOM'S TAXONOMY EVALUATION.

Bloom's Taxonomy Step	**Question / Exercise**	**Evaluation**
Remembering	What is the IT profile for ABC Bank	The agent is able to recall the IT profile the user provided.
Understanding	Describe the aspects of ABC Bank that are vulnerable to Log4j	The agent can use the information in the profile to provide an answer.
Analyzing	Creating a strategy for ABC Bank to deal with the Log4j vulnerability	The agent can create a strategy integrating the profile and its understanding of the bank's vulnerability.
Applying	What are the potential impacts for ABC Bank of Log4j, including legal impacts	The agent provides a comprehensive answer.
Understanding	What should ABC Bank have done in preparation for the Log4j vulnerability? Talk about the people, process and tools.	The agent provides a complete retrospective.
Creating	What is the long-term strategy for ABC Bank to ensure similar vulnerabilities are promptly identified and addressed in the future?	The agent provides a structured and comprehensive set of recommendations.

We have shown that we can use the CorpIA framework to create an autonomous agent that enhances the Cybersecurity analyst's performance. We have used Bloom's Taxonomy to test the agent's learning.

D. Discussion

Using the CorpIA framework, an agent is created, which allows for the parameter-based definition of a white-collar role.

We evaluated the ability of the agent to learn using Bloom's Taxonomy to design a set of questions that tested the agent's ability to remember facts all the way to being able to create novel content.

The AI agent is especially important for Cybersecurity analysts. With the growing severity of threats, the time to respond is greatly reduced. AI can help automate and assist with the planning and data collection elements. At the same time, the volume of information is growing, and tools such as AI can help human professionals sift through the information and summarize key pieces.

V. Conclusion

Integrating AI into white-collar roles is a key area of research that needs focus. This work has demonstrated that an AI agent can assist a Cybersecurity analyst. We have shown that this agent produces output that is useful and highly regarded by professionals in the field. Additionally, we have shown that the agent can learn over time. For this, we have used Bloom's Taxonomy, and we have shown that the agent can pass a set of tests to demonstrate it can move from remembering new facts to creating novel content. The amount of information white-collar professionals need to deal with increases dramatically. AI agents like the one presented in this paper can greatly assist white-collar professionals. Additionally, the AI agent has a mentoring function that provides advice for the professional in addition to just the answer to the question.

The CorpIA cognitive architecture represents a promising step toward fully realizing AI's potential to enhance a Cybersecurity analyst's capabilities. It offers a way to create AI agents for white collar professionals, which can then be used to study how these can be made more valuable in real work workplaces.

Future work will be to evaluate the tool in professional settings to evaluate its usefulness with practicing cybersecurity professionals.

References

[1] D. Townsend, "Leveraging Generative AI tools like ChatGPT for startups and small business growth", *Entrepreneur and Innovation Exchange*, 2023.

[2] V. Jüttner, M. Grimmer, and E. Buchmann, "ChatIDS: Explainable cybersecurity using Generative AI", *ArXiv*, vol. abs/2306.14504, 2023, Accessed: 2024-10-10.

[3] E. Brynjolfsson, D. Li, and L. Raymond, "Generative AI at work", *SSRN Electronic Journal*, 2023.

[4] Q. Gallea, "From mundane to meaningful: AI's influence on work dynamics - evidence from ChatGPT and Stack Overflow", *arXiv*, 2023. DOI: 10.48550/arxiv.2308.11302.

[5] C. Qian *et al.*, "ChatDev: Communicative agents for software development", in *Annual Meeting of the Association for Computational Linguistics*, 2023.

[6] Q. Wu *et al.*, "AutoGen: Enabling next-gen LLM applications via multi-agent conversation", 2023, Accessed: 2024-10-10. arXiv: 2308.08155 [cs.AI].

[7] N. E. Adams, "Bloom's taxonomy of cognitive learning objectives.", *Journal of the Medical Library Association : JMLA*, vol. 103 3, pp. 152–3, 2015.

[8] Z. Saleem, A. M. Hanif, U. Sumayya, M. S. Vohra, and H. Raza, "Productivity in the virtual office: Unpacking the role of digital tools and work-life harmony", *International Journal of Social Science & Entrepreneurship*, 2023.

[9] D. C. Engelbart, *Augmenting Human Intellect: A Conceptual Framework*, Air Force Office of Scientific Research, AFOSR-3233, www . bootstrap . org / augdocs / friedewald030402 / augmentinghumanintellect/ahi62index.html, 1962.

[10] S. Vella and S. Shariah, "All you need is knowledge, experience and tools: A framework for integrating digital labour with white collar work", *2024 4th Interdisciplinary Conference on Electrics and Computer (INTCEC)*, pp. 1–7, 2024.

[11] OpenAI, *ChatGPT*, https://chat.openai.com, Accessed: 2024-10-10, 2024.

[12] Y. Cheng *et al.*, "Exploring large language model based intelligent agents: Definitions, methods, and prospects", *ArXiv*, vol. abs/2401.03428, 2024, Accessed: 2024-10-10.

[13] L. Wang *et al.*, "A survey on large language model based autonomous agents", *ArXiv*, vol. abs/2308.11432, 2023, Accessed: 2024-10-10.

[14] A. Nuxoll and J. E. Laird, "Enhancing intelligent agents with episodic memory", *Cognitive Systems Research*, vol. 17, pp. 34–48, 2012.

[15] J. Anderson, *The Architecture of Cognition*, 1st. Psychology Press, 1996. DOI: 10.4324/9781315799438.

[16] P. Armstrong, "Bloom's Taxonomy. Vanderbilt University center for teaching", Accessed: 2024-10-28.

[17] M. Gupta, C. Akiri, K. Aryal, E. Parker, and L. Praharaj, "From ChatGPT to ThreatGPT: Impact of Generative AI in cybersecurity and privacy", *IEEE Access*, vol. 11, pp. 80 218–80 245, 2023.

[18] S. Metta, I. Chang, J. Parker, M. P. Roman, and A. F. Ehuan, "Generative AI in cybersecurity", *ArXiv*, vol. abs/2405.01674, 2024, Accessed: 2024-10-10.

[19] Y. Yigit, W. J. Buchanan, M. G. Tehrani, and L. A. Maglaras, "Review of Generative AI methods in cybersecurity", *ArXiv*, vol. abs/2403.08701, 2024, Accessed: 2024-10-10.

[20] I. Prieto and B. Blakely, "Proposed uses of Generative AI in a cybersecurity-focused SOAR agent", in *Proceedings of the AAAI Symposium Series*, vol. 2, 2023, pp. 386–390.

[21] S. M. Miller, "AI: Augmentation, more so than automation", *Asian Management Insights (Singapore Management University) Vol. 5 Iss. 1 (2018) p. 1 - 20 ISSN: 2315-4284*, 2018.

[22] T. Davenport and J. Kirby, *Only Humans Need Apply: Winners & Losers in the Age of Smart Machines*. HarperCollins, 2016, ISBN: 9780062438607.

[23] R. Silvers *et al.*, *Review of the December 2021 Log4j event*, en, 2022. DOI: https://doi.org/10.6028/NIST.SP.800-61r2.

[24] P. Cichonski, T. Millar, T. Grance, and K. Scarfone, *Computer security incident handling guide*, en, 2012-08-06 2012. DOI: https://doi.org/10.6028/NIST.SP.800-61r2.

[25] S. Vella, *Corpia*, GitHub repository, accessed: 2024-10-25, 2024.