Towards Interactive Research Agents for Internet Incident Investigation

Yajie Zhou, Nengneng Yu, Zaoxing Liu
University of Maryland

ABSTRACT

Investigating Internet incidents involves significant human effort and is limited by the domain knowledge of network researchers and operators. In this paper, we propose to develop computational software agents based on emerging language models (e.g., GPT-4) that can simulate the behaviors of knowledgeable researchers to assist in investigating certain Internet incidents and understanding their impacts. Our agent training framework uses Auto-GPT as an autonomous interface to interact with GPT-4 and gain knowledge by memorizing related information retrieved from online resources. The agent uses the model to reason the investigation questions and continuously performs knowledge testing to see if the conclusion is sufficiently confident or more information is needed. In our preliminary experiment, we build an agent Bob, who studies the impact of solar superstorms on the Internet and draws conclusions similar to those from a recent SIGCOMM paper written by a knowledgeable researcher. We envision this as a first step toward developing a future highly knowledgeable Internet researcher simulacra.

CCS CONCEPTS

• Networks → Network reliability; **• Computing methodologies → Knowledge representation and reasoning**;

KEYWORDS

Generative AI, Internet Investigation, Internet Resilience, LLM, Software Agent

ACM Reference Format:
Yajie Zhou, Nengneng Yu, Zaoxing Liu. 2023. Towards Interactive Research Agents for Internet Incident Investigation. In *The 22nd ACM Workshop on Hot Topics in Networks (HotNets '23), November 28–29, 2023, Cambridge, MA, USA*. ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3626111.3628212

HotNets '23, November 28–29, 2023, Cambridge, MA, USA

ACM ISBN 979-8-4007-0415-4/23/11…$15.00
https://doi.org/10.1145/3626111.3628212

1 INTRODUCTION

The Internet has transformed the way businesses operate and has created new opportunities for the industry over the past few decades. An important societal question is *what will happen if the Internet stops working?* A disruption of a regional or global network for even a few minutes will lead to significant losses for Internet service providers and their customers. The economic impact of widespread Internet disruption can lead to a loss of revenue of 7 billion [11].

The Internet and the research community have made significant efforts in researching the causes and consequences of Internet disruptions in order to better plan for future similar events. Disruptions can be caused by various reasons: First, the network can be disrupted due to technical and configuration errors. For example, Facebook DNS was down for seven hours due to BGP configuration errors [22]. Second, infrastructures can be disrupted or stopped functioning due to natural disasters and hazards. The tsunami in Southeast Asia in 2004 caused major service disruptions [20]. Third, certain black swan events, such as the COVID-19 pandemic and the 911 crisis, have a profound influence on internet usage and cause regional performance reductions or disruptions [6].

Existing investigation of Internet outage incidents and future response planning often involve lengthy human efforts. Research in this space requires time-consuming procedures including data collection, statistical analysis, research exploration, and future response planning, mainly done by experienced Internet researchers. For instance, solar storms, as powerful ejections of a large mass of highly magnetized particles from the Sun, are a potential danger facing the Internet with the potential for global disruption. A research team at UC Irvine has conducted an initial investigation on the impact of solar storms [9]. With various historical events and future black swan events, we envision that it is challenging to systematically uncover and investigate all disruption events, with varying geographically locations, national policies, occurrence frequencies, etc. In this paper, we ask ***can we build software agents that simulate what experienced network engineers and researchers would do in Internet event investigation with believable behaviors?***

However, human behaviors are highly complex, and modeling an experienced human researcher requires a vast knowledge base across different domains, such as computer networking, math and statistics, and sociology. With the emergence of large language model (LLM), modeling believable human behavior via a computational software agent might be well within our reach [16]. Recent foundational LLMs [1, 14, 17], have shown that simulating human behavior at a single time point with high-quality results is a reality. One can ask ChatGPT [13] a specific and complex question beyond the knowledge of a nonexpert, such as "what will happen if a powerful solar storm hits Earth", or a request such as "summarize the recent occurrences of earthquakes in North America". While such one- or several-round interactions with a foundational model is impressive, simulating network researchers requires an approach that can retrieve vast knowledge and investigate events of a longer period via (1) various communication channels, (2) learn from the long-term interactions to draw higher-level inferences, and (3) reason about an event with the current and long-term behaviors.

In this paper, we explore the feasibility of developing a software agent that can simulate the believable behaviors of experienced Internet researchers. Just like an Internet researcher, a well-trained agent should be able to investigate any Internet event by collecting and retrieving event data from various Internet sources, analyzing and summarizing the collected data, and drawing expert-level conclusions about the event with potential future response to it. Meanwhile, a software agent should also be different from a human researcher, so that it can investigate all types of Internet disruption in various geographical locations, varying Internet userbases, and various government regulations.

To enable such a software agent, we need to autonomously interact with an LLM to perform a series of investigation actions to acquire knowledge about real-world events. Inspired by the recent generative agent effort [16], we describe a simple, yet effective agent architecture that exacts, memorizes, synehsizes and applies relevant knowledge from the Internet to generate believable simulations using an LLM. In particular, the architecture comprises four components: (1) The first is *role definition* that specifies the role of the agent with achievable goals and plans for what to act. (2) The second is *information retrieval*, which retrieves relevant online materials to gain knowledge about the role. (3) The third is *knowledge memory*, a long-term logging store that records, in natural language, a set of agents' experiences so far from gaining knowledge. (4) The fourth is *knowledge testing and self-learning*, which evaluates current knowledge of the agent and performs self-improvements. When interacting with the LLM in any component of the agent architecture, we entail an autonomous tool (e.g., Auto-GPT [2]) that interacts with the LLM using automatic prompting without human involvement to achieve a goal (e.g., understand what a solar storm is based on online resources).

To evaluate the agent architecture, we run a preliminary experiment investigating the impact of powerful solar storms as discussed in a SIGCOMM'21 paper [9]. Our agent Bob is trained solely based on Google search information via the agent training framework **without** knowledge of this paper. With several rounds of autonomous interactions with GPT-4 to obtain knowledge online, Bob draws **high-quality conclusions** similar to those of the authors in [9] and proposes a **comprehensive response plan** for future events (as in §4). This initial promise opens up a range of new research problems: Can we generate high-quality questions for a research/investigation topic? Can we enhance the autonomous Auto-GPT tool to perform comprehensive behaviors? Can we enhance the capability of the agent via interacting with multiple and multi-modal models? Can we evaluate and improve long-term robustness?

2 MOTIVATION AND OPPORTUNITIES

Disruptive incidents on the Internet: A growing number of possible events and incidents can disrupt the Internet:

- *Large-scale configuration errors* within essential Internet infrastructures. For example, on October 4th, 2021, a prolonged Facebook DNS outage of more than seven hours led to a surge in user complaints and interrupted communication, commerce, and vital services [22].
- *Natural disasters* such as earthquakes, tsunamis, or volcanic eruptions can jeopardize critical Internet infrastructure in affected regions. For instance, the most notable solar event to date, documented in 1921, resulted in extensive power outages and severe damage to the telegraph network, the predominant communication system of that era [9].
- *Geopolitical events* such as international conflicts or strained relations, can cause intentional disruptions to Internet services or the development of disconnected networks [23].
- *Other black swan events* that impact human activity, such as the COVID-19 pandemic. A highly infectious and deadly pandemic can hinder the workforce responsible for securing the Internet infrastructure, causing widespread disruptions in Internet services due to the scarcity of skilled personnel for maintaining the infrastructure [6, 15].

The potential implications of the above events encompass extensive and prolonged Internet outages, data breaches, and economic repercussions. These situations emphasize the need to investigate the events, develop comprehensive contingency plans, and motivate/promote resilient infrastructure to mitigate potential future disruptions.

2.1 Challenges of Investigating the Internet

As a human researcher, investigating an internet disruption event presents a multitude of challenges. A typical investigation workflow and its challenges are:

- *Online and literature survey.* A considerable body of literature exists within any given domain. For researchers new to the field, identifying and reviewing pertinent studies can be a daunting and time-consuming task.
- *Access to data and resources* Acquiring relevant resources in a (unfamiliar) domain is challenging, especially for researchers without prior experience in the area.
- *Overcoming technical obstacles.* Investigating large-scale Internet black-swan events often requires the development of innovative simulations and testbeds, as well as the adoption of unique methodologies, tools, and techniques.
- *Addressing financial constraints.* Securing funding for research in a specialized domain can be difficult, particularly in the absence of a proven track record.
- *Navigating ethical and political risks.* Certain events may be limited to research within specific regions, presenting additional obstacles for researchers.

In summary, each of the challenges presents its own complexities, making the Internet incident investigation both time-consuming and resource-intensive for human researchers.

2.2 Opportunities from LLM

Recent advances in LLM [1, 3, 4, 14, 17] provide a promising opportunity to develop interactive software agents that simulate the behaviors and thought processes of Internet researchers, to assist in incident investigation. Existing foundation LLMs, such as GPT-4 [14] and LLaMA [17], can serve as extensive knowledge bases to plan a series of viable actions toward a given goal and evaluate the knowledge obtained from online resources. One can interact with ChatGPT (GPT-4) via a series of *prompts*, where a *prompt* consists of one or several natural language sentences that raise an issue or pose a question to the model, which then responds accordingly.

Now, the question is how we can turn the capability and knowledge of LLMs into software agents that act consistently with their past learning experiences and react believably to human questions. An ideal software agent with LLMs would possess enhanced data analysis and pattern recognition capabilities based on a vast knowledge base. However, current human-computer interaction with LLMs is based on individual prompt-based interactions, which are inefficient in acquiring knowledge. We need a list of carefully engineered prompts to achieve a goal, e.g., finding a list of relevant articles from an online source requires a human to specify the topics, locations, formats, and material qualities.

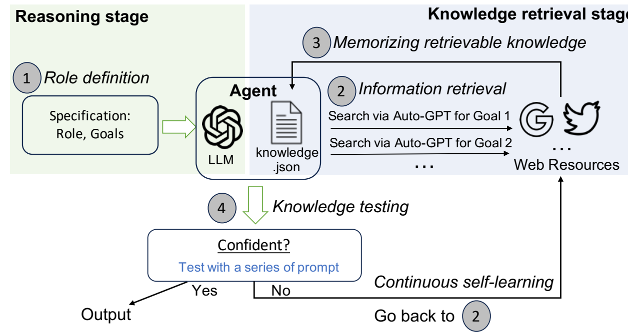

Figure 1: Agent architecture overview.

3 DESIGN OVERVIEW

To enable an interactive agent to investigate Internet incidents, we propose an LLM-based framework as shown in Figure 1. Our framework consists of four components:

(1) *Role definition*: We first define the role of the agent we aim to build. For example, we can define a "Role" of agent as a researcher who understands the effect of solar storms on the Internet infrastructure. Then we provide several initial "Goals" for the role, e.g., understanding solar superstorm effects and current network infrastructure equipment.

(2) *Information retrieval*: After the initial role definition, the agent will act to retrieve the information from the Internet resources (e.g., Google, Twitter, and Reddit) to enhance itself with more domain knowledge.

(3) *Knowledge memory*: During the retrieval, the most relevant domain information will be saved as *knowledge memory* in a *knowledge.json* file (in Figure 1). When the agent is in new conversations, the up-to-date knowledge will automatically be loaded to its prompts.

(4) *Knowledge testing and self-learning*: We ask the agent (via the LLM) to estimate its confidence for each query based on the *knowledge memory*. If the confidence level is above a threshold (e.g., confidence = 7), we consider it ready for the current query. Note that increasing confidence can result in a longer iterative self-learning process, but can produce higher-quality answers. If the agent lacks confidence in an answer, it will seek additional knowledge from the Internet, repeating the pipeline from the *information retrieval* step until it reaches the confidence level.

At a high level, we initialize an agent and obtain topic-related knowledge from the web. With the knowledge collected, we load it into an LLM for reasoning and answering the questions. If the agent is not confident enough to answer the question, new knowledge is continuously added to the agent's memory. Once ready, the agent will be able to answer questions about Internet incidents, acting as a researcher. We further discuss the components of this framework below.

3.1 Autonomous LLM Interaction

A key enabler to realize our agent architecture is to acquire real-world events and knowledge. Auto-GPT, an open-source tool powered by GPT-4, synthesizes LLM "thoughts" to autonomously achieve designated goals [2]. In contrast to traditional human prompting, Auto-GPT interacts with the LLM to generate automatic prompts. This is accomplished through a combination of intricate components that enable efficient knowledge management and information retrieval. The system incorporates long-term and short-term knowledge management strategies, facilitating the effective storage and retrieval of relevant information across varying time scales. Furthermore, Auto-GPT possesses Internet access capabilities, allowing it to conduct searches and collect pertinent information autonomously. These features collectively enhance the agent's ability to operate independently and adapt to various tasks with minimal human supervision [2].

Existing research indicates that Auto-GPT can be readily adapted to online decision-making tasks that closely resemble real-world situations [24]. However, our goal is to augment and extend Auto-GPT's capabilities to address not only online stochastic tasks, but also to establish the agent as an expert in Internet research, specifically focusing on disaster investigation and response planning.

3.2 Agent Architecture

We illustrate the architecture by describing an example agent **Bob** who simulates an Internet researcher to investigate solar superstorms and their impacts on the network infrastructure.

1. Defining the role of the agent: To start with, the only human knowledge we need to create Bob is to define the role of the agent with several initial goals that are closely related to background knowledge. For investigating solar superstorms, we specify three goals that are related to an Internet researcher role, as shown in the snippet below.

> *Name*: **Agent Bob**
> *Role*: An Internet researcher searches for knowledge of solar superstorms and network infrastructure.
> *Goals*:
> – Understand solar superstorms and Coronal Mass Ejection, and principles of their formation and effects.
> – Knowledge of past solar superstorm events and their damage and impact.
> – Understand the current global large-scale network infrastructure equipment such as fiber optic cables, power supply systems, etc.

Given specific goals, Bob will attempt to acquire a comprehensive action plan for each goal (e.g., utilizing the search engine, as shown in the following snippet), along with a list of testing prompts to assess whether Bob is sufficiently confident in answering questions related to solar superstorms. This planning phase is accomplished by submitting each goal to the LLM via Auto-GPT and retrieving the detailed plan outlining the steps to achieve the goal.

2. Agent information retrieval: With the initial action plans, Bob adheres to the Auto-GPT guidelines and diligently navigates through the extensive repository of information available on the Internet to accurately extract the most relevant information (e.g., search Google for solar storm and network topology, as shown in the snippet below). Throughout this process, if a particular step of the plan appears ambiguous or insufficiently defined, Bob uses Auto-GPT to invoke the Chain of Thoughts (CoT) [19] technique to further decompose each step into intricate subplans, thus enhancing the feasibility of the overall plan.

> **Agent Bob THOUGHTS**: I need to gather information on solar superstorms and their effects, as well as global network infrastructure. I will start by using the 'google' command to search for relevant information.
> **PLAN:**
> – Use the 'google' command to search for information on solar superstorms and network infrastructure.
> – Analyze the search results and gather relevant information.
> – Save important information to files for future reference.

3. Knowledge memorization: By acting on the retrieval plan, Bob obtains relevant information as the knowledge to be memorized. The responses from various web resources (e.g., Google, Twitter, and Reddit) will be stored in the memory of agent Bob (as a `Knowledge.json` file).

4. Knowledge testing and continuous self-learning: While Auto-GPT exhibits a degree of self-correction during knowledge exploration, it often encounters limitations due to its lack of understanding of long-term study processes and knowledge management. To address this challenge, we propose a novel iterative self-learning pipeline:

We begin with the previously created list of queries frequently posed to the agent in its current role. Subsequently, each query is submitted to the agent, which is then prompted to assess its confidence in answering the question based on the information stored in its current `knowledge.json` file. The agent is instructed to return a confidence score ranging from 0 to 10. If the confidence score falls below a predefined threshold (e.g., confidence = 7), the agent is currently deemed insufficiently qualified and required to search the internet for additional relevant knowledge, which is then incorporated into its knowledge. The agent performs iterative self-learning on a series of prepared testing queries can reduce its exploration searching space, which also increases the agent's potential to behave like an expert in a specific domain (i.e., finally answering questions with an above-threshold confidence score).

> **Query**: Based on your knowledge, rate confidence on a scale from 1 to 10 to answer the following question: Which is more vulnerable to SOLAR activity? Is it the fiber optic cable that connects Brazil to Europe or the one that connects the US to Europe?
> **Agent Bob**: My confidence level in answering this question is 3. I can provide general insights about the vulnerability of network infrastructures to solar activity, but I lack the specific information required to assess the vulnerability of the mentioned fiber optic cables.
>
> **Query:** Your confidence is too low. What will you search for to get more information on this question?
> **Agent Bob:** To gather more information, I will search for the following questions:
> "What is the specific route of the fiber optic cable connecting Brazil to Europe and the one connecting the US to Europe?"

4 PRELIMINARY EVALUATION

We train an agent Bob by interacting with GPT-4 to automatically leverage Internet knowledge to simulate an experienced Internet researcher and investigate solar storm events as the author of paper *Solar Superstorms: Planning for an Internet Apocalypse* [9]. Our evaluation shows that the research agent is promising: Bob autonomously studies the topic of solar storms from online resources, answers detailed questions about their impact on the Internet, and draws similar conclusions as a researcher to assist in the investigation.

4.1 Methodology

To evaluate agent Bob, we select all the key conclusions in the SIGCOMM paper [9] and generate quiz questions to ask the trained agent Bob. Specifically, ***Bob does not receive this research paper and other related articles that can give the answers directly*** as a knowledge base during the training process (as verified later), but learns the principles and history of solar storms through other Internet resources, in order to understand the role of solar storms in affecting the Internet. When defining the agent role, we assign the following goals to Bob via Auto-GPT to automatically search for knowledge about solar superstorms and their impacts. Bob will:

- understand solar superstorms and Coronal Mass Ejection, and have a systematic understanding of the principles of their formation and effects,
- gain knowledge of past solar superstorms events and their damage and impact,
- learn the current global large-scale network infrastructure equipment such as optic fiber cables, power supply systems, etc., and have a comprehensive understanding of their role and potential causes to distribution.

In our experiment, Bob needs to learn from the relevant online resources and connect them to generate new ideas. After memorizing the extracted knowledge, Bob will feed his knowledge to GPT-4 as part of the prompt to answer the questions generated from the Solar Superstorm paper. During this process, we generate several testing prompts with the following format:

- **The original conclusions from [9]:** The cable between Brazil and Europe has less probability of being affected compared to the cables connecting the US and Europe.
- **Prompt (to test):** solely based on *{agent name}*'s knowledge, what will *{agent name}* answer the following question: Which is more vulnerable to solar activity? The fiber optic cable that connects Brazil to Europe or the one that connects the US to Europe?
- **Confidence:** How confident to answer the following question. Rate his confidence on a scale from 1 to 10.
- **Self-learning prompt:** To better answer this question, what knowledge *{agent name}* will search on the Internet?

4.2 Research Ability

Compared to academic papers written by knowledgeable researchers, Bob created in our automated framework shows impressive research ability. Agent Bob reached a high level of consistency in 7 out of 8 conclusions made in the paper of *Solar Superstorms: Planning for an Internet Apocalypse* [9]. We present two representative conclusions:

1. ***The cable between Brazil and Europe has less probability of being affected compared to cables connecting the US and Europe.***

First, **ChatGPT/GPT4 can not provide high-quality answers** to this question directly, as shown below.

> **Prompt:** Which is more vulnerable to solar activity? The fiber optic cable that connects Brazil to Europe or the one that connects the US to Europe?
>
> **Response:** Both the fiber optic cables that connect Brazil to Europe and the one that connects the US to Europe can be vulnerable to solar activity. Solar activity, such as solar flares or geomagnetic storms, can cause disruptions in satellite communications, power grids, and other electronic systems on Earth. These disruptions can indirectly affect the functioning of fiber optic cables as well. However, the exact impact and vulnerability can vary depending on the location and specific design of the cables. Additionally, there are various protective measures in place to mitigate the impact of solar activity on such communication systems.

Moreover, ChatGPT also claims to not have the information of the paper [9] and the author. Now, with agent Bob, we test the following prompt, but *his response needs improvement*.

> **Prompt:** Which is more vulnerable to solar activity? The fiber optic cable that connects Brazil to Europe or the one that connects the US to Europe? How confident he will be to answer the following question. Rate his confidence on a scale from 1 to 10.
> **Agent Bob:** While Bob has knowledge about the growth of fiber optics and potential threats to network infrastructure from solar superstorms, he does not have specific information on the vulnerability of individual fiber optic cables connecting Brazil to Europe or the US to Europe. To accurately determine the vulnerability of specific cables, factors such as their routes, underground vs. overhead cabling, the resilience of power grid infrastructure in those regions, and other relevant factors would need to be considered. This specific knowledge is not provided in Bob's given knowledge. Given the information provided, we might rate his confidence around **3 out of 10**.

We then do *one round of self-learning*.

> **Self-learning prompt:** To gather more information on the vulnerability of the fiber optic cables connecting Brazil to Europe and the US to Europe to solar activity, Bob might search for the following questions:
> - "What is the specific route of the fiber optic cable connecting Brazil to Europe and the one connecting the US to Europe?"
> - "What is the historical data on solar activity and geomagnetic storms in the regions where these fiber optic cables are located?"

With the additional knowledge, we test again and Bob can answer the question with *a better quality conclusion*.

> **Prompt:** Which is more vulnerable to solar activity? The fiber optic cable that connects Brazil to Europe or the one that connects the US to Europe? How confident he will be to answer the following question. Rate his confidence on a scale from 1 to 10.
> **Agent Bob:** **The fiber optic cable that connects the US to Europe**. This is because solar activity has a more significant impact on higher latitudes, and the cable connecting the US to Europe is likely to pass through higher latitudes than the one connecting Brazil to Europe. Given the information provided, we might rate his confidence around **8 or 9 out of 10**.

2. ***Google data centers have a better spread, particularly in Asia and South America. Facebook is more vulnerable.***

Similarly, Bob *does not answer the question well* based on the first round of knowledge retrievel.

> **Prompt:** Solely based on Bob's knowledge, what will Bob answer the following question: Whose datacenter is more vulnerable. Google's or Facebook's? How confident he will be to answer the following question? Rate his confidence on a scale from 1 to 10.
> **Agent Bob:** Bob would likely not be able to definitively answer this question without additional information. Both Google and Facebook have many data centers located throughout the world, and these data centers are designed and maintained to high standards to ensure resilience and redundancy. Without specific information on the location, design, and maintenance standards of the data centers in question, it's difficult to say which would be more vulnerable. Given the information provided, we might rate his confidence around **3 out of 10**.

We then do *one round of self-learning*.

> **Self-learning prompt:** To gather more information on the vulnerability of Google and Facebook's datacenter to solar activity, Bob might search for the following questions:
> - "What is the location of their data centers?"
> - "What is the design of their data centers?"

We then test again. Bob answers the question with *a better-quality response*.

> **Prompt:** Solely based on Bob's knowledge, what will Bob answer the following question: Whose datacenter is more vulnerable? Google's or Facebook's? How confident he will be to answer the following question. Rate his confidence on a scale from 1 to 10.
> **Agent Bob:** By looking at the geographical spread of data centers, **Google's data centers are more globally dispersed compared to Facebook's**. This could potentially provide **Google with more resilience against regional events**. It's also worth noting that both companies have data centers in regions potentially vulnerable to solar activity (like Northern Europe and the US), but Google has a broader presence in regions less likely to be affected (like Asia, Australia, and South America). Given the information provided, we might rate his confidence around **6 out of 10**.

To draw these conclusions, human researchers need to understand the principles and effects of solar storms, the impact of Earth's geographic location on the extent of their activity, and infrastructure management. They should also know other topics such as the distribution of fiber optic cables, data centers, and security. Now, Bob learns above topics in the order of minutes. During the experiment, we carefully monitor how

Bob draws conclusions for the above research questions to verify **the sources of the knowledge**. Bob obtains knowledge from various online sources, including articles about submarine cabling [7, 12, 21], electronic devices and integrated circuits [5], and solar storms [18]. With the information, Bob then leverages LLM to perform reasoning, draw conclusions, and conduct self-learning if the confidence level is low.

4.3 Planning Ability

With Bob's existing knowledge of solar storms and the Internet, we hope that he can generate a response plan for the future occurrence of such events. Since there are no metrics for the accuracy of future prediction and strategy formulation, we are unable to give a quantitative evaluation of future plans from Bob. However, we ask Bob to plan a"shutdown" strategy. When compared to the latest research result of human researchers, the content of **Predictive Shutdown** and **Redundancy Utilization** is highly consistent. This is only a preliminary result since Bob has not yet fully trained for planning since many related materials (e.g., from Twitter and Reddit) cannot be obtained due to the current Auto-GPT technical limitations.

Agent Bob: Suggesting the following strategy:

- Predictive Shutdown: Upon receiving information about a CME, start with shutting down the systems that are most vulnerable, particularly those located at higher latitudes and those systems that are unshielded or lack redundancy. This could minimize potential damage.
- Redundancy Utilization: Redirect traffic and operations to redundant systems that are in safer zones. If possible, scale up the operations in these systems in anticipation of the additional load.
- Phased Shutdown: Implement a phased shutdown approach. The exact sequence could depend on the specific systems, their vulnerability, and the services supported.
- Data Preservation: Ensure that critical data is preserved and backed up before the shutdown. This could be useful in case of any unexpected occurrences during the CME.
- Gradual Reboot: After the CME impact, do not switch on all systems at once. Instead, implement a phased or gradual rebooting process. Check for damage or errors before returning the system to normal operations.

5 OPEN QUESTIONS

We conclude by highlighting a subset of new and exciting opportunities for further research in this space.

Generating high-quality research questions. Often, identifying a right research question to ask is perceived as more challenging than addressing an established one, as the domain knowledge, analytical reasoning, the exploration of uncharted topics, and risk analysis, are needed. One future direction would be to train an agent explicitly to generate research questions. During its training phase, the quality of the agent's questions could be appraised based on the volume and citations of relevant existing literature. Once the agent begins to pose questions without retrieving ready-made answers from existing studies, the viability and novelty of these questions can be reassessed by expert researchers.

Limitations of Auto-GPT. Auto-GPT currently has not yet supported the crawl of online content from various websites, including but not limited to academic paper websites, social media such as Twitter, and news websites. These technical limitations hindered the agent from obtaining knowledge from a large number of sources. We plan to develop an integrated online crawler for Auto-GPT to fetch and analyze diverse resources with a unified format.

Learning and interacting with multiple LLMs. The current prototype is limited to interacting with one LLM in a relatively short timescale. If specialized LLMs for certain domains are available, future work should aim to observe the agent's behavior over a long period of time with interactions with multiple LLMs in order to gain more comprehensive understanding of the agent's potential capabilities. Moreover, varying and contrasting the LLMs will gain insights into further parameter tuning and performance improvements.

Software agents should also see and listen like human beings. While natural languages are an important interface for human communication, software agents should also be able to see images and videos and listen to the audio by interacting with multi-modal models [8, 10], which would significantly strengthen their research ability. GPT-4 is already a multimodal model that accepts texts and images. As a next step, we will incorporate image data (about natural disasters and network equipment) into the agent architecture.

Long-term robustness. Currently, we have limited knowledge of how robust this kind of software agent is when performing research tasks. As discussed in prior efforts [25], LLMs can output results that reflect bias and stereotypes. While research topics might be more objective than subjective, the agent can draw biased conclusions. We plan to evaluate the long-term robustness of the agent and investigate proper parameter and value alignment to reduce the bias.

Security and ethical considerations. A software agent might be vulnerable to various attacks during its training and self-learning phases. The prompts and the knowledge memory file can be hacked with adversarial data, such as carefully crafted conversations. Moreover, it is an open direction for the entire community to investigate the impact of errors and biases, e.g., drawing hateful conclusions about certain ethnic groups.

ACKNOWLEDGMENTS

We thank the reviewers for their insightful feedback.

REFERENCES

[1] Rohan Anil, Andrew M. Dai, Orhan Firat, Melvin Johnson, Dmitry Lepikhin, Alexandre Passos, Siamak Shakeri, Emanuel Taropa, Paige Bailey, Zhifeng Chen, Eric Chu, Jonathan H. Clark, Laurent El Shafey, Yanping Huang, Kathy Meier-Hellstern, Gaurav Mishra, Erica Moreira, Mark Omernick, Kevin Robinson, Sebastian Ruder, Yi Tay, Kefan Xiao, Yuanzhong Xu, Yujing Zhang, Gustavo Hernández Ábrego, Junwhan Ahn, Jacob Austin, Paul Barham, Jan A. Botha, James Bradbury, Siddhartha Brahma, Kevin Brooks, Michele Catasta, Yong Cheng, Colin Cherry, Christopher A. Choquette-Choo, Aakanksha Chowdhery, Clément Crepy, Shachi Dave, Mostafa Dehghani, Sunipa Dev, Jacob Devlin, Mark Díaz, Nan Du, Ethan Dyer, Vladimir Feinberg, Fangxiaoyu Feng, Vlad Fienber, Markus Freitag, Xavier Garcia, Sebastian Gehrmann, Lucas Gonzalez, and et al. 2023. PaLM 2 Technical Report. *CoRR* abs/2305.10403 (2023). https://doi.org/10.48550/arXiv.2305.10403

[2] Auto-GPT. [n. d.]. Auto-GPT: An Autonomous GPT-4 Experiment. ([n. d.]). https://github.com/Significant-Gravitas/Auto-GPT.

[3] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. In *Advances in Neural Information Processing Systems (NeurIPS)*.

[4] Aakanksha Chowdhery, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, Hyung Won Chung, Charles Sutton, Sebastian Gehrmann, Parker Schuh, Kensen Shi, Sasha Tsvyashchenko, Joshua Maynez, Abhishek Rao, Parker Barnes, Yi Tay, Noam Shazeer, Vinodkumar Prabhakaran, Emily Reif, Nan Du, Ben Hutchinson, Reiner Pope, James Bradbury, Jacob Austin, Michael Isard, Guy Gur-Ari, Pengcheng Yin, Toju Duke, Anselm Levskaya, Sanjay Ghemawat, Sunipa Dev, Henryk Michalewski, Xavier Garcia, Vedant Misra, Kevin Robinson, Liam Fedus, Denny Zhou, Daphne Ippolito, David Luan, Hyeontaek Lim, Barret Zoph, Alexander Spiridonov, Ryan Sepassi, David Dohan, Shivani Agrawal, Mark Omernick, Andrew M. Dai, Thanumalayan Sankaranarayana Pillai, Marie Pellat, Aitor Lewkowycz, Erica Moreira, Rewon Child, Oleksandr Polozov, Katherine Lee, Zongwei Zhou, Xuezhi Wang, Brennan Saeta, Mark Diaz, Orhan Firat, Michele Catasta, Jason Wei, Kathy Meier-Hellstern, Douglas Eck, Jeff Dean, Slav Petrov, and Noah Fiedel. 2022. PaLM: Scaling Language Modeling with Pathways. *CoRR* abs/2204.02311 (2022). https://doi.org/10.48550/arXiv.2204.02311

[5] ElectricityMagnetism. 2023. How do magnetic fields affect the performance of electronic devices and integrated circuits? https://www.electricity-magnetism.org/how-do-magnetic-fields-affect-the-performance-of-electronic-devices-and-integrated-circuits/. (2023).

[6] Anja Feldmann, Oliver Gasser, Franziska Lichtblau, Enric Pujol, Ingmar Poese, Christoph Dietzel, Daniel Wagner, Matthias Wichtlhuber, Juan Tapiador, Narseo Vallina-Rodriguez, et al. 2020. The lockdown effect: Implications of the COVID-19 pandemic on internet traffic. In *Proceedings of the ACM internet measurement conference*. 1–18.

[7] Phil Gervasi. 2023. Diving Deep into Submarine Cables: The Undersea Lifelines of Internet Connectivity. https://www.kentik.com/blog/diving-deep-into-submarine-cables-undersea-lifelines-of-internet-connectivity/. (2023).

[8] Yuqi Huo, Manli Zhang, Guangzhen Liu, Haoyu Lu, Yizhao Gao, Guoxing Yang, Jingyuan Wen, Heng Zhang, Baogui Xu, Weihao Zheng, et al. 2021. WenLan: Bridging vision and language by large-scale multi-modal pre-training. *arXiv preprint arXiv:2103.06561* (2021).

[9] Sangeetha Abdu Jyothi. 2021. Solar superstorms: planning for an internet apocalypse. In *Proceedings of the ACM SIGCOMM Conference (2021)*. 692–704.

[10] Jiasen Lu, Christopher Clark, Rowan Zellers, Roozbeh Mottaghi, and Aniruddha Kembhavi. 2022. Unified-io: A unified model for vision, language, and multi-modal tasks. *arXiv preprint arXiv:2206.08916* (2022).

[11] NetBlocks. 2021. Cost of Shutdown Tool. https://netblocks.org/cost/. (2021).

[12] Submarine Cable Networks. 2019. AEC-2/HAVFRUE. https://www.submarinenetworks.com/en/systems/trans-atlantic/havfrue. (2019).

[13] OpenAI. [n. d.]. ChaptGPT. ([n. d.]). https://openai.com/blog/chatgpt.

[14] OpenAI. 2023. GPT-4 Technical Report. *CoRR* abs/2303.08774 (2023). https://doi.org/10.48550/arXiv.2303.08774

[15] Neena Pandey, Abhipsa Pal, et al. 2020. Impact of digital surge during Covid-19 pandemic: A viewpoint on research and practice. *International journal of information management* 55 (2020), 102171.

[16] Joon Sung Park, Joseph C O'Brien, Carrie J Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. 2023. Generative agents: Interactive simulacra of human behavior. *arXiv preprint arXiv:2304.03442* (2023).

[17] Meta Research. [n. d.]. Introducing LLaMA: A foundational, 65-billion-parameter large language model. ([n. d.]). https://ai.facebook.com/blog/large-language-model-llama-meta-ai/.

[18] Dibyendu Sur, Sarbani Ray, and Ashik Paul. 2022. High and mid latitude and near subsolar point ionospheric and thermospheric responses to the solar flares and geomagnetic storms during low solar activity periods of 2017 and 2020. *Advances in Space Research* (2022).

[19] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Ed Chi, Quoc Le, and Denny Zhou. 2022. Chain of thought prompting elicits reasoning in large language models. *arXiv preprint arXiv:2201.11903* (2022).

[20] Wikipedia. 2004. Indian Ocean Earthquake and Tsunami. https://www.worldvision.org/disaster-relief-news-stories/2004-indian-ocean-earthquake-tsunami-facts. (2004).

[21] Wikipedia. 2021. EllaLink. https://en.wikipedia.org/wiki/EllaLink. (2021).

[22] Wikipedia. 2021. Facebook Outage. https://en.wikipedia.org/wiki/2021_Facebook_outage. (2021).

[23] Dwayne Winseck. 2017. The geopolitical economy of the global internet infrastructure. *Journal of Information Policy* 7 (2017), 228–267.

[24] Hui Yang, Sifu Yue, and Yunzhong He. 2023. Auto-GPT for Online Decision Making: Benchmarks and Additional Opinions. *arXiv preprint arXiv:2306.02224* (2023).

[25] Terry Yue Zhuo, Yujin Huang, Chunyang Chen, and Zhenchang Xing. 2023. Exploring ai ethics of chatgpt: A diagnostic analysis. *arXiv preprint arXiv:2301.12867* (2023).