

# ChatGPT as a Copilot for Investigating Digital Evidence

Hans Henseler<sup>1,2</sup>, Harm van Beek<sup>2</sup>

<sup>1</sup>University of Applied Sciences Leiden, The Netherlands

<sup>2</sup>Netherlands Forensic Institute

## Abstract

In today's technology-driven legal landscape, practitioners must continually adapt to new tools and methods that aid not only in addressing cybercrime but also in managing traditional crimes with digital components. This paper explores the potential of advanced AI-powered solutions, such as ChatGPT, in enhancing the capabilities of investigators in various aspects of their investigations. We delve into three specific applications pertinent to legal professionals: (1) writing structured queries utilizing natural language and trace models, (2) summarizing, evaluating, and visualizing electronic communications, and (3) analysing search results. Our findings demonstrate that once ChatGPT is proficient in the query language and data model of the system containing the digital evidence, it holds significant promise in assisting legal professionals in conducting effective investigations.

## Keywords

digital forensics, eDiscovery, large language models, natural language processing, deep learning, chatgpt, gpt-4

## 1. Introduction

The legal profession is witnessing a significant surge in the adoption of artificial intelligence (AI) tools, with ChatGPT emerging as a prominent development since November 2022 [1]. Powered by OpenAI's advanced large language model, ChatGPT offers a natural and engaging conversational interface on an extensive array of topics encountered during its training. ChatGPT's web application provides users with access to various models, including the Default GPT-3.5 turbo (a refined and superior version of GPT-3), Legacy GPT-3.5 (the preceding ChatGPT model), and GPT-4 (the most sophisticated model, exclusively accessible to ChatGPT Plus subscribers). The experiments discussed in this paper employ the ChatGPT/GPT-4 model, which showcases its potential applications in the domain of digital evidence investigation.

ChatGPT has been fine-tuned with Reinforcement Learning from Human Feedback (RLHF), e.g., human reviewers have trained the model to be "helpful, honest and harmless", i.e., to not use undesirable language (violent, sexist, racist, etc.) in conversations. Also, in some cases reviewers were asked to write example responses for a chatbot. ChatGPT has learned what useful answers are, so that conversations flow smoothly. Employing people to review thousands of prompts is costly and seems to be one of ChatGPT's key success factors.

In: *Proceedings of the Third International Workshop on Artificial Intelligence and Intelligent Assistance for Legal Professionals in the Digital Workplace (LegalAIIA 2023)*, held in conjunction with ICAIL 2023, June 19, 2023, Braga, Portugal.

✉ [henseler.h@hsleiden.nl](mailto:henseler.h@hsleiden.nl) (H. Henseler); [harm.van.beek@nfi.nl](mailto:harm.van.beek@nfi.nl) (H. van Beek)

© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).  
CEUR Workshop Proceedings (CEUR-WS.org)

ChatGPT and large language models (LLMs) in general are often seen as statistical machines that have learned to predict the next word based on the sequence of preceding words [2]. The word 'statistical machine' might be misleading since there seems to be much more to a large language model than simply predicting the next word. For instance, Shanahan [3] talks about what LLMs actually do, how they compare to humans and about unexpected emerging behaviour. A more accurate description could be that ChatGPT has learned an algorithm to predict the next word.

Nonetheless, ChatGPT is certainly not flawless. It makes mistakes and sometimes hallucinates facts which is considered a dangerous aspect. However, when used as an assistant and when properly instructed, it can be a smart student that is able to help digital forensic experts more efficiently and effectively investigate cases with digital evidence it has never seen before.

In this paper we describe the rise of ChatGPT followed by prospective applications in digital investigation and an overview of related work. We then present empirical results from three experiments, followed by an evaluation, discussion and conclusions.

## 2. The rise of ChatGPT

Deep learning is a subset of machine learning that focuses on artificial neural networks with multiple layers. These networks can automatically learn representations from raw data and have demonstrated significant success in various domains, including computer vision, natural language processing, and speech recognition [4].

Large language models, such as GPT-3 [5], are a specific application of deep learning in the field of natural language processing. These models are trained on

vast amounts of text data, enabling them to generate human-like text and perform various language-related tasks with high accuracy. The development of large language models has significantly advanced the capabilities of AI systems, allowing them to generate text in a more context-aware and coherent manner.

The GPT-4 architecture that is underlying ChatGPT is one such large language model. It is trained using a diverse dataset, including books, articles and web pages, enabling it to generate contextually relevant responses to text inputs. Its architecture allows it to capture long-range dependencies and generate coherent responses across a wide range of topics.

Large language models have found success in several applications outside the field of digital evidence investigations. For instance, they have been used to generate coherent and context-aware text for creative writing [6, 7], assist with code generation and programming tasks [8], and provide relevant answers to questions in customer support scenarios [9, 10]. These models have also demonstrated capabilities in tasks such as machine translation [11] and summarization [12], showcasing their versatility and potential across various domains.

Finally we note that ChatGPT is also heavily debated in the field of law and legal services. For instance, in a panel discussion hosted by the Northwestern Law and Technology Initiative and the High Tech Law Society [13] panelists discussed the potential and limitations that ChatGPT has for various legal tasks, such as drafting contracts and briefs, providing legal advice, and automating legal reasoning. A few months later ChatGPT/GPT-4 passed the Uniform Bar Examination with a score of 297, which is higher than the passing threshold of any jurisdiction and nears the 90th percentile of human test-takers [14, 15]. The consensus in the field of law and legal services seems to be that ChatGPT is a powerful tool that could be transformative, but also requires careful evaluation and regulation to ensure its responsible use.

### 3. Prospective applications

Digital forensics and eDiscovery are vital fields that involve the identification, preservation, analysis, and presentation of digital evidence (often referred to as Electronic Stored Information or ESI) in legal proceedings. As technology continues to evolve and data volumes grow exponentially, the need for advanced tools and techniques to efficiently process, analyze, and interpret digital evidence becomes increasingly important.

In the past, we experimented with AI applications for eDiscovery related tasks and reported results and ideas at ICAIL and associated workshops DESI<sup>1</sup> and its successor

<sup>1</sup>Workshop on Discovery of Electronically Stored Information <http://users.umiacs.umd.edu/~oard/desi7/>

LegalAIIA<sup>2</sup>, e.g., network-based filtering for large email collections [16], semantic search in eDiscovery [17] and technology assisted analysis of timeline and connections in digital forensic investigations [18]. The experiments in this paper will touch on similar topics but are focused on the application of ChatGPT. We discuss the relevance of three specific tasks in digital evidence investigations.

#### 3.1. Writing structured queries utilizing natural language and trace models

One of the primary challenges faced by investigators is efficiently querying vast amounts of digital evidence. Typically, investigators need to have a thorough understanding of forensic data models and query languages, such as SQL, to retrieve relevant information from digital evidence. This can be a time-consuming and complex process, especially for investigators with limited technical expertise in this area.

The ability to convert natural language queries into structured queries can significantly streamline the investigative process. By allowing investigators to simply describe their query in everyday language, large language models can generate the corresponding structured query that retrieves the relevant information. This not only saves time and effort but also reduces the potential for errors caused by incorrect or incomplete queries.

#### 3.2. Summarize, evaluate and visualize chat messages

Chat conversations often play a critical role in digital evidence investigations, as they can provide valuable insights into the activities, intentions, and relationships of individuals involved in illicit activities. Analyzing chat conversations can be a labor-intensive and time-consuming task, particularly when dealing with lengthy or complex chat logs involving multiple participants.

Text summarization is an active research field since the 1950s [19]. It gained increased attention with the growth of the Internet and multimedia technology, which led to an exponential surge in text data. With the advent of emerging technologies, new challenges have arisen, including handling multimodal inputs, generating summaries across different domains and languages, ensuring coherence, readability, and objective and subjective evaluation of summaries. Traditional summarization technologies, such as statistical methods, rule-based methods, and neural network methods, had limitations in terms of hand-crafted features, data sparsity, and generating generic summaries. Large language models mitigate

[//users.umiacs.umd.edu/~oard/desi7/](http://users.umiacs.umd.edu/~oard/desi7/)

<sup>2</sup>International Workshop on AI and Intelligent Assistance for Legal Professionals in the Digital Workplace <https://sites.google.com/view/legalaiia2019>

some of these limitations by harnessing more data and computational resources, positioning themselves as a promising solution for summarization tasks [20], despite the ongoing challenges of ensuring factual consistency, diversity, and controllability.

By utilizing large language models to automatically summarize chat conversations, investigators can quickly gain an understanding of the key points and themes discussed within the conversations. This enables them to focus their attention on the most relevant content, making the investigation process more efficient. Moreover, summarization can also aid in the presentation of digital evidence in court, as it allows for a more concise and accessible representation of complex chat logs for non-technical audiences, such as judges and jurors.

### 3.3. Analysing search results

Since ChatGPT is capable of conducting a dialog and remembers previous prompts and responses, we perform an experiment where we introduce a fictitious case. We present different sets of search results to ChatGPT and ask for an analysis, e.g., see if traces are correlated and if ChatGPT can provide a narrative.

If ChatGPT can analyse search results and is able to correlate and group results from search queries with different types of digital evidence, this will show that ChatGPT is aware of the case context which will make it a much more effective copilot for digital investigations.

## 4. Related work

Digital evidence investigations have many things in common with other types of investigations and work-related tasks that are common place in all lines of work, such as report writing, researching open sources on the internet, writing software and taking notes. We covered the success of ChatGPT in these areas shortly in Section 2. In this section, we cover related work on the application of ChatGPT to tasks that are more specifically linked to digital evidence investigation.

### 4.1. eDiscovery tasks

In the realm of legal technology, specifically eDiscovery, other researchers have also recognized the potential of ChatGPT in assisting with various tasks. In “What Will eDiscovery Lawyers Do After ChatGPT?” [21], Tredenick and Webber discuss the capabilities and limitations of ChatGPT when applied to common eDiscovery tasks.

The authors conducted a series of experiments where they asked ChatGPT to construct complex boolean queries based on fairly simple questions. While ChatGPT cannot answer queries about information it has not

seen before, it is familiar with the well-known Enron case [22], which has been extensively discussed on the Internet and is frequently used in eDiscovery education. When asked to provide examples of how Enron violated U.S. Federal Government accounting standards, ChatGPT effortlessly and impressively answered, likely drawing from sources such as Wikipedia and numerous online publications that analyze Enron’s downfall.

Such analyses, of course, do not exist in new investigations. The key question raised by the authors is whether ChatGPT can also analyze documents in new investigations, as such analyses do not exist for cases it is unfamiliar with. In a limited test, ChatGPT was able to determine the relevance of an email from the Enron dataset when given a specific question. It could also explain the reason for its relevance. However, this approach may not be applicable to new cases, as the underlying GPT-3 model has not been trained on information about them. Such new information can be learned by GPT-3. OpenAI offers the ability to fine-tune models in the cloud. But that is costly, and lawyers are reluctant to upload their clients’ sensitive documents. That situation is likely to change as alternatives to ChatGPT /GPT-4 that can run on-premise are already emerging [23, 24, 25].

This related work demonstrates that large language models, like ChatGPT, have the potential to be valuable tools in the eDiscovery domain, particularly for tasks such as constructing complex boolean queries and determining the relevance of specific documents. However, as the authors have pointed out, the effectiveness of ChatGPT in new investigations remains a concern, as its knowledge is limited to the information it has been trained on. Nevertheless, the success of ChatGPT in the context of the Enron case serves as a promising indication that these models can contribute significantly to the field of digital forensics and legal technology, provided that their limitations are addressed and overcome.

### 4.2. Formulating SQL and Cypher queries

Recent advances in large language models, particularly those based on GPT-like architectures, have significantly improved the generation of SQL queries from natural language inputs. The study of neural text-to-SQL systems has become a “highly competitive and fast paced research field” [26]. Earlier approaches relied on sequence-to-sequence networks to generate SQL queries, but models such as GPT-3 and its descendants already demonstrated promising results. Mannelly (2022) [27] demonstrates how GPT-3 can be used to generate SQL queries from plain English text using NBA data, showcasing the potential of GPT-3 in transforming natural language inputs into structured SQL queries. In a more practical application, Alexander (2023) [28] provides a step-by-step guide on using ChatGPT to write SQL JOIN queries and

demonstrates the use of dbForge Studio for SQL Server to verify and test the generated queries. These recent developments highlight the potential of GPT-like models in revolutionizing the way SQL queries are generated from natural language inputs, opening up new possibilities for database management and investigating digital evidence.

The use of large language models to transform natural language queries into structured queries that are less well known than SQL has also gained attention, as exemplified by an online article that explores the application of GPT-3 in transforming English natural language questions into Cypher queries for the Neo4j graph database [29]. This work is closely related to our article, as it demonstrates the potential of large language models, such as GPT-3 and ChatGPT, in assisting digital forensic experts and investigators in converting natural language queries into a structured format, other than SQL, that takes advantage of specific data models and query languages.

In the article, the authors describe their experiments using GPT-3 to “understand” and process natural language queries and generate corresponding Cypher queries that can be used in the context of the Neo4j graph database. This is a powerful application, as it allows users to ask questions in everyday language and receive accurate responses from the database without requiring knowledge of the underlying query language or data model. By reducing the complexity and technical barriers associated with querying graph databases, large language models can empower a broader range of users to access and analyze the wealth of information contained within such databases.

### 4.3. Semantic search and relevance determination

At the ICDDF 2023<sup>3</sup> digital forensics lead scientist S. Cullen presented on the “Practical application of artificial intelligence to solve Digital Forensic challenges”. He described an experiment illustrating the capability of ChatGPT to search semantically and multilingual. In the example ChatGPT was asked about a car and Cullen demonstrated that ChatGPT would not only find ‘car’ but would also find ‘Mercedes’ and even was not bothered if the text was presented in Arabic.

In another experiment illustrating the use of ChatGPT, Cullen explained that in the UK victims of crime (including rape complainants) are asked for permission to view all data including messages, photographs, emails and social media accounts. Victims’ groups say this amounts to a “digital strip search” [30]. Cullen reported that ChatGPT is capable of protecting the privacy of a victim by

evaluating the contents of a chat conversation indicating if it is relevant for the criminal investigation or not.

Related to this, also at ICDDF 2023, senior responsible officer T. Baker, from the UK Police Digital Service<sup>4</sup>, referred to an experiment where technology similar to ChatGPT was successfully tested to support the disclosure process. Disclosure is a fundamental part of fair criminal investigations and trials, but Rape and Serious Sexual Offences cases can present particular challenges given the volumes of potentially relevant digital material generated. For example, the police and the Crown Prosecution Service may need to review messages, photos, videos, or location history from the phones or computers of the victims and suspects. This can be time-consuming and resource-intensive, and may also raise privacy and consent issues for the parties involved.

## 5. Experiments

In this section we first illustrate how ChatGPT can utilize a digital trace model and how it can then be instructed to formulate structured queries based on natural language input from an investigator. Next we present chat conversations to ChatGPT and ask for various types of summarization, including semantic analysis and network visualisation. In the third experiment we interact with ChatGPT via a series of prompts with facts from a fictitious case and query results containing browser history, chat messages and geo locations.

### 5.1. Writing structured queries utilizing natural language and trace models

Hansken<sup>5</sup> is a digital forensics as a service (DFaaS) [32] platform for law enforcement and intelligence agencies to search and analyze digital traces from seized digital devices such as phones and computers. Hansken has been around for more than 10 years and was developed by the Netherlands Forensic Institute. It is based on Hadoop and Elasticsearch, among others, making it capable of processing and storing petabytes of digital data.

Hansken processes digital traces using a trace model. Using the Hansken Query Language (HQL), these traces can be searched in the elasticsearch database with excellent performance, even in cases with hundreds of terabytes of seized data that have hundred million traces or more. HQL is a powerful language similar to other query languages such as SQL but is completely focused on the Hansken trace model. ChatGPT can start learning to transform questions in plain language into HQL syntax by prompting it with text from the Hansken HQL

<sup>3</sup>The annual conference on International Communications Data and Digital Forensics at the Renaissance Heathrow hotel, April 19th, 2023

<sup>4</sup>PDS Digital Forensics supports the implementation of the National Police Chiefs’ Council Digital Forensic Science Strategy [31]

<sup>5</sup>For more information, see <https://hansken.org/>



ChatGPT responds as follows:

```
[email.from,email.to,email.cc,email.bcc]:*Quan*
```

This is a valid HQL query that searches email fields from, to, cc, bcc for occurrences of \*Quan\*. We note that this syntax is not provided in the Hansken HQL manual but it is present as an example in the HQL cheat sheet.

## 5.2. Summarize, evaluate and visualize chat messages

The previous experiment showed that ChatGPT can assist with writing HQL queries. The next logical step is to test if ChatGPT can process the results from a query. In this experiment we ask ChatGPT to assist us with analysing chat conversations that have been extracted from multiple phones belonging to different suspects and/or witnesses in an investigation.

At the start of an investigation, very little is known (e.g., only the name of a suspect and a suspicion of a crime). Using contacts in a phone and laptop, an investigator can gain insight into the suspect's network, certain events, and locations linked to date and time. Digging out all that information is time-consuming and doesn't always lead to relevant information. ChatGPT can summarize and organize transcripts of chats. Then ChatGPT can also answer questions.

We manually copied 12 chat conversations (a chronologically ordered list of chat messages between two persons) from the Crystal Clear training case<sup>6</sup>. Initially we presented all chats in a single prompt to ChatGPT but that exceeded the capacity of the current version of ChatGPT. Then, we split the conversations over two and three files respectively. In both cases it was difficult to get summaries that were consistent. We then decided to present the chat conversations separately, i.e., as single prompts, but in the same ChatGPT session. Here is the initial prompt asking ChatGPT to summarize a chat between two persons as well as the first conversation:

```
You are SleuthGPT. You are a detective that needs to find out what has been going on in a group of people. They want to help you and gave transcripts of their conversations. Most of them via whatsapp and some of them via Telegram and SMS text messages. You study the conversations and as a sleuth it is your job to figure out who is the leader of the group, who is the buyer, who is transporting and who are helping. It would be helpful if you identify the relationship between persons and if they are friendly, close, if they like each other or not. Dates are in dd-mm-yyyy format. For example 5-7-2022 is July 5th 2022. A name may have a identifier associated. This identifier is shown at the start of the conversion between brackets () following the name if it occurs for the first time. Summarize the following conversation including names of persons communicating, date of first message, date of last message and a summary including a sentiment analysis and characterization of the conversation.
```

<sup>6</sup>The Crystal Clear case is a fictitious case that was prepared by the Hansken Academy for Hansken e-learning and class room trainings

Person	Role
Quan	Seeker of information and guidance; involved in group activities; interacts with Figo, Sheila, Joseph, and Nerijus
Figo	Provider of instructions and assistance; possible leader or coordinator in the group; interacts with Quan
Sheila	Friend or acquaintance; part of the same social circle as Quan; no specific involvement in group activities; interacts with Quan
Joseph	Helps with transportation; possibly involved in group activities; interacts with Quan, Nerijus, Liss, and Rosie
Nerijus	Supports and coordinates tasks; possibly involved in group activities; interacts with Joseph, Antonio Gotta, Rosie, Liss, and Quan
Antonio Gotta	Business partner of Nerijus; involved in financial transactions and deals; interacts with Nerijus
Rosie	Trusted associate of Nerijus; involved in group operations and plans; interacts with Nerijus
Liss	Joseph's estranged partner; threatened by Nerijus; seeks to distance herself from group activities; interacts with Joseph and Nerijus

**Table 1**  
Summary of persons and their role

Conversation 1 between Quan (owner) and Figo (31644891202@es.whatsapp.net):

```
5-7-2022 12:08:53 - 31644891202@es.whatsapp.net: Hi Quan, do you want to meet again or do you have enough information?
5-7-2022 12:09:22 - owner: i know enough!
5-7-2022 12:09:31 - 31644891202@es.whatsapp.net: Excellent
5-7-2022 12:09:32 - owner: Thanks you !
5-7-2022 12:11:50 - owner: What's the fee for the party?
5-7-2022 12:13:41 - owner: Can you shine a light on that?
5-7-2022 12:30:48 - 31644891202@es.whatsapp.net: Check your e-mail
11-7-2022 08:34:23 - 31644891202@es.whatsapp.net: Hi Quan, I have some instructions for your family visit: - buy suitcase via Marktplaats for transfer (make sure to pay cash to leave no traceble receipt) - Sent me a photo of the suitcase via Snapchat - Sent me a photo of the drugs and the lab via Snapchat - Get a signal account, I will contact you there today
```

ChatGPT's response to this prompt is as follows:

```
Conversation 1:
Participants: Quan (owner) and Figo (31644891202@es.whatsapp.net)
Date of First Message: 5-7-2022
Date of Last Message: 11-7-2022

Summary:
Quan and Figo discuss meeting up and sharing information. Quan seems satisfied with the current information and thanks Figo. Quan inquires about the fee for the party, and Figo directs Quan to check their email. Later, Figo provides instructions for Quan regarding a family visit, including purchasing a suitcase for transfer, taking photos of the suitcase and drugs /lab, and setting up a Signal account.
```

Sentiment Analysis:

Participants	Dates	Summary	Sentiment	Characterization
Quan, Figo	5-11/7/2022	Meeting coordination and info sharing. Party fee, family visit instructions.	Cooperative, efficient	Leader/coordinator
Quan, Sheila	5-8/7/2022	Discuss daily activities, coffee, dinner, college, clubbing, McDonald's and Club Thouma invites.	Casual, friendly	Socializing friends
Quan, Joseph	12-21/7/2022	Coordinate airport pickup, meet at large AKO store.	Polite, cooperative	Transporting Quan
Joseph, Quan	21/7/2022	Coordinate airport pickup, meet at large AKO store.	Polite, cooperative	Transporting Quan
Joseph, Nerijus	13-18/7/2022	Boxing gala tickets, Gervonta Davis vs. Artur Beterbiev.	Enthusiastic, friendly	Shared interest in boxing
Joseph, Nerijus	11-20/7/2022	Airport pickup, Lizz's knowledge concern, Nerijus handles Lizz.	Focused, serious	Leadership, supporting role
Joseph, Liss	4-22/7/2022	Package delivery, police stop, car on watchlist, divorce papers, Joseph arrested.	Negative, confrontational	Caring partner, responsible
Nerijus, Antonio	5-25/7/2022	Business deals, UK lab inspection, club payments, Patek watch, quality inspection issue.	Positive, professional	Business-focused
Nerijus, Joseph	11-20/7/2022	Airport pickup, Lizz concern, address in IJmuiden.	Neutral, concerned	Casual, coordinating
Nerijus, Rosie	4-22/7/2022	Urgent contact, dinner, new car, transport routes, trouble at port, Liss spotted, Graphene OS.	Mixed, casual	Personal, business
Nerijus, Liss	20/7/2022	Warning not to contact police, defiance from Liss.	Negative, tense	Threats, defiance
Nerijus, Quan	7/7/2022	Expected arrival on July 21, Joseph will pick up.	Neutral	Information exchange

**Table 2**  
Summary of conversations

The conversation appears to be cooperative and efficient. Both participants seem to be on good terms and focused on getting tasks done.

Characterization:  
Quan seems to be the one seeking information and guidance, while Figo appears to be the one providing instructions and assistance. Figo's role in this conversation suggests that they could be a leader or coordinator in the group.

Next we present the remaining 11 chat conversations one by one, but in the same ChatGPT session. After 8 conversations we had to remind ChatGPT of the format by prefixing the following prompt on chat conversation number 9:

Summarize the following conversation including names of persons communicating, date of first message, date of last message and a summary including a sentiment analysis and characterization of the conversation.

We concatenate the summaries for all chat conversations in a single text that we refer to as the combined summaries. Using the combined summaries, we perform three more experiments that illustrate how ChatGPT can be used to further evaluate and visualise chat conversations.

### 5.2.1. List the persons and describe their roles

In the first experiment we ask ChatGPT who are involved in the conversations and about their role. We create a prompt by prefixing the combined summaries with the following sentence:

Here is a document containing multiple conversations. Create a table with a row for each person involved in these conversations and describe their role.

ChatGPT responds with the information in a tabular layout. Since ChatGPT "understands" how to produce  $\text{\LaTeX}$  code, we asked it to present the table in  $\text{\LaTeX}$  by prompting "Can you format this as a Latex table?". This results in Table 1.

### 5.2.2. Summarize the conversations in a table

In the next experiment, we ask ChatGPT to create a  $\text{\LaTeX}$  table that summarizes the chat conversations. We do this by prefixing the combined summaries (same as before) with the following instruction and using that as a prompt:

Here is a document containing multiple conversations. Create a Latex table showing participants, dates, summary in max 15 words, sentiment in max 5 words and characterization in max 5 words.

The result is presented in Table 2. ChatGPT shortened the summaries according to the instructions. Overall, the entries in the table are accurate and informative for someone that wants to get an overview of what the conversations are about.

### 5.2.3. Visualize interactions between persons

As a final experiment, we ask ChatGPT to produce a diagram depicting the social network of the group in TikZ (a language for creating graphics in  $\text{\LaTeX}$ ). This idea was inspired by a paper from Microsoft Research [33] in which many early experiments with GPT-4 are reported.

We split this experiment into two smaller tasks so that we can keep more control. In the first task, we present

the following prompt plus the combined summaries to ChatGPT:

Here is a document containing multiple conversations. List all names as nodes. List conversations as connections between nodes. Label connections with the topic of the conversation including an emoticon to reflect the sentiment of conversation. Avoid duplicate topics.

ChatGPT presents two tables. The first table contains the nodes, see Table 3.

Nodes	Names
1.	Quan
2.	Figo
3.	Joseph
4.	Sheila
5.	Nerijus
6.	Liss
7.	Antonio
8.	Rosie

**Table 3**  
List of names as nodes

Participants	Topic
Quan-Figo	Meeting up and sharing information 😊
Quan-Joseph	Coordinating airport pick-up 🚗
Quan-Sheila	Casual conversation and socializing 😊
Joseph-Nerijus	Boxing gala tickets and shared interest 🎟
Joseph-Nerijus	Picking up Quan and handling Lizz situation 😞
Joseph-Liss	Relationship conflict and divorce 💔
Nerijus-Antonio	Business arrangements and transactions 📄
Nerijus-Joseph	Coordinating Quan's pick-up and discussing Lizz issue 🗣
Nerijus-Rosie	Personal and business-related discussions 😊
Nerijus-Liss	Threats and defiance ⚠
Nerijus-Quan	Arrival and pick-up arrangements 🚗

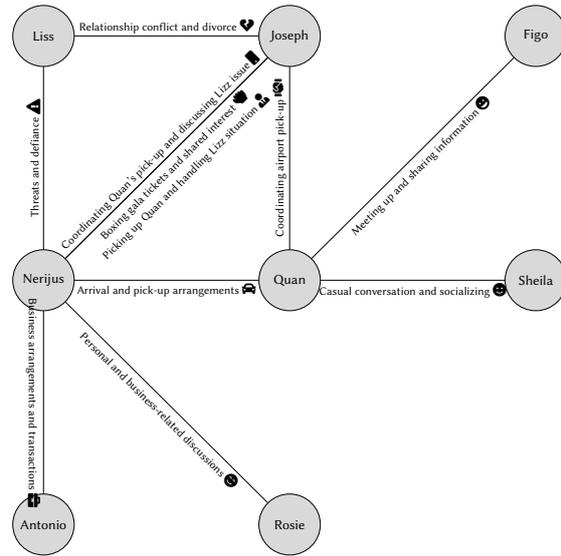
**Table 4**  
List of conversations as connections

The other table containing connections is presented in Table 4. We note that ChatGPT managed to provide emoticons for the sentiment description that is included in the combined summaries. We then instruct ChatGPT to use fontawesome5 to translate the emoticons in the HTML to  $\LaTeX$  codes. Next we prompt ChatGPT with the following text concatenated with the HTML versions of Tables 3 and 4:

Format the following information using TikZ in a Latex figure. Use the fontawesome5 package for the emoticons. Use the

connection text as labels that run along the edges. Use fixed size small circles with gray color and black contour to depict the nodes. Can you define a separate style for the edges and reference this in the draw sections. Use a smaller font for the edge labels. Arrange the nodes so that edges are not crossing.

The code for the TikZ diagram that is produced by ChatGPT is illustrated in Figure 4 below. Note that to get this particular layout it was necessary to provide additional prompts to get the end result.



**Figure 4:** Network of individuals involved in illegal activities based on chat conversations

### 5.3. Analysing search results

If ChatGPT can assist with writing structured queries and is able to summarize, evaluate and visualize chat messages, the next interesting question is if ChatGPT can analyse results from multiple search queries across different types of evidence. In this experiment we ask ChatGPT to assist us with analysing and cross referencing search results containing browser history, chats and locations from a single device.

We start with introducing the Crystal Clear case to ChatGPT. The prompt also instructs ChatGPT it should act as a sleuth and asks to analyse browser history url's containing the word 'search' that are found on the phone from Quan (HQL 'type:browserHistory search'). Quan is one of the suspects in this case. We only include the first three browser history entries but in total there are 50 unique url's.

You are SleuthGPT. You are a detective that is investigating a case. Quan and Joseph were arrested on July 24, 2022 because they were carrying a large sum of cash money. Quan's Motorola

phone and Josephs iPhone 6s were seized as well as the MacBook from Joseph. These devices have been loaded in Hansken and digital traces have been extracted. Here is the browserhistory from Quan's phone. Can you summarize his internet activity including dates and try to understand his motive and possible activities related to money he was carrying with Joseph?

```
"browserHistory.pageTitle", "browserHistory.accessedOn", "
browserHistory.application", "browserHistory.visitCount", "
browserHistoryLog.application", "total duplicates", "url.host"
"whatsapp group gambling - Google Zoeken", "2022-07-04T13
:29:18.598Z", "Google Chrome", "4", "", "1", "www.google.com"
"google - Google Zoeken", "2022-07-04T13:29:14.057Z", "Google
Chrome", "3", "", "1", "www.google.com"
"whatsapp group fitness - Google Zoeken", "2022-07-04T13
:29:10.843Z", "Google Chrome", "3", "", "1", "www.google.nl"
...
```

ChatGPT replies with the following analysis:

Based on Quan's browser history, here's a summary of his internet activity:

1. On July 4, 2022, Quan searched for "WhatsApp group gambling" and "WhatsApp group fitness."
2. On July 5, 2022, he searched for "uu.nl blackboard," "risks drug smuggling," "sentence checker," and "drug trafficking punishment UK."
3. On July 9, 2022, Quan looked up information about Class B drugs, London City Airport, flights from Amsterdam to London, and Saowapa, a restaurant in Utrecht. He also searched for "friends finder Utrecht," "computing science uu," and "gambling addiction."
4. On July 20, 2022, he searched for "transpofit for London."

Based on this information, it seems that Quan might be involved in gambling activities, as he searched for a gambling WhatsApp group and information about gambling addiction. He also showed interest in drug smuggling and trafficking, possibly researching the risks and potential punishments involved. Quan's searches about London City Airport, flights from Amsterdam to London, and London's transportation system indicate that he could have been planning a trip to London, which may be related to the large sum of cash he was carrying with Joseph when they were arrested.

We note that ChatGPT "understands" the dates in the input and automatically groups searches per date. At the end ChatGPT tries to infer how these searches may be related to the large sum of cash he was carrying.

Next we take chat conversations that were extracted from Quan's phone and present these with the following prompt (only listing the first three lines of the first conversion, there are 3 conversations in total).

Here is a summary of chats that we found on Quan's phone. Does this give you a better idea of what Quan has been doing. Do you see a connection between the chats and the browser history both in topic as well as date times?

Conversation 1 between Quan (owner) and Figo (31644891202@.whatsapp.net):

```
5-7-2022 12:08:53 - 31644891202@.whatsapp.net: Hi Quan, do
you want to meet again or do you have enough information?
5-7-2022 12:09:22 - owner: i know enough!
5-7-2022 12:09:31 - 31644891202@.whatsapp.net: Excellent
...
```

ChatGPT replies as follows:

Based on the chats found on Quan's phone, it appears that Quan may have been involved in some potentially illegal activities. The conversations with Figo (31644891202@.whatsapp.net) mention buying a suitcase for a transfer, paying cash to leave no traceable receipt, and sending photos of drugs and a lab via Snapchat. These activities could be related to the browser history, particularly the searches on transferring money and drug production.

The conversations with Sheila (+31685197340) and Joseph (31647730075@.whatsapp.net) seem to be more casual and focused on social activities, such as going to McDonald's, visiting a club, and meeting family members in Germany. However, the conversation with Joseph does mention Quan's flight details and airport meet-ups, which could be relevant to the investigation.

The connection between the chats and the browser history is not only in the topics discussed but also in the dates. The conversation with Figo about the suitcase and drugs takes place on 5-7-2022, while the browser history shows searches related to transferring money and drug production on 6-7-2022. This suggests that Quan may have been researching these topics after discussing them with Figo.

From the previous subsection we know that ChatGPT can summarize chat messages. We now see that ChatGPT can also relate such chat messages to the topics and dates of Quan's browser history and is able to note that Quan searches on July 6th occurred after his conversation with Figo on July 5th.

Next we present ChatGPT with GPS location traces (HQL `gps.latlong:\*`). There are in total 20 pictures that contain GPS traces. The prompt looks as follows (showing only the first three lines):

We found the following gps locations on Quan's phone. Can you give a summary of the locations and dates Quan has been assuming that he took the pictures. Can you interpret the geo location and give a name in stead? Is there any relation to activities mentioned in the browser history and chats that you found earlier?

```
"name", "gps.createdOn", "gps.latlong", "gps.misc", "gpsLog.
application", "gps.application"
"IMG_20220715_072808635_HDR.jpg", "2022-07-15T07:28:09.274Z
", "+52.30876+004.76278", "", "", ""
"IMG_20220715_072847128_HDR.jpg", "2022-07-15T07:28:47.824Z
", "+52.30864+004.76151", "", "", ""
...
```

and the reply from ChatGPT is:

Based on the GPS coordinates found in the images, here is a summary of the locations and dates Quan has been:

1. 2022-07-15: Two locations close to each other in Amsterdam, Netherlands, near the intersection of Parnassusweg and Strawinskyaan.  
IMG\_20220715\_072808635\_HDR.jpg: 52.30876, 4.76278  
IMG\_20220715\_072847128\_HDR.jpg: 52.30864, 4.76151
2. 2022-07-21: Two locations in Amsterdam, Netherlands, near the intersection of Parnassusweg and Gustav Mahlerlaan.  
IMG\_20220721\_094819167\_HDR.jpg: 52.30436, 4.75076  
IMG\_20220721\_092831675.jpg: 52.30995, 4.76099
3. 2022-07-09: One location in Amstelveen, Netherlands, near the intersection of Keizer Karelweg and Graaf Aelbrechtlaan.  
IMG\_20220709\_132100583\_HDR.jpg: 52.25898, 4.55369  
IMG\_20220709\_132021308\_HDR.jpg: 52.25897, 4.55371  
IMG\_20220709\_132050518\_HDR.jpg: 52.25898, 4.55369  
IMG\_20220709\_132058716\_HDR.jpg: 52.25898, 4.55369

These locations do not show a direct connection to the browser history and chats, but they provide a general idea of Quan's movements during that time period. The images were taken in Amsterdam and Amstelveen, which are in the Netherlands. The chat with Sheila (+31685197340) on 7-7-2022 mentioned going to Club Thouma, which could be located in or around Amsterdam. The conversation with Joseph (31647730075@.whatsapp.net) on 21-7-2022 mentioned meeting at the airport, which could be Amsterdam Airport Schiphol (also in the Netherlands). However, these connections are speculative and would require further investigation to confirm.

Interestingly, ChatGPT “understands” that .jpg refers to pictures, can translate GPS locations to places and associates with Amsterdam Airport Schiphol while the name Schiphol is not mentioned in the prompts that we provided. ChatGPT knows that Joseph and Quan talked about meeting at the airport and suggests this could be Schiphol because pictures have been taken in Amstelveen which is close to Schiphol. ChatGPT also states that this connection is speculative and does require further investigation.

Finally we ask ChatGPT to analyse the browser history, chats, locations including date times as a whole with the following prompt. Note that we added the fact that Quan and Joseph were arrested and carried a large sum of cash money. We did this to keep ChatGPT focused on the investigation:

```
Can you analyse the date times, locations, chats and browser history and suggest any relations between them? As a sleutgpt, what do these facts tell you about suspect Quan and if he has been involved in illicit activities? Remember that Quan and Joseph were arrested on July 24, 2022 because they were carrying a large sum of cash money.
```

This results in the following response from ChatGPT:

```
Based on the available information from the date times, locations, chats, and browser history, we can attempt to establish some possible connections and speculate about Quan's potential involvement in illicit activities. However, these connections are not definitive proof and would require further investigation.

1. On July 7, 2022, Quan had a chat with Sheila discussing going to Club Thouma. The location of this club is not specified, but it could be in or around Amsterdam, considering Quan's GPS coordinates from July 9, 15, and 21 are all in the Amsterdam area. However, there is no direct evidence connecting Quan's visit to the club with any illicit activities.

2. On July 21, 2022, Quan had a conversation with Joseph about meeting at the airport. Quan's GPS coordinates from the same day (IMG_20220721_094819167_HDR.jpg and IMG_20220721_092831675.jpg) indicate that he was in Amsterdam, not far from Amsterdam Airport Schiphol. The fact that Quan and Joseph were arrested on July 24, 2022, with a large sum of cash could suggest that they were involved in some form of illicit activity, possibly related to their meeting at the airport.

3. Quan's browser history includes searches for "how to transport large amounts of cash" and "money laundering techniques." These searches suggest that Quan might have been involved in or considering involvement in illicit activities related to handling large sums of money. The arrest on July 24, 2022, with a large sum of cash further supports this possibility.

4. The browser history also shows that Quan searched for "buying a burner phone." A burner phone is typically a cheap, disposable mobile phone often used for illicit activities to avoid being tracked. This search could indicate Quan's involvement in or intent to engage in illegal activities.

While the available data hints at the possibility of Quan being involved in illicit activities, it is important to note that these connections are speculative and not conclusive evidence. Further investigation would be required to determine if Quan and Joseph were engaged in illegal activities, and if so, the nature of those activities.
```

This response provides an accurate representation of the information that was presented to ChatGPT. It appears that ChatGPT is able to cross reference different types of

evidence which can be very helpful when searching for digital evidence.

## 6. Evaluation and Discussion

The experiments reflect ChatGPT’s potential in assisting investigators with digital evidence analysis. ChatGPT can interact with Hansken through a well defined and rich query language that is related to a digital trace model. However, limitations such as hallucinations, increased size requirements, and the need for larger context storage are present. Augmented Language Models (ALMs) offer a solution to these issues, improving LLMs’ performance in digital forensics tasks [34].

ALMs combine LLMs with additional components, such as external tools and reasoning strategies, to enhance their effectiveness [35]. They can access more relevant context or information, providing more accurate results in tasks like converting natural language queries into structured queries or summarizing chat conversations. ALMs can also use external tools to augment their current context, improving their performance in complex tasks like reverse engineering obfuscated JavaScript [36].

Autonomous agents like BabyAGI, Auto-GPT, and HuggingGPT leverage GPT-4’s capabilities, offering a new dimension for digital forensics investigations [37, 38, 39]. These agents use ChatGPT to create a strategic plan which they will execute autonomously (e.g., search the internet, and write/debug code, streamlining investigative tasks and allowing investigators to focus on more critical aspects of a case).

Visual foundation models (VFMs) represent a significant advancement in AI, integrating multimodal inputs (text and images) and generating text-based outputs [40, 14, 41]. GPT-4’s multimodal capabilities are not yet available in ChatGPT but once they are they will be a great addition for digital forensic tools, enhancing AI-assisted analysis of multi-modal data, including text, images, voice and video.

In summary, ALMs, autonomous agents, and GPT-4’s multimodal capabilities each hold significant potential for enhancing digital forensics. They can efficiently manage tasks, search for information, generate new agents, and incorporate external tools and reasoning strategies, providing invaluable support to digital forensic experts. However, in addition to technical experiments as described in this paper it will also be necessary to evaluate user experiences and establish objectively if investigators become more effective in real investigations that contain much more data and thus also much more irrelevant traces.

## 7. Conclusions

ChatGPTs' adaptability to the Hansken trace model and query language, strengthened by similar experiments in related work, suggest it can be applied to a wide range of databases and systems used in digital investigations. As the field evolves, integrating LLMs like GPT-4 and ChatGPT into digital forensic tools has the potential to enhance investigators' capabilities, ultimately contributing to more effective legal outcomes.

However, ChatGPT's fine-tuning to be "helpful, honest, and harmless" can limit its effectiveness when analyzing conversations between criminals. Additionally, providing manuals, trace model types, and HQL examples alone is insufficient; it is crucial to correct ChatGPT when it makes mistakes, sometimes requiring a trained digital forensic expert's input.

ChatGPT is not yet the ideal sleuth for digital evidence analysis, but LLMs can be invaluable. A SleuthGPT model could be trained by fine-tuning the original GPT-4 model on a dataset specifically focused on investigative and detective work, enabling it to generate more informed and accurate responses about investigative processes and techniques [14].

Future research will explore the synergistic application of these technologies in digital evidence investigations, such as providing examples to ChatGPT on how to interact with the Hansken REST API and automatically process the outcome of queries. By combining the strengths of ALMs, autonomous agents, and GPT-4's multimodal capabilities, we aim to develop a digital forensics copilot, which includes comprehensive AI-assisted tools capable of processing multi-modal data and addressing diverse challenges faced in investigating digital evidence.

## Acknowledgments

The authors thank Edwin Rijgersberg from the Forensic Big Data Analysis team of the NFI for his valuable comments on earlier versions of this paper.

## References

- [1] Y. Liu, T. Han, S. Ma, J. Zhang, Y. Yang, J. Tian, H. He, A. Li, M. He, Z. Liu, Z. Wu, D. Zhu, X. Li, N. Qiang, D. Shen, T. Liu, B. Ge, Summary of chatgpt/gpt-4 research and perspective towards the future of large language models, 2023. [arXiv:2304.01852](https://arxiv.org/abs/2304.01852).
- [2] E. M. Bender, T. Gebru, A. McMillan-Major, S. Shmitchell, On the dangers of stochastic parrots: Can language models be too big?, 2021, pp. 610–623. doi:10.1145/3442188.3445922.
- [3] M. Shanahan, Talking about large language models, 2023. [arXiv:2212.03551](https://arxiv.org/abs/2212.03551).
- [4] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, *Nature* 521 (2015) 436–444.
- [5] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, et al., Language models are few-shot learners, *Advances in Neural Information Processing Systems* 33 (2020).
- [6] G. Franceschelli, M. Musolesi, On the creativity of large language models, 2023. [arXiv:2304.00008](https://arxiv.org/abs/2304.00008).
- [7] J. Roach, Microsoft outlines framework for building ai apps and copilots, 2023. <https://news.microsoft.com/source/features/ai/microsoft-outlines-framework-for-building-ai-apps-and-copilots-expands-ai-plugin-ecosystem/> (Accessed on June 1, 2023).
- [8] P. Vaithilingam, T. Zhang, E. L. Glassman, Expectation vs. experience: Evaluating the usability of code generation tools powered by large language models, *CHI EA '22*, 2022, pp. 1–7. doi:10.1145/3491101.3519665.
- [9] B. Farrukh, Gpt-4: Openai's game-changing ai advancement, 2023. Available at: <https://medium.com/ai-in-plain-english/gpt-4-openais-game-changing-ai-advancement-18eed5a9aa49>, Accessed: 2023-04-14.
- [10] C. Lamanna, Introducing microsoft dynamics 365 copilot, 2023. Available at: <https://blogs.microsoft.com/blog/2023/03/06/introducing-microsoft-dynamics-365-copilot/>, Accessed: 2023-04-14.
- [11] W. Zhu, H. Liu, Q. Dong, J. Xu, L. Kong, J. Chen, L. Li, S. Huang, Multilingual machine translation with large language models: Empirical results and analysis, 2023. [arXiv:2304.04675](https://arxiv.org/abs/2304.04675).
- [12] T. Zhang, F. Ladhak, E. Durmus, P. Liang, K. McKeown, T. B. Hashimoto, Benchmarking large language models for news summarization, 2023. [arXiv:2301.13848](https://arxiv.org/abs/2301.13848).
- [13] M. Mohny, How chatgpt could impact law and legal services delivery, *Northwestern Engineering News* (2023). Available at: <https://www.mccormick.northwestern.edu/news/articles/2023/01/how-chatgpt-could-impact-law-and-legal-services-delivery/>. Accessed: 2023-05-01.
- [14] OpenAI, Gpt-4 technical report, 2023. [arXiv:2303.08774](https://arxiv.org/abs/2303.08774).
- [15] D. C. Weiss, Latest version of chatgpt aces bar exam with score nearing 90th percentile, *ABA Journal* (2023). Available at <https://www.abajournal.com/web/article/latest-version-of-chatgpt-aces-the-bar-exam-with-score-in-90th-percentile> Accessed: 2023-05-01.
- [16] H. Henseler, Network-based filtering for large email collections in e-discovery, *Artificial Intelligence and Law* 18 (2010) 413–430. doi:10.1007/s10506-010-9099-3.
- [17] D. van Dijk, D. Graus, Z. Ren, H. Henseler, M. de Ri-

- jke, Who is involved? semantic search for ediscovery, in: DESI VI Workshop, ICAIL 2015, 2015.
- [18] H. Henseler, J. Hyde, Technology assisted analysis of timeline and connections in digital forensic investigations, in: LegalAIIA Workshop, ICAIL 2019, volume 2484 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2019, pp. 32–37. URL: <https://ceur-ws.org/Vol-2484/>.
- [19] S. Alhojely, J. Kalita, Recent progress on text summarization, in: 2020 International Conference on Computational Science and Computational Intelligence (CSCI), 2020, pp. 1503–1509. doi:10.1109/CSCI51800.2020.00278.
- [20] C. Ma, W. E. Zhang, M. Guo, H. Wang, Q. Z. Sheng, Multi-document Summarization via Deep Learning Techniques: A Survey, *arXiv e-prints (2020)* arXiv:2011.04843. doi:10.48550/arXiv.2011.04843. arXiv:2011.04843.
- [21] J. Tredennick, D. W. Webber, What will ediscovery lawyers do after chatgpt?, *Legaltech News* (2023). Available online at <https://www.law.com/legaltechnews/2023/01/25/what-will-ediscovery-lawyers-do-after-chatgpt/>. Accessed on April 14, 2023.
- [22] B. Klimt, Y. Yang, Introducing the enron corpus., in: *CEAS*, volume 45, 2004, pp. 92–96.
- [23] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar, A. Rodriguez, A. Joulin, E. Grave, G. Lample, Llama: Open and efficient foundation language models, 2023. arXiv:2302.13971.
- [24] T. V. Team, Vicuna: An open-source chatbot impressing gpt-4 with 90%\* chatgpt quality, 2023. Available at: <https://vicuna.lmsys.org/> (Accessed: 2023-05-01).
- [25] R. Taori, I. Gulrajani, T. Zhang, Y. Dubois, X. Li, C. Guestrin, P. Liang, T. B. Hashimoto, Alpaca: A strong, replicable instruction-following model, 2023. Available at: <https://crfm.stanford.edu/2023/03/13/alpaca.html> (Accessed: May 1, 2022).
- [26] G. Katsogiannis-Meimarakis, G. Koutrika, A survey on deep learning approaches for text-to-sql, *The VLDB Journal* 32 (2023) 905–936. URL: <https://doi.org/10.1007/s00778-022-00776-8>. doi:10.1007/s00778-022-00776-8.
- [27] J. Mannelly, How to generate sql queries from text with gpt3, *Medium* (2022). Available at: <https://jman4190.medium.com/how-to-generate-sql-queries-from-text-with-gpt3-69ef7c44f47a>, Accessed: 2023-04-17.
- [28] H. Alexander, How to use chatgpt to write sql join queries, *Devart Blog* (2023). Available at: <https://blog.devart.com/how-to-use-chatgpt-to-write-sql-join-queries.html>, Accessed: 2023-04-17.
- [29] S. Huang, Using gpt-3 to transform english natural language questions to cypher queries in neo4j graph database, 2022. Available at: <https://towardsdatascience.com/using-gpt-3-to-transform-english-natural-language-questions-to-cypher-queries-in-neo4j-graph>, Accessed: 2023-04-14.
- [30] BBC News, Scrap 'digital strip search' say police bosses, 2019. Available at: <https://www.bbc.com/news/uk-48166182>. Accessed: 2023-04-25.
- [31] National Police Chiefs' Council (NPCC), Digital Forensic Science Strategy, Technical Report, National Police Chiefs' Council, 2020. Available at: <https://www.npcc.police.uk/our-work/Digital-policing-and-technology/>. Accessed: 2023-05-01.
- [32] H. van Beek, E. van Eijk, R. van Baar, M. Ugen, J. Bodde, A. Siemelink, Digital forensics as a service: Game on, *Digital Investigation* 15 (2015) 20–38. doi:10.1016/j.diin.2015.07.004, special Issue: Big Data and Intelligent Data Analysis.
- [33] S. Bubeck, V. Chandrasekaran, R. Eldan, J. Gehrke, E. Horvitz, E. Kamar, P. Lee, Y. T. Lee, Y. Li, S. Lundberg, H. Nori, H. Palangi, M. T. Ribeiro, Y. Zhang, Sparks of artificial general intelligence: Early experiments with gpt-4, 2023. arXiv:2303.12712.
- [34] G. Mialon, R. Dessi, M. Lomeli, C. Nalmpantis, R. Pasunuru, R. Raileanu, B. Rozière, T. Schick, J. Dwivedi-Yu, A. Celikyilmaz, E. Grave, Y. LeCun, T. Scialom, Augmented language models: a survey, 2023. arXiv:2302.07842.
- [35] T. Schick, J. Dwivedi-Yu, R. Dessi, R. Raileanu, M. Lomeli, L. Zettlemoyer, N. Cancedda, T. Scialom, Toolformer: Language models can teach themselves to use tools, 2023. arXiv:2302.04761.
- [36] G. Brockman, The inside story of chatgpt's astonishing potential, 2023. URL: [https://www.ted.com/talks/greg\\_brockman\\_the\\_inside\\_story\\_of\\_chatgpt\\_s\\_astonishing\\_potential](https://www.ted.com/talks/greg_brockman_the_inside_story_of_chatgpt_s_astonishing_potential), accessed: 2023-04-26.
- [37] Significant-Gravitas, Auto-gpt: An autonomous gpt-4 experiment, 2023. Available at: <https://github.com/Significant-Gravitas/Auto-GPT>, Accessed: 2023-04-26.
- [38] Y. Nakajima, Babyagi: A pared-down version of the original task-driven autonomous agent, 2023. Available at: <https://github.com/yoheinakajima/babyagi>, Accessed: 2023-04-26.
- [39] Y. Shen, K. Song, X. Tan, D. Li, W. Lu, Y. Zhuang, Hugginggpt: Solving ai tasks with chatgpt and its friends in huggingface, 2023. arXiv:2303.17580.
- [40] C. Wu, S. Yin, W. Qi, X. Wang, Z. Tang, N. Duan, Visual chatgpt: Talking, drawing and editing with visual foundation models, 2023. arXiv:2303.04671.
- [41] D. Zhu, J. Chen, X. Shen, X. Li, M. Elhoseiny, Minigpt-4: Enhancing vision-language understanding with advanced large language models, 2023. arXiv:2304.10592.