OSINT Clinic: Co-designing AI-Augmented Collaborative OSINT Investigations for Vulnerability Assessment

Anirban Mukhopadhyay
Department of Computer Science
Virginia Tech
Blacksburg, Virginia, USA
anirban@vt.edu

Kurt Luther
Department of Computer Science
Virginia Tech
Arlington, Virginia, USA
kluther@vt.edu

Abstract

Small businesses need vulnerability assessments to identify and mitigate cyber risks. Cybersecurity clinics provide a solution by offering students hands-on experience while delivering free vulnerability assessments to local organizations. To scale this model, we propose an Open Source Intelligence (OSINT) clinic where students conduct assessments using only publicly available data. We enhance the quality of investigations in the OSINT clinic by addressing the technical and collaborative challenges. Over the duration of the 2023-24 academic year, we conducted a three-phase co-design study with six students. Our study identified key challenges in the OSINT investigations and explored how generative AI could address these performance gaps. We developed design ideas for effective AI integration based on the use of AI probes and collaboration platform features. A pilot with three small businesses highlighted both the practical benefits of AI in streamlining investigations, and limitations, including privacy concerns and difficulty in monitoring progress.

CCS Concepts

• **Human-centered computing** → **Collaborative and social computing systems and tools**; **Empirical studies in HCI**.

Keywords

OSINT, Cybersecurity Vulnerability Assessment, Co-Design, Matchmaking, Generative AI, Collaborative AI Platform

ACM Reference Format:

Anirban Mukhopadhyay and Kurt Luther. 2025. OSINT Clinic: Co-designing AI-Augmented Collaborative OSINT Investigations for Vulnerability Assessment. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 22 pages. https://doi.org/10.1145/3706598.3713283

1 Introduction

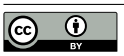
The 2019 Verizon Data Breach Investigations Report highlights that small businesses are prime targets for cybercriminals, accounting for 43% of all breaches [5]. However, many small businesses are unaware of the cyber risks they face or how to mitigate them. Cyber vulnerability assessments can provide a view of their attack surface to address weaknesses before malicious actors can exploit them. Comprehensive vulnerability assessments include tasks like controls interviews, physical facility tours, and internal network scanning [1]. However, these require technical expertise and time to execute. Even if the small business has an IT staff member, they are often overwhelmed with tasks and lack specific training in cybersecurity.

Cybersecurity clinics offer a potential solution [11]. They provide essential cybersecurity services to community organizations — including small businesses, nonprofits, and local governments — while giving students real-world experience. Modeled after legal and medical school clinics, these programs are typically housed at colleges and universities under the direction of clinical professors [21]. Students from diverse backgrounds and disciplines receive training to offer free vulnerability assessments and cybersecurity assistance to clients who may lack the resources to secure these services otherwise [11]. This model not only offers a valuable skills-based learning environment for students but also strengthens the cybersecurity resilience of communities. The cybersecurity clinic model has recently seen exponential growth thanks to support from private charities, government, and industry [69].

However, these clinics also face scalability challenges. Training a diverse group of students to conduct high-quality vulnerability assessments is time-consuming and resource-intensive. Even for trained students, a single investigation can take months. Further, some components of vulnerability assessments, such as controls interviews and facility tours, require significant time investments from clients.

In this paper, we propose the concept of an *OSINT Clinic*, where students can learn and apply OSINT skills in a real-world context to address the scalability challenges of cybersecurity clinics. Open Source Intelligence, or OSINT, refers to investigations that rely entirely on publicly available data. In a cybersecurity context, OSINT tasks include attack surface mapping for identifying all publicly accessible digital assets such as websites, servers, webcams, and social media accounts, as well as their vulnerabilities; brand monitoring for examining review sites and social media for mentions of the businesses, including negative rumors and disinformation campaigns; and breach data discovery for finding leaked credentials on the dark web or other sources [34, 58, 91]. While comprehensive vulnerability assessments sometimes include an OSINT component, cyber clinics rarely, if ever, focus on OSINT.

OSINT provides unique advantages for a student team doing vulnerability assessments. Dealing with publicly available data is a lower risk for both the students and the target businesses; this passive, "zero-touch" approach minimizes accidental damage and



CHI '25, Yokohama, Japan

ACM ISBN 979-8-4007-1394-1/25/04
https://doi.org/10.1145/3706598.3713283

saves time for business owners. Second, the ethos of the field dictates that the process for finding, verifying, and documenting each piece of evidence must be transparent and repeatable, providing a rich learning environment for students. Despite the growing interest and promise of OSINT investigations [46, 58, 72, 81, 98], only a few studies have taken a human-centered approach to investigate specific challenges involved in applying these techniques [81]. We address this gap by focusing on identifying and addressing the technical and collaborative challenges with OSINT tasks for vulnerability assessments in the clinic.

OSINT investigations for cybersecurity are complex, requiring sifting through vast amounts of unstructured public data, which can be noisy, overwhelming, and time-consuming [29, 53, 90, 91]. Analysts often face difficulties in data verification due to the unreliable nature of sources and misinformation [29, 56]. We argue that generative AI can be effective in addressing these challenges based on previous research on its role in narrowing the skill gap in creative and knowledge work [44, 89, 104]. Examples from OSINT practitioners and trainers like Matt Edmondson and Chris Poulter [48, 79] also back this up. However, reflecting Ackerman's "socio-technical gap" [23], the technical capabilities of AI currently lag behind the nuanced social requirements of trust and confidentiality for sensitive tasks like OSINT. Following a growing thread of HCI and CSCW research, our study aims to effectively integrate generative AI to augment existing workflows — in this case, those of an OSINT clinic — and help them scale up.

This research is driven by the following guiding questions:

(1) How can we scale up cybersecurity vulnerability assessments using OSINT?
(2) What challenges arise when performing vulnerability assessments in the OSINT clinic?
(3) How can generative AI, particularly large language models like ChatGPT, address the technical and collaboration challenges of OSINT investigations among non-experts?
(4) How can generative AI support real-world OSINT investigations for vulnerability assessments?

To explore these questions, we designed and launched an OSINT clinic with six undergraduate students. While the small number of participants may limit generalizability, it allowed us to closely examine the development of OSINT skills in a real-world investigative context. Through regular practice sessions and workshops, we trained the students to build their OSINT expertise in preparation for real-world investigations. Over the 2023–24 academic year, we conducted a longitudinal co-design study by adapting the Matchmaking for AI method [70] with three phases (Studies 1–3) to support the work of students in the clinic. Study 1 identified key challenges across the Intelligence Cycle phases through workshops, revealing performance gaps in student investigations. In Study 2, we utilized generative AI and collaboration platform probes to explore how AI could help novices address the challenges and develop expertise across the stages of planning, data collection, processing, analysis, and dissemination. We analyzed the user experience and workshopped design ideas to show how generative AI can support training and be integrated into workflows. Finally, Study 3 piloted OSINT-based vulnerability assessments with three small businesses, integrating curated AI prompts into real-world investigations. This study demonstrated AI's practical benefits and also revealed limitations, such as privacy concerns and difficulties in monitoring progress.

Together, these studies make the following contributions:

(1) We make a conceptual contribution by developing the OSINT clinic program to scale up cybersecurity vulnerability assessments based on publicly available information.
(2) We make a methodological contribution by extending the Matchmaking for AI co-design process to include learning goals, collaboration challenges, and real-world deployments.
(3) We make an empirical contribution by studying an OSINT clinic with real-world investigations and clients, identifying emergent challenges and learning goals, and subsequently exploring generative AI-driven design solutions.
(4) We develop design considerations for practical, AI-augmented, and collaborative investigations based on the use of generative AI.

2 Related Work

Our paper bridges two key areas of recent HCI research: Clinical Computer Security, which defines protocols for addressing targeted digital security threats (Section 2.1), and CSCW research focused on exploring generative AI tools to facilitate sensemaking and collaboration (Section 2.2). By situating our work at this intersection, we highlight how AI-augmented systems can empower novice investigators in navigating complex OSINT tasks, supporting both their learning goals and the effectiveness of real-world investigations. In terms of method, we extend the Matchmaking for AI co-design framework with a longitudinal study design (Section 2.3), focusing on collaboration and real-world validation. We situate the work within the context of cybersecurity vulnerability assessments using OSINT (Section 2.4).

2.1 Clinical Computer Security in HCI

Clinical Computer Security (CCS) [50] is an emerging approach in HCI that provides personalized, hands-on support to individuals facing technology-based security threats. CCS methods involve pairing clients with trained technologists who collaboratively investigate and address digital security vulnerabilities. This process is guided by the Understand-Investigate-Advise (UIA) framework, in which consultants conduct interviews to map a client's digital footprint, investigate potential security threats (e.g., spyware or compromised accounts), and offer tailored mitigation strategies [57]. CCS draws inspiration from existing support paradigms such as the Citizen Lab [6] and Cybersecurity Clinics [21], which focus on assisting victims of nation-state digital attacks. These clinics not only provide critical services but also serve as research sites for developing new security practices. For example, Tseng et al. co-designed data stewardship approaches by collaborating with consultants and clients in CCS settings [95].

The OSINT Clinic builds on CCS research by applying its foundational principles — such as personalized, collaborative investigations and the UIA framework — to cybersecurity vulnerability assessments for small businesses using OSINT. While CCS has primarily focused on intimate partner violence (IPV) contexts [50, 57, 96],

the OSINT Clinic demonstrates how these principles can be extended to broader cybersecurity challenges. Beyond tackling a new domain, our work explores the HCI and CSCW challenges of leveraging generative AI to support clinical work. We propose a sociotechnical system where AI not only assists with technical tasks but also fosters learning, coordination, and leadership among student analysts.

OSINT investigations face numerous challenges that are amplified in the context of clinical work, where student teams analyze public information. Investigators often work with limited resources such as funds, time, and access to specialized software [66, 75]. Issues of data quality, quantity, integration, and ethical considerations further complicate the process [54]. Mukhopadhyay et al. found training on technical skills and ethics was essential to get novices to collaborate successfully with OSINT experts on real-world investigations [73]. However, comprehensive training is difficult to scale. Effective leadership in OSINT is also challenging as it requires navigating legal and ethical boundaries while applying investigative techniques across various domains [24]. In this paper, we explore how generative AI can help scale this clinic model by understanding and addressing these challenges.

2.2 CSCW Research and Generative AI Opportunities in OSINT

OSINT investigations are inherently complex, requiring the collection and analysis of large volumes of data as part of a sensemaking process to extract meaningful insights [26, 39, 102]. Collaboration can enhance these efforts by dividing up workloads and integrating diverse perspectives [47]. Previous HCI and CSCW research has studied various approaches to collaborative and crowdsourced investigative work, including top-down strategies [25], bottom-up explorations [61], and hybrid methods [100]. Belghith et al. studied "social OSINT" — i.e., the growing landscape of competitive and cooperative OSINT-themed events — and characterized it as a community of practice with a culture of transparency and collaboration [29]. This model with overlapping goals of training and delivering real-world value also closely relates to growing HCI research on how trained and well-supported crowds can handle complex investigations [73, 100, 101]. Existing systems like Newstrition [3] and CAPER [27] facilitate expert collaboration in specific contexts like information verification and prevention of organized crime, but there is potential to expand to non-experts through the use of AI tools [42, 73].

Generative AI presents a promising solution for augmenting OSINT investigations, particularly when compared to custom-built AI solutions [34]. Current AI and machine learning tools in OSINT face challenges due to the lack of focus on real-world applications and inadequate integration with existing intelligence tools [35, 65, 111]. Moreover, AI solutions fail to address key aspects of intelligence work, such as planning and collaboration [34, 43]. Following the definition by Sarkar et al., we define generative AI as "an end-user tool whose technical implementation includes a generative model based on deep learning" [83]. Unlike specialized, narrowly focused AI tools, LLMs are general-purpose and capable of performing a range of complex tasks, such as reasoning, summarization, and web searches, without the need for extensive fine-tuning. Their flexibility makes them well-suited for dynamic and high-uncertainty contexts like OSINT investigations. Prior studies, such as PentestGPT, have demonstrated the effectiveness of LLMs in cybersecurity applications, including interpreting tool outputs and suggesting next steps during penetration testing [40].

Generative AI also offers advantages in training and skill development for novice OSINT analysts [37, 87]. By supporting iterative learning processes through demonstrations, feedback, and self-evaluation [45, 73], LLMs can help bridge knowledge gaps in areas like tool usage, data collection, and verification. Unlike custom AI solutions, which often require specialized technical expertise to design and deploy, generative AI tools are accessible to a broader audience and can be seamlessly integrated into existing workflows.

Research has begun exploring how generative AI, like LLMs, can enhance collaboration by serving as active participants in group settings [59, 114]. Shifting from single human-AI interactions to multi-human and AI environments could improve performance in complex scenarios, providing more context-aware collaboration [28, 59, 74]. Generative AI can facilitate dynamic teamwork and decision-making, improving collaborative sensemaking in high-uncertainty situations [43, 59, 89]. Based on related work, we posit that generative AI can significantly improve the effectiveness and scalability of OSINT investigations across the dimensions of practical integration, training, and collaboration.

2.3 Co-Designing AI for OSINT

Despite the growing importance of OSINT in cybersecurity, there is limited research on incorporating stakeholder values in developing OSINT technologies [60, 81]. Co-design methods applied to the design of AI tools enable stakeholders to brainstorm ideas where the design landscape is underexplored, translating high-level user needs into actionable AI features [82]. However, a gap remains in applying co-design methodologies with OSINT investigators to address their specific challenges and needs.

Current AI design paradigms often struggle to identify innovative and feasible design concepts that effectively integrate human values. The Matchmaking for AI approach, introduced by Liu et al. [70], addresses this by borrowing from the earlier, more general Matchmaking method [31, 99]. It maps AI techniques to user activities using probes, fostering mutual learning between stakeholders and AI researchers. This method extends traditional co-design practices by focusing on practical feasibility and aligning existing AI capabilities with domain-specific tasks.

We extend the Matchmaking for AI method by incorporating a longitudinal aspect, allowing us to study the use of generative AI tools over an extended period and across multiple phases of co-design. This enables AI solutions to be practically integrated into workflows and adaptable to evolving team dynamics [38, 85]. By embedding the method within a real-world context, we validated design ideas through longitudinal observation of participants conducting vulnerability assessments with actual clients.

2.4 Cybersecurity Vulnerability Assessment using OSINT

The study deals with tasks related to cybersecurity vulnerability assessment, which is a systematic process of identifying, analyzing, and prioritizing potential security weaknesses in an organization. These assessments can have many components, including internal network discovery and vulnerability scanning with tools like Nmap [17] and Nessus [16]; security controls interviews based on frameworks like CIS [9]; and physical security reviews looking for access control, fire suppression, and water detection [1]. Passive information-gathering techniques in OSINT achieve a subset of these tasks to assess cybersecurity vulnerabilities without direct interaction with the target. Prior work informs our use of OSINT tools and techniques relevant to vulnerability assessment [34, 63, 75].

Urban et al. found that over 83% of the analyzed companies leak enough employee attributes to enable highly sophisticated phishing attacks [98]. From a defense perspective, organizations can leverage OSINT techniques to assess their attack surfaces or identify social engineering opportunities [46, 58, 77]. As the use of social media and other online resources has grown, the availability of open-source information has expanded, making it a key tool for organizations to understand the threat landscape, identify the techniques and tools used by adversaries, and mitigate potential cyberattacks [77, 97, 107]. OSINT-based vulnerability assessments therefore shift the focus of cyber defense from reactive measures to proactive strategies [58, 91]. OSINT enables fast, real-time information collection, and clear sourcing, and is both convenient and cost-effective [52]. Despite its critical need during the reconnaissance phase, there is a lack of understanding of challenges involved in effective OSINT investigations for vulnerability assessments.

3 Study Design

3.1 Study Protocol

We adopted the Matchmaking for AI method [70] to conduct our co-design study. The method allows stakeholders — students in our case — to actively contribute to the design and development of effective tools for OSINT investigations. The iterative workshops and use of AI probes provide a structured yet flexible framework for identifying user needs and aligning them with existing AI capabilities. This method ensures that the solutions we develop are not only technically feasible but also usable in practical scenarios, particularly in the context of cybersecurity vulnerability assessments. We discuss our methodological contributions to the framework by supporting real-world investigations and collaboration among students in Section 7.1.

We ran the co-design study with a group of six undergraduate students during the 2023–24 academic year at our institution, a large public university in the US. Previous research has demonstrated that groups of student analysts can help characterize intelligence investigations [36, 66], effectively experiment with new crowdsourcing methods, and provide valuable process recommendations [108, 110]. Students have also securely engaged in real-world OSINT investigations, including cybersecurity vulnerability assessments and human rights investigations, as part of experiential learning [2, 4].

As a part of the clinic, the authors, who have relevant industry experience and certifications, trained students on different aspects of OSINT investigations like (1) practicing operational security (OPSEC) and being aware of privacy, legal, and ethical issues; (2) mining different public sources of information like social media for brand monitoring and employee profiles; (3) using satellite imagery for geolocating physical assets; (4) identifying network infrastructure for identifying open ports, IoT devices; (5) inspecting website and network infrastructure to identify vulnerabilities and unpatched systems; and (6) looking for breach data and leaked credentials. Table 1 shows the seven key areas for vulnerability assessments along with relevant OSINT tools and techniques.

Our longitudinal study spanned two semesters. Study 1 (Design Workshop 1) was conducted after three months of OSINT training. Right after, we ran Study 2 (Design workshops 2–6) over two months during the end of the first semester and the beginning of the second. Finally, Study 3 involved two focus group interviews conducted over a span of one month at the end of the second semester. The longitudinal design of this study was essential for us to capture the use of generative AI across different phases of the Intelligence Cycle. This extended engagement provided insights into how AI is integrated, how collaboration challenges develop, and how leadership roles evolve in response to AI support. Figure 1 represents the flow of the research study with the objective, data source(s), and probe(s) for each of Studies 1-3.

3.2 Participant Recruitment

This study was approved by our university's IRB. Six undergraduate students (P1–P6) were recruited based on their interest in gaining practical OSINT skills as part of a team focused on cybersecurity vulnerability assessments. The group represented a range of academic backgrounds, including majors in Cybersecurity, Computer Science, National Security, and Engineering. The participants' ages ranged from 18 to 22. Three participants identified as women, two as men, and one as non-binary. Some students had prior experience with OSINT-related activities, such as capture-the-flag (CTF) challenges, OSINT counter-disinformation projects, and research involving fake identity generation. Others were newer to the field but eager to learn.

The participants met twice a week for two-hour sessions to learn and practice OSINT skills. They were paid an hourly wage of $18 for a maximum of 10 hours per week to attend the OSINT clinic. The design workshops were integrated into the clinic to help them reflect and augment their workflows with generative AI.

4 Study 1: Identifying Challenges and Learning Goals with OSINT Tasks

4.1 Method

In this study, we sought to understand participant workflows related to the OSINT tasks in Table 1 during a recorded 2-hour long co-located workshop (design workshop 1). We adapted the Domain Speciality Canvas [70] from Step 1 of the Matchmaking for AI process and added another column for collaboration challenges to facilitate data collection (shown in Figure 4). Before this study, participants were trained on and practiced common OSINT tools and techniques mentioned in Table 1. They also worked in teams

Table 1: OSINT tasks in cybersecurity vulnerability assessments

Category	Task Description	OSINT Tools and Techniques
Network Discovery	Identify all digital assets (websites, email, servers, webcams) associated with the business.	DNS enumeration, IP mapping, Reverse IP lookup, WHOIS lookup [22], Shodan [20]
Website Vulnerabilities	Discover vulnerabilities on the business's website such as outdated software, insecure admin panels, open ports, files that should not be public, etc.	BuiltWith [7], CMS identification, port scanning, mail server scanner, Google dorking [12]
Physical Assets	Identify physical assets like buildings, properties, or vehicles from online resources or images.	Google Maps, Google Earth, Street View, Image metadata analysis, Geolocation techniques
Third-Party Services	Document third-party services (vendors, etc.) that the client does business with to identify possible supply chain vulnerabilities.	External service provider evaluation based on reviews, past breaches, and terms and conditions
Employee Public Information	Identify key employees from a company using social media platforms, gathering information like job titles, email addresses, or sensitive personal data accidentally shared online.	LinkedIn search; Advanced searches on social media platforms like Instagram, Twitter, YouTube, TikTok; Censys [8], other phone number and email verification tools
Brand Monitoring	Find instances of negative mentions, disinformation, inappropriate content, legal filings, and settlements related to the business.	Online reviews, social media monitoring, court case databases
Breach Data	Search for a company's leaked passwords and data in breach data dumps and on the dark web.	Have I Been Pwned [13], IntelX [14], other breach databases

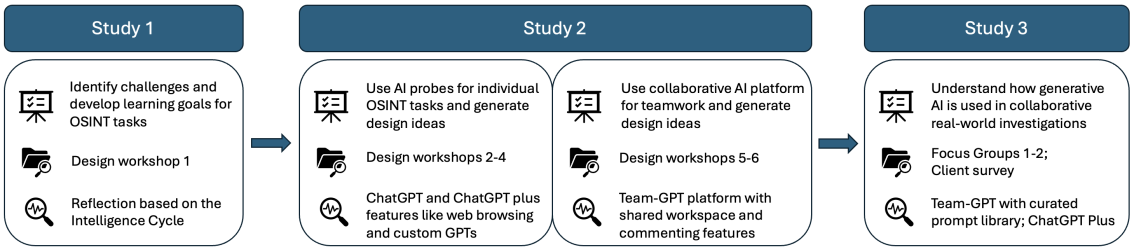


Figure 1: Study Diagram: (1) Study 1 was formative and aimed at identifying the key challenges in the Intelligence Cycle phases of OSINT investigations. We asked participants to reflect on previous investigations and document the steps and challenges faced during Design Workshop 1. (2) In Study 2, we introduced generative AI probes in Design Workshops 2–4 to explore how AI could address challenges experienced earlier. During the workshops, we identified areas where generative AI was effective and also its limitations. We focused on the emergent challenges of collaboration and leadership during Design Workshops 5 and 6, with a collaborative AI platform (Team-GPT) as a probe. (3) In Study 3, we piloted OSINT-based vulnerability assessments with three small businesses, integrating curated generative AI prompts into ongoing investigations. This real-world application highlighted the practical utility and surfaced challenges of using foundational generative AI models like ChatGPT in vulnerability assessments, especially in collaborative settings.

of three to investigate a few local businesses. To reflect on the work, we asked participants to discuss their most recent investigation, one successful team investigation, and one challenging team investigation. Corresponding to the normative workflow used in Step 1 by Liu et al. [70], we presented the Intelligence Cycle (IC) [34], widely used to structure investigations, to guide participants to add notes. We mapped out the different challenges on a digital whiteboard (FigJam) and asked follow-up questions to clarify participants' responses.

4.2 Analysis

The first author transcribed the recorded workshop and, in collaboration with the rest of the research team, conducted an inductive thematic analysis [32] of the transcript to identify the challenges. We aligned the identified challenges with the five phases of the OSINT Intelligence Model (OIM) presented by Hwang et al. [63] (see Figure 2), which closely follows the IC. Hwang et al's model

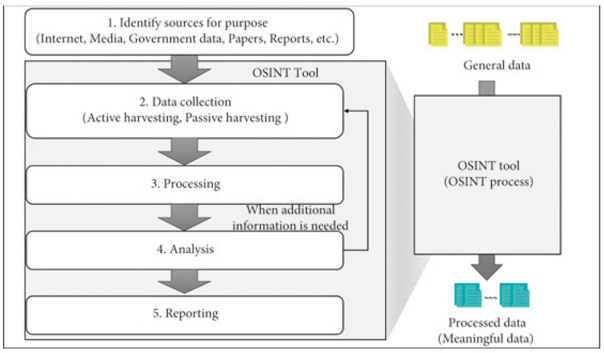


Figure 2: OSINT Intelligence Model in the context of cybersecurity as described by Hwang et al. [63]

focuses on continuous iteration and practical applications to cybersecurity [63]. We denote the first stage of "identifying sources for purpose" as "planning" to match the IC, as it entails the same task of figuring "where and how to get information."

4.3 Results

We identified major challenges and learning goals across the five phases of planning, collection, processing, analysis, and dissemination in OSINT investigations. In the following sections, we describe the challenges students faced while completing the tasks, and develop learning goals based on existing guidelines [73, 91] and our own experiences as OSINT researchers. Table 2 collates the challenges and learning goals across the five phases.

4.3.1 *Planning.*

Navigating the ethical and legal boundaries. Participants were trained on the ethics and safety of investigations, but they often faced dilemmas around the use of tools and techniques. In OSINT investigations, active techniques commonly used beyond the reconnaissance phase of penetration testing are typically avoided. P6 highlighted the need to understand the limitations and restrictions on using such tools. P1 thought investigations were complicated by the need to stay clearly within legal boundaries while investigating vulnerabilities, particularly when trying to gain access to leaked data. P6 also highlighted practical concerns such as private browsing on platforms like LinkedIn to avoid alerting the target of the investigation.

We argue that understanding which OSINT data, tools, and techniques are both private and trustworthy can help alleviate the challenge. Participants also need to implement strong operational security practices, such as creating and using sock puppet accounts, to protect both the investigation and the investigator. These two aspects can form a learning goal to address the ethical challenges.

Learning Goal 1 (LG1): Develop appropriate trust for OSINT data, tools, and techniques; Implement operational security.

Defining investigation objectives and overlapping work. Participants found it difficult to define clear objectives and scoping for the OSINT investigations, which sometimes led to inefficiencies and miscommunication. P1 emphasized the need for better procedural guidelines, stating, "... just more of an understanding of format and kind of a procedure that we should follow as to what things should look like". There were instances where a lack of communication led to redundant efforts, and P1 found multiple members focusing on a specific platform reduced efficiency. P3 and P4 reported that working on a shared document to highlight information collected wasn't effective in minimizing overlap.

These insights suggest that clearer division of work and structured guidelines could enhance the effectiveness of team-based OSINT investigations. A learning goal could be to develop the ability to effectively break down investigations into specific subtasks. This includes identifying relevant data sources and facilitating a structured approach for task division.

Learning Goal 2 (LG2): Break down investigations into subtasks with data sources.

4.3.2 *Collection.*

Skill gaps in data collection. Participants faced significant challenges in identifying effective techniques for data collection, compounded by varying skill levels among team members in using advanced OSINT tools and platforms. Many struggled with the technical aspects, such as understanding code, creating effective queries for search, and navigating tools like the Linux command line interface, which are not always user-friendly for beginners. P2 highlighted the steep learning curve: "I don't really know coding languages, so I'm trying to mess with an inspect element, and it just looks like a bunch of text to me." Participants also struggled to identify the right OSINT tool(s) for the job among potentially dozens of possibilities. P3 and P4 mentioned the difficulty of staying updated with changing techniques.

Participants need to select appropriate tools among the many available options to improve performance. Finding alternate sources of information becomes essential when stuck.

Learning Goal 3 (LG3): Identify and utilize the most effective OSINT tools.

Lack of clarity in information presentation. Participants mentioned difficulties in collecting and organizing data due to the lack of a clear structure and formatting for tracking different intelligence sources. Inconsistencies in format and overall disorganization of documents where they collected data often hindered the utility of the gathered information. For example the "data dumps" or raw results from tools made it hard for other members of the team to parse the information. As P6 put it, "It's been told we have to interpret the information, analyze it, and then put it in a format to be understandable. But I often notice it comes out very variable, just because some people put it in some ways, or it has to be very messy in the [working] document." P1 emphasized the need for a clearer structure, stating, "Just more of an understanding of format and kind of a procedure that we should follow as to what things should look like".

Implementing templates can significantly improve the organization and clarity of OSINT investigations. Templates provide a structured approach to document and cover multiple sources of information, ensuring consistency and comprehensiveness.

Table 2: Challenges and Learning Goals across phases of OSINT investigations

OSINT Cycle Step	Challenges	Learning Goals
Planning	Navigating ethical and legal boundaries	Develop appropriate trust on OSINT data, tools, and techniques; Implement operational security
	Defining investigation objectives and overlapping work	Breakdown investigations into subtasks with data sources
Collection	Skill gaps in data collection	Identify and utilize the most effective OSINT tools
	Lack of clarity in information presentation	Develop and utilize standardized templates to systematically collect and organize information
Processing	Handling high volume of raw information	Structure raw information into parsable format and improve documentation
	Verifying public information	Find multiple sources for OSINT verification
Analysis	Identifying vulnerabilities	Interpret results from tools and identify vulnerabilities
	Developing recommendations from vulnerabilities identified	Generate actionable recommendations
Dissemination	Not knowing "what a good end product will look like"	Iterate on recommendations to tailor it to business needs

Learning Goal 4 (LG4): Develop and utilize standardized templates to systematically collect and organize information.

4.3.3 *Processing.*

Handling high volume of raw information. The sheer volume of unstructured information often made it difficult for participants to identify and focus on the most pertinent details. The team spent a lot of time summarizing, sorting, and categorizing information and eliminating irrelevant data. Additionally, there was a need for better documentation practices, like P5 wanted to, "…document better how I found certain information so somebody else can follow the steps I took". The effectiveness of collaboration was often hindered when team members failed to properly document their work or explain their processes to others, leading to potential oversights like duplication of work or inaccurate findings.

To tackle the overwhelming amount of data, participants emphasized the importance of identifying relevant details that can be verified and used for further analysis. P4 thought, "AI could be very useful, especially if you have a large data set that you wanna maybe format in the best way". Improved documentation could help with getting other members to leverage the collected information.

Learning Goal 5 (LG5): Structure raw information into parsable format and improve documentation.

Verifying public information. A key challenge for participants was accurately verifying information, especially in relation to smaller companies where multiple sources were harder to find. Participants wanted to follow best practices and ensure the accuracy of important details by cross-referencing multiple sources. P2 highlighted the importance of "getting at least two sources confirming something," such as verifying an employee's identity. This process often involved checking social media profiles and company websites or finding official state records to double-check the accuracy of the information. However, the difficulty of finding multiple sources, especially for verifying website vulnerabilities, made this task more challenging.

Developing the ability to thoroughly verify information by consistently finding and cross-referencing multiple sources is essential for the accuracy of findings. Supporting this process, particularly when dealing with less transparent or hard-to-verify sources, can improve the investigation results.

Learning Goal 6 (LG6): Find multiple sources to enable OSINT verification.

4.3.4 *Analysis.*

Identifying vulnerabilities. Participants lacked a deep understanding of common vulnerabilities and how different data sources can be combined for vulnerability assessments. There were difficulties in interpreting and analyzing complex technical information from tools. P6 wanted to be "able to pull the information that you have readily available, and be able to interpret that in a way that can help the small business." Participants highlighted the challenge of not only gathering relevant data, but also interpreting it in a meaningful way that aligns with the investigation's objectives.

Supporting the ability to interpret results from tools and accurately identify vulnerabilities is crucial in overcoming the challenges. By helping bridge the skill gaps, participants can more effectively analyze complex data and generate reliable vulnerability assessments.

Learning Goal 7 (LG7): Interpret results from tools and effectively identify vulnerabilities.

Developing recommendations from vulnerabilities identified. Participants found it difficult to develop recommendations for mitigating the vulnerabilities based on analyzed data. For example, P1 mentioned that it was hard to identify vulnerabilities for websites that are well-maintained by third-party vendors; in this case, the recommendation could be to continue using these services. Participants needed guidance to effectively present their recommendations, integrating the identified vulnerabilities with supporting data.

Understanding different types of recommendations helps the students in iterating over their investigations and making them actionable. Providing a starting point and directions for recommendations to the students can improve this process.

Learning Goal 8 (LG8): Generate actionable recommendations.

4.3.5 Dissemination.

Not knowing "what a good end product will look like". We observed that the issue of developing appropriate recommendations was exacerbated when presenting findings to non-technical clients, as they require clear explanations and simplifications. Participants also lacked clarity on the best format to use for presenting information. P1 thought their deliverables "…come out very disorganized and the way information is presented is highly variable."

Creating standardized formats for reporting findings will ensure clarity and consistency. Iterating on recommendations to tailor it to the needs of the small business will make the results of investigations highly impactful.

Learning Goal 9 (LG9): Iterate on recommendations to tailor them to business needs.

Study 1 findings highlighted performance gaps in OSINT investigations. To address these challenges and achieve the learning goals (LG 1–9), Study 2 explores the use of generative AI probes across the five phases of IC.

5 Study 2: Using Generative AI Probes to Achieve Learning Goals for OSINT Investigation

5.1 Method

In Study 2, we conducted three co-located design workshops (Design workshops 2–4) to address the challenges identified in Study 1 and achieve the learning goals. These two-hour workshops allowed participants to individually explore how AI could enhance various phases of OSINT investigations. Here, we closely followed Step 2 (playing with AI probes) and Step 3 (brainstorming AI design elements) of the Matchmaking for AI method [70]. Participants spent time working on OSINT tasks using generative AI probes before coming up with design ideas during the workshops. The ideas were represented on a digital whiteboard (FigJam) using the "Co-Design AI Canvas" from the Matchmaking for AI method (see Figure 5). We included columns for generative AI ideas, challenges faced, and function and form, to gather what added customizations or controls they would like beyond the probe. Participants were asked to copy over the purpose of the prompt and the prompt used into a shared prompt library document for all the workshops.

Design Workshop 2 focused on planning and collection-related challenges. These two phases go hand-in-hand, as there needs to be a clear plan before starting with data collection. Participants performed two sets of tasks or two activities; after each activity, there was a focus group to brainstorm design ideas and potential improvements on a digital whiteboard. The activities and brainstorming lasted 25 minutes each with a break of 10 minutes between the sets. Participants created subtasks and assigned them to their team members, and also identified the right tools and techniques for the tasks during the first activity. They developed templates using ChatGPT and did some preliminary data collection to fill out the template during the second activity. This workshop utilized the standard out-of-the-box ChatGPT 3.5 with no web browsing, as this was the latest version at the time.

Design Workshop 3 targeted challenges in the processing, analysis, and dissemination phases. Similar to the previous workshop, there were two sets of tasks. In the first activity, the tasks included verifying information, documenting the investigation process, and interpreting the results as to how they could be vulnerabilities. For the second activity, participants had to create a report based on their findings so far that could be sent to small businesses. Participants could use the templates developed in the earlier workshop and the data collected. This workshop also relied on the standard ChatGPT 3.5 without additional enhancements.

Design Workshop 4 addressed all identified challenges using a new set of AI probes, including web browsing, image analysis, custom GPTs, and plugins (now discontinued) made available through ChatGPT Plus [15]. These features are relevant to OSINT investigations as they help utilize publicly available information from websites and provide access to external tools for analysis. Participants were first presented with specific goals related to the challenges. They played around with each of the four AI probes for 10 minutes and then brainstormed how these tools could be used to help overcome each challenge. This workshop aimed to explore more advanced AI capabilities to support participants in each phase of OSINT investigations.

We used deductive thematic analysis [32] to code transcripts from the workshops and notes from the FigJam boards, corresponding to each learning goal. We coded (1) how participants used ChatGPT; (2) what worked well and what challenges faced; and (3) desired features to improve the experience. We organized our findings for each challenge into three main areas to provide a clear understanding of how generative AI could be leveraged in OSINT investigations. First, we discuss the tasks performed and prompting methods used. Second, we explore the utility of generative AI, highlighting what worked well and the challenges encountered. Finally, we present new ideas and suggested features that participants proposed to enhance the effectiveness of AI in their investigations.

5.2 Results

5.2.1 Planning.

LG1: Developing appropriate trust on OSINT data, tools, and techniques; Implementing operational security. Participants used AI to analyze privacy measures of various tools and identify trustworthy resources. P1 sought guidance on whether their investigative methods stayed within ethical boundaries. P3 and P6 mentioned

using prompting techniques to work around restrictions in ChatGPT's responses, asking for ethical means to obtain information. P2 highlighted the importance of finding tools commonly used in real OSINT investigations and then understanding how they work before applying them.

Participants found that while ChatGPT was useful in creating and updating sock puppet accounts and providing general guidance on privacy and trust, it often failed to provide detailed and specific information about the trustworthiness of tools. P3 noted, "Finding specific details about tools and their trustworthiness was challenging. ChatGPT's responses were sometimes too general to be useful." P1 echoed this idea, pointing out that the AI's comments on trust were often based on popularity rather than a deep analysis of the tool's security features. Participants realized the need to ask more targeted questions to extract useful information.

LG2: Break down investigations into subtasks with data sources. Participants used ChatGPT to break down complex tasks into manageable subtasks tailored to their team's capabilities. For example, P1 mentioned, "I gave it one of the OSINT tasks, and I asked it to divide the work into subtasks…trying to make it adapt to our team by providing the number of members and experience levels". The templates served as conceptual models to guide the investigations. Participants also used specific prompting techniques to receive more structured and relevant answers, such as asking ChatGPT to provide references, using chain-of-thought by asking for steps to get to the subtasks and tools, and framing questions in a scenario-based manner like "if I were to tell my friend how to [specific action] what should I say?" (P2).

Participants found generative AI to be effective in assisting with task division and resource identification, making it easier to manage team-based OSINT investigations. P5 reflected, "I basically just asked it to come up with a list of subtasks and a list of tools and techniques…and it worked perfectly". However, there were challenges with the AI's limitations in providing specific details related to smaller businesses as P4 observed, "ChatGPT provides little to no information on smaller businesses". There were also difficulties crafting prompts to avoid "guardrails" against inappropriate content or eliciting new, non-repetitive information. Participants proposed custom GPT models trained with OSINT-specific databases and plugins to make the AI more tailored to the unique needs of OSINT investigations.

5.2.2 *Collection.*

LG3: Identify and utilize the most effective OSINT tools. Participants used ChatGPT to gather information sources and tools relevant to their tasks, along with descriptions of how to use them effectively. P3 tried to obtain specific "website links to look into data breach sources" as he found broad directions for investigation unhelpful. Prompt structures varied, like role-playing as a cybersecurity student or asking for queries for techniques like Google dorking. P1 and P5 used creative approaches to bypass ChatGPT guardrails, such as asking it to suggest websites similar to those known for posting leaked passwords rather than directly requesting such information.

Participants found ChatGPT to be particularly effective for discovering resources when they asked for specific tools. They could generate alternative options through follow-ups when the initial suggestions did not meet their needs. However, challenges were noted, particularly regarding the AI's outdated knowledge base, as it only contained information on tools up to January 2022. Furthermore, the tools suggested were not always available. P1 described the process of verifying each tool suggested by ChatGPT: "You have to painstakingly copy and paste each one [tool] and research it yourself." Participants noted that while ChatGPT's web browsing capabilities could cover a wide range of knowledge, the credibility of the sources provided was questionable and emphasized the need for careful cross-verification.

LG4: Develop and utilize standardized templates to systematically collect and organize information. Participants used ChatGPT to create comprehensive templates that could be applied in future investigations. P6 said, "I specifically used ChatGPT to help construct a table so all the information could be organized and easy to read." The process often involved iterative prompting, where participants refined their queries to add information like data sources and OSINT techniques. P2 and P6 mentioned that these details helped divide up work within the team.

Participants particularly appreciated the AI's ability to quickly lay out different sections and specify what should go under each. Participants faced challenges in reformatting the AI-generated templates to make them more presentable. P1 explained, "ChatGPT is great for getting your data quickly into format but honestly it doesn't look very good. With a little bit of editing and using a format made by a human I think it would look much better." Importantly, the AI's responses varied depending on the prompts, leading to inconsistencies in the templates generated across different users. Participants proposed the idea of a prompt library, where successful prompts for templates could be shared and rated by others.

5.2.3 *Processing.*

LG5: Structure raw information into parsable format and improve documentation. Participants employed ChatGPT to organize raw information into readable formats and eliminate duplicate or less relevant information. For instance, P6 mentioned, "I asked it to look for different verification sources mentioned throughout the data…it was a lot of information just written, but it's hard to pick through and find all the little details." Participants created specific formats for documenting sources and then used ChatGPT to follow these formats, though some manual adjustments were still required.

While the AI was able to organize data into a list or by topics, it often failed to follow requests like placing data directly into tables. The organized lists were still helpful in making the raw data easier to process and document. Another issue was with ChatGPT's overwhelming amount of information, making it difficult to identify the most relevant details.

LG6: Support OSINT verification based on multiple sources. Participants noted that ChatGPT was useful in identifying potential verification sources and organizing data. However, tool suggestions often fell short of providing the detailed, specific information needed to trust the AI's recommendations, as sources or tools were unfamiliar. P2 cautioned about relying too heavily on AI's web browsing capabilities, as they still needed to be double-checked for accuracy.

5.2.4 Analysis.

LG7: Interpret results from tools and identify vulnerabilities. Participants often began by attempting to interpret the results on their own and then used ChatGPT to further analyze the data. This approach helped participants gain a broader understanding of the vulnerabilities they were investigating, even if the AI's feedback was sometimes generalized.

Participants found ChatGPT's explanations to be somewhat helpful. As P3 noted, "It was useful for elaborating on vulnerabilities and explaining in simpler terms…especially for a non-technical audience.". However, P6 found output adjustments to be hard, "Using ChatGPT for interpretation requires specific prompts for better detail. It can be helpful, but you need to be clear about what you're asking for". P5 found that ChatGPT did not provide detailed security analysis when using web browsing capabilities, often citing limitations in its scope.

LG8: Generate actionable recommendations. Participants used ChatGPT to develop tailored recommendations for non-technical small business owners. Prompting techniques included asking ChatGPT to structure recommendations into a table and develop a timeline for implementing these actions to present information professionally.

ChatGPT helped present supporting data for reports and generate recommendations. Participants found it difficult to identify the most relevant actions from a large number of recommendations. Similar to previous challenges, ChatGPT and custom GPTs often failed to generate specific and actionable recommendations tailored to the unique needs of a given situation, sometimes refusing a detailed analysis for being out of scope or unethical.

5.2.5 Dissemination.

LG9: Iterate on recommendations to tailor it to business needs. Participants iterated on recommendations by refining them through a feedback loop. P3 described the process as "using ChatGPT to create specific recommendations based on vulnerabilities and then simulating client feedback to further improve the recommendations." They also wanted outputs to have severity ratings for vulnerabilities.

Participants found that using ChatGPT to simulate client feedback was particularly useful in preparing for client interactions. P1 emphasized that iterative feedback from ChatGPT helped in making the recommendations more actionable and better tailored to client needs. However, P4 observed that the iterative process was often limited to removing, or reformatting content rather than delving into more detailed analysis. Participants also worried about privacy issues in involving the AI at this final stage, when security flaws were at their most distilled and contextualized.

5.3 Follow-up Study: Collaboration and Leadership

As participants kept performing OSINT investigations and incorporated generative AI in their workflow, we observed new collaboration challenges emerge:

Difficulty in achieving consistent formatting and style. Participants saw that despite using similar prompts, the outputs varied widely, making it difficult to maintain uniformity across documents. P6 summarized it as, "ChatGPT is pretty different depending on the prompt you give, and it always gives different answers for different people. So it is a little difficult to make our templates match". This inconsistency required additional time and effort to ensure consistency across the team's work.

Difficulty in developing shared resources. Participants found it challenging to build on each other's work and learn from prompts used by others. P6 reflected, "I didn't end up testing out anyone else's prompts, but I might do that in the future just to see what [comes] as a result of it". P3 and P4 were interested in finding successful prompts to use them as starting points. In terms of improvements, P6 said that "the collaborative process could be improved if we used a different platform to document our investigation notes and sources". This platform would help to share editable results from generative models.

5.3.1 Method.

As a follow-up study, we brainstormed design ideas around these two collaboration challenges. We conducted an online workshop (design workshop 5) to focus on collaboratively forming templates for data collection and dividing up work, with P1 serving as the team leader. We used Team-GPT [93], a commercially available web-based platform that enables teams to collaboratively interact in real-time with AI models like ChatGPT, as a probe. Figure 3 shows a screenshot of the Team-GPT workspace with its core features.

Effective leadership is crucial for guiding the team, assigning roles, and facilitating the efficient use of collaboration tools [71]. P6 emphasized that having a leader helped streamline tasks and prevent overlap. We argue that while Team-GPT offers valuable collaboration features, integrating strong leadership could significantly improve the collaboration process. However, we observed during design workshop 5 and a team-based practice session, that assigned leaders found it overwhelming to help get the team started and keep track of all the information. To address the issue, we wanted students to take on the role of leaders and imagine how they would use ChatGPT to display task-oriented leadership behaviors [71] during a 30-minute focus-group interview (design workshop 6).

Analysis. For collaboration, we performed a deductive thematic analysis [32] of the workshop 5 transcript based on (1) Team-GPT features used (mentioned in Figure 3) and (2) how they impacted collaboration.

For leadership, we performed deductive thematic analysis [32] of the workshop 6 transcript based on the themes of four leadership behaviors [71]: (1) dividing and assigning tasks; (2) clarifying requirements by thinking about the end goal and client presentation; (3) monitoring progress and identifying areas to dig deeper; and (4) problem-solving by providing feedback and iterating on results.

5.3.2 Results.

Collaboration. Participants found several of Team-GPT's collaboration features helpful. The shared workspace was particularly useful for visibility and access, allowing team members to build on each other's contributions and avoid duplication. P4 mentioned, "I thought it was really useful for collaboration…being able to just all contribute to one was really useful". However, when multiple active

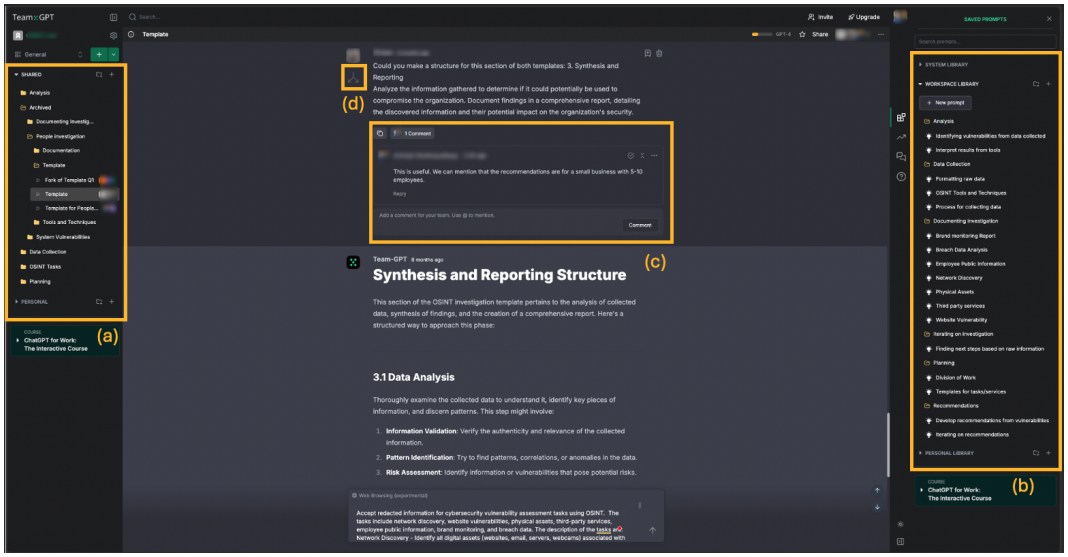


Figure 3: Team-GPT platform [93] with core features: (a) personal chats for individual work and shared workspaces for team collaboration. Chats were grouped by the different phases of OSINT investigation in our case; (b) a team library to store shared prompt templates which can be invoked with a click after filling in the placeholders. We used the library to curate a set of useful prompts for the participants in Study 3; (c) directly comment on prompts and responses to provide feedback; (d) fork chats to new personal or shared chats to continue the conversation.

chats were created, it sometimes led to confusion and a lack of coordination, making it feel as though team members were working in silos. Similarly, excessive forking of chats, while flexible, resulted in a fragmented workflow that complicated team collaboration. To mitigate these challenges, participants suggested staying within the same chat more consistently and marking successful strategies within the chat for others to build upon.

Commenting on prompts and responses allowed for immediate feedback from team members, which helped refine outputs and guide the next steps based on that information. Participants found the prompt library useful for storing and reusing effective prompts but were skeptical about its practicality in diverse, case-specific scenarios. P5 said, "I'm not sure how often that would be used…a lot of this is just a case-by-case basis". Participants mentioned that placeholders within prompt templates allowed for easy customization, which could increase the library's utility.

Leadership. Participants suggested that AI could use team members' skills and past performance to propose optimal task assignments to leaders, but human judgment would still be crucial for considering team members' preferences. P6 thought AI-generated templates could help leaders ensure that the final outputs were thorough and aligned with client needs. P1 suggested AI could create monitoring frameworks, such as tables or templates with detailed timelines and task descriptions. Participants reiterated the need for real-time communication and human oversight along with AI for effective monitoring.

AI could help identify gaps in the data for the leader, who can then follow up on it, as P6 stated, "I'm sure if you put like the info from a shared document to ChatGPT and ask like which areas need more research, it might tell you". As a leader, P1 wanted suggestions for visualization to enhance the presentation's impact. They also noted the importance of the human touch in finalizing decisions and being accountable for presenting results to clients.

In conclusion, Study 2 explored how generative AI could address key challenges in OSINT investigations and developed design ideas. The findings show how AI supported learning and helped achieve the learning goals. Building on the insights from using it in different parts of the investigation, participants used generative AI in conducting OSINT-based vulnerability assessments on small businesses during Study 3.

6 Study 3: Performing Real-World Investigations Augmented By Generative AI

6.1 Method

In this study, our six student participants of the OSINT Clinic offered free cyber vulnerability assessments to three small businesses in our region. These assessments provided small businesses with a view of their attack surface to address vulnerabilities before malicious actors could exploit them. Unlike a comprehensive vulnerability assessment, which might include control interviews with IT staff, physical facility tours and inspections, and internal network enumeration and scanning with tools like Nmap and Nessus / OpenVAS, ours were limited to vulnerabilities that could be detected via OSINT techniques.

We structured the consultation process using the Understand-Investigate-Advise (UIA) framework introduced by Havron et al.[50, 57]. In the initial Understand phase, the researchers met with the clients to explain the consultation process, set expectations regarding timelines and involvement, and answer any questions. Clients

were then asked to select relevant services from a menu of tasks based on the categories outlined in Table1. Those who wished to proceed signed a letter of agreement consenting to the investigation and provided basic information about their digital infrastructure.

In the Investigate phase, participants drew on their semester-long training in OSINT tradecraft and their experience with generative AI platforms like ChatGPT and Team-GPT (as discussed in Studies 1 and 2). For the first investigation, all six participants collaborated as a single team. In subsequent investigations, they split into smaller teams of 2–3 members, each with a designated leader. Over approximately 2.5 weeks per investigation, the teams collected and analyzed data, identifying potential vulnerabilities in the clients' digital systems.

Finally, during the Advise phase, the teams synthesized their findings into presentations. These presentations detailed the identified vulnerabilities and offered tailored recommendations for mitigation. By delivering clear and actionable advice, the teams ensured clients received practical strategies to improve their cybersecurity posture.

We curated a prompt library on Team-GPT [93] based on highly-rated prompts from the previous study workshops across the different phases of the Intelligence Cycle. This library was provided to participants as an instrument to reuse the ChatGPT capabilities they found useful. The prompt templates allowed user customization for variables like the OSINT task type, output from tools, team size, and data collection details. We include these templates in Appendix A.1 in Tables 4, 5, 6, and 7. In addition to the prompt library, we gave teams access to a separate group chat on Mattermost (an open-source Slack clone) and individual subscriptions to ChatGPT Plus.

We followed best practices for security, privacy, and ethics established in previous research on student-led real-world OSINT investigations [10, 73]. The students signed Non-Disclosure Agreements (NDAs) to protect the confidentiality of their findings except for sharing with the customer. The students also enhanced their safety by taking operational security measures like using Virtual Private Networks (VPNs), Virtual Machines (VMs), role-based access control and encryption for storing and sharing data, and sock puppet accounts.

We conducted two 30-minute focus group interviews, one after the first investigation and then after the final one. The goal was to understand the use of generative AI and document challenges faced during real-world investigations. We added a reflection survey for participants to rate their use of generative AI across the different phases of the investigations and for supporting leadership. We also distributed a survey to the small business clients asking them to rate the vulnerability assessment service in order to measure the real-world value of the service and how actionable the recommendations were.

In terms of analysis, first, we performed deductive thematic analysis [32] on the focus group transcripts to understand (1) the use of generative AI, and (2) the challenges that emerged and how participants circumvented them. We also analyzed the shared chats on the Team-GPT workspace to monitor the use of the prompt library. Then, we aggregated the three client evaluations of vulnerability assessments to understand their utility.

6.2 Results

6.2.1 Utility of generative AI during investigations. Participants employed AI to structure project plans, identify tasks, and allocate responsibilities. P4, the leader for the first investigation, used templates for streamlining brand monitoring and employee information management. This initial structuring helped teams set clear objectives and assign roles.

AI was particularly useful in identifying tools and generating templates to gather relevant information. P2 mentioned, "I used the prompt library for generating report templates for brand monitoring and data breach exposure." These templates allowed team members to collect data in a structured manner, ensuring consistency and thoroughness across different team members' efforts. AI also assisted them in identifying gaps in data and making suggestions on how to add data from other sources for verification.

Teams used AI to interpret data from tools and draft vulnerability mitigation strategies. They organized findings and added explanations, making it easier to present findings to non-technical stakeholders. P1 emphasized, "I used it a lot when it came to rephrasing my thoughts. So sometimes I would just dump all the information I got, and then I would say, hey, can you organize this in ChatGPT?". During dissemination, AI helped participants format their findings concisely, improve the written report, and create polished client presentations.

There was limited usage of collaborative features of Team-GPT, as participants preferred to work separately on their personal ChatGPT instances. This was to avoid getting information "clogged up" on a shared workspace as P4 mentioned, "[Shared chat] helped to get started and you could take a look at others, but then you did not really want to go in and like be in a shared chat". Leaders accessed the prompt library on Team-GPT more than other members to create templates to guide the team's work and structure results for presentation.

6.2.2 Challenges and limitations faced during investigations. A significant challenge was the concern over privacy and data security. Participants were cautious about feeding sensitive information into AI tools, as there was uncertainty about how this data might be handled or stored. One participant reflected on this issue, stating, "We should keep this in mind while we are using ChatGPT…we do not want to put in the raw data in the form that breaches the privacy of the company." The researchers provided a Python script that participants could run locally to scrub Personal Identifiable Information (PII), but it wasn't used as it required extra setup. P4 and P5 navigated privacy concerns by avoiding including specific or sensitive data in their prompts, focusing instead on general scenarios. For example, P4 said, "like keeping it general enough, and not having specific information, but having, like the scenario where I asked, what are the possible cybersecurity risks knowing the family members of an owner of a company or high-level employee".

Generative AI in the form of a large language model like ChatGPT, showed limitations in data collection and analysis. A recurring theme in Study 2, participants again found that AI often provided broad, generalized responses that required further refinement, as P5 mentioned, "It tends to give broad responses and doesn't always provide the specific details needed for verification". This may be partly due to the use of abstracted prompts to bypass privacy issues.

Consequently, participants spent the majority of their investigation outside the Team-GPT or ChatGPT platforms, using external tools and searching websites or social media accounts.

Another challenge was the difficulty in monitoring team progress and maintaining coordination, especially when using AI in asynchronous collaboration settings. Participants rated monitoring progress as the leadership behavior where generative AI was the least useful. While AI tools were used to create structured templates and organize information, keeping track of who was doing what and ensuring that all tasks were aligned with the overall investigation goals was not straightforward.

6.2.3 Usefulness of vulnerability assessments based on client survey. Based on the survey responses, all clients valued the insights into their cybersecurity risks provided by the student teams from the OSINT clinic. The consistently highest possible ratings for the value and actionability of the assessments indicate that the clients found the services beneficial, not only for understanding their current vulnerabilities but also for guiding their future cybersecurity efforts. All three clients wanted to return for a similar exercise and recommended the program to other organizations.

One respondent noted that the information was "very actionable" and that they now have "a game plan to address the issues and mitigate the potential risks". Another client emphasized the usefulness of the report as they were in the growth phase and had already started implementing some of the recommendations. Table 3 shows some of the common vulnerabilities identified and recommendations made.

7 Discussion

7.1 Reflection on Co-designing via Matchmaking for AI

7.1.1 Adapting the method for OSINT investigations. We adapted and extended the Matchmaking for AI method [70] for an OSINT context to structure design workshops and generate innovative yet feasible ideas. We followed the three key steps of the original method. First, for mapping stakeholder domain expertise and AI needs, we focused on challenges and learning goals across the OSINT Intelligence Model for OSINT investigations. For Step 2, playing with AI probes to make participants familiar with "affordances of the technology," we used ChatGPT and Team-GPT as probes. For the third step, brainstorming AI design elements, we identified ways to tailor general-purpose LLMs in developing and applying OSINT skills.

We also extended this method by addressing two limitations revealed by our study context. First, our brainstorming sessions in a group setting, with a mix of co-located and online setups, alleviated the limits of interpersonal engagement and creative expression among participants in a remote, individual study setting. Second, we used ChatGPT and Team-GPT as functional technology probes [62], compared to Wizard-of-Oz simulations, to help participants envision more diverse and concrete AI solutions.

7.1.2 Extending the method for collaboration and learning. Our goal was to support investigations of student analysts, as they are essential for the cybersecurity clinic model [11]. Their needs, expertise, and skill levels differ significantly from expert practitioners, who are the primary stakeholders for previous matchmaking studies [70, 103]. Our adaptations of the Matchmaking for AI method cater to learners and the clinic model, characterized by achieving learning goals and performing collaborative real-world investigations.

Chin Jr. et al. highlight collaborative analysis as a critical approach for intelligence analysis, and we report on its evolution over time by observing AI interactions and teamwork across workshops and real investigations [38]. In Study 2 (Section 5.3), we investigate new challenges and design ideas that emerge as the team collaborates and incorporates generative AI tools over time. Our longitudinal and cyclical study design promotes fluidity between the phases of co-design and helps to study collaboration.

As Stahl suggests, understanding groupware requires groups to use it in real-world conditions over time [85]. In Study 3 (Section 6), we enhance the matchmaking process by validating design ideas in a natural setting. Team-GPT along with other content collaboration tools like Google Docs and Mattermost serve as a research tool to observe team behaviors and use of generative AI in ongoing investigations. The insights gained from real-world investigations help prioritize the design ideas, providing a deeper understanding of how to integrate AI tools effectively into collaborative OSINT practices. Our study goes beyond the typical use of AI for data collection and analysis [34], focusing on how student teams learn new skills and close skill gaps. Our study design refines the matchmaking process to support collaboration and real-world needs, making it more flexible and useful for areas like crowdsourcing complex tasks [80, 100] and co-creation [33].

7.2 Role of Generative AI in Learning and Collaboration for Investigations

Studies 1 and 2 helped identify and achieve a set of learning goals (Table 2) for performing vulnerability assessments as a part of the OSINT Clinic. Generative AI, particularly ChatGPT, was instrumental in helping students plan investigations (LG2) and select effective OSINT tools (LG3). For example, AI simplified the process of breaking down complex tasks into structured subtasks tailored to the team's capabilities. This scaffolding allowed students to navigate the vast array of OSINT techniques without feeling overwhelmed, supporting novice learners in developing systematic approaches to investigations. These results contribute to research on using AI to structure complex tasks [30, 86].

In terms of data collection and verification (LG4, LG6), AI-generated templates provided consistent structures for organizing collected data, reducing redundancy, and enhancing collaboration. AI's ability to suggest alternative sources for verification encouraged students to adopt rigorous cross-referencing practices. However, the AI's generalization tendencies and outdated information posed challenges, requiring students to manually verify AI-suggested tools and sources. AI-assisted interpretation of tool outputs enabled participants to better identify vulnerabilities, enhancing their analytical skills (LG7). By simplifying technical findings into clear, understandable language, AI helped students translate complex data into actionable insights. This was particularly beneficial in learning how to present findings to non-technical clients (LG9). Similar benefits like getting new insights and support [94] have

Table 3: Common vulnerabilities identified and recommendations for mitigation

Common Vulnerabilities	Recommendations
Out-of-date tech stacks like JavaScript libraries on client websites	Regularly update software versions to avoid potential security risks
Company social media reveals questionable personal data or connections	Separate personal and business social media accounts to maintain professionalism and protect privacy
Employee emails found in data breach	Regularly change passwords and use a password manager to prevent unauthorized access through compromised emails
Personal home address registered as business address can be easily found in public records	Register your business at a P.O. box, or use virtual office space for registration
Third-party services have vulnerabilities and predatory policies	Switch to more secure alternatives and thoroughly read any agreement before signing and research their reputation beforehand

been demonstrated in applications of group ideation [59], coding [76], and general creativity [44].

We found that Team-GPT supported sharing and content collaboration, which are collaborative behaviors first identified in Kang et al.'s longitudinal study of teams of intelligence analysts [66]. Teams used the shared workspace to finalize templates and structure workflows, which minimized duplication of effort and streamlined task division. Content collaboration was supported by the prompt library and saved chats, and led to the development of a repository of useful prompts (see Appendix A.1). However, challenges emerged around maintaining consistency in formatting and style across different team members' outputs. This required students to develop stronger communication and coordination skills, reflecting real-world team dynamics in cybersecurity investigations [81].

While generative AI supported the development of OSINT skills and collaboration, challenges such as privacy concerns, generalization, and inconsistent outputs underscore the need for careful integration. The students' ability to critically engage with AI, verify its suggestions, and adapt their workflows reflects a key outcome: they not only learned OSINT techniques but also developed meta-skills in evaluating AI outputs and incorporating them responsibly.

7.3 Design Considerations for Using Generative AI in Real-World Investigations

In our study, we structured OSINT-based vulnerability assessments into key areas (see Table 1) that helped evaluate threats such as outdated tech stacks, leaked credentials, and social engineering vulnerabilities. Building on Clinical Computer Security (CCS) [50], we showed how its core approach — collaborative, client-focused security investigations — can be adapted for broader cybersecurity assessments. While CCS traditionally supports individual victims, our study applied the UIA framework to proactively assess organization-level cybersecurity risks.

We integrated generative AI to support clinical investigations. In Study 3, tools like ChatGPT and Team-GPT assisted across multiple phases of the Intelligence Cycle during real-world investigations. Students used AI to create templates for structuring data collection, generate actionable recommendations for clients, and simplify technical findings into more digestible reports. Leaders in the team relied on AI to break down complex tasks, assign subtasks, and scaffold the investigation process, ensuring the team maintained a structured workflow. Generative AI also supported the verification process by suggesting alternative data sources, enhancing the rigor of the investigations. However, participants used these tools selectively, often choosing to work independently with OSINT tools and personal ChatGPT instances to maintain efficiency.

The results underscored the need for balancing AI automation with human judgment, aligned with findings from previous studies [59, 67]. While generative AI can handle routine tasks like template generation and summarization, more nuanced activities — such as verifying sensitive information or providing context-aware leadership — still require human expertise. These results build on previous research that has highlighted several challenges in using generative AI, such as misalignments in understanding its capabilities [86, 92], prompting techniques [86, 113], and evaluating its outputs [68, 86]. Integrating generative AI into real-world OSINT workflows revealed design considerations related to privacy, workflow integration, and collaboration. We combined participants' design ideas with recent technological advancements to propose future directions and mitigation strategies across these challenges.

7.3.1 Privacy Challenges. Privacy concerns were a significant limitation in fully leveraging generative AI tools like ChatGPT and Team-GPT. Although clients did not specifically mention concerns about AI tools or privacy, our NDA explicitly ensured that all data would remain within the clinic and not be shared externally. Participants were cautious about feeding sensitive client data into AI platforms due to uncertainty about how the data might be stored or used by these tools. This caution was more pronounced during the later stages of the investigations, particularly when analysis and dissemination involved sensitive and identifiable information. Even though participants attempted to anonymize data or use generalized scenarios, the fear of exposing confidential information

remained prevalent. While a Python script was provided for this purpose, the extra setup was a barrier to its adoption.

The reliance on commercially available, cloud-based AI platforms creates a tension between efficiency and privacy. This issue becomes more critical with OSINT clinics as students may be more prone to taking "shortcuts" or inadvertently breaching confidentiality. This tension highlights the need for privacy-preserving AI frameworks that can support OSINT investigations without risking client confidentiality. For instance, AI models that run locally on secure machines, rather than cloud-based platforms, could allow for sensitive data processing without risking data leakage [19, 55]. Features that let users toggle privacy modes, disable chat history, or automatically redact sensitive information can help build trust in these tools. Educational guidelines and training for students on responsible AI use can reduce the likelihood of shortcuts that compromise privacy.

7.3.2 Integration with Existing Workflows. Integrating generative AI into the existing workflows of student analysts presented practical challenges. There is a lack of support for collecting OSINT information with general-purpose models, even though this is the most time-consuming part of the investigation. Analysts also faced workflow disruptions due to context-switching between multiple tools and platforms (ChatGPT, Team-GPT, Google Docs, Mattermost, browser, command line tools).

To address this, future design efforts should focus on creating seamless integrations between generative AI tools and commonly used collaboration platforms. Features like embedded AI prompts within investigation tools or plugin-based extensions could streamline workflows and reduce context-switching. Agentic workflows could help automate data collection by orchestrating the use of reputable tools [18, 41, 49]. Trust in results can be achieved by integrating well-known and reliable OSINT tools [64, 106].

7.3.3 Collaboration and Leadership Issues. Generative AI tools introduced both opportunities and challenges for collaboration and leadership within the student teams. Team-GPT features facilitated initial brainstorming and task organization but led to fragmented workflows when used excessively. Forking conversations and managing multiple active chats often created silos, reducing the efficiency of team coordination. As a result, participants preferred to use personal ChatGPT instances to maintain focus and avoid cluttered workspaces.

Effective leadership was crucial in mitigating these challenges. Leaders used AI to structure tasks and generate templates, but monitoring team progress remained difficult. Here, AI can assist leaders by suggesting task assignments based on team members' skills and past performance [30, 105]. Features that maintain group awareness, for example, real-time dashboards showing team progress, task ownership, and AI interactions could help leaders and members stay synchronized [78].

Moreover, while AI facilitated some aspects of task-oriented leadership, it lacked the contextual understanding needed for decision-making. This reinforces the importance of human oversight and interpersonal skills in leadership roles, suggesting that AI should be seen as a complement rather than a replacement for leadership behaviors [84, 112].

7.4 Limitations and Future Work

A limitation of this study is the small participant size of six undergraduate students, which may limit the generalizability of the results. However, this is consistent with prior HCI and intelligence analysis research, where studies often rely on small groups due to the sensitive nature of the work [38, 51, 58, 60, 66, 81, 109]. This challenge reflects broader issues in accessing cybersecurity professionals for research. As Sundaramurthy et al. found, barriers such as intensive workloads, sensitive work environments, and mistrust of outsiders make it difficult to recruit larger, more diverse groups [88].

Another limitation is that the findings are primarily reflective of a group with limited professional experience. While only half of our student participants were computer science majors, the group represented diverse academic backgrounds, including business and national security majors. Despite this limitation, the insights from this group remain valuable. The participants' status as novice OSINT analysts mirrors the demographics of many cybersecurity clinics and educational programs, where students are the primary workforce. Their challenges, skill gaps, and learning goals offer a critical lens for understanding how generative AI can support early-stage analysts, who represent the future of the OSINT and cybersecurity workforce. Addressing the needs of novices can lead to the development of AI tools that are accessible, educational, and scalable for broader use in training environments and real-world applications.

The current study provides initial findings on the integration of generative AI in OSINT clinics. The developed prompt library, along with examples of generative AI use across the different phases of OSINT investigations, offers a starting point for refining design ideas in future research. Expanding this work to include larger and more varied participant pools — such as students from diverse backgrounds, professional analysts, or interdisciplinary teams — will help assess the broader applicability and effectiveness of these AI-driven approaches. Including students from various disciplines (e.g., social sciences, law, journalism) could provide a more holistic understanding of how different backgrounds influence the adoption and effectiveness of AI tools in OSINT. Additionally, longitudinal studies that follow participants into professional settings could reveal how AI-supported skills translate to real-world practice.

8 Conclusion

This paper introduced the OSINT clinic, which was aimed at providing students with training and practical experience in conducting cybersecurity vulnerability assessments using publicly available data, addressing the scalability challenge of helping small businesses protect themselves against cyber attacks. Through a three-phase co-design study, we identified key challenges, such as information management, skill gaps, and collaboration inefficiencies. We developed design ideas to address these gaps by integrating generative AI into various stages of the OSINT investigation process, including planning, data collection, processing, analysis, and dissemination. The design ideas can benefit novices and experts to leverage large language models like ChatGPT and collaborative AI platforms like Team-GPT in their workflows. A pilot with three small businesses demonstrated the benefits of using generative AI

to streamline OSINT investigations, as clients found value in the resulting vulnerability assessment reports. It also revealed challenges, including privacy concerns and difficulties in monitoring progress. We propose future research on human-AI collaboration for OSINT tasks, informed by our design requirements and recommendations.

Acknowledgments

This work was partly supported by the Virginia Commonwealth Cyber Initiative. We sincerely thank the students and clients who participated in our study and contributed valuable insights to our research.

References

[1] 2020. NIST SP 800-115. *NIST* (Jan. 2020). Retrieved 2024-09-13 from https://www.nist.gov/privacy-framework/nist-sp-800-115 Last Modified: 2021-04-23T09:13-04:00.

[2] 2021. *J298 OSINT Seminar — Open Source Investigations.* Retrieved 2024-09-10 from https://journalism.berkeley.edu/course-section/j298-human-rights-center-seminar-f21/

[3] 2022. Join us in pushing back on misinformation. Retrieved 2024-09-10 from https://web.archive.org/web/20200602035809/https://our.news/

[4] 2023. *Urban Cyber Defense: Cybersecurity Clinic.* http://urbancyberdefense.mit.edu/cybersecurityclinic Accessed: 2023-10-02.

[5] 2024. 2019 Data Breach Investigations Report Summary of Findings. Retrieved 2024-09-08 from https://enterprise.verizon.com/resources/reports/dbir/2019/summary-of-findings/

[6] 2024. *About the Citizen Lab.* Technical Report. Citizen Lab, University of Toronto. Retrieved 2024-12-10 from https://citizenlab.ca/about/ Section: Uncategorized.

[7] 2024. BuiltWith. Retrieved 2024-09-08 from https://builtwith.com/

[8] 2024. Censys: Exposure Management and Threat Hunting Solutions. Retrieved 2024-09-08 from https://censys.com/

[9] 2024. CIS Controls Navigator v8.1. Retrieved 2024-09-08 from https://www.cisecurity.org/controls/cis-controls-navigator/

[10] 2024. Commonly Asked Questions About Cybersecurity Clinics – Consortium of Cybersecurity Clinics. Retrieved 2024-09-12 from https://cybersecurityclinics.org/blog/commonly-asked-questions-about-cybersecurity-clinics/

[11] 2024. Consortium of Cybersecurity Clinics. Retrieved 2024-09-07 from https://cybersecurityclinics.org/

[12] 2024. Google Dorking: An Introduction for Cybersecurity Professionals. Retrieved 2024-09-08 from https://www.splunk.com/en_us/blog/learn/google-dorking.html

[13] 2024. Have I Been Pwned: Check if your email has been compromised in a data breach. Retrieved 2024-09-08 from https://haveibeenpwned.com/

[14] 2024. Intelligence X. Retrieved 2024-09-11 from https://intelx.io/

[15] 2024. Introducing ChatGPT Plus. Retrieved 10-09-2024 from https://openai.com/index/chatgpt-plus/

[16] 2024. Nessus Vulnerability Scanner: Network Security Solution | Tenable®. Retrieved 2024-09-08 from https://www.tenable.com/products/nessus

[17] 2024. Nmap: the Network Mapper - Free Security Scanner. Retrieved 2024-09-08 from https://nmap.org/

[18] 2024. OpenAI Platform. Retrieved 2024-09-11 from https://platform.openai.com

[19] 2024. PrivateGPT: Use Cases in OSINT. Retrieved 2024-09-07 from https://www.osintcombine.com/post/privategpt-use-cases-in-osint

[20] 2024. Shodan. Retrieved 2024-09-08 from https://www.shodan.io

[21] 2024. UC Berkeley Cybersecurity Clinic. Retrieved 2024-12-10 from https://cltc.berkeley.edu/program/cybersecurity-clinic/

[22] 2024. Whois.com - Free Whois Lookup. Retrieved 2024-09-08 from https://www.whois.com/whois/

[23] Mark S Ackerman. 2000. The intellectual challenge of CSCW: the gap between social requirements and technical feasibility. *Human–Computer Interaction* 15, 2-3 (2000), 179–203.

[24] Babak Akhgar, P Saskia Bayerl, and Fraser Sampson. 2017. *Open source intelligence investigation: from strategy to implementation.* Springer.

[25] Joelle Alcaidinho, Larry Freil, Taylor Kelly, Kayla Marland, Chunhui Wu, Bradley Wittenbrook, Giancarlo Valentin, and Melody Jackson. 2017. Mobile Collaboration for Human and Canine Police Explosive Detection Teams. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) *(CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 925–933. https://doi.org/10.1145/2998181.2998271

[26] Sultan A. Alharthi, Nicolas James LaLone, Hitesh Nidhi Sharma, Igor Dolgov, and Z O. Toups. 2021. An Activity Theory Analysis of Search and; Rescue Collective Sensemaking and Planning Practices. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 146, 20 pages. https://doi.org/10.1145/3411764.3445272

[27] Carlo Aliprandi, Juan Arraiza Irujo, Montse Cuadros, Sebastian Maier, Felipe Melero, and Matteo Raffaelli. 2014. CAPER: Collaborative Information, Acquisition, Processing, Exploitation and Reporting for the Prevention of Organised Crime. *Communications in Computer and Information Science* (2014), 6.

[28] Gagan Bansal, Besmira Nushi, Ece Kamar, Eric Horvitz, and Daniel S. Weld. 2021. Is the Most Accurate AI the Best Teammate? Optimizing AI for Teamwork. http://arxiv.org/abs/2004.13102 arXiv:2004.13102 [cs].

[29] Yasmine Belghith, Sukrit Venkatagiri, and Kurt Luther. 2022. Compete, Collaborate, Investigate: Exploring the Social Structures of Open Source Intelligence Investigations. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–18. https://doi.org/10.1145/3491102.3517526

[30] Ananya Bhattacharjee, Yuchen Zeng, Sarah Yi Xu, Dana Kulzhabayeva, Minyi Ma, Rachel Kornfield, Syed Ishtiaque Ahmed, Alex Mariakakis, Mary P. Czerwinski, Anastasia Kuzminykh, Michael Liut, and Joseph Jay Williams. 2023. Understanding the Role of Large Language Models in Personalizing and Scaffolding Strategies to Combat Academic Procrastination. http://arxiv.org/abs/2312.13581 arXiv:2312.13581 [cs].

[31] Sara Bly and Elizabeth F. Churchill. 1999. Design through matchmaking: technology in search of users. *Interactions* 6, 2 (March 1999), 23–31. https://doi.org/10.1145/296165.296174

[32] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (Jan. 2006), 77–101. https://doi.org/10.1191/1478088706qp063oa Publisher: Routledge _eprint: https://www.tandfonline.com/doi/pdf/10.1191/1478088706qp063oa.

[33] Shannon Briggs, Matthew Peveler, Jaimie Drozdal, Lilit Balagyozyan, Jonas Braasch, and Hui Su. 2020. Brainstorming for sensemaking in a multimodal, multiuser cognitive environment. In *HCI International 2020-Late Breaking Papers: User Experience Design and Case Studies: 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings 22*. Springer, 66–83.

[34] Thomas Oakley Browne, Mohammad Abedin, and Mohammad Jabed Morshed Chowdhury. 2024. A systematic review on research utilising artificial intelligence for open source intelligence (OSINT) applications. *International Journal of Information Security* (June 2024). https://doi.org/10.1007/s10207-024-00868-2

[35] Thomas Oakley Browne, Mohammad Abedin, and Mohammad Jabed Morshed Chowdhury. 2024. A systematic review on research utilising artificial intelligence for open source intelligence (OSINT) applications. *International Journal of Information Security* (2024), 1–28.

[36] Dong Chen. 2021. *Supporting Collaborative Information Analysis Using Interactive Visualization: A Design Research.* The Pennsylvania State University.

[37] Xiang 'Anthony' Chen, Jeff Burke, Ruofei Du, Matthew K. Hong, Jennifer Jacobs, Philippe Laban, Dingzeyu Li, Nanyun Peng, Karl D. D. Willis, Chien-Sheng Wu, and Bolei Zhou. 2023. Next Steps for Human-Centered Generative AI: A Technical Perspective. http://arxiv.org/abs/2306.15774 arXiv:2306.15774 [cs].

[38] George Chin Jr, Olga A Kuchar, and Katherine E Wolf. 2009. Exploring the analytical processes of intelligence analysts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 11–20.

[39] Dharma Dailey and Kate Starbird. 2015. "It's Raining Dispersants": Collective Sensemaking of Complex Information in Crisis Contexts. In *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work and; Social Computing* (Vancouver, BC, Canada) *(CSCW'15 Companion)*. Association for Computing Machinery, New York, NY, USA, 155–158. https://doi.org/10.1145/2685553.2698995

[40] Gelei Deng, Yi Liu, Víctor Mayoral-Vilches, Peng Liu, Yuekang Li, Yuan Xu, Tianwei Zhang, Yang Liu, Martin Pinzger, and Stefan Rass. 2024. PentestGPT: Evaluating and Harnessing Large Language Models for Automated Penetration Testing. In *33rd USENIX Security Symposium (USENIX Security 24)*. USENIX Association, Philadelphia, PA, 847–864. https://www.usenix.org/conference/usenixsecurity24/presentation/deng

[41] Alan R. Dennis, Akshat Lakhiwal, and Agrim Sachdeva. 2023. AI Agents as Team Members: Effects on Satisfaction, Conflict, Trustworthiness, and Willingness to Work With. *Journal of Management Information Systems* 40, 2 (April 2023), 307–337. https://doi.org/10.1080/07421222.2023.2196773 Publisher: Routledge _eprint: https://doi.org/10.1080/07421222.2023.2196773.

[42] Shayan Doroudi, Ece Kamar, Emma Brunskill, and Eric Horvitz. 2016. Toward a Learning Science for Complex Crowdsourcing Tasks. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 2623–2634. https://doi.org/10.1145/2858036.2858268

[43] Stephen L. Dorton and Rob A. Hall. 2021. Collaborative Human-AI Sensemaking for Intelligence Analysis. In *Interacción*. https://api.semanticscholar.org/CorpusID:236150803

[44] Anil R. Doshi and Oliver Hauser. 2023. Generative artificial intelligence enhances creativity but reduces the diversity of novel content. https://doi.org/10.2139/ssrn.4535536

[45] Steven Dow, Anand Kulkarni, Scott Klemmer, and Björn Hartmann. 2012. Shepherding the Crowd Yields Better Work. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work* (Seattle, Washington, USA) *(CSCW '12)*. Association for Computing Machinery, New York, NY, USA, 1013–1022. https://doi.org/10.1145/2145204.2145355

[46] Matthew Edwards, Robert Larson, Benjamin Green, Awais Rashid, and Alistair Baron. 2017. Panning for gold: Automatically analysing online social engineering attack surfaces. *Computers & Security* 69 (Aug. 2017), 18–34. https://doi.org/10.1016/j.cose.2016.12.013

[47] Kristie Fisher, Scott Counts, and Aniket Kittur. 2012. Distributed Sensemaking: Improving Sensemaking by Leveraging the Efforts of Previous Users. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA) *(CHI '12)*. ACM, New York, NY, USA, 247–256. https://doi.org/10.1145/2207676.2207711

[48] Flashpoint. 2024. Generative AI for OSINT: Next Level Techniques for ChatGPT and Beyond. Retrieved 2024-09-08 from https://www.youtube.com/watch?v=7zzubjKEUW4

[49] Christopher Flathmann, Beau G. Schelble, and Nathan J. McNeese. 2021. Fostering Human-Agent Team Leadership by Leveraging Human Teaming Principles. In *2021 IEEE 2nd International Conference on Human-Machine Systems (ICHMS)*. 1–6. https://doi.org/10.1109/ICHMS53169.2021.9582649

[50] Diana Freed, Sam Havron, Emily Tseng, Andrea Gallardo, Rahul Chatterjee, Thomas Ristenpart, and Nicola Dell. 2019. " Is My Phone Hacked?" Analyzing Clinical Computer Security Interventions With Survivors of Intimate Partner Violence. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–24.

[51] Matylda Gerber, B.L. William Wong, and Neesha Kodagoda. 2016. How Analysts Think: Intuition, Leap of Faith and Insight. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 60, 1 (Sept. 2016), 173–177. https://doi.org/10.1177/1541931213601039 Publisher: SAGE Publications Inc.

[52] Helen Gibson. 2016. Acquisition and preparation of data for OSINT investigations. *Open source intelligence investigation: From strategy to implementation* (2016), 69–93.

[53] Michael Glassman and Min Ju Kang. 2012. Intelligence in the internet age: The emergence and evolution of Open Source Intelligence (OSINT). *Computers in Human Behavior* 28, 2 (March 2012), 673–682. https://doi.org/10.1016/j.chb.2011.11.014

[54] Devu Govardhan, Grandhi Guna Sai Hari Krishna, V. Charan, Sribhashyam Venkata Anantha Sai, and Radhika Rani Chintala. 2023. Key Challenges and Limitations of the OSINT Framework in the Context of Cybersecurity. In *2023 2nd International Conference on Edge Computing and Applications (ICECAA)*. 236–243. https://doi.org/10.1109/ICECAA58104.2023.10212168

[55] Suchin Gururangan, Ana Marasović, Swabha Swayamdipta, Kyle Lo, Iz Beltagy, Doug Downey, and Noah A. Smith. 2020. Don't Stop Pretraining: Adapt Language Models to Domains and Tasks. https://doi.org/10.48550/arXiv.2004.10964 arXiv:2004.10964 [cs].

[56] Nihad A Hassan, Rami Hijazi, Nihad A Hassan, and Rami Hijazi. 2018. The evolution of open source intelligence. *Open Source Intelligence Methods and Tools: A Practical Guide to Online Intelligence* (2018), 1–20.

[57] Sam Havron, Diana Freed, Rahul Chatterjee, Damon McCoy, Nicola Dell, and Thomas Ristenpart. 2019. Clinical Computer Security for Victims of Intimate Partner Violence. 105–122. https://www.usenix.org/conference/usenixsecurity19/presentation/havron

[58] Darren R. Hayes and Francesco Cappa. 2018. Open-source intelligence for risk assessment. *Business Horizons* 61, 5 (Sept. 2018), 689–697. https://doi.org/10.1016/j.bushor.2018.02.001

[59] Jessica He, Stephanie Houde, Gabriel E. Gonzalez, Darío Andrés Silva Moran, Steven I. Ross, Michael Muller, and Justin D. Weisz. 2024. AI and the Future of Collaborative Work: Group Ideation with an LLM in a Virtual Canvas. In *Proceedings of the 3rd Annual Meeting of the Symposium on Human-Computer Interaction for Work*. ACM, Newcastle upon Tyne United Kingdom, 1–14. https://doi.org/10.1145/3663384.3663398

[60] Sam Hepenstal, Leishi Zhang, and BL William Wong. 2021. An analysis of expertise in intelligence analysis to support the design of Human-Centered Artificial Intelligence. In *2021 ieee international conference on systems, man, and cybernetics (smc)*. IEEE, 107–112.

[61] Y. Linlin Huang, Kate Starbird, Mania Orand, Stephanie A. Stanek, and Heather T. Pedersen. 2015. Connected Through Crisis: Emotional Proximity and the Spread of Misinformation Online. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Vancouver, BC, Canada) *(CSCW '15)*. Association for Computing Machinery, New York, NY, USA, 969–980. https://doi.org/10.1145/2675133.2675202

[62] Hilary Hutchinson, Wendy Mackay, Bo Westerlund, Benjamin B. Bederson, Allison Druin, Catherine Plaisant, Michel Beaudouin-Lafon, Stéphane Conversy, Helen Evans, Heiko Hansen, Nicolas Roussel, and Björn Eiderbäck. 2003. Technology probes: inspiring design for and with families. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03)*. Association for Computing Machinery, New York, NY, USA, 17–24. https://doi.org/10.1145/642611.642616

[63] Yong-Woon Hwang, Im-Yeong Lee, Hwankuk Kim, Hyejung Lee, and Donghyun Kim. 2022. Current status and security trend of osint. *Wireless Communications and Mobile Computing* 2022, 1 (2022), 1290129.

[64] Xisen Jin, Bill Yuchen Lin, Mohammad Rostami, and Xiang Ren. 2022. Learn Continually, Generalize Rapidly: Lifelong Knowledge Accumulation for Few-shot Learning. https://doi.org/10.48550/arXiv.2104.08808 arXiv:2104.08808 [cs].

[65] Márcio Romero João Rafael Gonçalves Evangelista, Renato José Sassi and Domingos Napolitano. 2021. Systematic Literature Review to Investigate the Application of Open Source Intelligence (OSINT) with Artificial Intelligence. *Journal of Applied Security Research* 16, 3 (2021), 345–369. https://doi.org/10.1080/19361610.2020.1761737 arXiv:https://doi.org/10.1080/19361610.2020.1761737

[66] Youn-ah Kang and John Stasko. 2014. Characterizing the intelligence analysis process through a longitudinal field study: Implications for visual analytics. *Information Visualization* 13, 2 (April 2014), 134–158. https://doi.org/10.1177/1473871612468877 Publisher: SAGE Publications.

[67] Jihyun Kim, Kelly Merrill Jr., and Chad Collins. 2021. AI as a friend or assistant: The mediating role of perceived usefulness in social AI vs. functional AI. *Telematics and Informatics* 64 (Nov. 2021), 101694. https://doi.org/10.1016/j.tele.2021.101694

[68] Chinmay Kulkarni, Stefania Druga, Minsuk Chang, Alex Fiannaca, Carrie Cai, and Michael Terry. 2023. A Word is Worth a Thousand Pictures: Prompts as AI Design Material. https://doi.org/10.48550/arXiv.2303.12647 arXiv:2303.12647 [cs].

[69] Doug Lederman. [n. d.]. The Latest Campus Clinic: Cybersecurity. Retrieved 2024-09-11 from https://www.insidehighered.com/news/tech-innovation/2023/06/27/cybersecurity-clinics-blend-hands-training-and-community-services

[70] Houjiang Liu, Anubrata Das, Alexander Boltz, Didi Zhou, Daisy Pinaroc, Matthew Lease, and Min Kyung Lee. 2024. Human-centered NLP Fact-checking: Co-Designing with Fact-checkers using Matchmaking for AI. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW2 (2024), 1–44.

[71] Kurt Luther, Casey Fiesler, and Amy Bruckman. 2013. Redistributing leadership in online creative collaboration. In *Proceedings of the 2013 conference on Computer supported cooperative work (CSCW '13)*. Association for Computing Machinery, New York, NY, USA, 1007–1022. https://doi.org/10.1145/2441776.2441891

[72] Cláudio Martins and Ibéria Medeiros. 2022. Generating Quality Threat Intelligence Leveraging OSINT and a Cyber Threat Unified Taxonomy. *ACM Transactions on Privacy and Security* 25, 3 (Aug. 2022), 1–39. https://doi.org/10.1145/3530977

[73] Anirban Mukhopadhyay, Sukrit Venkatagiri, and Kurt Luther. 2024. OSINT Research Studios: A Flexible Crowdsourcing Framework to Scale Up Open Source Intelligence Investigations. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (2024), 1–38.

[74] Michael Muller and Justin Weisz. 2022. Frameworks for Collaborating Humans and AIs: Sequence and Sociality in Organizational Applications. In *CHIWORK*.

[75] Javier Pastor-Galindo, Pantaleone Nespoli, Félix Gómez Mármol, and Gregorio Martínez Pérez. 2020. The not yet exploited goldmine of OSINT: Opportunities, open challenges and future trends. *IEEE access* 8 (2020), 10282–10304.

[76] Sida Peng, Eirini Kalliamvakou, Peter Cihon, and Mert Demirer. 2023. The Impact of AI on Developer Productivity: Evidence from GitHub Copilot. https://doi.org/10.48550/arXiv.2302.06590 arXiv:2302.06590 [cs].

[77] Muhammad Hasban Pervez, Mert Ilhan Ecevit, Najiba Zainab Naqvi, Reiner Creutzburg, and Hasan Dag. 2023. Towards Better Cyber Security Consciousness: The Ease and Danger of OSINT Tools in Exposing Critical Infrastructure Vulnerabilities. In *2023 8th International Conference on Computer Science and Engineering (UBMK)*. 438–443. https://doi.org/10.1109/UBMK59864.2023.10286573 ISSN: 2521-1641.

[78] Sebastian Porsdam Mann, Brian D. Earp, Sven Nyholm, John Danaher, Nikolaj Møller, Hilary Bowman-Smart, Joshua Hatherley, Julian Koplin, Monika Plozza, Daniel Rodger, Peter V. Treit, Gregory Renard, John McMillan, and Julian Savulescu. 2023. Generative AI entails a credit–blame asymmetry. *Nature Machine Intelligence* 5, 5 (May 2023), 472–475. https://doi.org/10.1038/s42256-023-00653-1 Publisher: Nature Publishing Group.

[79] Chris Poulter. 2023. OSINT Workflow with ChatGPT: Tips, Risks, and Benefits. https://www.osintcombine.com/post/osint-workflow-with-chatgpt-tips-risks-and-benefits

[80] Daniela Retelny, Michael S. Bernstein, and Melissa A. Valentine. 2017. No Workflow Can Ever Be Enough: How Crowdsourcing Workflows Constrain Complex Work. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article 89 (Dec. 2017), 23 pages. https://doi.org/10.1145/3134724

[81] Thea Riebe, Julian Bäumler, Marc-André Kaufhold, and Christian Reuter. 2023. Values and Value Conflicts in the Context of OSINT Technologies for Cybersecurity Incident Response: A Value Sensitive Design Perspective. *Computer Supported Cooperative Work (CSCW)* (April 2023). https://doi.org/10.1007/s10606-022-09453-4

[82] Elizabeth B.-N. Sanders and Pieter Jan Stappers. 2008. Co-creation and the new landscapes of design. *CoDesign* 4 (2008), 18 – 5. https://api.semanticscholar.

org/CorpusID:111148589
[83] Advait Sarkar, Ian Drosos, Rob Deline, Andrew D Gordon, Carina Negreanu, Sean Rintel, Jack Williams, and Benjamin Zorn. 2023. Participatory prompting: a user-centric research method for eliciting AI assistance opportunities in knowledge workflows. *arXiv preprint arXiv:2312.16633* (2023).
[84] Ben Shneiderman. 2022. Ensuring Human Control over AI-Infused Systems | National Academies. https://www.nationalacademies.org/news/2022/04/ensuring-human-control-over-ai-infused-systems
[85] Gerry Stahl. 2005. Groups, group cognition and groupware. In *International Conference on Collaboration and Technology*. Springer, 1–16.
[86] Hariharan Subramonyam, Christopher Lawrence Pondoc, Colleen Seifert, Maneesh Agrawala, and Roy Pea. 2023. Bridging the Gulf of Envisioning: Cognitive Design Challenges in LLM Interfaces. http://arxiv.org/abs/2309.14459 arXiv:2309.14459 [cs].
[87] Minhyang (Mia) Suh, Emily Youngblom, Michael Terry, and Carrie J Cai. 2021. AI as Social Glue: Uncovering the Roles of Deep Generative AI during Social Music Composition. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3411764.3445219
[88] Sathya Chandran Sundaramurthy, John McHugh, Xinming Simon Ou, S. Raj Rajagopalan, and Michael Wesch. 2014. An Anthropological Approach to Studying CSIRTs. *IEEE Security & Privacy* 12, 5 (Sept. 2014), 52–60. https://doi.org/10.1109/MSP.2014.84
[89] Siddharth Suri, Scott Counts, Leijie Wang, Chacha Chen, Mengting Wan, Tara Safavi, Jennifer Neville, Chirag Shah, Ryen W. White, Reid Andersen, Georg Buscher, Sathish Manivannan, Nagu Rangan, and Longqi Yang. 2024. The Use of Generative Search Engines for Knowledge Work and Complex Tasks. https://doi.org/10.48550/arXiv.2404.04268 arXiv:2404.04268 [cs].
[90] Sabina Szymoniak and Kacper Foks. 2024. Open Source Intelligence Opportunities and Challenges: a Review. *Advances in Science and Technology Research Journal* 18, 3 (June 2024), 123–139. https://doi.org/10.12913/22998624/186036
[91] Fahimeh Tabatabaei and Douglas Wells. 2017. OSINT in the Context of Cyber-Security. *Open Source Intelligence Investigation: From Strategy to Implementation* (2017), 213–231.
[92] Lev Tankelevitch, Viktor Kewenig, Auste Simkute, Ava Elizabeth Scott, Advait Sarkar, Abigail Sellen, and Sean Rintel. 2024. The Metacognitive Demands and Opportunities of Generative AI. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–24. https://doi.org/10.1145/3613904.3642902
[93] Team-GPT. 2024. Enterprise AI software for teams between 2 and 5,000. Retrieved 2024-09-08 from https://team-gpt.com/
[94] Jakob Tholander and Martin Jonsson. 2023. Design Ideation with AI - Sketching, Thinking and Talking with Generative Machine Learning Models. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference (DIS '23)*. Association for Computing Machinery, New York, NY, USA, 1930–1940. https://doi.org/10.1145/3563657.3596014
[95] Emily Tseng, Rosanna Bellini, Yeuk-Yu Lee, Alana Ramjit, Thomas Ristenpart, and Nicola Dell. 2024. Data Stewardship in Clinical Computer Security: Balancing Benefit and Burden in Participatory Systems. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (2024), 1–29.
[96] Emily Tseng, Mehrnaz Sabet, Rosanna Bellini, Harkiran Kaur Sodhi, Thomas Ristenpart, and Nicola Dell. 2022. Care infrastructures for digital security in intimate partner violence. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–20.
[97] Andrea Tundis, Samuel Ruppert, and Max Mühlhäuser. 2022. A feature-driven method for automating the assessment of osint cyber threat sources. *Computers & Security* 113 (2022), 102576.
[98] Tobias Urban, Matteo Große-Kampmann, Dennis Tatang, Thorsten Holz, and Norbert Pohlmann. 2020. Plenty of Phish in the Sea: Analyzing Potential Pre-attack Surfaces. In *Computer Security – ESORICS 2020 (Lecture Notes in Computer Science)*, Liqun Chen, Ninghui Li, Kaitai Liang, and Steve Schneider (Eds.). Springer International Publishing, Cham, 272–291. https://doi.org/10.1007/978-3-030-59013-0_14
[99] Bart van Dijk and John Zimmerman. 2019. Discovering Users for Technical Innovations through Systematic Matchmaking. *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (2019). https://api.semanticscholar.org/CorpusID:144207663
[100] Sukrit Venkatagiri, Aakash Gautam, and Kurt Luther. 2021. CrowdSolve: Managing Tensions in an Expert-Led Crowdsourced Investigation. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 118 (April 2021), 30 pages. https://doi.org/10.1145/3449192
[101] Sukrit Venkatagiri, Anirban Mukhopadhyay, David Hicks, Aaron Brantly, and Kurt Luther. 2023. CoSINT: Designing a Collaborative Capture the Flag Competition to Investigate Misinformation. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (Pittsburgh, PA, USA) *(DIS '23)*. Association for Computing Machinery, New York, NY, USA, 2551–2572. https://doi.org/10.1145/3563657.3595997
[102] Sukrit Venkatagiri, Jacob Thebault-Spieker, Rachel Kohler, John Purviance, Rifat Sabbir Mansur, and Kurt Luther. 2019. GroundTruth: Augmenting Expert Image Geolocation with Crowdsourcing and Shared Representations. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 107 (Nov. 2019), 30 pages. https://doi.org/10.1145/3359209
[103] Froukje Sleeswijk Visser, Pieter Jan Stappers, Remko van der Lugt, and Elizabeth B.-N. Sanders. 2005. Contextmapping: experiences from practice. *CoDesign* 1 (2005), 119 – 149. https://api.semanticscholar.org/CorpusID:279110
[104] Qian Wan, Siying Hu, Yu Zhang, Piaohong Wang, Bo Wen, and Zhicong Lu. 2024. "It Felt Like Having a Second Mind": Investigating Human-AI Co-creativity in Prewriting with Large Language Models. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (April 2024), 1–26. https://doi.org/10.1145/3637361 arXiv:2307.10811 [cs].
[105] Qiaosi Wang, Koustuv Saha, Eric Gregori, David Joyner, and Ashok Goel. 2021. Towards Mutual Theory of Mind in Human-AI Interaction: How Language Reflects What Students Perceive About a Virtual Teaching Assistant. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3411764.3445645
[106] Justin D. Weisz, Jessica He, Michael Muller, Gabriela Hoefer, Rachel Miles, and Werner Geyer. 2024. Design Principles for Generative AI Applications. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–22. https://doi.org/10.1145/3613904.3642466
[107] Heather J. Williams and Ilana Blum. 2018. *Defining Second Generation Open Source Intelligence (OSINT) for the Defense Enterprise*. Technical Report. RAND Corporation. https://www.rand.org/pubs/research_reports/RR1964.html
[108] Joseph Jay Williams, Juho Kim, Anna Rafferty, Samuel Maldonado, Krzysztof Z Gajos, Walter S Lasecki, and Neil Heffernan. 2016. Axis: Generating explanations at scale with learnersourcing and machine learning. In *Proceedings of the Third (2016) ACM Conference on Learning@ Scale*. 379–388.
[109] B.L. William Wong and Neesha Kodagoda. 2015. How Analysts Think: Inference Making Strategies. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 59, 1 (Sept. 2015), 269–273. https://doi.org/10.1177/1541931215591055 Publisher: SAGE Publications Inc.
[110] Anbang Xu, Huaming Rao, Steven P Dow, and Brian P Bailey. 2015. A classroom study of using crowd feedback in the iterative design process. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*. 1637–1648.
[111] Ashok Yadav, Atul Kumar, and Vrijendra Singh. 2023. Open-source intelligence: a comprehensive review of the current state, applications and future perspectives in cyber security. *Artificial Intelligence Review* 56, 11 (2023), 12407–12438.
[112] Meng Yin, Shiyao Jiang, and Xiongying Niu. 2024. Can AI really help? The double-edged sword effect of AI assistant on employees' innovation behavior. *Computers in Human Behavior* 150 (Jan. 2024), 107987. https://doi.org/10.1016/j.chb.2023.107987
[113] J.D. Zamfirescu-Pereira, Richmond Y. Wong, Bjoern Hartmann, and Qian Yang. 2023. Why Johnny Can't Prompt: How Non-AI Experts Try (and Fail) to Design LLM Prompts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 437, 21 pages. https://doi.org/10.1145/3544548.3581388
[114] Rui Zhang, Wen Duan, Christopher Flathmann, Nathan McNeese, Guo Freeman, and Alyssa Williams. 2023. Investigating AI teammate communication strategies and their impact in human-AI teams for effective teamwork. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (2023), 1–31.

A Appendix

A.1 Prompt templates for OSINT investigsations for cybersecurity vulnerability assessment

A.2 FigJam canvases showing categories and ideas during workshops

Table 4: Curated prompt templates for Planning in Team-GPT. Users are prompted to fill in {{}} placeholders before running the prompts

Phase	Prompt name and purpose	Prompt details
Planning	Division of work: Takes the team size and divides up tasks to perform network discovery, website vulnerabilities, physical assets, third-party services, employee public information, brand monitoring, and breach data in two weeks as a team.	Create a planning template for an OSINT cybersecurity investigation involving 7 services: network discovery, website vulnerabilities, physical assets, third-party services, employee public information, brand monitoring, and breach data. Description of tasks are as follows: Network Discovery - Identify all digital assets (websites, email, servers, webcams) associated with the business. Website Vulnerabilities - Discover vulnerabilities on the business's website such as outdated software, insecure admin panels / logins, files that should not be public, etc. Physical Assets - Identify physical assets like buildings, properties, or vehicles from online resources or images. Third Party Services - Document third party services (vendors, etc.) that the client does business with to identify possible supply chain vulnerabilities. Employee Public Information - Identify key employees from a company using social media platforms, gathering information like job titles, email addresses, or sensitive personal data accidentally shared online. Brand Monitoring - Find instances of negative mentions, disinformation, inappropriate content, legal filings and settlements related to the business. Breach Data - Search for a company's leaked passwords and data in breach data dumps and on the dark web. My team of {{team_size}} is looking to perform the 7 tasks. Divide our workflow into the most efficient subtasks between the {{team_size}} of us. Create a timeline for two weeks for a team of {{team_size}} to work on. The first week is for data collection and preliminary analysis to identify areas to dig deeper. The second week should focus on completing the analysis, incorporating any new information and feedback from reviewers. At the end of the second week, the goal is to generate a set of vulnerabilities and recommendations to present to the small business. Take into account the member's expertise while dividing up the task. Member 1 has preference for the following tasks {{Member_1_preference}}. Member 2 has preference for the following tasks {{Member_2_preference}}. Member 3 has preference for the following tasks {{Member_3_preference}}. Create internal deadlines for the tasks.
	Templates for tasks: Create template for data collection for a particular OSINT task/subtask	You are an OSINT investigator specializing in cybersecurity vulnerability assessments. I am developing a template for an Open Source Intelligence (OSINT) investigation focused on identifying cybersecurity risks within an organization. The investigation requires answering the following OSINT question: {{task_description_or_guiding_question}}. Create a comprehensive template that outlines the steps for data collection, analysis, and documentation related to these questions. The template should include sections for different data sources, how to use the data sources, methods for analyzing the information, risk assessment techniques, and protocols for reporting findings. Additionally, please suggest OSINT tools and resources that can aid in collecting and analyzing information effectively. Format the template and information so that it is within a table.

Figure 4: FigJam board showing the canvas for brainstorming during Design Workshop 1

Table 5: Curated prompt templates for Data Collection in Team-GPT. Users are prompted to fill in {{}} placeholders before running the prompts

Phase	Prompt name and purpose	Prompt details
Data Collection	OSINT Tools and Techniques: Learning about related tools by providing a known tool	What are other examples of OSINT tools like {{tool/website}} ? Provide them in a list format with a description of each tool/website and how to utilize them for the {{OSINT_task}}?
	Process for collecting data: Given a subtask, the prompt generates a plan to collect relevant data	Develop a step-by-step guide for leveraging OSINT (Open Source Intelligence) techniques to gather data on {{OSINT_subtask}}. Include methods for identifying relevant sources and analyzing the gathered data.
	Data collection templates: Each task has a detailed description based on tools and how to document that data	Example for data breach: Develop a detailed template for analyzing and documenting a small business's exposure in data breaches. Focus on searching public breach databases and the dark web for leaked data related to the company. The template should include sections for tool usage (e.g., Have I Been Pwned, BuiltWith, Censys), specific techniques for finding and verifying breaches, and the handling of sensitive information. Describe the process for updating the report with new data and organizing findings (including raw data descriptions and archived links) for effective communication with clients.

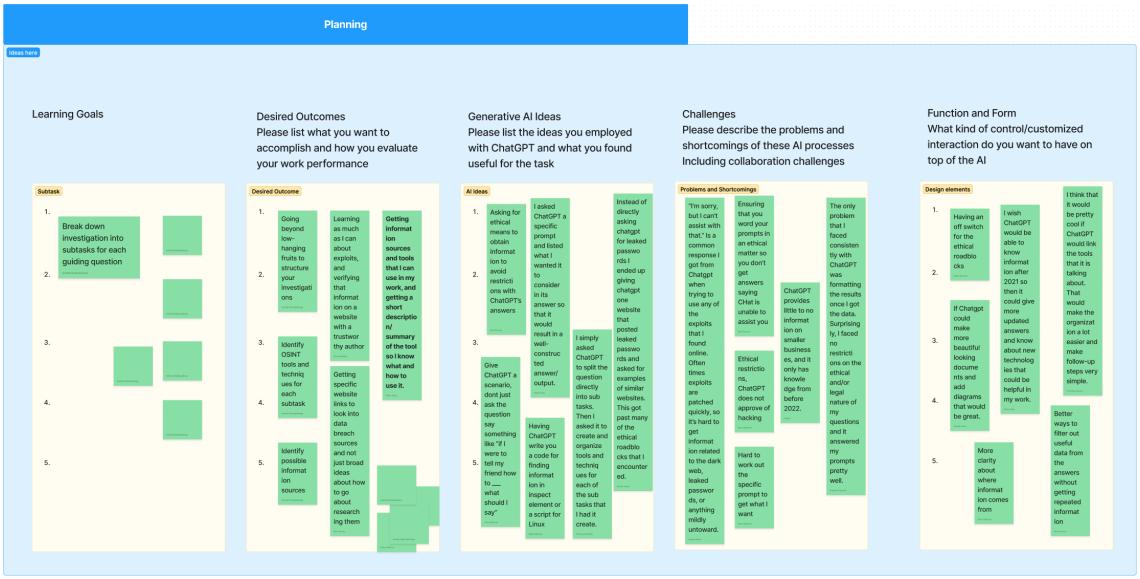


Figure 5: FigJam board showing the canvas and results for Learning Goal 2 during Design Workshop 2

Table 6: Curated prompt templates for Data Processing and Analysis in Team-GPT. Users are prompted to fill in {{}} placeholders before running the prompts

Phase	Prompt name and purpose	Prompt details
Data Processing	Formatting raw data: Put the raw data into a table based on template	Create a table using this template {{template_for_organizing_data}} and organize the information I collected {{raw_data}}
	Finding next steps based on raw information: Enter the broad task like network discovery, website vulnerabilities, physical assets, third-party services, employee public information, brand monitoring, and breach data. Then input the raw data from different tools with a heading for what the data is about. The results describe the information and find areas to dig deeper.	You are an OSINT investigator specializing in cybersecurity vulnerability assessments. Accept redacted information for cybersecurity vulnerability assessment tasks using OSINT. The tasks include network discovery, website vulnerabilities, physical assets, third-party services, employee public information, brand monitoring, and breach data. The description of the tasks are: Network Discovery - Identify all digital assets (websites, email, servers, webcams) associated with the business. Website Vulnerabilities- Discover vulnerabilities on the business's website such as outdated software, insecure admin panels / logins, files that should not be public, etc. Physical Assets -Identify physical assets like buildings, properties, or vehicles from online resources or images. Third Party Services -Document third party services (vendors, etc.) that the client does business with to identify possible supply chain vulnerabilities. Employee Public Information- Identify key employees from a company using social media platforms, gathering information like job titles, email addresses, or sensitive personal data accidentally shared online. Brand Monitoring - Find instances of negative mentions, disinformation, inappropriate content, legal filings and settlements related to the business. Breach Data - Search for a company's leaked passwords and data in breach data dumps and on the dark web. The particular task is {{OSINT task}}. The raw information is {{raw_information}}. Use raw information to find other sources of relevant OSINT information and describe data collection methods that can enhance current results for the task. Describe the data explaining the significant aspects. What are some interesting insights that might be useful for the business client who wants to know about their cybersecurity vulnerabilities. Finally pinpoint areas to look into for future research for that task.
Data Analysis	Interpret results from tools: Describe the results and identify possible vulnerabilities by providing the tool name and results from the tool	Given the {{result_from_tool}} from tool {{tool_name}}, can you explain the concepts and significance of each heading? What are some vulnerabilities that can be reported based on this data?
	Identifying vulnerabilities from data collected: Provide results from tools and raw data to identify possible vulnerabilities	Interpret the redacted OSINT data {{collected_data}} to document how this information can be used to identify vulnerabilities. What vulnerabilities can be presented to a small business based on this information?

Table 7: Curated prompt templates for Dissemination in Team-GPT. Users are prompted to fill in {{}} placeholders before running the prompts

Phase	Prompt name and purpose	Prompt details
Dissemination	Develop recommendations from vulnerabilities: Provide the list of vulnerabilities to generate recommendations for the business	Given the list of website vulnerabilities :{{list_of_vulnerabilities_or_key_findings}}, what are some recommendations you would make as a cybersecurity professional? Remember the client is a small business with no dedicated personnel for cybersecurity and operating on a small budget. Add a page at the end for summarized vulnerabilities and corresponding recommendations. Provide a timeline and more detailed explanations on these recommendations. Also add any other suggestions/disclaimers you would put in a document for an OSINT investigator who is giving this to a stakeholder.
	Iterating on recommendations: Simulate feedback from a small business on the recommendations	Act as a small business owner responding to the information I just gave them as a cybersecurity analyst they hired to look into vulnerabilities in their business. Here's a set of vulnerabilities and recommendations: {{vulnerabilities_and_recommendations}}. Provide feedback that can help with making the recommendations more actionable and complete.