Like a Hammer, It Can Build, It Can Break: Large Language Model Uses, Perceptions, and Adoption in Cybersecurity Operations on Reddit

Souradip Nath[1], Chih-Yi Huang[1], Aditi Ganapathi[1],
Kashyap Thimmaraju[2], Jaron Mink[1], Gail-Joon Ahn[1]
[1]*Arizona State University* [2]*Technische Universität Berlin*

Abstract

Large language models (LLMs) have recently emerged as promising tools for augmenting Security Operations Center (SOC) workflows, with vendors increasingly marketing autonomous AI solutions for SOCs. However, there remains a limited empirical understanding of how such tools are used, perceived, and adopted by real-world security practitioners. To address this gap, we conduct a mixed-methods analysis of discussions in cybersecurity-focused forums to learn how a diverse group of practitioners use and perceive modern LLM tools for security operations. More specifically, we analyzed 892 posts between December 2022 and September 2025 from three cybersecurity-focused forums on Reddit, and, using a combination of qualitative coding and statistical analysis, examined how security practitioners discuss LLM tools across three dimensions: (1) their stated tools and use cases, (2) the perceived pros and cons of each tool across a set of critical factors, and (3) their adoption of such tools and the expected impacts on the cybersecurity industry and individual analysts. Overall, our findings reveal nuanced patterns in LLM tools adoption, highlighting independent use of LLMs for low-risk, productivity-oriented tasks, alongside active interest around enterprise-grade, security-focused LLM platforms. Although practitioners report meaningful gains in efficiency and effectiveness in LLM-assisted workflows, persistent issues with reliability, verification overheads, and security risks sharply constrain the autonomy granted to LLM tools. Based on these results, we also provide recommendations for developing and adopting LLM tools to ensure the security of organizations and the safety of cybersecurity practitioners.

USENIX Symposium on Usable Privacy and Security (SOUPS) 2026.
August 23–26, 2026, Hannover, Germany.

1 Introduction

SOCs play a critical role in protecting organizations against an increasingly complex and fast-moving threat landscape. They combine people, processes, and technology to support a wide range of defensive functions, including continuous monitoring, real-time detection, alert triage, and incident response [1–3]. While many SOC processes have become increasingly automated through tools like SIEM, SOAR [3], and traditional machine-learning (ML)–based techniques [4–8], evidence from both industry and academia indicates that SOC teams remain under significant strain. Industry reports show that SOCs receive thousands of alerts daily, with persistent false positives contributing to severe alert fatigue, stress, and burnout among analysts [9, 10]. Academic studies also highlight the limits of existing automation and the continued need for more effective decision support in the SOC [1, 11, 12].

Against this backdrop, LLMs' generative reasoning capabilities have recently emerged as promising tools for augmenting SOC workflows [13–15]. Over the past years, research efforts have increasingly explored the application of LLM-powered automation to support a range of SOC tasks, from routine analysis to more advanced investigative workflows [16–20]. Capitalizing on this momentum, cybersecurity vendors have begun introducing autonomous solutions, such as Microsoft Copilot for Security [21] and CrowdStrike's Charlotte AI [22], to augment alert triage and incident response processes.

While these advances consistently highlight the transformative potential of LLM tools within SOCs [23], understanding how these tools are used in practice and perceived by practitioners remains critical for assessing their real-world impact and for guiding the effective integration of LLMs into SOCs. Prior to the rise of LLMs, researchers examined the perceptions and challenges of security practitioners regarding traditional ML–based security tools [24, 25]. More recent work has shifted toward task-focused investigations of practitioners' interactions with LLM tools, examining issues such as explainability, autonomy, trust, and human-AI collaboration within specific SOC contexts [13, 14, 26, 27]. While these

studies provide valuable insights into particular tools, tasks, or LLM capabilities, a broader understanding is still needed regarding how LLM tools are used across diverse SOC use cases, what motivating factors and barriers shape their adoption, and the expected impacts on the cybersecurity industry and individual analysts.

To address this gap, we turn to large-scale discourse in online cybersecurity forums. Reddit has been widely used as a rich source of community discussion in security and privacy research [28–31], hosting several large and active cybersecurity-focused communities, with the largest `r/cybersecurity` having over one million members as of January 2026. In light of this, to capture natural, practitioner-driven discussions at scale around the uses, perceptions, and adoption of LLM tools in security operations, we conducted a mixed-method analysis of 892 posts drawn from three active cybersecurity-focused forums on Reddit between December 2022 and September 2025 (inclusive), with the majority originating from `r/cybersecurity`. Based on this analysis, we seek to answer the following research questions:

RQ1 *What LLM tools and use cases are mentioned in security practitioners' discussions of SOC work?*
RQ2 *What benefits and drawbacks of LLM tools do security practitioners discuss in the context of SOC workflows?*
RQ3 *How do security practitioners discuss LLM adoption in SOCs and its potential implications for future practice?*

Our findings reveal key patterns in how security practitioners discuss uses, perceptions, and adoption of LLM tools in cybersecurity operations:

First, practitioner discussions reveal a fragmented awareness of the rapidly expanding ecosystem of security-focused LLM platforms with a small set of popular general-purpose models. Notably, tools such as ChatGPT and Microsoft Copilot, not explicitly designed for security operations, are referenced far more frequently than security-specific commercial tools, such as Security Copilot, Dropzone, or Intezer.

Second, LLM use cases span a broad range of operational SOC activities, with incident response and investigation support emerging as the most prevalent, followed by productivity-oriented tasks such as scripting and reporting. Furthermore, security-specific LLMs are more strongly associated with more autonomous triage and incident response workflows, whereas general-purpose LLMs are more commonly discussed for analyst-driven productivity tasks that afford greater control and easier verification.

Third, practitioner discussions reveal clear gradients of autonomy: LLMs are most commonly used as decision support tools, less frequently embedded in human-in-the-loop triage pipelines, and only rarely granted fully autonomous control over end-to-end mitigation.

Fourth, practitioners' perceptions of LLM tools are shaped by a balance of benefits and concerns across multiple factors. While many report positive experiences with the *capabilities* and *efficiency* of LLM tools in augmenting analyst workflows and reducing workload in specific contexts, these benefits are tempered by strong concerns related to *insufficient reliability*, *security and privacy risks*, *limited autonomy*, and *unjustifiable cost*. Overall, practitioners view LLM tools as *promising yet insufficiently trustworthy* to warrant broad delegation of high-stakes SOC tasks without sustained human oversight.

Fifth, practitioner discussions reveal a nuanced view of the workforce implications of LLMs within SOCs. While many anticipate that LLMs will reduce repetitive entry-level roles, practitioners overwhelmingly reject the idea of fully autonomous SOCs, emphasizing that higher-skill responsibilities, human oversight, governance, and accountability will remain fundamentally human-driven.

Contributions. Our study makes the following contributions:

- We present the **first large-scale mixed-methods discourse analysis of practitioner discussions surrounding LLM adoption in SOCs**. Through a multi-dimensional analysis spanning tools, use cases, adoption patterns, practitioner opinions, and experiences, our study provides a broad empirical view of how practitioners perceive and operationalize LLMs within real-world security operations.
- We surface several **previously unexplored aspects**, including the fragmented ecosystem of security-specific LLM tools, emerging use cases such as knowledge discovery and learning, varying levels of autonomy in LLM-assisted workflows, practitioner curiosity and skepticism, and workforce implications in increasingly AI-augmented SOCs.
- Based on these findings, we derive **broader sociotechnical implications for LLM adoption in SOCs** and identify **research opportunities** surrounding trustworthy AI-assisted workflows, human-AI co-learning, and the long-term sustainability of the SOC workforce.

2 Related Work and Background

Human Factors in Security Operations. Recent work has increasingly examined the human and organizational dimensions of security operations, using complementary methodologies, including qualitative interviews [1, 11, 32], surveys [11, 33], and anthropological analyses of analyst activities and tool usage [25, 34–36]. Collectively, these studies highlight SOCs as complex sociotechnical environments and identify persistent operational issues, including alert fatigue, burnout, stress, and cognitive overload due to high alert volumes and false positives, along with mismatched priorities between analysts and organizational leadership.

Our work revisits these concerns in the context of LLM adoption within SOCs, examining both where LLMs may help alleviate existing burdens and what new challenges emerge as such systems become more widely adopted.

Practitioner Perceptions of ML-based Security Tools. Building on this foundation, subsequent research examined

how ML-based security tools interact with operational SOC dynamics [24, 25]. For instance, Mink et al. [24] conducted a qualitative study examining how practitioners perceive ML-driven detection systems and explanation mechanisms, identifying where such tools are considered effective compared to traditional rule-based approaches. Similarly, Oesch et al. [25] analyzed analysts' interactions with ML-assisted SOC tooling and highlighted challenges related to usability, analyst mental models, and the interpretability of ML-generated outputs. Together, these studies demonstrate that the adoption of AI-assisted security tooling deeply depends on practitioners' ability to understand, calibrate, and operationalize AI outputs.

Our work builds on these prior studies, particularly Mink et al. [24], whose findings inform our initial codebook development (§ 3.2). However, our study extends the *scope* and *technological focus* by studying practitioners' perceptions of LLM-based tools, which introduce new forms of interaction, workflows, and capabilities that fundamentally reshape how AI systems are integrated into security operations.

LLMs and Human-AI Collaboration in SOCs. More recently, research has begun examining the role of LLMs and human-AI collaboration within SOCs [37–39], touching upon specific areas, such as autonomy, trust, interaction patterns, and explainability [14, 26, 27]. For example, Roch et al. [26] conducted in-depth interviews with cybersecurity leaders to explore autonomy and the evolving division of labor between analysts and AI. Similarly, Rastogi et al. [27] examined the role of explainability, highlighting how LLM-generated explanations influence trust and actionability in high-stakes SOC environments. Other recent studies have explored task-specific uses of LLMs within SOCs, such as evaluations of LLM-assisted incident response summarization using real-world incidents [13]. Among these, Singh et al. [14] is particularly relevant to our work, as it empirically examines analyst interactions with LLM systems in SOC settings. However, they focus on general-purpose LLMs such as ChatGPT, whereas our study additionally examines security-specific LLM platforms and more autonomous operational workflows.

Our work complements and extends this growing literature through its *methodology*, *scale*, and *analytical breadth*. Rather than focusing on a specific tool, task, or AI capability, we conduct a large-scale multi-dimensional discourse analysis of practitioner discussions across online security communities. This approach captures focused yet candid conversations surrounding LLM adoption, operational use, organizational concerns, autonomy, trust, and perceived risks across a diverse range of tools and SOC use cases. By analyzing practitioner discourse at scale, our work provides a broader view of how security practitioners collectively negotiate the emerging role of LLM systems within real-world security operations.

Terminology. Next, we define the key terms related to data analytics and insights that are used throughout the paper.

LLM Tools. In this paper, the term 'LLM tools' is used as an umbrella term to refer to a group of standalone tools, compound systems, or commercial platforms in which LLMs play a central role, enabling generative to agentic capabilities, such as reasoning, planning, and learning.

Practitioners. In this paper, we refer to Reddit posters collectively as 'security practitioners,' as the analyzed communities are cybersecurity-focused. When available, we report more specific, self-identified roles (e.g., SOC manager) stated in posts or user flairs. Because these roles cannot be independently verified (§ 3.3), we report them only when explicitly self-identified; otherwise, we use the term 'practitioner.'

Reddit Terms. Throughout this paper, we use three Reddit-specific terms: *subreddit*, *thread*, and *post*. Reddit is organized into topic-specific communities called 'subreddits,' within which users participate in discussion threads. A 'thread' refers to a user-initiated discussion consisting of a title, a conversation prompt, and all subsequent replies. Following prior works [29, 30], we use the term 'post' to refer to any individual contribution within a thread, including both the original submission and any subsequent replies.

3 Methodology

To investigate how LLMs are used and perceived in practice, we perform a large-scale qualitative analysis across online cybersecurity forums discussing LLM tools. We detail the ethical considerations for our IRB-approved study in Section 8.

3.1 Data Collection

To accurately capture real-world security practitioners' uses and perceptions of LLM tools, we gathered and analyzed 1,703 posts made across online cybersecurity forums.

Discovering Subreddits. Inspired by prior work analyzing Reddit data [28–31], we selected a set of popular and active subreddits that enable discussions among security practitioners. To discover security-centric discussions on emerging LLM tools, we began with informal keyword-based searches on Reddit to identify subreddits that yielded relevant practitioner discussions, and then expanded this set by including closely related, active communities with similar topical focus. We selected active forums covering general cybersecurity discussions (`r/cybersecurity`, `r/ComputerSecurity`), cybersecurity leadership (`r/ciso`), blue-team and SOC operations (`r/blueteamsec`, `r/SIEM`), and cyberspace Q&A (`r/CyberSecurityAdvice`). Furthermore, we read and adhered to each forum's posted community guidelines and avoided any communities that prohibited the use of forum data for academic research. In total, we gathered a set of nine subreddits that spanned between 4.5K–1.4M users (Table 1).

Gathering Threads and Posts. Within each subreddit, we looked at how LLMs were perceived and used by its members. First, we performed a keyword-based search to collect and filter threads; with a set of 13 AI-focused terms

Table 1: **Collected Reddit Dataset** – number of threads retrieved per subreddit, number and share of relevant threads, and the number (and share) of posts/comments coded as relevant.

Subreddits	Threads (Pulled)	Threads (Relevant)	Posts (Pulled)	Posts (Relevant)
r/cybersecurity	830	69 (8.31%)	1,665	863 (51.83%)
r/Information_Security	96	5 (5.21%)	21	15 (71.43%)
r/ciso	25	2 (8.00%)	17	14 (82.35%)
r/blueteamsec	187	0	–	–
r/CyberSecurityAdvice	76	0	–	–
r/cybersecurity_help	36	0	–	–
r/ComputerSecurity	35	0	–	–
r/SecurityBlueTeam	15	0	–	–
r/SIEM	10	0	–	–
Total	**1,310**	**76 (5.80%)**	**1,703**	**892 (52.38%)**

including '`SOC AI`', '`AI Security Operations`', and '`LLM in cybersecurity`' (full list in [40]), we retrieved 1,310 threads on September 30, 2025, collecting the submission's title, content, anonymized username, posting date, and number of posts. We further refined the set of collected threads using AI-assisted classification. Using `GPT-4.1-mini` API, we created a few-shot prompt to guide the LLM in performing a relevance classification task, supplying it with the thread title and the original post content [40]. To establish reliability, the primary author manually reviewed a random sample of 100 threads, labeled each as 'relevant' or 'irrelevant'. Out of 100 threads, only 2 were incorrectly coded, and no relevant threads were missed, yielding a high inter-rater reliability (κ=0.96). Having found reliability, we then used the LLM to classify the remaining threads at scale, yielding 80 relevant and 1,230 irrelevant threads. All relevant threads were also further manually reviewed, and 4 were deemed irrelevant, giving us 76 relevant threads. We also randomly sampled 50 threads classified as irrelevant, and manually verified that all the classifications were correct; these threads focused on topics outside the scope of this study, including frameworks for securing AI systems, cybersecurity career pathways, and interview preparation, or syntactically similar but unrelated topics, e.g., SoC (System-on-chip) Security.

Within each thread, we retrieved all posts underneath it, producing a dataset of 1,703 posts as shown in Table 1. These posts were then manually coded for relevance following our coding protocol (§ 3.2). Specifically, posts were double-coded in batches of 50 with iterative inter-rater reliability (IRR) checks and conflict resolution until a high level of agreement was achieved (α=0.88). A post was deemed irrelevant if it had been deleted by the creator, removed by moderators or the platform, or if it did not pertain to our research questions. Additionally, redundant replies from the same user that repeated already shared opinions without contributing new information were also labeled irrelevant. Of the 1,703 posts coded, 892 were considered relevant (see Table 1).

Data Availability Statement. Consistent with research practices involving Reddit data [28–31], we will not publicly release the dataset to protect the privacy of Reddit posters. Instead, interested researchers may contact the author to request an archived copy. To promote reproducibility, we provide the list of keywords and the few-shot prompt used in our data collection process in our replication package [40].

3.2 Data Analysis

To analyze online discussions, we employed a mix of qualitative and quantitative methods [41] to gain insight into the discussions and statistically compare the prevalence of different LLM tools use cases and opinions.

Qualitative Analysis of Posts. To analyze the gathered posts, we first conducted a hybrid coding methodology [42]. To ensure proper interpretation, all posts were analyzed in the context of any posts that they were in reply to. If any posts contained available external links, as in other works [28], we reviewed all linked content as well.

To construct the codebook, a single coder began with a set of codes informed by prior work [24], from which initial use cases (alert triage, threat analysis, documentation) and AI factors (efficiency, cost, security) were derived. This codebook was then expanded through inductive coding of 350 posts to capture LLM-specific practitioner discussions, which introduced emergent use cases, such as scripting, query support, knowledge support, and training, as well as additional LLM factors, including capabilities, reliability, autonomy, and privacy. To assess IRR for this initial codebook, a second coder was introduced and familiarized with the codebook through co-coding of 25 posts. After developing a shared understanding, the two coders then independently coded a new set of 50 posts, after which IRR was computed using Krippendorff's alpha (α) [43, 44] for every code. The authors then resolved coding disagreements and adjusted the codebook to clarify and refine definitions. If at least one code did not obtain substantial agreement ($\alpha \geq 0.8$), this process was repeated. After eight rounds of double coding, agreement for all codes was achieved, and the primary coder then independently coded the remaining 1,278 posts. The final codebook and resulting reliability are presented in our replication package [40].

Following codebook development, the primary coder, along with the full research team, conducted a reflexive thematic analysis [45]. Through data exploration and routine meetings with the research team, we distilled patterns and trends in the codes into large-scale themes that we present in our results.

Quantitative Analysis of Codes. Having established the reliability of codes, we also conducted quantitative analysis to measure and statistically compare their prevalence.

In particular, to assess the prevalence of certain LLM tools (§ 4.1) and use cases (§ 4.2) discussed, we conducted pairwise comparisons using two-sample tests for equality of proportions [46], which assume sufficiently large sample sizes for

the asymptotic approximation underlying the chi-squared test statistic to be valid [47]. Also, to examine associations between categorical dimensions in our dataset, we conducted a series of chi-square tests of independence across multiple analyses, including relationships between tool categories and use cases (§ 4), LLM factors and opinions, tool categories and opinions (§ 5), and adoption stages over time (§ 6). For statistically significant results, we conducted post hoc analyses using adjusted standardized residuals (z) [48] to identify disproportionately represented categories. To account for multiple comparisons, we applied the Bonferroni correction to the post hoc tests [49].

3.3 Limitations

Similar to other works that analyze anonymous community forums [28–31], our study shares a number of limitations that we account for when interpreting our results. *First*, our selection of forums and keywords, while carefully curated, may reflect biases and capture only a subset of relevant discussions about practitioners' experiences with LLM tools in SOCs. *Second*, we rely on self-reported information provided by Reddit users. Although we report specific roles when explicitly stated in posts, we cannot independently verify the roles, professional identities, or organizational contexts of the contributors, which may substantially influence how certain tools or factors are perceived and prioritized. *Third*, Reddit users represent only a subset, and possibly a biased section of the security community. Prior studies have noted that Reddit users tend to be more engaged with emerging technologies than the general practitioner population [50]. As such, these discussions may reflect an upper bound on awareness of LLM tools rather than typical practice.

Despite these limitations, our work surfaces nuanced practitioner perspectives and timely insights into emerging LLM adoption trends in SOCs, which can be further validated through empirical and design-oriented methodologies.

4 Tools and Use Cases in Practice (RQ1)

We find that security practitioners discuss a wide range of LLM tools and uses within SOC workflows. In particular, these discussions reflect a dominant mention of general-purpose LLMs alongside a fragmented ecosystem of commercial tools for security. Moreover, LLM use cases span a diverse set of SOC activities, with discussions most heavily concentrated on incident response and triage, followed by scripting and reporting tasks, and less frequently on knowledge support, threat analysis, and training-related functions.

4.1 LLM Tools

Across 410 posts, practitioners discussed a wide range of LLM tools used within SOCs. To understand the tools used, we code each by whether it is (1) general-purpose, such as ChatGPT, or security-specific, such as Microsoft Security Copilot, and (2) commercially available or built in-house.

Table 2: **Reported LLM Tools** – Categorized by purpose and origin. Note that individual posts may mention multiple tools.

Categories of LLM Tools	# (Out of 410)
By Purpose	
General-Purpose LLM Tools	248 (60.49%)
Security-Specific LLM Tools	180 (43.90%)
By Origin	
Commercially Available	398 (97.07%)
In-House Built	14 (3.42%)

General-Purpose LLMs Are Referenced More Than Security-Focused LLMs. As shown in Table 2, while not explicitly designed for security operations, 60.5% (n=248) of posts referenced general-purpose LLMs, such as ChatGPT, Microsoft Copilot, Claude, along with broader terms such as "LLMs," or "Generative AI." Conversely, only 43.9% (n=180) of tools references mentioned security-specific tools, including Microsoft Security Copilot, Dropzone AI, and Intezer. To evaluate whether these differences are significant, we ran a two-sample test for equality of proportions, and found that general LLM tools are significantly more likely to be discussed by security practitioners ($\chi^2(1) = 21.94$, $p < .001$). Across these discussions, practitioners described diverse integration patterns (§ 4.2), alongside sharing hands-on reflections on their benefits and limitations (§ 5).

A Long Tail of Security-Specific Tools Exists. While general-purpose LLMs were centered around a few key players, a long list of security-specific tools was referenced (see Appendix B.1). Practitioners mentioned only *nine* distinct general-purpose tools, with discussion dominated by ChatGPT (n=90), followed by MS Copilot (n=30), and others. In contrast, security-focused LLMs had a long list of 30 distinct commercial platforms, yet only four of these tools were mentioned more than five times: Security Copilot (n=40), Dropzone AI (n=10), Intezer (n=8), and Cortex XSIAM (n=6). The remaining appeared three times or fewer, with half of the 30 tools mentioned only once, indicating fragmented awareness across a rapidly expanding AI-for-cybersecurity vendor ecosystem. Notably, half of these tools are marketed as "AI SOC analysts" (further discussed in § 6.1) positioned as autonomous assistants capable of performing Tier 1/2 tasks [23].

In-House Custom Workflows Are Rare, but Used. In-house developments accounted for only 3.4% (n=14) of posts, compared to the commercially available solutions referenced in 97% (n=398) of posts. Across this smaller subset, practitioners described building custom LLM workflows internally to augment security workflows. For example, in a thread about LLMs for SIEM, one practitioner (P310) described leveraging open-source agent frameworks to enhance SIEM search

Table 3: **Reported LLM Tools Use Cases and Descriptions Across Tool Categories (General-Purpose, Security-Specific).**

Types of LLM Use Cases	Description	# General	# Security	p-value
Triage & Incident Response (n=139)	*Support for alert triage, investigation, correlation, mitigation, and response workflows.*	**30 (40.54%)**	**44 (59.46%)**	**< 0.001***
Scripting & Query Support (n=88)	*Generation and refinement of security scripts and detection queries.*	**51 (89.47%)**	**6 (10.53%)**	**< 0.01***
Reporting & Documentation (n=84)	*Drafting and summarizing investigation reports, policies, and threat briefings, etc.*	**43 (89.58%)**	**5 (10.42%)**	**< 0.05***
Threat Analysis (n=54)	*Support for risk and vulnerability analysis, threat hunting, and modeling workflows.*	16 (57.14%)	12 (42.86%)	0.948
Knowledge Support (n=53)	*Concept explanation, information retrieval, document-based knowledge extraction.*	31 (88.57%)	4 (11.43%)	0.119
Training, Compliance, Others (n=18)	*Drafting training exercises, reviewing compliance materials, and other non-core tasks.*	9 (81.82%)	2 (18.18%)	–

Counts and percentages are reported row-wise and represent the relative distribution of tool types within each use case category. Statistical significance was assessed using a chi-square test of independence followed by Bonferroni-corrected post-hoc analysis of adjusted residuals. Rows denoting significant associations are bolded.

workflows: *"I have developed custom Python scripts using open-source agent frameworks to feed SIEM query results into an agent to iteratively process the data and generate a consolidated review of the search output."* This suggests that although off-the-shelf solutions are popular, such custom workflows may be valuable for closing integration gaps and supporting organization-specific security processes.

4.2 Uses of LLM Tools in Security Operations

As shown in Table 3, across 325 posts, LLM use cases fell into six broad categories: triage and incident response, scripting support, reporting, threat analysis, knowledge support, and miscellaneous activities. To understand the relative prevalence across use cases, we conducted pairwise two-sample tests of proportions (full statistical results in Appendix B.2) and found that discussions surrounding triage and IR were significantly more prevalent than all other use cases, followed by scripting and documentation. Threat Analysis and knowledge support were discussed less frequently, but were more prominent than training, compliance, and other categories.

Furthermore, we conducted a cross-dimensional analysis linking the use cases with LLM tool categories (general-purpose or security-specific) and found that practitioners mentioned a specific tool across 60% of use case discussions (n=193). A chi-square test of independence (Appendix B.2) revealed a significant association ($\chi^2(4) = 58.289$, $p < 0.01$); security-specific tools were more strongly associated with triage and incident response workflows, where applications tended to be more autonomous, whereas general-purpose tools were more commonly discussed for analyst-driven productivity tasks such as scripting, query support, and reporting. These findings suggest a functional distinction between productivity-oriented uses of general-purpose LLMs and operational automation expectations surrounding security-specific systems.

Incident Response and Investigation Support. Across all LLM cybersecurity use cases, 42.77% (n=139) of posts focused on improving or automating tasks relating to incident response, including triaging, investigating, responding, and mitigating alerts and security threats, and was thus mentioned significantly more often than all other reported use cases ($\chi^2(1) \geq 16.92$, all $p < .001$). Practitioners described a varied set of ways for LLM tools to aid in reducing their workload and making investigations more effective.

LLM Investigation Buddies: Similar to common uses of LLMs in non-cybersecurity contexts, practitioners described using LLMs as decision support tools during investigations, operating in a 'think and assist' mode that performs subtasks such as correlation, hypothesis generation, and sensemaking, while analysts retain full control over interpretation and final decision making. As P452 shared: *"During a complex investigation, I utilize LLMs as a companion, throwing ideas at them. I let them examine the data to help with correlation, and I handle the critical thinking."* Similarly, P757 described providing contextual artifacts to LLMs for exploratory analysis, noting: *"I provide logs, screenshots, or event timelines to LLMs to help me piece together what might be happening, either to validate my findings, or help me zoom in on a problem area."*

LLM-Driven Triaging: Beyond providing investigative support, practitioners described using LLMs to partially automate alert triage through human-in-the-loop (HITL) workflows. In these settings, HITL means workflows where LLMs may perform an initial analysis of incoming alerts, proposing likely scenarios and potential mitigation actions, but human operators review and approve or reject LLM decisions before any action is executed. Such LLM-driven triage was most commonly discussed for high-volume 'Tier-1' alerts, where rapid filtering decisions are required, though some accounts also described its use in more in-depth analytical contexts. For instance, one L3 analyst (P725) described an LLM-driven pipeline that automatically extracts context from EDR alerts and presents a structured breakdown for analysis to ultimately review and take action on: *"While we use it to augment the analysis and get more clarity on things, we do not allow it to take actions. Ultimately, trust but verify!"* Others, however, described allowing LLMs to handle and resolve low-impact alerts, escalating to humans only when necessary: *"Our team has been using LLMs to assess high-noise/high-volume alerts that are low-payoff/low-impact. Only if there are any outliers, the alert is flagged for human evaluation"* (P864).

Fully Autonomous LLM Mitigation: Lastly, a small fraction of posts (n=5) described delegating entire classes of routine high-confidence tasks to fully autonomous LLM pipelines. Commonly integrated with SOAR systems, these workflows

independently triage alerts, correlate incidents, enrich findings, and initiate actions with minimal human involvement (P013, P066, P247). In a few cases, this autonomy extended into active responses; for instance, a CISO (P086) detailed how their LLM-powered correlation engine *"determines incident severity and autonomously dispatches remediation tasks, like isolating endpoints, running full scans, or disabling suspicious accounts."* These use cases demonstrate that, although limited, some organizations might be experimenting with fully autonomous workflows to scale high-volume alert workloads.

Scripting & Query Support. Consistent with recent advances in LLM-based code generation [51–53], 27.08% (n=88) of use cases described practitioners using LLMs to generate scripts and queries to augment their daily workflows.

Code Generation: Several posts (n=46) discussed using tools such as ChatGPT to draft scripts in Python, PowerShell, or Bash, typically to create boilerplate code or outlining logic that the analyst then adapts. As P753 explained, *"I find "write code to do XYZ in language A" prompts are useful for getting started with boilerplate code instead of writing it manually."* Others mentioned leveraging LLMs to enhance or debug scripts they wrote themselves (P742).

Query Support: Practitioners also mentioned LLM-assisted query generation and refinement (n=51), particularly for SIEM and detection engineering workflows. These cases included assistance with crafting or troubleshooting KQL, SQL, or regex queries. As P309 noted, *"I've explored a number of LLM chatbots to assist me fix SIEM queries, offer recommendations, or craft highly targeted queries for specific needs."*

Reporting & Documentation. Across 25.85% (n=84) of mentioned tasks, practitioners described incorporating LLMs into their technical reporting and documentation workflows, supporting tasks ranging from *general writing assistance* (n=20) to SOC-specific activities such as *policy and SOP development* (n=19), *investigation summaries* (n=13), *threat intelligence briefings* (n=11), and *risk assessment reports* (n=9). As one mentioned using LLMs to offload time-intensive reporting tasks: *"I use ChatGPT to ingest articles and generate summaries of CTI briefings. This allows me to save time and concentrate on threat hunting for any IOCs"* (P823).

Learning & Knowledge Support. Practitioners also described using LLMs as versatile knowledge-support tools (n=53, 16.31%), reflecting a broader shift in how analysts retrieve, learn, and navigate complex technical information.

Learning and Explanation: Across 20 posts, practitioners mentioned relying on LLMs to clarify obscure command-line behavior or provide context for vague error messages, highlighting the value of LLM-generated explanations for learning something new or complex. As P921 noted, *"having an LLM explain something as if you were five may seem trivial, but can be incredibly helpful when approaching complex topics."*

Information Retrieval: In 17 posts, LLMs were framed as a more efficient path from questions to actionable directions, often contrasted with a conventional Google search. As P442, a threat hunter, described *"the role of an analyst is to know the right question to get the answer. Google has been a part of our lives for many years. LLMs are simply far more efficient."*

Knowledge Extraction: Finally, practitioners discussed utilizing LLMs to extract knowledge from documents (n=13) to receive targeted responses. For example, P740 described embedding internal materials into an LLM agent: *"I created an agent loaded with our client policies, risk reports, and acronym lists to allow our analysts to ask questions or perform light analysis"* (P740).

Threat Analysis. Around 16.62% (n=54) of posts discussed the use of LLMs to support analytical reasoning in activities such as risk and vulnerability assessment, threat modeling, and threat hunting. For instance, P879 described using LLMs to help plan and review risk assessments, noting: *"We have been augmenting LLMs extensively to our risk assessment planning to ensure completeness."* Others reported using LLMs to sanity-check penetration testing findings, such as asking: *"What description and CVSS score would you assign to my pentest discovery?"* (P751), to save time by offloading the analytical thinking to the LLM.

Training, Compliance, and Others. A smaller subset of posts (n=18, 5.54%) described LLM use for miscellaneous tasks, including drafting cyber training exercises and developing scenarios to test defensive controls (n=6), reviewing policies or guideline reports (n=6), and other non-technical activities. As P823 noted: *"I train junior analysts on IR, I use ChatGPT to generate training scenarios and practice questions."*

5 Perceptions of LLM Tools (RQ2)

Across 406 posts, practitioners shared opinions and experiences with LLM tools for cybersecurity operations. Based on our analysis, we find that practitioners framed their experiences primarily around *six* factors: capabilities, efficiency, reliability, security and privacy, autonomy, and cost (Table 4). To further examine whether sentiment toward LLM tools differed across the discussed factors, we conducted a chi-square test of independence followed by post-hoc analysis using adjusted residuals (full statistical results in Appendix B.2). Our findings provide both statistical and qualitative evidence that, although practitioners are increasingly satisfied with the capabilities and efficiency of LLM tools, persistent concerns around reliability, autonomy, security, and cost temper their trust and limit their willingness to delegate sensitive tasks.

5.1 Capabilities

Practitioners' discussions around LLMs' capabilities were significantly more positive than negative ($z = 8.11, p < .001$). Notably, nearly three-quarters of these posts (n=130) also

Table 4: **Reported LLM Factors, Descriptions and Opinions.**

Factors	Description	# Positive	# Negative	p-value
Capabilities	*Specific tasks and SOC use cases where LLM tools are perceived as helpful*	**174 (64.21%)**	**97 (35.79%)**	**< 0.001***
Efficiency	*Time-related aspects, such as speed of analysis, deployment, and workload reduction*	**54 (84.38%)**	**10 (15.62%)**	**< 0.001***
Reliability	*Trustworthiness, accuracy, and consistency of LLM-generated insights*	**3 (5.17%)**	**55 (94.83%)**	**< 0.001***
Security & Privacy	*Security of LLM tools, organizational data governance, and privacy considerations*	**6 (11.54%)**	**46 (88.46%)**	**< 0.001***
Autonomy	*Ability of LLM tools to operate independently with minimal human supervision*	**6 (12.77%)**	**41 (87.23%)**	**< 0.001***
Cost	*Financial aspects, such as operational costs and perceived return on investment*	**5 (14.29%)**	**30 (85.71%)**	**< 0.001***

Counts and percentages are reported row-wise and represent the relative proportion of opinions within each LLM factor. Statistical significance was assessed using a chi-square test followed by Bonferroni-corrected post-hoc analysis of adjusted residuals. Rows denoting significant associations are bolded.

included specific use cases (§ 4.2), highlighting that LLM-powered tools can effectively augment real-world SOC tasks.

Better Contextualization and Visibility of Alerts. Practitioners frequently described LLM tools as effective for incident enrichment and surfacing low-visibility behaviors (n=13). While existing ML tools can correlate well-defined alerts, practitioners noted that LLM tools can effectively interpret a multitude of varied signals to better answer an investigation, which can also provide essential context for analyst interpretation. For instance, reflecting on experiences with Agentic LLM tools such as Purple AI, one practitioner noted that, *"LLMs are going to transform incident enrichment. Compared to scripting or googling, LLMs can rapidly synthesize key contextual details surrounding an incident"* (P044). Similarly, P038 noted that these tools dramatically reduce their workload, and provide important context and remain easily accessible with natural language queries: *"I use Agentic LLMs, and they're incredible. Not only does it turn half a million alerts/events a month into 1-3 relevant daily alerts, but also gives important context. During an investigation, I can ask, 'Does this event fire every time a user logs in or is this a new alert?' or, 'Did they get challenged for MFA?'... It's not just reducing my workload, it's finding things I couldn't see."*

Improved Interpretability of Signals. In addition to better contextualization of signals, practitioners (n=10) also noted that LLMs made previously hard-to-understand signals readily interpretable. In particular, practitioners emphasized how LLMs can translate verbose or opaque log data into human-readable narratives through querying, summarization, and explanations. As P294 explained, *"When you feed ChatGPT raw logs, it can assist in explaining what's happening in simple language. Windows events can be difficult to interpret due to the event IDs and combinations you need to memorize."*

Reduced False Positives. Perhaps due to their effective use and analysis of many signals, practitioners also consistently framed LLM tools as effective in performing accurate classification decisions and automating alert triage and management pipelines (n=18). In particular, while concerns of high false negatives were notable in traditional alert classification systems [1, 11, 32], and thought to be further exacerbated in traditional ML systems [24], LLM tools were surprisingly perceived by analysts as broadly effective in detecting and re-emptively removing false positive alerts. For instance, P017 noted: *"The 'autonomous' aspect of LLM tools is most evident in their ability to eliminate high-confidence false positives that are not worth wasting an analyst's time."*

Inabilities of LLMs. Beyond the predominantly positive discussions of LLM capabilities, 97 posts (38.65%) expressed reservations about their practical value in SOC contexts. A recurring concern was whether the probabilistic nature of LLM tool outputs was fundamentally susceptible to incorrect conclusions, particularly around novel and unseen examples. As P910 argued, *"Cybersecurity mostly depends on dealing with outliers and anomalies... LLMs are awful at even comprehending these things, let alone acting upon."* Skepticism also stemmed from firsthand experiences with performance breakdowns in complex or structured tasks, including failures to correctly interpret alert data or generate functional detection queries in complex languages, such as KQL. Some posts emphasized systemic data challenges affecting LLM use at both training and inference stages. As P263 emphasized that alert data already suffers from poor signal-to-noise ratios, suggesting that, *"...with AI, the problem of garbage in, garbage out still persists."* Furthermore, these issues were compounded by vendors overpromising LLM tools' abilities, the perception that existing tools may be good enough (§ 6.2), and concerns that even outputs for tasks that LLMs can readily perform can also still suffer from hallucinations and inaccuracies (§ 5.3).

5.2 Efficiency

Practitioners were significantly more likely to speak positively about the efficiency of LLM tools ($z = 6.38, p < .001$), frequently noting that LLM tools speed up their analysis.

LLM Tools Speed Up Analysis and Deployment. Within core SOC workflows such as alert triage, investigation, and IR, practitioners frequently reported measurable efficiency gains from using LLM tools in their workflow. For instance, an L3 analyst (P725) described how using an LLM tool in their workflows helped them automate their triaging and *"reduced MTTT (mean-time-to-triage) from around 45 minutes to less than 2 minutes."* Similarly, the use of Agentic LLM tools drastically reduced how long investigations took by pre-emptively

filtering and analyzing alerts they needed to respond to: *"By delivering end-to-end investigations, these tools condense the investigation from "here's 500 things to look at", to "here's what happened and what you should probably do""* (P085). Practitioners also noted that, compared with human analysts in particular, the detection and response time with LLM tools were often much faster for processing and resolving large volumes of alerts, with P870 noting: *"AI can process massive volumes of data quickly and detect threats that would take a human analyst hours to identify."* Furthermore, beyond the speed of the tool's operation, practitioners also noted that the time required to effectively deploy LLM tools was lower than that of others. This was discussed in both how fast analysts could learn to use (P002) and set up (P001) such tools.

LLM Tools Introduce Verification Overhead. In contrast to their time-saving abilities, a small set of posts (n=10) highlighted that LLM tools introduce new overhead as analysts may often need to correct, tune, or otherwise verify their outputs. For example, P206 expressed frustration after their SOAR was replaced with an LLM tool-driven workflow: *"We spend nearly four times the amount of effort correcting and changing LLM behavior as we do actually addressing incidents."* Similarly, P289 reflected after using LLMs for generating regular expressions for firewall rules: *"I had to verify and correct the output, so did I end up saving time? Not sure."*

5.3 Reliability

When discussing the trustworthiness of LLMs' predictions, practitioners were significantly more likely to discuss negative aspects ($z = -6.77, p < .001$) such as inconsistent behavior and false confidence in hallucinated answers.

Hallucinations and Non-determinism Cause Concern. Most posts discussing reliability focused on LLM tools' tendencies to hallucinate [54] (n=25). Often, these were discovered through personal experiences; for instance, P171 noted that an LLM tool's confident but incorrect conclusion still makes them worry about relying on such tools: *"Whenever I've tried LLMs for security work, it's produced pure garbage. It made up descriptions of an imaginary malware I named. I do not trust any of it."* Beyond arriving at incorrect conclusions, practitioners noted that LLMs could also fabricate justifications and *"hallucinate evidence to prove it"* (P120). For many, this felt antithetical to the idea of cybersecurity; as noted by P924, *"LLMs can be so confident in wrong answers... in the security space, these 'small' mistakes may create cascading security risks by weakening multiple layers of defense."* Furthermore, for others (n=4), their concerns were not just based around incorrect answers, but the lack of determinism around LLM outputs. Consistent with previous studies highlighting this issue [55, 56], practitioners such as P058 noted: *"The problem with implementing autonomous AI solutions for security is the unpredictable nature of the outputs."* Practitioners also described how reliability can vary based on the specific task and its context; for instance, the particular language in which the code is generated: *"When writing code, outcomes get increasingly unreliable as the code language gets more complex, like with PowerShell"* (P175).

5.4 Privacy and Security

When discussing privacy and security risks of LLM tools, practitioners were significantly more likely to bring up negative viewpoints ($z = -5.41, p < .001$). These concerns often focused on unintentional data leakage (n=31), and the expanded attack surfaces LLMs may introduce (n=17).

Difficulties in Data Governance. Practitioners frequently expressed concern that users may inadvertently enter organizational information into commercial LLMs, raising the risk that public models could retain or learn from sensitive inputs (e.g., P783, P827, P789). As P294 cautioned: *"LLMs may learn from the information you provide, be cautious about what you prompt; otherwise, sensitive organizational information may be inadvertently exposed."* These concerns were particularly pronounced around integrated tools that require access to internal enterprise data. As P863 shared, while their teams were excited by the success of early LLM tools, they remained *"uncertain about the tool's access to internal data"* and felt a tension between giving the tool access to be *"useful without that access being a huge risk to themselves."* Some practitioners questioned the necessity of such access and became worried that it may be misused or stolen: *"It sounds like what all AI companies are trying to do: get as much data to train models that they can sell to other customers"* (P885).

LLMs Increase Attack Surfaces. Practitioners also raised concerns that LLM tools themselves become new attack surfaces. In the excitement to produce and sell LLM tools for security, practitioners noted that security concerns of the tools themselves may not be prioritized: *"Be careful, security with new technology often lags"* (P705). Indeed, using simple techniques like jailbreaking [57] and prompt injections [58], practitioners, like P926, became worried that LLM tools can be easily attacked: *"I play with around LLMs a lot–it is alarmingly easy to get a model violate their guardrails."* After breaking these boundaries, practitioners became worried that LLM tools themselves could be used to conduct attacks against their own company: *"you could keep rephrasing open-ended questions to bypass its built-in safety restrictions"* (P496).

5.5 Autonomy

In discussing autonomous task execution without human supervision, practitioners were significantly more likely to hold negative perceptions ($z = -4.94, p < .001$), noting that, despite vendor claims, LLM tools still required heavy oversight.

While Improving, Human Oversight Is Still Needed. Participants noted that while LLM tools are often marketed as fully

autonomous and can act as replacement analysts, they were not often reliable enough to fully delegate tasks (P047, P257, P925). Most practitioners believed that current systems were better suited to decision support than to autonomous operations. As P047 mentioned, *"I work with a wide range of tools that span the Gartner Quadrant; they are still deeply flawed and require human intervention and oversight."* SOC work, in particular, was viewed as too nuanced, context-dependent, and constantly evolving for LLMs to fully handle (P011). Thus, while LLMs were viewed as useful and in some cases essential, practitioners would not allow them to perform actions autonomously: *"In my professional circle, we all agree that we will need to leverage AI, but only for advice, not for taking autonomous actions"* (P089). Instead, they recommended to others that they should *"Use it like an intern. Allow it to gather information and raise its hand when it notices something, but don't let it touch anything critical"* (P089).

5.6 Cost

Practitioners were significantly more likely to be negative ($z = -4.02, p < .001$) when discussing the cost of LLM tools. While practitioners do not typically control organizational budgets, their negative perceptions around LLM costs reflect operational feasibility concerns within SOC workflows.

Query Costs Add Up Fast. Several practitioners noted that LLM inference costs [59] can become prohibitively expensive, raising concerns about the economics of scaling LLM-driven analysis to SOC-sized datasets. As P227 commented, *"Each inference will cost a few cents, it might not be the most economical option for processing a large number of events."* Moreover, some believed that privacy-concerned SOCs will need to invest heavily in computing infrastructure: *"It's going to be so expensive to locally train and operate these high-quality LLMs–you will need data centers"* (P895).

Returns Are Limited. Several participants were skeptical that, given this high cost, LLM tools were practical. Costs were so high that practitioners, such as P790, questioned whether they could simply hire just as many workers with the same money: *"Although Security Copilot has several interesting features, it is still not worth the price. Even minimal usage can be equivalent to one full-time employee's salary."* Because of this, several participants, such as P362, believed that, *"At this stage of its development, LLM tools may not provide enough value or a real return on investment."* Other practitioners, such as P048 and P919, explicitly noted that they did not adopt LLM tools solely because of the cost.

6 Adoption of LLM Tools (RQ3)

We now analyze how practitioners discuss adopting LLM tools within their operations and the concerns they hold.

Table 5: **Reported Adoption of LLM Tools Over Time.**

	Counts			p1 vs p2		p2 vs p3	
Adoption Stage	**p1**	**p2**	**p3**	**Delta**	***p*-val**	**Delta**	***p*-val**
Evaluating	20	16	87	+4.40%	1.000	**-33.65%**	**0.002***
Using	12	7	181	-7.36%	1.000	**+31.09%**	**0.013***
Not Using	3	3	44	+2.97%	1.000	+2.56%	1.000

p1: Dec 2022–Oct 2023, **p2**: Nov 2023–Sep 2024, **p3**: Oct 2024–Aug 2025
Pairwise comparisons were conducted using a chi-square test followed by Bonferroni-corrected post-hoc analysis of adjusted residuals. Bold values indicate statistically significant differences. Delta percentages are based on the proportions of adoption stages.

6.1 Trajectories of LLM Adoption

By analyzing posts (n=373) for whether practitioners adopted LLM tools within their work, we found that 200 posts (53.62%) reported actively using, 123 posts (32.98%) were curious or in the process of evaluating LLM tools for adoption, and 50 posts (13.40%) were not yet adopting or evaluating.

LLM Adoption Shifts from Curiosity to Operational Use. To examine whether practitioner attitudes toward LLM adoption changed over time, we divided the dataset timeline into three approximately equal intervals: (**p1**) **Dec 2022–Oct 2023**, (**p2**) **Nov 2023–Sep 2024**, and (**p3**) **Oct 2024–Aug 2025**, and compared the relative distribution of posts discussing each adoption stage (full statistical results in Appendix B.2).

As shown in Table 5, between **p1** and **p2**, we observed no statistically significant changes across adoption stages ($p = 1.00$). Discussions in these earlier phases were similarly dominated by evaluating LLM tools, while active adoption remained comparatively limited, with only a few posts describing early operational use of ChatGPT (n=6), Microsoft Copilot (n=2), and Gemini in Google SecOps (n=1). In contrast, statistically significant changes emerged between **p2** and **p3**, characterized by a significant increase in active operational use (from 26.92% to 58.01%, $p < 0.05$) and a corresponding decline in exploratory discussions (from 61.54% to 27.88%, $p < 0.01$), suggesting a shift from curiosity and experimentation toward concrete workflow integration. Notably, discussions expressing non-adoption remained comparatively stable across periods ($p = 1.00$), indicating that increasing operational use did not eliminate practitioner skepticism toward LLM tools. Instead, many cases of non-adoption stemmed from negative experiences during early evaluations (§ 6.2).

General-Purpose LLMs Are Often Adopted, But Commercial Tools Drive Interest. Across the posts describing active use of LLM tools, general-purpose LLMs were referenced most frequently and appeared in 103 posts (51.5%), compared to only 37 posts (18.5%) mentioning security-specific LLMs. However, when looking at posts that discuss either actively evaluating or considering LLM use, security-focused LLM tools were referenced in 58 posts (47.16%), more than double the frequency of general LLMs, which appeared in only 28 posts (22.76%). In general, this exploratory discussion

came from several SOC managers (e.g., P001, P094, P305) actively exploring whether security-specific tools that promise near-automation of tasks were practical: *"We have started looking into AI SOC Analysts. Our team still spends a significant amount of time on L1/L2-type work, which should have been automated by now"* (P094). Furthermore, practitioners sought community input on several related questions, including whether these LLM tools delivered value (P008, P055, P094, P860), introduced overhead (P055), integrated well with existing technology (P786), compared favorably with traditional SOAR solutions (P001, P008, P316, P411, P890), and justified their cost (P305, P890).

Tools Are Often Adopted For Non-Core Security Operations. While broad discussions around use cases were often focused on core security operations (§ 4.2), posts describing active adoption often focused on productivity-oriented tasks. Of posts that described adopted uses for LLM tools, 28% focused on scripting & query support and 26% on reporting and documentation; in comparison, 18% discussed triage & IR, and only 6% discussed threat analysis. This may indicate that while core operational tasks dominate overall discussions, active adoption is likely to concentrate on self-scoped tasks that afford greater analyst control and easier verification.

6.2 Cautions against LLM Adoption

Despite reported adoption and curiosity, practitioner discussions highlighted recurring barriers that shaped how they evaluated the appropriateness of LLM tools for SOC.

Inflated Vendor Promises. Practitioners frequently interpreted LLM adoption through the lens of a long-standing history of overpromising technologies in cybersecurity, leading to skepticism toward marketing claims of "autonomous SOC" capabilities. As P859 emphasized, *"The marketing people will claim their product can do everything. However, without hands-on experiments, such products can be anything and everything at once, or nothing."* Consistent with this view, across 20 posts, practitioners reported that they or their organizations had previously evaluated tools, such as Security Copilot, that did not result in adoption, primarily due to inefficient performance and unjustifiable costs.

Sufficiency of Traditional Solutions. Practitioner skepticism toward LLM adoption was also shaped by a perceived tendency to overuse, or *"look for places to cram in AI"* (P226). As P463 argued that *"many organizations lack the use case, scale, or resources to justify LLMs or agentic security solutions, yet they are forcing AI into workflows."* Across 11 posts, practitioners expressed a clear preference for traditional automation, emphasizing that many SOC tasks are already effectively addressed by existing approaches: *"It's like bringing a tank to a knife fight when using LLMs to deterministic security problems. Stick to traditional automation, it's simpler, cheaper, and gets the job done"* (P460).

Organizational Restrictions and Policies. Another cluster of narratives (n=10) highlighted that organizations often restrict the use of public LLMs primarily due to data-loss prevention or confidentiality concerns (§ 5.4). As P922 explained, *"Since people continued to carelessly pour company data into public LLMs, we incorporated them into our company block list."* One practitioner also described broader uncertainty within organizations about how to assess the risks associated with LLMs, driven by the lack of established frameworks for evaluating these tools. As P784 noted, *"Since Gen AI is so new, nobody even understands how to discuss it... the CISO is unsure what the security risks are, the Chief Risk Officer doesn't know how to characterize them on the risk matrix."*

Anti-LLM Sentiment. Lastly, a smaller subset of posts (n=9) reflected a clear reluctance toward LLM adoption, rooted in personal or ideological opposition to the technology. These posts were rarely supported by detailed reasoning; instead, they expressed dismissiveness through statements such as *"No, I have a brain"* (P768), and *"If people could stop talking about AI, I would pay for it"* (P005). P840, however, articulated inhibition grounded in the broader implications of LLMs for the SOC workforce (further discussed in § B.3). As they explained, *"Organizations are eager to replace analysts with AI, making it difficult to be enthusiastic about technologies that may ultimately reduce the need for human roles."*

7 Discussion

Next, we reflect on our findings and outline future research directions to understand the nuances of LLM adoption, design trustworthy systems, and sustain workforce development.

LLM Adoption is Shaped by Control and Commitment. Our findings suggest that LLM adoption in SOCs is not monolithic but rather a layered process that differs across tools and stakeholders. General-purpose LLMs are often adopted by analysts through independent experimentation on self-scoped tasks, whereas much of the curiosity surrounding security-specific enterprise tools is expressed by self-identified decision-makers, reflecting organizational priorities (§ 6.1). This split helps explain why SOCs may exhibit high reported adoption of general LLMs while still showing friction around enterprise adoption. LLMs afford analysts greater control and reversibility, allowing them to selectively apply the tools to low-risk tasks without deep integration or organizational commitment. In contrast, adopting enterprise security-focused platforms is a procurement and governance-level decision [60, 61] that often entails broader data access and tighter coupling with existing SOC tooling.

Research Opportunity. While recent work has begun to explore the adoption of LLMs for specific SOC tasks [13, 14], future research should examine adoption as a *multi-stakeholder phenomenon*. Prior work has already identified inherent mismatches between how managers and analysts evaluate SOC

work [1]; adopting a stakeholder-specific lens is therefore critical to understand the divergent priorities that shape adoption decisions. Future work should also investigate how informal *analyst-level LLM use can be secured under governed deployments*, such as [62], without sacrificing the low-friction interaction patterns that make LLMs inherently valuable.

Reliability is the Hard Ceiling for Autonomy. Our findings indicate that practitioners' reluctance to grant autonomy to LLM tools is not rooted in abstract distrust, but in hands-on experience with unreliable system behavior. Reliability issues (§ 5.3) necessitate manual verification before LLM-generated outputs can be applied, directly constraining the autonomy these systems can be granted. When LLM judgments cannot be confidently trusted, delegation becomes a source of risk rather than benefit, echoing prior work linking low trust in automation to reversion to manual control [63].

Research Opportunity. This framing highlights the need for future research that treats *trust and autonomy as interdependent properties*, and examines how LLM systems communicate uncertainty through *reliability metrics* that allow analysts to assess the reliability of LLM tools in situ. Prior work shows that generic uncertainty disclaimers, such as, "this model may make mistakes," have minimal impact on the credibility threshold in high-stakes expert settings [64]. Reliability metrics are therefore intended not to unconditionally increase trust, but to provide task-specific, actionable signals (e.g., via external hallucination detectors) helping practitioners decide when to act on LLM output and when to fall back to deterministic approaches, as a necessary step toward more trustworthy integration of LLMs within SOC workflows.

The SOC Workforce Development Crisis. Our findings surface an important tension with the implications for the long-term sustainability of the SOC workforce. Practitioners widely expect LLMs to reduce or replace entry-level responsibilities, while analysts are increasingly positioned as reviewers and governors of LLM-mediated workflows that presuppose substantial domain expertise (§ B.3). However, prior research emphasized that such expertise develops incrementally through hands-on operational exposure [65–67]. This creates a circular dependency: effective oversight of LLMs requires experienced analysts, yet the experiential learning pathways that produce such expertise through entry-level tasks are now being automated. As a result, questions emerge about how future analysts will acquire the experiential knowledge needed to critically evaluate LLM decisions.

Research Opportunity. Sustaining expertise in LLM-augmented SOCs may require *rethinking training as a process of co-learning* between humans and LLMs. While machines have traditionally been learning from humans, recent work has begun to formalize co-learning paradigms through interaction, feedback, and shared problem-solving [68, 69]. Early industry efforts are beginning to explore this space; COACH by Dropzone [70] is an LLM-powered security mentor that provides junior analysts with real-time investigative guidance [71]. Recent academic work also explored integrating LLM-powered tutoring into cybersecurity training environments [72–74]. Future research can explore how co-learning approaches can be grounded in cyber workforce development principles [67], while supporting LLM-mediated skill development.

8 Conclusion

In this paper, we present a large-scale mixed-methods analysis of practitioner discussions on Reddit, providing a multi-dimensional view of how LLM tools are used, perceived, and adopted within security operations. Our findings reveal that while practitioners increasingly view LLMs as valuable for augmenting SOC workflows, particularly for self-scoped productivity tasks, concerns surrounding reliability, governance, and workforce sustainability continue to constrain broader operational delegation. Together, these findings highlight that the future of LLM adoption in SOCs is fundamentally sociotechnical, shaped by organizational constraints, practitioner experiences, and long-term workforce considerations.

Ethics Statement

This study does not involve direct interactions with human subjects, as it relies exclusively on publicly available data [75, 76]. Nevertheless, we consulted the Institutional Review Board (IRB) at the leading institution and obtained an IRB exemption. That said, we actively recognize the ethical implications of analyzing online public discourse and took steps to mitigate potential harms, adhering to the ethical standards set in prior published works [28–31]. *First*, we acknowledge that Reddit users may not have anticipated their posts being used for research, and that neither community members nor moderators explicitly consented to such use. To address this, we followed best practices for Reddit data collection and limited our corpus to publicly accessible subreddits that do not prohibit research use under their terms of service. Unlike some subreddits that explicitly restrict academic research due to the sensitivity of their content [77], the communities we studied impose no such restrictions. *Second*, although the analyzed data is likely not sensitive, we implemented additional safeguards to protect user privacy. All identifying metadata (e.g., usernames and organization names) were removed prior to analysis, and contributors are referenced using anonymized identifiers (e.g., `PXXX`). To further reduce the risk of re-identification through reverse search, all excerpts included in this paper were carefully paraphrased in accordance with established ethical guidelines [78, 79].

LLM Usage Considerations. During the preparation of this manuscript, LLMs were used solely for editorial assistance, including language refinement and editing. All generated content was reviewed by the authors, who take full responsibility for the manuscript's accuracy, originality, and integrity.

Acknowledgment

We thank the anonymous reviewers and our shepherd for their thoughtful and constructive feedback.

This work was partly supported by the U.S. National Science Foundation (NSF) under grant NSF-CICI-2232911, the Institute of Information & Communications Technology Planning & Evaluation (IITP) through grants RS-2024-004398199 and RS-2024-00442085, and the German Federal Ministry of Research, Technology, and Space through the Q-Fiber project (No. 16KISQ124). We also gratefully acknowledge the support of Jean-Pierre Seifert.

References

[1] Faris Bugra Kokulu, Ananta Soneji, Tiffany Bao, Yan Shoshitaishvili, Ziming Zhao, Adam Doupé, and Gail-Joon Ahn. Matched and Mismatched SOCs: A Qualitative Study on Security Operations Center Issues. In *Proc. of CCS*, pages 1955–1970, 2019.

[2] Manfred Vielberth, Fabian Böhm, Ines Fichtinger, and Günther Pernul. Security Operations Center: A Systematic Study and Open Challenges. *IEEE Access*, 8:227756–227779, 2020.

[3] Jenny Hofbauer and Kevin Mayer. Blue Team Fundamentals: Roles and Tools in a Security Operations Center. In *Proc. of SECURWARE*, pages 176–184, 2024.

[4] S Sreelakshmi, A Aalan Babu, C Lakshmipriya, LA Anto Gracious, M Nalini, and R Siva Subramanian. Enhancing Intrusion Detection Systems with Machine Learning. In *Proc. of ICSSAS*, pages 557–564, 2024.

[5] Giuseppina Andresini, Feargus Pendlebury, Fabio Pierazzi, Corrado Loglisci, Annalisa Appice, and Lorenzo Cavallaro. INSOMNIA: Towards Concept-Drift Robustness in Network Intrusion Detection. In *Proc. of AISec@CCS*, pages 111–122, 2021.

[6] Nicola Capuano, Giuseppe Fenza, Vincenzo Loia, and Claudio Stanzione. Explainable Artificial Intelligence in CyberSecurity: A Survey. *IEEE Access*, 10:93575–93600, 2022.

[7] Farid Binbeshr, Muhammad Imam, Mustafa Ghaleb, Mosab Hamdan, Mussadiq Abdul Rahim, and Mohammad Hammoudeh. The Rise of Cognitive SOCs: A Systematic Literature Review on AI Approaches. *IEEE Open Journal of the Computer Society*, 2025.

[8] Jalal Ghadermazi, Ankit Shah, and Sushil Jajodia. A Machine Learning and Optimization Framework for Efficient Alert Management in a Cybersecurity Operations Center. *Digital Threats: Research and Practice*, 5(2):1–23, 2024.

[9] Tines. 2023 Voice of the SOC. https://www.tines.com/reports/voice-of-the-soc-2023/. [Accessed: 05-26-2026].

[10] Vectra. 2023 State of Threat Detection. https://www.vectra.ai/resources/2023-state-of-threat-detection. [Accessed: 05-26-2026].

[11] Bushra A Alahmadi, Louise Axon, and Ivan Martinovic. 99% False Positives: A Qualitative Study of SOC Analysts' Perspectives on Security Alarms. In *Proc. of USENIX Security Symposium*, pages 2783–2800, 2022.

[12] Wajih Ul Hassan, Shengjian Guo, Ding Li, Zhengzhang Chen, Kangkook Jee, Zhichun Li, and Adam Bates. NoDoze: Combatting Threat Alert Fatigue with Automated Provenance Triage. In *Proc of NDSS*, 2019.

[13] Diana Kramer, Lambert Rosique, Ajay Narotam, Elie Bursztein, Patrick Gage Kelley, Kurt Thomas, and Allison Woodruff. Integrating Large Language Models into Security Incident Response. In *Symposium On Usable Privacy and Security (SOUPS)*, pages 133–148, 2025.

[14] Ronal Singh, Shahroz Tariq, Fatemeh Jalalvand, Mohan Baruwal Chhetri, Surya Nepal, Cecile Paris, and Martin Lochner. LLMs in the SOC: An Empirical Study of Human-AI Collaboration in Security Operations Centers. *arXiv preprint arXiv:2508.18947*, 2025.

[15] Alsharif Abuadbba, Chris Hicks, Kristen Moore, Vasilios Mavroudis, Burak Hasircioglu, Diksha Goel, and Piers Jennings. From Promise to Peril: Rethinking Cybersecurity Red and Blue Teaming in the Age of LLMs. *arXiv preprint arXiv:2506.13434*, 2025.

[16] Xihuan Lin, Jie Zhang, Gelei Deng, Tianzhe Liu, Xiaolong Liu, Changcai Yang, Tianwei Zhang, Qing Guo, and Riqing Chen. IRCopilot: Automated Incident Response with Large Language Models. *arXiv preprint arXiv:2505.20945*, 2025.

[17] Xinye Tang, Amir H Abdi, Jeremias Eichelbaum, Mahan Das, Alex Klein, Nihal Irmak Pakis, William Blum, Daniel L Mace, Tanvi Raja, Namrata Padmanabhan, and Ye Xing. NL2KQL: From Natural Language to Kusto Query. *arXiv preprint arXiv:2404.02933*, 2024.

[18] Saleha Muzammil, Rahul Reddy, Vishal Kamalakrishnan, Hadi Ahmadi, and Wajih Ul Hassan. Towards Small Language Models for Security Query Generation in SOC Workflows. *arXiv preprint arXiv:2512.06660*, 2025.

[19] Charupriya Bisht and Anurag Jain. Improving Cybersecurity Decision-Making Through Text Summarization Addressing Key Applications and Overcoming Challenges. In *Proc. of International Conference on Networks and Cryptology*, pages 1544–1550, 2025.

[20] Walaa Saber Ismail. Threat Detection and Response Using AI and NLP in Cybersecurity. *Journal of Internet Services and Information Security*, 14(1):195–205, 2024.

[21] Vasu Jakkal. Microsoft Copilot for Security is generally available on April 1, 2024, with new capabilities. https://www.microsoft.com/en-us/security/blog/2024/03/13/microsoft-copilot-for-security-is-generally-available-on-april-1-2024-with-new-capabilities/, 2024. [Accessed: 05-26-2026].

[22] Charlotte AI: Agentic Analyst for Cybersecurity. https://www.crowdstrike.com/en-us/platform/charlotte-ai/. [Accessed: 05-26-2026].

[23] Nir Kshetri. Transforming cybersecurity with agentic AI to combat emerging cyber threats. *Telecommunications Policy*, 49(6):102976, 2025.

[24] Jaron Mink, Hadjer Benkraouda, Limin Yang, Arridhana Ciptadi, Ali Ahmadzadeh, Daniel Votipka, and Gang Wang. Everybody's Got ML, Tell Me What Else You Have: Practitioners' Perception of ML-Based Security Tools and Explanations. In *Proc. of IEEE Symposium on Security and Privacy*, pages 2068–2085, 2023.

[25] Sean Oesch, Robert Bridges, Jared Smith, Justin Beaver, John Goodall, Kelly Huffer, Craig Miles, and Dan Scofield. An Assessment of the Usability of Machine Learning Based Tools for the Security Operations Center. In *Proc. of iThings*, pages 634–641, 2020.

[26] Neele Roch, Hannah Sievers, Lorin Schöni, and Verena Zimmermann. Navigating Autonomy: Unveiling Security Experts' Perspectives on Augmented Intelligence in Cybersecurity. In *Symposium On Usable Privacy and Security (SOUPS)*, pages 41–60, 2024.

[27] Nidhi Rastogi, Shirid Pant, Devang Dhanuka, Amulya Saxena, and Pranjal Mairal. Too Much to Trust? Measuring the Security and Cognitive Impacts of Explainability in AI-Driven SOCs. *arXiv preprint arXiv:2503.02065*, 2025.

[28] Elijah Bouma-Sims, Hiba Hassan, Alexandra Nisenoff, Lorrie Faith Cranor, and Nicolas Christin. "It was honestly just gambling": Investigating the Experiences of Teenage Cryptocurrency Users on Reddit. In *Symposium On Usable Privacy and Security (SOUPS)*, pages 333–352, 2024.

[29] Rajvardhan Oak and Zubair Shafiq. Victims, Vigilantes, and Advice Givers: An Analysis of Scam-Related Discourse on Reddit. In *Symposium On Usable Privacy and Security (SOUPS)*, pages 57–71, 2025.

[30] Jaakko Väkevä, Perttu Hämäläinen, and Janne Lindqvist. "Don't You Dare Go Hollow": How Dark Souls Helps Players Cope with Depression, a Thematic Analysis of Reddit Discussions. In *Proc. of the CHI Conference on Human Factors in Computing Systems*, pages 1–20, 2025.

[31] Elijah Bouma-Sims, Mandy Lanyon, and Lorrie Faith Cranor. "Is this a scam?": The Nature and Quality of Reddit Discussion about Scams. In *Proc. of CCS*, pages 2444–2458, 2025.

[32] Kashyap Thimmaraju, Sybe Izaak Rispens, and Gail-Joon Ahn. Human Performance in Security Operations: A Survey on Burnout, Wellbeing and Flow State Among Practitioners. In *Workshop on Security Operations Center Operations and Construction*, pages 2–4, 2025.

[33] Subigya Nepal, Javier Hernandez, Robert Lewis, Ahad Chaudhry, Brian Houck, Eric Knudsen, Raul Rojas, Ben Tankus, Hemma Prafullchandra, and Mary Czerwinski. Burnout in Cybersecurity Incident Responders: Exploring the Factors that Light the Fire. *Proceedings of the ACM on Human-Computer Interaction*, 8(CSCW1):1–35, 2024.

[34] Sathya Chandran Sundaramurthy, Alexandru G Bardas, Jacob Case, Xinming Ou, Michael Wesch, John McHugh, and S Raj Rajagopalan. A Human Capital Model for Mitigating Security Analyst Burnout. In *Symposium On Usable Privacy and Security (SOUPS)*, pages 347–359, 2015.

[35] Sathya Chandran Sundaramurthy, Jacob Case, Tony Truong, Loai Zomlot, and Marcel Hoffmann. A Tale of Three Security Operation Centers. In *Proc. of ACM Workshop on Security Information Workers (SIW)*, pages 43–50, 2014.

[36] Limin Yang, Zhi Chen, Chenkai Wang, Zhenning Zhang, Sushruth Booma, Phuong Cao, Constantin Adam, Alexander Withers, Zbigniew Kalbarczyk, Ravishankar K Iyer, and Gang Wang. True Attacks, Attack Attempts, or Benign Triggers? An Empirical Measurement of Network Alerts in a Security Operations Center. In *Proc. of USENIX Security Symposium*, pages 1525–1542, 2024.

[37] Mohan Baruwal Chhetri, Shahroz Tariq, Ronal Singh, Fatemeh Jalalvand, Cecile Paris, and Surya Nepal. Towards Human-AI Teaming to Mitigate Alert Fatigue in Security Operations Centres. *ACM Transactions on Internet Technology*, 24(3):1–22, 2024.

[38] Ahmad Mohsin, Helge Janicke, Ahmed Ibrahim, Iqbal H Sarker, and Seyit Camtepe. A Unified Framework for Human AI Collaboration in Security Operations

Centers with Trusted Autonomy. *arXiv preprint arXiv:2505.23397*, 2025.

[39] Masike Malatji. Augmented Intelligence Framework for Human–Artificial Intelligence Teaming in Cybersecurity. *Human-Centric Intelligent Systems*, 5(2):151–180, 2025.

[40] Souradip Nath, Chih-Yi Huang, Aditi Ganapathi, Kashyap Thimmaraju, Jaron Mink, and Gail-Joon Ahn. [Replication Package] Like a Hammer, It Can Build, It Can Break: Large Language Model Uses, Perceptions, and Adoption in Cybersecurity Operations on Reddit. https://github.com/confusedDip/Reddit-Analysis-on-LLMs-for-SOC-Replication-Package. [Accessed: 05-26-2026].

[41] James Mattei, Christopher Pellegrini, Matthew Soto, Marina Sanusi Bohuk, and Daniel Votipka. "I'm trying to learn... and I'm shooting myself in the foot": Beginners' Struggles When Solving Binary Exploitation Exercises. In *Proc. of USENIX Security Symposium*, pages 2867–2886, 2025.

[42] Theophilus Azungah. Qualitative research: deductive and inductive approaches to data analysis. *Qualitative Research Journal*, 18(4):383–400, 2018.

[43] Andrew F Hayes and Klaus Krippendorff. Answering the Call for a Standard Reliability Measure for Coding Data. *Communication Methods and Measures*, 1(1):77–89, 2007.

[44] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. Reliability and Inter-rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–23, 2019.

[45] Virginia Braun and Victoria Clarke. *Thematic Analysis: A Practical Guide*. SAGE publications Ltd, 2021.

[46] Comparing Two Proportions. https://online.stat.psu.edu/stat415/lesson/9/9.4. [Accessed: 05-26-2026].

[47] Rdocumentation. prop.test: Test of Equal or Given Proportions. https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/prop.test. [Accessed: 05-26-2026].

[48] Eric Naioti and Erika Mudrak. Using Adjusted Standardized Residuals for Interpreting Contingency Tables. https://cscu.cornell.edu/wp-content/uploads/conttableresid.pdf, 2020. [Accessed: 05-26-2026].

[49] Eric W Weisstein. Bonferroni Correction. https://mathworld.wolfram.com/, 2004. [Accessed: 05-26-2026].

[50] Jan H Klemmer, Stefan Albert Horstmann, Nikhil Patnaik, Cordelia Ludden, Cordell Burton Jr, Carson Powers, Fabio Massacci, Akond Rahman, Daniel Votipka, Heather Richter Lipford, Awais Rashid, Alena Naiakshina, and Sascha Fahl. Using AI Assistants in Software Development: A Qualitative Study on Security Practices and Concerns. In *Proc. of CCS*, pages 2726–2740, 2024.

[51] Juyong Jiang, Fan Wang, Jiasi Shen, Sungju Kim, and Sunghun Kim. A Survey on Large Language Models for Code Generation. *ACM Transactions on Software Engineering and Methodology*, 35(2):1–72, 2026.

[52] Jianxun Wang and Yixiang Chen. A Review on Code Generation with LLMs: Application and Evaluation. In *Proc. of IEEE International Conference on Medical Artificial Intelligence (MedAI)*, pages 284–289, 2023.

[53] Andrei Sobo, Awes Mubarak, Almas Baimagambetov, and Nikolaos Polatidis. Evaluating LLMs for Code Generation in HRI: A Comparative Study of ChatGPT, Gemini, and Claude. *Applied Artificial Intelligence*, 39(1):2439610, 2025.

[54] Jia-Yu Yao, Kun-Peng Ning, Zhen-Hui Liu, Mu-Nan Ning, Yu-Yang Liu, and Li Yuan. LLM Lies: Hallucinations are not Bugs, but Features as Adversarial Examples. *arXiv preprint arXiv:2310.01469*, 2023.

[55] Berk Atil, Sarp Aykent, Alexa Chittams, Lisheng Fu, Rebecca J Passonneau, Evan Radcliffe, Guru Rajan Rajagopal, Adam Sloan, Tomasz Tudrej, Ferhan Ture, Zhe Wu, Lixinyu Xu, and Breck Baldwin. Non-Determinism of "Deterministic" LLM Settings. *arXiv preprint arXiv:2408.04667*, 2024.

[56] Yifan Song, Guoyin Wang, Sujian Li, and Bill Yuchen Lin. The Good, The Bad, and The Greedy: Evaluation of LLMs Should Not Ignore Non-Determinism. In *Proc. of Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 4195–4206, 2025.

[57] Junjie Chu, Yugeng Liu, Ziqing Yang, Xinyue Shen, Michael Backes, and Yang Zhang. JailbreakRadar: Comprehensive Assessment of Jailbreak Attacks Against LLMs. In *Proc. of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 21538–21566, 2025.

[58] Yi Liu, Gelei Deng, Yuekang Li, Kailong Wang, Zihao Wang, Xiaofeng Wang, Tianwei Zhang, Yepang Liu,

Haoyu Wang, Yan Zheng, Leo Yu Zhang, and Yang Liu. Prompt Injection attack against LLM-integrated Applications. *arXiv preprint arXiv:2306.05499*, 2023.

[59] LLM Inference Benchmarking: How Much Does Your LLM Inference Cost? https://developer.nvidia.com/blog/llm-inference-benchmarking-how-much-does-your-llm-inference-cost/, 2025. [Accessed: 05-26-2026].

[60] Jens Opdenbusch, Jonas Hielscher, and M Angela Sasse. "Where Are We On Cyber?" – A Qualitative Study On Boards' Cybersecurity Risk Decision Making. In *Proc of NDSS*, 2025.

[61] Stef Schinagl and Abbas Shahim. What do we know about information security governance? "From the basement to the boardroom": towards digital security governance. *Information & Computer Security*, 28(2):261–292, 2020.

[62] Georgios Syros, Anshuman Suri, Jacob Ginesin, Cristina Nita-Rotaru, and Alina Oprea. SAGA: A Security Architecture for Governing AI Agentic Systems. *arXiv preprint arXiv:2504.21034*, 2025.

[63] John D Lee and Katrina A See. Trust in Automation: Designing for Appropriate Reliance. *Human Factors*, 46(1):50–80, 2004.

[64] Yavuz Selim Kıyak, Özlem Coşkun, and Işıl İrem Budakoğlu. 'ChatGPT can make mistakes' warnings fail: A randomized controlled trial. *Medical Education*, 60(2):138–142, 2025.

[65] Jessica Dawson and Robert Thomson. The Future Cybersecurity Workforce: Going Beyond Technical Skills for Successful Cyber Performance. *Frontiers in Psychology*, 9:744, 2018.

[66] John R Goodall, Wayne G Lutters, and Anita Komlodi. Developing expertise for network intrusion detection. *Information Technology & People*, 22(2):92–108, 2009.

[67] Marie Baker. Striving for Effective Cyber Workforce Development. https://www.sei.cmu.edu/documents/475/2016_019_001_473577.pdf, 2016. [Accessed: 05-26-2026].

[68] Jinwei Lu, Yikuan Yan, Keman Huang, Ming Yin, and Fang Zhang. Do We Learn From Each Other: Understanding the Human-AI Co-Learning Process Embedded in Human-AI Collaboration. *Group Decision and Negotiation*, 34:235–271, 2025.

[69] Yi-Ching Huang, Yu-Ting Cheng, Lin-Lin Chen, and Jane Yung-jen Hsu. Human-AI Co-Learning for Data-Driven AI. *arXiv preprint arXiv:1910.12544*, 2019.

[70] COACH: AI-Powered Security Alert Mentor for SOC Analysts. https://www.dropzone.ai/coach. [Accessed: 05-26-2026].

[71] AI Is Automating SOC, But Can It Train the Next Generation of Analysts? https://www.dropzone.ai/blog/ai-soc-training-junior-analysts. [Accessed: 05-26-2026].

[72] Connor Nelson, Adam Doupé, and Yan Shoshitaishvili. SENSAI: Large Language Models as Applied Cybersecurity Tutors. In *Proc. of ACM Technical Symposium on Computer Science Education*, pages 833–839, 2025.

[73] Tianyu Wang, Nianjun Zhou, and Zhixiong Chen. CyberMentor: AI Powered Learning Tool Platform to Address Diverse Student Needs in Cybersecurity Education. *arXiv preprint arXiv:2501.09709*, 2025.

[74] Chola Chhetri. Exploring Large Language Model-Powered Pedagogical Approaches to Cybersecurity Education. In *Proc. of Annual Conference on Information Technology Education*, pages 163–166, 2024.

[75] Ivor A Pritchard. Searching for "Research Involving Human Subjects": What Is Examined? What Is Exempt? What Is Exasperating? *IRB: Ethics & Human Research*, 23(3):5–13, 2001.

[76] Shruti Sannon, Billie Sun, and Dan Cosley. Privacy, Surveillance, and Power in the Gig Economy. In *Proc. of the CHI Conference on Human Factors in Computing Systems*, pages 1–15, 2022.

[77] r/Drugs Subreddit. https://www.reddit.com/r/Drugs/. [Accessed: 05-26-2026].

[78] Casey Fiesler, Michael Zimmer, Nicholas Proferes, Sarah Gilbert, and Naiyan Jones. Remember the Human: A Systematic Review of Ethical Considerations in Reddit Research. *Proceedings of the ACM on Human-Computer Interaction*, 8(GROUP):1–33, 2024.

[79] Joseph Reagle. Disguising Reddit sources and the efficacy of ethical research. *Ethics and Information Technology*, 24(3):41, 2022.

[80] Kamal Shah. Prophet Security launches with an Agentic AI SOC Analyst. https://www.prophetsecurity.ai/blog/announcing-prophet-security, 2024. [Accessed: 05-26-2026].

[81] Gemini in Google SecOps. https://docs.cloud.google.com/chronicle/docs/secops/release-notes#March_26_2024, 2024. [Accessed: 05-26-2026].

A Overview of Forum Threads and Posts

The threads in our dataset came from three subreddits: `r/cybersecurity` (n=69), `r/Information_Security` (n=5), and `r/ciso` (n=2). These relevant threads were posted between December 2022 and August 2025 (Figure 1). We can see that conversations around LLM tools have increased over time, with 75% of posts made within the last year of data collection (Sep 2024–Aug 2025). This aligns with the recent emergence and broader visibility of SOC-focused LLM tools around mid-2024, including Microsoft Copilot for Security [21], Prophet Security's AI SOC Analyst [80], and the introduction of Gemini in Google SecOps [81].

Relevant threads contained an average of 22 ($\pm$32.8) posts, ranging from 1–193 posts. Across these 76 threads, we analyzed 892 (52.38%) relevant posts. As shown in Table 6, these relevant posts touched on our topics of inquiry: Tools (n=410), Use Cases (n=325), Opinions (n=406), LLM Factors (n=459), LLM Adoption (n=373), and Vision for the Future (n=276).

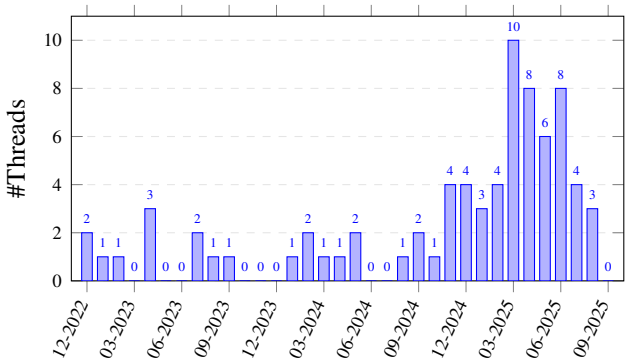

Figure 1: **Number of Threads in Our Dataset Over Time.**

B Additional Results and Discussion

B.1 Commercial LLM Tools

Table 7 provides a comprehensive list of all commercial LLM tools, along with their purpose (general-purpose or security-specific), observed frequencies, and brief descriptions, mentioned in the practitioner discussions.

B.2 Statistical Results

Opinions on LLM Tools. To further examine whether practitioner sentiment differed across categories of LLM tools, we analyzed positive and negative opinions associated with general-purpose and security-specific tools (Table 8). Overall, discussions surrounding the broader categories exhibited relatively balanced sentiment distributions, with no statistically significant differences observed at the category level ($\chi^2(1) = 0.06$, $p = 0.81$). However, tool-specific patterns reveal more nuanced perceptions. Among general-purpose systems, generic references to LLMs and GenAI systems were more commonly associated with negative opinions, whereas discussions of ChatGPT were comparatively more positive. Notably, while overall sentiment toward security-specific platforms remained slightly more positive, discussions surrounding Microsoft Security Copilot were substantially more negative than positive, suggesting a disconnect between broader excitement around AI SOC systems and practitioner experiences with specific enterprise deployments. Together, these findings suggest that practitioner sentiment toward LLM tools is shaped less by broad tool categories and more by hands-on experiences with specific platforms and their perceived operational value within SOC workflows.

Table 6: **Distribution of Discussion Topics** – Identified across 892 relevant posts. Individual posts often addressed multiple themes.

Topics of Discussion	**# Total Posts (Out of 892)**
LLM Uses (§ 4)	
Tools Mentioned	410 (45.96%)
Use Cases Mentioned	325 (36.44%)
Perceptions of LLM Tools (§ 5)	
Opinions Shared	406 (45.52%)
LLM Factors	459 (51.46%)
Implications of LLM Adoption (§ 6)	
LLM Adoption	373 (41.82%)
Vision for the Future	276 (30.94%)

Relative Prevalence of Reported Use Cases. As briefly discussed in Section 4.2, to understand the relative prevalence across the six types of use cases, we conducted pairwise two-sample tests of proportions with a Bonferroni correction (Table 3). We observed that Triage and IR were significantly more prevalent than all other reported uses ($\chi^2(1) \geq 16.92$, all $p < 0.001$), followed by writing-oriented tasks, including scripting, query generation, and reporting. Statistically, scripting (n=88) and reporting (n=84) did not differ significantly from one another ($\chi^2(1) = 0.07, p = 0.79$), but were both significantly higher than threat analysis and knowledge support ($\chi^2(1) \geq 7.74, p < 0.05$). Lastly, the discussions around threat analysis and knowledge support did not differ significantly ($\chi^2(1) = 0.00, p = 1.00$), but both were greater than misc.

Correlating Tools with Use Cases. To examine whether the distribution of discussed use cases differed between general-purpose and security-specific LLM tools, we conducted a chi-square test of independence followed by Bonferroni-corrected post-hoc analysis of adjusted residuals (Table 3). Due to its small sample size, the "Training, Compliance, Others" category was excluded from this analysis. The chi-square test revealed a significant association between use case category and tool category ($\chi^2(4) = 58.289$, $p < 0.01$). Post-hoc analysis showed that triage and IR discussions were significantly over-represented among security-specific tools (z=6.83, $p < 0.001$).

Table 7: **Comprehensive List of Reported LLM Tools** – grouped by general-purpose and security-specific platforms, along with their observed frequencies.

Tool Name	Freq.	Description
General-Purpose LLM Tools		
ChatGPT	90	A general-purpose conversational LLM developed by OpenAI
MS Copilot	30	A generative AI assistant integrated across Microsoft products
Claude	9	A conversational LLM developed by Anthropic
Gemini	6	A multimodal LLM developed by Google
Llama	5	A family of open-source large language models released by Meta
Perplexity	3	A generative AI-powered web search assistant
NotebookLM	3	A research and note-taking online tool powered by Google Gemini
Grok	2	A conversational LLM developed by xAI
Amazon Q	1	Amazon's enterprise generative AI assistant for AWS customers
Security-Specific LLM Tools		
Security Copilot	40	Microsoft's LLM-powered agentic security automation platform
Dropzone AI	10	Autonomous LLM-powered agentic "AI SOC Analyst" platform
Intezer	8	Autonomous LLM-powered agentic "AI SOC Analyst" platform
Cortex XSIAM	6	Extended security intelligence automation management platform
Prophet Security	4	Autonomous LLM-powered agentic "AI SOC Analyst" platform
Purple AI	4	Autonomous LLM-powered agentic "AI SOC Analyst" platform
CMD Zero	3	Autonomous & AI-assisted cyber investigation platform
Abnormal	3	AI-native platform for human behavior security
Google SecOps	3	Google's intelligence-driven security operations platform
Darktrace	3	AI-powered proactive platform for enterprise security
Torq Socrates	2	Autonomous LLM-powered agentic "AI SOC Analyst" platform
Qevlar AI	2	Autonomous LLM-powered agentic "AI SOC Analyst" platform
Arcanna AI	2	Trustworthy agentic "AI SOC Analyst" platform
Vectra AI	2	AI-powered platform for network, identity, and cloud security
WhiterabbitNeo	2	Cybersecurity model built for offensive reasoning
D3 Morpheus	1	Autonomous LLM-powered agentic "AI SOC Analyst" platform
TandemTrace	1	Autonomous LLM-powered agentic "AI SOC Analyst" platform
Radiant Security	1	Autonomous LLM-powered agentic "AI SOC Analyst" platform
Charlotte AI	1	Autonomous LLM-powered agentic "AI SOC Analyst" platform
Exaforce	1	Autonomous LLM-powered agentic "AI SOC Analyst" platform
7ai	1	Autonomous LLM-powered agentic "AI SOC Analyst" platform
Rapid7	1	AI-powered MDR platform for business resilience
Whistic	1	AI-first platform for comprehensive third-party risk management
Splunk ES	1	AI-powered threat detection, investigation, and response platform
SIRP	1	Autonomous LLM-powered agentic "AI SOC Analyst" platform
HackerAI	1	LLM-powered penetration testing platform
XBOW	1	LLM-powered penetration testing platform
Nebula AI	1	LLM-powered penetration testing platform
Gradient Cyber	1	AI-assisted MXDR designed for mid-market organizations
ReliaQuest	1	Autonomous LLM-powered agentic "AI SOC Analyst" platform

Table 8: **Practitioner Sentiment Across LLM Tools.**

LLM Tools	# Positive	# Negative
General-Purpose Tools	**88 (50.57%)**	**86 (49.43%)**
Unnamed (LLMs, GenAI, etc.)	36 (42.35%)	49 (57.65%)
ChatGPT and OpenAI models	38 (58.46%)	27 (41.54%)
Microsoft Copilot	10 (50.00%)	10 (50.00%)
Claude	4 (66.67%)	2 (33.33%)
Gemini	0 (0.00%)	5 (100.00%)
Security-Specific Tools	**56 (52.83%)**	**50 (47.17%)**
Unnamed (AI SOC Analyst, Agentic SOC)	37 (59.68%)	25 (40.32%)
Microsoft Security Copilot	8 (24.24%)	25 (75.76%)
Intezer	4 (100.00%)	0 (0.00%)
Dropzone	3 (100.00%)	0 (0.00%)
Cortex XSIAM	3 (100.00%)	0 (0.00%)

Counts and percentages are reported row-wise and represent the relative proportion of opinions within each LLM tool category. Only tools with three or more total opinion mentions are included.

In contrast, scripting and query support (z=3.57, $p < 0.01$) and reporting and documentation (z=3.22, $p < 0.05$) discussions were significantly overrepresented among general-purpose tools. No statistically significant differences were observed for knowledge support or threat analysis.

Statistical Analysis of Factors and Perceptions. To understand whether sentiment toward LLM tools differed across the discussed factors, we constructed a contingency table capturing positive versus negative mentions for each factor (Table 4), and conducted a chi-square test of independence. The test revealed a significant association between factor type and sentiment ($\chi^2(5) = 172.21, p < 0.001$), indicating that practitioners' perceptions differed significantly across factors. Post-hoc tests with adjusted residuals find that comments around LLMs' capabilities ($z = 8.11, p < .001$), and efficiency ($z = 6.38, p < .001$) were significantly more likely to be positive than negative, while reliability ($z = -6.77, p < .001$), security and privacy ($z = –5.41, p < .001$), level of independent autonomy ($z = -4.94, p < .001$), and cost ($z = -4.02, p < .001$) were significantly more likely to be negative.

Temporal Analysis of LLM Adoption Stages. To examine whether practitioner discussions surrounding LLM adoption changed over time, we conducted chi-square tests of independence between adjacent temporal intervals: p1 vs p2 and p2 vs p3 (Table 5), and measured delta values to represent proportional changes in the relative prevalence of each adoption stage between adjacent periods. No statistically significant difference was observed between the first two phases, p1 and p2 ($\chi^2(2) = 0.44$, $p = 0.802$). Relative proportions suggested that discussions during the earlier periods remained similarly centered around evaluating and exploring LLM tools. In contrast, a significant shift was observed between p2 and p3 ($\chi^2(2) = 13.20$, $p = 0.001$). Post-hoc analysis using Bonferroni-corrected adjusted residuals revealed a transition in practitioner discourse toward active operational use: exploratory discussions became significantly underrepresented in p3 ($z = -3.58$, $p = 0.002$), whereas discussions describing active use became significantly overrepresented ($z = 3.07$, $p = 0.013$). No statistically significant temporal differences were observed for discussions expressing non-adoption.

B.3 Impact of LLMs on Security Workforce

Building on observed adoption patterns and barriers (§ 6.1, § 6.2), practitioners' discussions also reflected on the implications of LLMs for the SOC workforce. Despite vendor claims of autonomy and replacement, practitioner discussions revealed a more nuanced view of how LLM tools reshape roles, responsibilities, and skill demands within SOCs.

Where LLMs Can Meaningfully Augment Humans. Across practitioner discussions (n=28), a widely shared belief was that if LLM tools were to replace any SOC role, L1 responsibilities are the most vulnerable. Practitioners ar-

gued that L1 positions were already being reduced prior to the LLM hype and that the widespread adoption of LLMs is likely to accelerate this trend. For example, as one practitioner, P250, explained how their company *"let go of all eight L1 SOC members, because the SOAR playbooks handled almost everything, and phishing, the only task manually reviewed, was later handed off to an AI tool."* Moreover, several posts frame L1 workflows as *"button-clicking"* (P916), *"brain-dead work"* (P244), *"barely a security role"* (P237), or *"simply processing routine tasks shown by the SIEM"* (P304), justifying that LLM tools need not human-level reasoning to affect L1 staffing. Building on this reasoning, while some practitioners predicted substantial reductions in entry-level positions, *"AI could impact headcount by 15–20%"* (P227).

Where Human Expertise Remains Critical. Despite concerns about the future of L1 roles, practitioners overwhelmingly rejected the idea that SOCs are close to becoming fully autonomous. Practitioner discussions consistently emphasized that even as LLM tools become deeply embedded in workflows, human oversight remains indispensable.

High-skill SOC Responsibilities: Across 20 posts, practitioners emphasized that many high-skill SOC responsibilities, including digital forensics and IR, threat hunting, and penetration testing (P711, P367, P539, P927), typically handled by L2–L3 teams, inherently require humans. As P279 emphasized, *"Any task that requires complex reasoning, logical synthesis and judgment, I see people having a strong presence in handling."* This perspective was further reinforced by stressing the criticality of human expertise: *"Even the most advanced AI tools are ineffective without a skilled security team to implement them or without strong executive backing from a CISO (or equivalent)"* (P631).

LLM Supervision and Governance: Another set of posts (n=18) highlighted that with LLMs in the picture, humans are needed more than ever. Practitioners consistently pointed out that organizations will still need humans to *"supervise and verify LLM outputs"* (P005, P243, P401, P390). As P243 summarized, *"Even with LLMs, there will still be a need for certain levels of verification, which in itself could be an L1 responsibility."* Practitioners further stressed that *"certain responsibilities, such as governance and compliance, cannot be delegated to AI, as doing so risks allowing the system to effectively oversee itself."* (P241). They also argued that within SOC, AI cannot operate without human-provided context. As P354 explained, even someone skilled at prompting requires substantial prior experience to guide the LLM:

> *If a company hires someone who can generate solutions through effective prompting, that alone does not make them a replacement for skilled analysts. Meaningful use of LLMs still requires domain knowledge and expertise of an experienced analyst.*

Accountability: Lastly, across a small set of posts (n=6), practitioners highlighted humans will still be needed for accountability, arguing that organizations cannot solely rely on LLM tools for decisions that may have legal, regulatory, or financial consequences (P048, P237, P285). This was especially evident for incident response workflows. As P347 noted, *"Unsure what a breach response would look like if the 'AI employee' overlooked anything and it caused harm to people... some 'human' will ultimately need to be held responsible."*

How Analysts Must Adapt. Beyond the debate between replacement and augmentation, a few posts (n=11) highlighted that early adopters of LLM tools are already integrating them into daily workflows, and therefore avoiding LLMs or dismissing their relevance should be an untenable stance: *"We would be naïve to overlook the scale and speed of change underway, no one can predict with certainty how the field will evolve"* (P275). Consequently, practitioners offered explicit advice to peers on how to navigate this shift.

Developing AI Literacy: A commonly repeated suggestion was developing AI literacy, followed by the sentiment that *"AI itself is not the reason for job-threat; rather, it is the peers who learn to use it effectively"* (P248). Several, including P246, P818, and P843, described *"AI as a bell that cannot be unrung,"* (P246) stressing that analysts who fail to build fluency will be the first to fall behind as organizations increasingly seek people who can work confidently with AI-enabled tooling. P292 contextualized this urgency by pointing out the unprecedented pace of LLM adoption, noting that, *"only a few technologies in recent history have achieved such rapid global familiarity and enterprise adoption."*

More Depth in Security Reasoning and Response: Another category of advice emphasized the importance of strengthening fundamentals and continuous upskilling. As one practitioner, P304, advised, *"To anyone aspiring, make upskilling a part of your DNA. Go beyond simply responding to alerts to actually understand why detections fire and how systems operate."* Others also underscored the ongoing importance of *"hands-on experience with traditional SOC tools such as SIEM and EDR"* (P034), as well as familiarity with emerging areas such as AI security threats, e.g., ISO 27090 (P271).