Collegiate Aviation Review International

Volume 44 | Issue 1 Peer Reviewed Article #6

02-18-2026

Comparative Analysis of Human and Independent Large Language Model (LLM) Perspectives on the Top Ten Cybersecurity Issues in Aviation

Sean Course
Embry-Riddle Aeronautical University

Isabel La Coss
Embry-Riddle Aeronautical University

Connor Rice
University of Florida

Stephen Rice
Embry-Riddle Aeronautical University

This study compares cybersecurity threat prioritizations produced by aviation subject-matter experts and ten independently developed large language models (LLMs). Using Borda ranking methods and Kendall's W to evaluate agreement, we analyze aligned themes and divergences in aviation-specific threat emphasis. SMEs prioritized risks at integration boundaries, safety-critical navigation interference, and supply-chain and cloud-to-aircraft trust paths. LLMs successfully identified broad risk categories but emphasized generic IT attack surfaces more heavily than domain-specific vectors. Findings suggest LLMs provide value for horizon scanning and taxonomy scaffolding but require aviation context to support operational prioritization. Limitations include a small SME sample (n=4) and rapidly evolving AI model capabilities. Future work should expand stakeholder representation across operators, ANSPs, airport authorities, and regulators.

Recommended Citation:
Course, S., Coss, I. L., Rice, C., & Rice, S. (2026). Comparative analysis of human and independent large language model (LLM) perspectives on the top ten cybersecurity issues in aviation. *Collegiate Aviation Review International*, *44*(1), 115–135. Retrieved from https://ojs.library.okstate.edu/osu/index.php/CARI/article/view/10409/9315

Introduction

Digital transformation across aviation's operational technology (OT) and information technology (IT) systems has expanded cybersecurity vectors in flight operations, maintenance, ground infrastructure, and satellite-based navigation. Prior work in aviation cybersecurity and natural language processing (NLP) applications demonstrates the value of Artificial Intelligence (AI) assisted reasoning (Mishra et al., 2022; Rios-Campos et al., 2024). However, comparative studies evaluating how Large Language Models (LLMs) prioritize cybersecurity threats in safety-critical aviation environments remain limited, particularly in relation to expert-driven assessments. The current study addresses that gap by systematically comparing top-ten cyber-risk rankings from four aviation cybersecurity Subject Matter Experts (SMEs) and ten independently developed LLMs. We analyze convergence and divergence in threat themes and discuss implications for operational decision-making.

The aviation sector faces escalating cybersecurity challenges as digital interconnectivity expands across avionics, communication systems, and OT. Vulnerabilities in satellite communications, navigation signals, and supply-chain software have demonstrated that aviation is not only susceptible to conventional IT threats but also to domain-specific hazards with direct safety implications. Although regulators such as the Federal Aviation Administration (FAA) and the European Union Aviation Safety Agency (EASA) have issued guidance, gaps persist in enforcement, resilience planning, and protection of legacy systems. At the same time, LLMs have emerged as powerful tools for knowledge synthesis and structured analysis in cybersecurity. Capable of generating taxonomies, classifying threats, and proposing mitigation strategies, LLMs hold a promise for augmenting traditional expert-driven assessments. However, the role of LLMs in safety-critical contexts, such as aviation, remains underexplored, particularly in comparison with human SMEs. This study bridges that gap by comparing cybersecurity threat rankings and proposed solutions from SMEs and independently developed LLMs.

Study Objectives and Research Questions

The primary objective of this study is to compare how human experts and independently developed LLMs perceive and prioritize the most pressing cybersecurity issues in aviation. By collecting ranked top ten lists from both groups, the study aims to identify areas of consensus, divergence, and potential blind spots. In addition to rankings, the study examines the types of solutions each entity proposes to address these cybersecurity threats. The goal is not only to understand how LLMs perform in structured risk assessment tasks but also to evaluate the originality, feasibility, and relevance of their proposed interventions. The following research questions guide this investigation: (1) What are the top ten cybersecurity concerns in aviation according to human experts versus independent LLMs? (2) How do the rankings compare across entities?

Literature Review

General Literature on Cybersecurity

Cybersecurity is the discipline focused on protecting digital systems, networks, and data from unauthorized access, disruption, or damage (National Institute of Standards and Technology [NIST], n.d.). Traditional cybersecurity practices centered on antivirus software, firewalls, and patch management, but the scope has grown significantly with the proliferation of cloud computing, mobile devices, and remote connectivity (Alam, 2022). Cybersecurity now encompasses areas such as identity and access management, encryption protocols, endpoint protection, and behavioral analytics. Foundational models like the confidentiality, integrity, and availability (CIA) triad continue to guide system design and threat evaluation (IBM, n.d.-a).

The cybersecurity threat landscape is increasingly diverse and dynamic. Common threats include phishing, ransomware, zero-day vulnerabilities, and denial-of-service (DoS) attacks, often carried out by a mix of independent hackers, cybercriminal organizations, and nation-state actors (IBM, n.d.-b). Research has shown that adversaries are becoming more organized, leveraging artificial intelligence and automation to scale attacks (Mirsky et al., 2021). This has prompted increased attention to proactive threat intelligence, security orchestration, and cyber resilience frameworks across critical industries (Abbas, 2025).

In parallel with technical defenses, scholars have emphasized the importance of human factors and policy (Triplett, 2022). Studies highlight the role of user behavior, training, organizational culture, and regulatory compliance in mitigating risk (Delso-Vicente et al., 2025). Recent work in cybersecurity governance explores how regulatory frameworks (e.g., General Data Protection Regulation (GDPR), NIST, International Organization for Standardization (ISO) 27001) influence defense strategies and shape how organizations report, manage, and recover from cyber incidents (Lokare et al., 2025).

Cybersecurity in Aviation

The aviation industry is uniquely vulnerable to cybersecurity threats due to its reliance on highly interconnected and safety-critical systems (Ukwandu et al., 2021). Commercial aviation networks encompass everything from onboard avionics and navigation systems to ground-based air traffic control, airline IT infrastructure, and passenger service systems (Government Accountability Office [GAO], 2021). As these components increasingly rely on digital communication and automation, the potential attack surface expands (Musa & Osazuwa, 2024). Foundational work in this field has mapped out vulnerabilities in satellite communications, aircraft datalinks like Aircraft Communications Addressing and Reporting System (ACARS), and surveillance systems such as Automatic Dependent Surveillance-Broadcast (ADS-B), which currently lack robust encryption or authentication protocols (Strohmeier et al., 2013).

Real-world incidents and test scenarios have validated these risks. Researchers have demonstrated the feasibility of spoofing or jamming navigation signals, accessing in-flight entertainment systems, and exploiting unsecured maintenance interfaces (Ukwandu et al., 2021). Regulatory bodies like the FAA and EASA have acknowledged these threats and issued evolving guidance, but the literature suggests gaps remain in enforcement, threat modeling, and vendor accountability. Concerns are especially acute in older legacy systems, which were not designed with cybersecurity in mind (GAO, 2021). Recent literature emphasizes the growing need for a unified cybersecurity approach ac aviation that spans both OT and IT (Ukwandu et al., 2021).

Scholars call for better coordination between regulators, airlines, manufacturers, and cybersecurity professionals, stressing that aviation's safety-first culture must now integrate security as a core component of that mission (World Economic Forum, 2021). While earlier studies (e.g., Ukwandu et al., 2021; GAO, 2021) identified vulnerabilities in datalinks and navigation systems, more recent analyses (Musa & Osazuwa, 2024) emphasized the rapid convergence of OT and IT risks. Together, these studies establish a growing recognition of systemic interdependence in aviation cybersecurity but provide limited insight into how such risks are prioritized across different stakeholder perspectives.

Emergence and Function of LLMs

LLMs are advanced AI systems that generate human-like text based on deep neural architectures trained on vast textual corpora (Naveed et al., 2023). Their ability to summarize information, explain technical concepts, and generate strategic recommendations has made them increasingly relevant in domains such as cybersecurity (Xu et al., 2024). Despite the increasing maturity of aviation cybersecurity research, most existing studies remain descriptive, mapping vulnerabilities without empirically comparing how different entities (e.g., human experts, AI systems, or regulatory bodies) prioritize risks. Moreover, LLM-related research (Xu et al., 2024; Zhang et al., 2024) tends to focus on technical model performance or general cybersecurity datasets rather than domain-specific applications in regulated, safety-critical contexts such as aviation. This gap highlights the need for comparative analyses that evaluate alignment and divergence between algorithmic reasoning and expert judgment. The current study specifically focuses on independently developed LLMs, excluding direct derivatives or commercial clones of GPT models, to examine how diverse underlying architectures and training methodologies influence outputs on complex, domain-specific tasks.

Research Gaps and Rationale for the Study

As aviation systems grow increasingly digitized and interconnected, cybersecurity has become a critical concern for maintaining safety and operational continuity (Bentley, 2025). Traditional risk assessments rely heavily on expert judgment, yet emerging technologies, particularly LLMs, offer a new avenue for structured analysis and prioritization. Despite their growing use in security and policy research, little is known about how independent LLMs compare to human experts in identifying and ranking cyber threats in aviation (Mezzi et al., 2025). Existing literature often focuses either on the technical capabilities of LLMs or on cybersecurity risks in isolation, leaving a gap at the intersection of AI and domain-specific risk evaluation (Zhang et al., 2024).

The literature reveals robust understanding of aviation cybersecurity architecture and emerging AI-assisted analytical methods, yet little integration between the two domains. Existing studies primarily describe vulnerabilities or evaluate isolated AI applications rather than examining how algorithmic models and human experts converge or diverge in their perception of aviation cybersecurity priorities. This absence of comparative analysis constitutes a critical gap, motivating the present study to assess how independently developed large language models align with, or differ from, expert judgments in identifying and ranking cyber threats within the aviation ecosystem. This study seeks to fill that gap by comparing the ranked top ten cybersecurity issues

in aviation as perceived by human experts and by a range of independent, non-GPT-based LLMs. The study further investigates how each group proposes solutions, offering insight into the practical and conceptual value LLMs may bring to future aviation cybersecurity strategy and providing new insights into the intersection of human and machine reasoning in aviation cybersecurity.

Methods

Participants

This study involved two participant groups: human experts (n = 4) and LLMs (n = 10). The expert group consisted of four aviation cybersecurity professionals, including researchers and industry practitioners with at least five years of domain experience. All interviews took place in the Summer of May 2025. Table 1 details the SMEs, focus area, and years of experience.

Table 1
Details of SMEs used in the Current Study

Participant	Role	Years of Experience
Participant 1	Academia & Product Cybersecurity	38+
Participant 2	OEM Manufacturing	20+
Participant 3	OEM Manufacturing	25+
Participant 4	OEM Manufacturing	30+

A total of 10 LLMs were used in the current study. All models were accessed in their unmodified forms via official APIs or platforms during one week in May 2025. Models were carefully selected to represent non-OpenAI architectures to ensure architectural diversity in reasoning patterns. Table 2 highlights the name and model number used for the study.

Table 2
Details of LLMs used in the Current Study

Name	Model	URL
ChatGPT	GPT-3.5	https://chatgpt.com
Google Gemini	Gemini 2.5 Pro	https://gemini.google.com/app
xAI Grok	Grok 3	https://x.ai/grok
Meta Llama	Llama 4 Maverick	https://www.llama.com
Anthroopic Claud	Claud 3.5 Sonnet	https://claude.ai
DeepSeek	DeepSeek-V3	https://chat.deepseek.com
Mistral	Mistrial 8x7B	https://chat.mistral.ai/chat
Qwen	Qwen3	https://chat.qwen.ai
Adept Fuyu	Fuyu-8b	https://huggingface.co/adept/fuyu-8b
TII Falcon	Falcon-180B	https://huggingface.co/tiiuae/falcon-180B-chat
Databricks DBRX	DBRX Base	https://huggingface.co/databricks/dbrx-base

LLMs used in the Study

ChatGPT (OpenAI)

ChatGPT, developed by OpenAI, is one of the most widely recognized and commercially adopted large language models (Paris, 2025). Built on the GPT-3.5 and GPT-4 architectures, ChatGPT has been fine-tuned extensively for instruction following, conversational fluency, and broad general knowledge. It combines reinforcement learning from human feedback (RLHF) with continual updates from user interactions to improve performance and alignment. While not the focus of this study, ChatGPT serves as a benchmark in the LLM landscape, offering insight into the capabilities and limitations of large-scale proprietary models. Its exclusion from the primary comparison is intentional, to emphasize models developed independently from the OpenAI ecosystem and explore how a diverse range of architectures interpret domain-specific cybersecurity challenges.

Claude (Anthropic)

Anthropic's Claude series (Bai et al., 2022) is built with safety and alignment as core principles, using a technique called Constitutional AI. Instead of relying solely on human reinforcement learning, Claude is trained to follow a set of guiding ethical principles, enhancing both consistency and interpretability. Claude models are known for their structured reasoning and caution, often producing balanced and articulate responses, particularly valuable in high-risk domains like aviation cybersecurity.

Gemini (Google DeepMind)

Gemini, developed by DeepMind, is a multimodal LLM trained on large-scale textual and programming data (Google DeepMind, 2023). It evolved from prior models such as Chinchilla and PaLM. Emphasizing grounded reasoning and real-time knowledge retrieval, Gemini integrates capabilities in both language and logic. Its design favors factual precision, and its performance in technical fields, including cybersecurity, has been benchmarked as strong across a range of reasoning tasks.

Mistral (Mistral.ai)

Mistral is a European initiative focused on efficiency and transparency in open-weight models (Mistral AI, 2025). Using sparse mixture-of-expert architectures, Mistral delivers high performance with relatively fewer parameters. It is optimized for speed and reproducibility, making it attractive for academic research. Its responses are often terse and analytical, suitable for scenarios requiring clarity over verbosity, though sometimes less context-aware than models with more behavioral fine-tuning.

LLaMA (Meta)

Meta's LLaMA series is widely used in research due to its open-access model weights and documentation (Meta AI, 2024). LLaMA's training emphasizes scale, multilingual data, and

low-resource adaptability. Instruction-tuned versions show strong performance in structured tasks and technical reasoning. Because LLaMA has been adapted in many third-party applications, care is taken in this study to use an unmodified version to preserve consistency in output evaluation.

Grok (xAI)

Grok is developed by xAI, an independent AI lab founded by Elon Musk (Business Insider, 2024). Grok is distinct in that it integrates with real-time data sources (e.g., via X/Twitter) and emphasizes humor, contrarian perspectives, and conversational engagement. While not as widely studied in academia, Grok offers a novel lens on cybersecurity through its informal tone and creative framing, which may introduce unique but less conventional insights.

Falcon (Technology Innovation Institute)

Falcon is an open-source LLM developed in the UAE, trained on a large corpus of web data with a strong emphasis on multilingual content (Technology Innovation Institute, 2023). Falcon-40B and its successors have been designed for accessibility and research reproducibility. Though less prominent in mainstream usage, Falcon has demonstrated strong benchmark performance and is known for its flexibility in downstream fine-tuning, making it a valuable inclusion in comparative analyses.

Qwen (Alibaba Cloud)

Qwen is a family of large language models developed by Alibaba Cloud, encompassing both dense and sparse architectures (Yang et al., 2025). The latest release, Qwen 3, includes models ranging from 0.6B to 235B parameters, trained on 36 trillion tokens across 119 languages and dialects. Notably, Qwen2.5-Omni supports multimodal inputs—text, images, audio, and video—and can generate both text and audio outputs, facilitating real-time voice interactions. These features make it valuable for research purposes.

Deepseek

DeepSeek, a Chinese AI startup founded in 2023, has developed a series of open-weight large language models, including the 67B parameter DeepSeek LLM and the multimodal DeepSeek-VL. Their flagship model, DeepSeek-R1, has been recognized for its performance (DeepSeek-AI, 2025). DeepSeek's models are trained on extensive datasets and are designed to be cost-effective, allowing them to efficiently complete research tasks.

DBRX (Databricks)

DBRX is an open-source, general-purpose large language model developed by Databricks (Databricks, 2024). It employs a fine-grained mixture-of-experts (MoE) architecture with 132B total parameters, of which 36B are active during inference. Trained on 12 trillion tokens, DBRX excels in language understanding, programming, math, and logic tasks. The

model is available under an open license, allowing for customization and fine-tuning for aviation and cybersecurity applications.

Procedure

The study followed a standard data collection process. Both human experts and LLMs were independently asked to list and rank what they considered the top 10 cybersecurity issues currently facing the aviation industry. Human experts completed this task via a short online questionnaire, while LLMs were prompted using standardized language to ensure consistency across platforms. Each LLM was queried in a new session to avoid contextual contamination, and the prompts were phrased to encourage a ranked response reflecting perceived severity or urgency. Each LLM was queried twice on two different computers using two different user accounts. They will be referred to as Run 1 and Run 2 in the results section.

To preserve fairness in comparison, care was taken not to steer LLMs toward any specific framework, terminology, or risk model. The entire process was conducted within a one-week period to minimize temporal variability in threat awareness, particularly for LLMs that may have access to live or recently updated data (e.g., ChatGPT and Gemini). The following question was given to both the SEMs and LLMs: *Please list the top 10 most critical cybersecurity threats facing the commercial aviation industry today. Provide the items in a ranked list of importance, with 1 being the most critical and 10 being the least.*

Data Analysis

The ranked cybersecurity issue lists from both human experts and LLMs were compiled and analyzed using both quantitative and qualitative methods. To compare rankings, each issue was coded and mapped across participants, and a Borda count method was applied to generate aggregated rankings for each group (Martin, n.d.). The Borda count method converts rankings into points (1st = 10 points…10th = 1 point) and aggregates across participants to determine consensus priority. Inter-rater agreement among human experts and among LLMs was measured using Kendall's W correlation to assess consistency within and across entities (University of Florida, n.d.). Kendall's W measures agreement among raters (0 = no agreement, 1 = perfect agreement). Lower values reflect inherent diversity in aviation cyber-risk views and LLM sampling variability.

For the solution proposals, responses were thematically coded using inductive content analysis (Vears & Gillam, 2022). Each proposed solution was categorized based on type (e.g., technical, regulatory, procedural, educational) and evaluated for novelty, specificity, and feasibility. Emergent themes were then compared between human and LLM-generated responses to identify commonalities, gaps, and outliers. Representative quotes or outputs from each LLM and the human group were selected to illustrate key themes. All data were anonymized prior to analysis. Responses from LLMs were treated as individual contributors to maintain comparability with human inputs, and no weighting was applied based on model size or architecture (Yan et al., 2025).

Results

The results are organized into three main sections. The first section presents the outcomes of the LLM assessments conducted using the Borda Count Method. The second section showcases the results from the Human SMEs, which were evaluated using a Normalized Borda Count Method. The LLMs utilized the Borda Count method without the need for scaling, as their lists were equivalent. In contrast, the SMEs employed the normalized Borda Count to ensure that each respondent's list contributed equally, with the results then aggregated at the theme level. The final section is a thematic analysis of the LLMs vs. the SMEs.

LLM Results

Table 3 presents the Borda Count Method results. Each cybersecurity issue was ranked from 1-10, with points assigned to each ranking. These points were then summed in the final column to show the rankings of all 27 issues that were listed by the LLMs.

Table 3
Borda Count Method for LLMs

RANKING	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th		Total Score
Ransomware	10	3	0	1	0	0	2	0	2	0		146
Supply Chain Attack	0	6	6	3	2	1	1	0	0	0		144
Insider Threat	0	0	2	4	5	2	3	2	1	1		105
Aircraft Attack	6	1	2	2	0	0	0	0	0	1		100
Data Breach	1	0	0	4	3	5	2	1	0	1		93
ATC Attack	0	3	3	1	0	0	0	0	1	0		60
Phishing	0	1	0	1	2	1	4	1	1	4		58
Critical Systems Attack	3	1	0	0	1	1	0	1	0	1		54
Airport Attack	0	1	1	1	0	1	2	3	2	0		50
DDoS	0	0	0	0	1	5	1	3	2	1		49
Spoofing	0	0	1	1	2	0	1	1	1	0		36
Advanced Persistent Threat	0	1	1	0	0	2	0	0	1	1		30
Digital Infrastructure Attack	0	0	1	0	0	0	1	2	3	4		28
Cyber Espionage	0	2	0	0	0	0	1	0	0	0		22
WiFi Attack	0	0	1	0	0	0	0	2	1	2		18
SATCOM Attack	0	0	0	0	1	0	2	0	0	0		14
AI-Powered Cyberattack	0	0	0	1	1	0	0	0	0	0		13
Malware	0	0	1	0	0	0	0	1	1	0		13
Poor Regulations	0	0	1	0	0	0	0	0	1	0		10
Privacy Attack	0	0	0	0	0	2	0	0	0	0		10
Information Sharing	0	1	0	0	0	0	0	0	0	0		9
Outdated Tech	0	0	0	0	0	0	0	2	0	3		9
Lack of Training	0	0	0	1	0	0	0	0	0	1		8
Cyber-Physical Attack	0	0	0	0	1	0	0	0	0	0		6
Emerging Threat	0	0	0	0	0	0	0	0	3	0		6
Man-in-the-middle Attack	0	0	0	0	1	0	0	0	0	0		6
Mobile App Attack	0	0	0	0	0	0	0	1	0	0		3

In Run 1, the interrater reliability was very low, Kendall's W = 0.03. In Run 2, the interrater reliability was also very low, Kendall's W = 0.054. These results indicate very little agreement between the different LLMs.

Human SME Results

Fourteen industry SMEs were contacted for the current study. Several declined to provide a list of current cyberthreats as their companies would not allow them to share the information publicly. We received four responses: three OEMs and one aviation cybersecurity expert with over 35 years of experience. A normalized Borda count per list was used to rank order the threats identified by the SMEs. Borda scoring converts ranks into points and then sums them across lists. To keep the lists of different lengths comparable, we scaled the points to 0, 1 within each

list and renormalized them so each SME list contributes a total weight equal to 1. We assigned tied items to the average of the $\hat{s}$ values for their occupied ranks.

The normalized Borda count analysis first involved cleaning the SME lists (deduping, expanding acronyms) and mapping each item to the categories used for the LLM analysis. We then calculated the score ($\hat{s}$ values) for each ranked item in the list. For duplicates, we kept the highest rank and dropped the duplicate or treated it as a tie if it was truly equal. Next, we aggregated the sum categories across the SME list to produce the final combined score for category (S_c). The final S_c was then used to create the ranked list with the contribution from SMEs. Table 4 shows the normalized Borda Count analysis of SMEs.

Table 4
Normalized Borda Count Analysis of SMEs

	Rank	Item	Mapped Category	$\hat{s}$	Weight
SME 1 - OEM	1	Network security (segmentation, OT separation)	Digital Infrastructure Attack	0.250	0.250
	2	Asset management/monitoring	Digital Infrastructure Attack	0.214	0.214
	3	Removable media usage controls	Malware	0.179	0.089
	3	Removable media usage controls	Ransomware	0.179	0.089
	4	Business continuity/resiliency	Critical Systems Attack	0.143	0.143
	5	Secure remote access to equipment	Digital Infrastructure Attack	0.107	0.107
	6	Aging systems (patches/spares)	Outdated Tech	0.071	0.071
	7	Industry 4.0 connectivity & vendor security gaps	Digital Infrastructure Attack	0.036	0.018
	7	Industry 4.0 connectivity & vendor security gaps	Cyber-Physical Attack	0.036	0.018
	8	Role-based training for factory staff	Lack of Training	0.000	0.000
SME 2 - OEM	1	Connect new equipment to legacy systems	Digital Infrastructure Attack	0.333	0.167
	1	Connect new equipment to legacy systems	Outdated Tech	0.333	0.167
	2	Users: awareness of actual risks	Lack of Training	0.267	0.267
	3	Expense: need to quantify security value	Information Sharing	0.200	0.200
	4	Domain complexity: right-time/right-place protections	Digital Infrastructure Attack	0.133	0.067
	4	Domain complexity: right-time/right-place protections	Critical Systems Attack	0.133	0.067
	5	Speed of attacker capability advancement	Emerging Threat	0.067	0.067
	6	Speed of change in certification environment	Poor Regulations	0.000	0.000
SME 3 - OEM	1	Customers/OEMs not understanding risk	Information Sharing	0.333	0.333
	2	Misapplied security due to conservative risk thinking	Poor Regulations	0.267	0.267
	3	Software development supply chain quality	Supply Chain Attack	0.200	0.200
	4	Data distribution supply chain (off-aircraft to on-aircraft)	Supply Chain Attack	0.133	0.133
	5	Cloud-hosted services connecting to aircraft	Digital Infrastructure Attack	0.067	0.067
	6	Detect HW/SW changes to airborne systems	Critical Systems Attack	0.000	0.000
SME 4 - Expert	1	GNSS interference (spoofing/jamming/meaconing)	Spoofing	0.200	0.200
	2	TCAS spoofing	Spoofing	0.178	0.178
	3	Aircraft software supply-chain threats	Supply Chain Attack	0.156	0.156
	4	ATM software supply-chain threats	Supply Chain Attack	0.133	0.067
	4	ATM software supply-chain threats	ATC Attack	0.133	0.067
	5	ATC/ATM network cyber attacks	ATC Attack	0.111	0.111
	6	Airport systems cyber attacks	Airport Attack	0.089	0.089
	7	Airline operational IT (booking/operations/dispatch)	Digital Infrastructure Attack	0.067	0.067
	8	Factory cybersecurity (additive manufacturing, composites)	Cyber-Physical Attack	0.044	0.044
	9	Human vulnerability / social engineering	Lack of Training	0.022	0.022
	10	Insider threat across OEMs, MROs, operators	Insider Threat	0.000	0.000

A compiled list of the category totals was then analyzed to determine rank order. This analysis generated a top 10 list, mapped to the LLM models for further analysis. The ranked ordered list can be found in Table 5.

Table 5
Ranked-Ordered Top 10 List for SMEs

Rank	Category	S_c	Share of Total	SME Support
1	Digital Infrastructure Attack	0.956	23.90%	4/4
2	Supply Chain Attack	0.556	13.90%	2/4
3	Information Sharing	0.533	13.30%	2/4
4	Spoofing	0.378	9.40%	1/4
5	Lack of Training	0.289	7.20%	2/4
6	Poor Regulations	0.267	6.70%	1/4
7	Outdated Tech	0.238	6.00%	2/4
8	Critical Systems Attack	0.210	5.20%	2/4
9	ATC Attack	0.178	4.40%	1/4
10	Malware	0.089	2.20%	1/4

Thematic Analysis

Once the ranked-ordered top 10 list was completed, a thematic analysis was conducted of the LLMs vs. the SME responses. We used the LLM-generated Borda Count threat list and four SME lists combined with the Normalized Borda Count to conduct the analysis. Eight themes emerged from the lists: (1) Supply & Data-Chain; (2) Legacy OT/Segmentation; (3) Network/Comms; (4) Human/Insider; (5) Malware & Extortion; (6) Cyber-Physical/Safety-Critical; (7) Governance/Certification; and (8) Business Resilience/Continuity. Each listed threat was mapped to the most specific theme. If an item spanned multiple themes (e.g., ATM software supply-chain is both Supply-Chain and ATC), its contribution was split evenly across those themes. When wording was broad, we mapped by primary attack surface (e.g., data/software chain, network/comms, OT plant, etc.). Aviation-specific subcases were captured as subcodes (e.g., GNSS/TCAS within Network/Comms; additive-manufacturing within Cyber-Physical). We then explored whether LLM outputs contained the theme label and the aviation-specific instantiations emphasized by SMEs (e.g., GNSS/TCAS spoofing, cloud-to-aircraft data services, AM tamper). A table was generated with a theme-level comparison table (LLM items vs. SME-only specifics) with SME weights, and the LLM coverage was labeled High (explicit), Partial (generic only), or Gap (missing). The results are shown below in Table 6.

Table 6
Thematic Analysis of LLM vs. SME

Theme	LLM threats captured	SME-only specifics (from responses)	Total S_c	Total weight	LLM coverage
Legacy OT, Smart-Factory & Segmentation	Outdated Tech; Digital Infrastructure Attack	Industry 4.0 connectivity risks; asset-inventory gaps	1.194	29.90%	High (exact)
Human Factors & Insider Activity	Insider Threat; Phishing; Lack of Training; Information Sharing	Social engineering of maintenance staff; credential phishing of cloud dashboards	0.822	20.60%	High (exact)
Supply & Data-Chain Compromise	Supply Chain Attack; Data Breach; (LLM also listed Malware/Ransomware)	Aircraft & ATM *software* supply-chain tampering; cloud-hosted ground-to-air data services	0.556	13.90%	Partial (LLM covered the category, not the aviation-specific flows)
Cyber-Physical / Safety-Critical Systems	Aircraft Attack; Cyber-Physical Attack; SATCOM Attack; ATC Attack; Airport Attack	Additive-manufacturing sabotage; 3-D-printed part tampering	0.538	13.50%	Partial (LLM covered categories; SMEs added manufacturing-specifics)
Network / Comms Attacks	DDoS; Spoofing; Man-in-the-Middle; Wi-Fi; Mobile App Attack	GNSS jamming/spoofing; TCAS spoofing; SATCOM hijack	0.378	9.40%	Partial (LLM generic "spoofing"; SMEs emphasized PNT/TCAS)
Governance, Regulation & Certification	Poor Regulations; Privacy Attack	Certification pace/clarity; self-attestation gaps	0.267	6.70%	Partial (LLM named it; SMEs added concrete fixes/gaps)
Malware & Extortion	Ransomware; Malware; AI-Powered Cyberattack	RansomOps targeting supply-chain providers	0.179	4.50%	High (exact)
Business Resilience & Continuity	Emerging Threat (LLM listed); Critical Systems Attack sometimes conflated	Incident-response & recovery planning for ops continuity	0.067	1.70%	Partial (LLM didn't call out IR/BCP explicitly)

Discussion

Summary of Key Findings

First, the SMEs concentrated risk at the integration boundaries where old, new, and external systems meet. The highest-weighted themes are legacy OT / segmentation (29.9%) and human/insider (20.6%). While LLMs named these areas, the SMEs emphasized the mechanics that make them hazardous in practice (e.g., asset discovery at the line-of-business level, micro-segmentation patterns for mixed-vintage tooling, secure remote-maintenance paths, and role-based training tied to specific factory and airline operations roles). This reveals that the LLMs were able to identify the category labels, whereas the SMEs provided actionable details.

Second, within Network/Comms themes, LLMs listed generic items (DDoS, Wi-Fi, MITM), but SMEs elevated PNT/TCAS spoofing to a top network risk (9.4%). This reflects the SMEs' knowledge of safety-of-flight consequences unique to aviation. GNSS interference and

TCAS spoofing create operational hazards that are much more critical than common IT nuisances. Similarly, the SMEs identified cloud-to-aircraft service trust boundaries and software/data supply-chain paths as core to digital infrastructure and supply-chain (together ~38% of total weight). LLMs referenced supply chain, but the SMEs added specific information relevant to aviation flows (e.g., signed updates, flight-plan and performance-data distribution, gatelink/SATCOM exposure).

Third, in the cyber-physical / safety-critical group, LLMs named broad categories (e.g., aircraft, SATCOM, ATC, airport attacks), while SMEs highlighted factory/manufacturing sabotage and additive-manufacturing tampering (13.5%). These nuances tie directly to certification, part traceability, and the integrity of long-lived fleets, again highlighting the lack of detail that generic LLM outputs rarely surface without domain prompts. A further consideration is that industry experts are often unable to discuss specific cybersecurity threats tied to current operational environments due to nondisclosure agreements, proprietary systems, and export-control restrictions. This limited openness contributes to the small sample size and helps explain the contrast between detailed but restricted human knowledge and the broader, publicly available information accessible to LLMs. Because domain-specific expertise is rarely published or shared online, LLMs can approximate general categories but not the nuanced operational challenges understood by practitioners. This observation helps contextualize both the study's limited SME participation and the generality of the LLM outputs.

The over-/under-weight pattern explains the apparent gap. Several LLM-listed threats drew little or no SME weight (DDoS, Wi-Fi, mobile-app, MITM, phishing, privacy, SATCOM-generic, aircraft-generic, data breach, "AI-powered attack"). That doesn't mean that the threats are nonexistent; rather, SMEs deem them managed, lower consequence, or insufficiently specific to aviation safety and continuity relative to higher-impact concerns. Conversely, the SME-specific results emphasize that LLMs underspecified threats (e.g., GNSS/TCAS spoofing, cloud-to-aircraft trust, AM/3-D part tamper paths), showcasing exactly where domain knowledge and operational experience matter most. Overall, the LLMs captured the breadth of aviation-cyber themes. However, SMEs supplied the aviation-specific instantiations that drive operational risk and thus carry most of the weight in the normalized Borda analysis.

Real-world events further validate the specific threat areas emphasized by the SMEs. For example, recurrent GPS jamming incidents across Eastern Europe between 2022 and 2024 disrupted commercial flight navigation, underscoring the operational consequences of spoofing and interference within aviation's Position, Navigation, and Timing (PNT) systems. Similarly, the 2023 FAA NOTAM system outage (FAA, 2023), traced to an internal data-handling failure with cybersecurity implications, highlighted the dependency between IT and OT systems. Ongoing research into ACARS has demonstrated the feasibility of message spoofing and the need for stronger authentication protocols (Choudhary et al., 2022; Smith et al., 2016). In parallel, independent assessments of satellite communication (SATCOM) terminals, including the IOActive study, revealed exploitable weaknesses in aircraft connectivity infrastructure (Milvich et al., 2022; Santamarta, 2019). Together, these examples illustrate that the threats prioritized by SMEs are not hypothetical but reflect observed vulnerabilities with direct implications for flight safety and operational continuity.

Interpretation and Implications

The comparison indicates that LLMs are effective at covering the breadth of aviation-cyber themes, while subject-matter experts (SMEs) provide the aviation-specific instantiations that ultimately dominate priority when weighted with the normalized Borda method. Areas of alignment include the presence of broad categories (e.g., legacy/OT exposure, human/insider risk, and supply-chain compromise) appearing in both LLM and SME outputs. However, misalignment emerges in the granularity and salience of threats. LLM lists tend to emphasize generic IT vectors (e.g., DDoS, Wi-Fi/MITM, "AI-powered attacks"), whereas SMEs concentrate weight on operationally consequential flows: GNSS/TCAS spoofing (safety-of-flight), cloud→aircraft service trust boundaries, software/data-distribution chains, and manufacturing integrity (e.g., additive-manufacturing tamper). In short, LLMs reproduce recognizable taxonomic labels; SMEs elevate where and how those labels manifest in aviation.

The normalized Borda analysis reinforces this pattern. The highest weighted SME themes coalesce around Legacy OT/segmentation and Human/insider—not simply as labels, but as implementation contexts (asset discovery, micro-segmentation of mixed-vintage tooling, secure remote maintenance, role-specific training). Likewise, within Network/Comms, SMEs shift attention from generic attacks to PNT/TCAS spoofing because of its disproportionate safety impact. This suggests that, as used here, LLMs are strong for horizon scanning and taxonomy scaffolding, but priority setting in safety-critical domains still depends on domain knowledge to anchor threats to concrete aviation data flows, certification constraints, and operational consequences.

Limitations

This study has several limitations. First, the comparative design employs different aggregation schemes across LLMs and SMEs. Specifically, SME rankings were converted via a normalized Borda method. In contrast, LLM outputs were initially summarized with a plain Borda-style presentation, which complicates direct magnitude comparisons, even though patterns are still interpretable. Second, the SME sample is small and role-skewed toward OEM/manufacturing perspectives. The small sample size ($n = 4$) was constrained by proprietary restrictions and export controls on aviation cybersecurity knowledge. While small, it reflects deep domain expertise. Future studies should expand to airlines, ANSPs, airports, and regulators to capture ecosystem diversity. Third, the thematic coding itself introduces judgment. Some items span multiple themes and require proportional splitting, and alternative taxonomies or coders could shift marginal weights. Fourth, LLM results are sensitive to prompt framing, model/version, and sampling nondeterminism; low inter-model agreement can depress the apparent LLM consensus, and the prompts may have favored generic IT threats over aviation-specific instantiations. Finally, the analysis is time-bound; evolving architectures, regulations, and attacker tradecraft may change the outcomesReal-world incidents of themes after the period studied.

Conclusion

LLMs and SMEs converge on the major categories, but diverge on where risk concentrates in aviation practice. LLMs provide comprehensive taxonomic coverage, whereas SMEs inject domain-anchored detail that shifts priority toward navigation integrity, cloud-to-aircraft trust, software/data-distribution chains, manufacturing integrity, and the concrete realities of legacy OT and human/insider risk. The normalized Borda results make this explicit: broad themes alone do not determine priority, aviation-specific instantiation does. Accordingly, LLMs are best understood as complements to expert judgment. They are useful for breadth and hypothesis generation, but require domain grounding full understand the domain. Future work should broaden SME sampling beyond OEM/manufacturing to include airlines, ANSPs, MROs, airport operators, and regulators, and then analyze subgroup differences to see how priorities shift by role

References

Abbas, F. (2025). *Enhancing cyber resilience in critical sectors through robust security frameworks*. ResearchGate. https://www.researchgate.net/publication/388823746_Enhancing_Cyber_Resilience_in_Critical_Sectors_Through_Robust_Security_Frameworks

Alam, S. (2022). *Cybersecurity: Past, Present and Future*. arXiv. https://arxiv.org/abs/2207.01227

Bai, Y., Kadavath, S., Kundu, S., Askell, A., Kernion, J., Jones, A., Chen, A., Goldie, A., Mirhoseini, A., McKinnon, C., Brown, T. B., & Amodei, D. (2022). *Constitutional AI: Harmlessness from AI Feedback*. arXiv. https://arxiv.org/abs/2212.08073

Bavishi, R., Elsen, E., Hawthorne, C., Nye, M., Odena, A., Somani, A., & Taşırlar, S. (2023, October 17). Fuyu-8B: A multimodal architecture for AI agents. Adept AI. https://www.adept.ai/blog/fuyu-8b

Bentley, S. (2025, January 29). EASA Aviation Cyber Security Overview. Sofema Aviation Services. https://sassofia.com/blog/easa-aviation-cyber-security-overview/

Business Insider. (2024, December). *Meet the power players at Elon Musk's xAI, the startup taking on OpenAI with its Grok chatbot*. https://www.businessinsider.com/xai-power-players-elon-musk-startup-grok-2024-12

Choudhary, G., Sihag, V., Gupta, S., & Shandilya, S. K. (2022). Aviation attacks based on ILS and VOR vulnerabilities. *Journal of Surveillance, Security and Safety*, *3*(2), 27–40. https://doi.org/10.20517/jsss.2021.17

Databricks. (2024, March 27). *Introducing DBRX: A new state-of-the-art open LLM*. https://www.databricks.com/blog/introducing-dbrx-new-state-art-open-llm

DeepSeek-AI. (2025). DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning. arXiv. https://arxiv.org/abs/2501.12948

Delso-Vicente, A.-T., Díaz-Marcos, L., Aguado-Tevar, O., & García de Blanes-Sebastián, M. (2025). Factors influencing employee compliance with information security policies: A systematic literature review of behavioral and technological aspects in cybersecurity. *Future Business Journal, 11*(1), Article 452. https://fbj.springeropen.com/articles/10.1186/s43093-025-00452-7

FAA. (2023, January 19). *FAA NOTAM Statement*. Federal Aviation Administration. https://www.faa.gov/newsroom/faa-notam-statement

Google DeepMind. (2023). Introducing Gemini: our largest and most capable AI model. Retrieved from https://blog.google/technology/ai/google-gemini-ai/blog.google+2

Government Accountability Office. (2021). *Aviation cybersecurity: FAA should fully implement key practices to strengthen its oversight of passenger airline cybersecurity*. https://www.gao.gov/assets/gao-21-86.pdf

IBM. (n.d.-a). *What is information security?*. IBM. Retrieved from https://www.ibm.com/think/topics/information-security

IBM. (n.d.-b). *Types of cyberthreats*. IBM. Retrieved from https://www.ibm.com/think/topics/cyberthreats-types

Lokare, A., Bankar, S., & Mhaske, P. (2025). *Integrating cybersecurity frameworks into IT security: A comprehensive analysis of threat mitigation strategies and adaptive technologies*. arXiv. https://arxiv.org/abs/2502.00651

Martin, J. L. (n.d.). *The Borda Count Method (Tannenbaum, §1.3)*. University of Kansas. https://jlmartin.ku.edu/courses/math105-F11/Lectures/chapter1-part2.pdf

Meta AI. (2024, April 18). *Introducing Meta Llama 3: The most capable openly available LLM to date*. https://ai.meta.com/blog/meta-llama-3/

Mezzi, E., Massacci, F., & Tuma, K. (2025). Large Language Models are Unreliable for Cyber Threat Intelligence. arXiv. https://arxiv.org/abs/2503.23175

Milvich, M., Hammond, J., Bowman, G., & Shackelford E. (2022). *Cyberattacks on SATCOM: Understanding the Threat.* IO Active White Paper. https://www.ioactive.com/wp-content/uploads/2025/05/IOA-SATCOM-22.pdf

Mirsky, Y., Demontis, A., Kotak, J., Shankar, R., Gelei, D., Yang, L., Zhang, X., Lee, W., Elovici, Y., & Biggio, B. (2021). The Threat of Offensive AI to Organizations. arXiv. https://arxiv.org/abs/2106.15764

F, A., Alzoubi, Y. I., Gill, A. Q., & Anwar, M. J. (2022). Cybersecurity enterprises policies: a comparative study. *Sensors, 22*(2), 538. https://doi.org/10.3390/s22020538

Mistral AI. (2025). *Models*. Retrieved from https://mistral.ai/models

Musa, M., & Osazuwa, M. C. (2024). The expanding attack surface: Securing AI and machine learning systems in security operations. *International Journal of Innovative Science and Research Technology, 9*(5), 2498–2505. https://doi.org/10.38124/ijisrt/IJISRT24MAY1613

National Institute of Standards and Technology. (n.d.). *Cybersecurity*. NIST Computer Security Resource Center. Retrieved from https://csrc.nist.gov/glossary/term/cybersecurity

Naveed, H., Khan, A. U., Qiu, S., Saqib, M., Anwar, S., Usman, M., Akhtar, N., Barnes, N., &

Mian, A. (2023). *A comprehensive overview of large language models*. arXiv. https://arxiv.org/abs/2307.06435

Paris, M. (2025, April 12). *ChatGPT hits 1 billion users? 'Doubled in just weeks' says OpenAI CEO Sam Altman*. Forbes. https://www.forbes.com/sites/martineparis/2025/04/12/chatgpt-hits-1-billion-users-openai-ceo-says-doubled-in-weeks/

Ríos-Campos, C., Paz, S. C. V., Vilema, G. O., Díaz, L. M. R., Zambrano, D. P. F., Zambrano, G. M. M., … & Anchundia-Gómez, O. (2024*). South Florida Journal of Development, 5*(8), e4276. https://doi.org/10.46932/sfjdv5n8-021

Santamarta, R. (2019). *Arm IDA and cross check: Reversing the 787s core network.* IOActive White Paper. https://www.ioactive.com/wp-content/uploads/2025/05/Arm-IDA-and-Cross-Check_-Reversing-the-787s-Core-Network.pdf

Smith M., Strohmeier M., Lenders V., & Martinovic I. (2016). *On the security and privacy of ACARS*. 2016 Integrated Communications Navigation and Surveillance (ICNS). IEEE, 2016. pp. 1–27. https://pdfs.semanticscholar.org/7ad9/83c960d531968355647d68e3f116cccbe55c.pdf

Strohmeier, M., Lenders, V., & Martinovic, I. (2013). On the security of the Automatic Dependent Surveillance-Broadcast protocol. *arXiv*. https://arxiv.org/abs/1307.3664

Technology Innovation Institute. (2023). *Falcon LLM: Open-source multilingual large language model*. Retrieved from https://falconllm.tii.ae/falcon-models.htmlHugging Face+4

Triplett, W. J. (2022). Addressing human factors in cybersecurity leadership. *Journal of Cybersecurity and Privacy, 2*(3), 573–586. https://doi.org/10.3390/jcp2030029

Ukwandu, E., Ben Farah, M. A., Hindy, H., Bures, M., Atkinson, R., Tachtatzis, C., & Bellekens, X. (2021). Cyber-security challenges in aviation industry: A review of current and future trends. *arXiv*. https://arxiv.org/abs/2107.04910

University of Florida. (n.d.). *Statistics Cheat Sheet for Inter-Rater Reliability*. Retrieved from https://fycs.ifas.ufl.edu/swisher/6802_22/Statistics_Inter-Rater_Reliability.pdf

Vears, D. F., & Gillam, L. (2022). Inductive content analysis: A guide for beginning qualitative researchers. *Focus on Health Professional Education, 23*(1), 111–118. https://fohpe.org/FoHPE/article/view/544

World Economic Forum. (2021). *Cybersecurity in aviation: Building a resilient future*. Retrieved from https://www.weforum.org/stories/2021/06/cybersecurity-in-aviation/

Xu, H., Wang, S., Li, N., Wang, K., Zhao, Y., Chen, K., Yu, T., Liu, Y., & Wang, H. (2024). *Large language models for cyber security: A systematic literature review*. arXiv. https://arxiv.org/abs/2405.04760

Yan, B., Li, K., Xu, M., Dong, Y., Zhang, Y., Ren, Z., & Cheng, X. (2025). *On protecting the data privacy of large language models (LLMs): A survey*. arXiv. https://arxiv.org/abs/2403.05156

Yang, A., Li, A., Yang, B., Zhang, B., Hui, B., Zheng, B., Yu, B., Gao, C., Huang, C., Lv, C., ... & Zhou, J. (2025). *Qwen3 Technical Report*. arXiv. https://arxiv.org/abs/2505.09388

Zhang, J., Bu, H., Wen, H., Liu, Y., Fei, H., Xi, R., Li, L., Yang, Y., Zhu, H., & Meng, D. (2024). When LLMs Meet Cybersecurity: A Systematic Literature Review. arXiv. https://arxiv.org/abs/2405.03644

Appendix A – Glossary of Key Terms

- **Aircraft Communications Addressing and Reporting System (ACARS)**—a digital system that allows for the transmission of messages between aircraft and ground stations.
- **Air navigation service provider (ANSP)**—An organization that manages air traffic. Many ANSPs may also develop and validate instrument flight procedures.
- **Air traffic control (ATC)**—A service that directs aircraft on the ground and through controlled airspace with the aim of preventing collisions and organizing the flow of air traffic.
- **Air traffic management (ATM)**—A system that aims to ensure the safe and efficient flow of air traffic.
- **Automatic Dependent Surveillance–Broadcast (ADS-B)**—Technology that tracks the position of aircraft.
- **Global Navigation Satellite System (GNSS)**—satellite-based systems that provide global positioning, navigation, and timing (PNT) services.
- **Information technology (IT)**—the use of computers to manage data and information.
- **Maintenance, Repair, and Overhaul (MRO)**—services required to keep an aircraft airworthy.
- **Man-in-the-middle attack (MITM)**—a cyberattack where an attacker secretly intercepts and possibly alters communication between two other parties.
- **Notice to Airmen (NOTAM)**—a time-sensitive message that alerts pilots and personnel to potential hazards or changes to flight operations.
- **Operational Technology (OT)**—the hardware and software that monitor industrial equipment.
- **Original Equipment Manufacturer (OEM)**—a company that originally built an aircraft or its components.
- **Positioning, Navigation, and Timing (PNT)**—services that provide accurate location, direction of travel, and precise time information.
- **Satellite communications (SATCOM)**—communications that use an artificial satellite to relay information.
- **Traffic collision avoidance system (TCAS)**—a system designed to reduce mid-air collisions between aircraft.