LLM-Assisted CPSTRIDE Threat Modeling for Critical Water Infrastructure

Dallas Elleman, Amorita Christian and John Hale
The University of Tulsa, USA

dallas-elleman@utulsa.edu
amorita-christian@utulsa.edu
john-hale@utulsa.edu

Abstract: Critical infrastructures face hybrid threats that threat modeling frameworks like STRIDE, MITRE ATT&CK, and Cyber Kill Chain are ill-suited to capture. These frameworks focus on cybersecurity, leaving blind spots, and their reliance on human expertise limits scalability. Attacks on critical water infrastructure underscore the importance of cyber-physical threat modeling, as does the emergence of autonomous vehicles as hybrid attack vectors. This research presents CPSTRIDE, a framework for cyber-physical threat modeling that extends Microsoft's STRIDE. CPSTRIDE defines security properties for cyber-physical systems and exposes vulnerabilities, threats, and attack vectors that conventional approaches miss. We also introduce an LLM-assisted methodology, leveraging Anthropic's Claude Sonnet 4.5 as a domain expert. We apply this approach to construct a comprehensive threat landscape for a water treatment facility, articulating hybrid attack scenarios involving unmanned aerial and underwater vehicles.

Keywords: Critical infrastructure, Cyber-physical systems, LLM-assisted security analysis, Threat modeling, Water treatment facilities, Unmanned aerial vehicles (UAVs), Unmanned underwater vehicles (UUVs)

1. Introduction

Water and wastewater systems (W/WWS) play a foundational role among the critical infrastructure (CI) sectors vital to U.S. national security and societal well-being (Bello et al., 2023; Parsons, 2024). Other sectors are interdependent on W/WWS -- relying on and enabling water services. Failures in W/WWS can cascade across multiple sectors (Zechman Berglund et al., 2020).

W/WWS and other CI are cyber-physical systems (CPS) because they rely on sensors, actuators, computation, and networks to monitor and control physical processes (Duo, Zhou and Abusorrah, 2022). These technologies promote efficiency but introduce threats to essential services (Joint Cybersecurity Advisory, 2021). UcedaVelez and Morana (2015) define threat modeling as a process for identifying attack scenarios and vulnerabilities to assess risk. Systematic threat modeling is imperative to build secure and resilient CPS (Duo, Zhou and Abusorrah, 2022).

Three frameworks have helped defenders assess cyber-threats: Microsoft's STRIDE, MITRE's ATT&CK, and Lockheed Martin's Cyber Kill Chain. However, none explicitly represents physical aspects, leaving blind spots around physical vulnerabilities and cyber-physical interactions in modern CI (Alexander, Belisle and Steel, 2020; Assante and Lee, 2015; Yampolskiy et al., 2012). These frameworks also depend on domain expertise and extensive human effort to identify threats in complex systems (Huang, Poskitt and Shar, 2024).

STRIDE's data flow diagram (DFD), for instance, includes informational stores, flows, and processes but excludes material and energetic counterparts, as well as physical components, media, channels, and signals (Yampolskiy et al., 2012). MITRE's ATT&CK framework covers a range of cyber-attack tactics, but even its industrial control systems (ICS) variant does not explicitly represent physical threats or system aspects (Alexander, Belisle and Steel, 2020). Lockheed Martin's Cyber Kill Chain acknowledges the contrast between operational technology (OT) and IT but includes no physical examples of reconnaissance, weaponization, delivery, persistence, or other kill chain phases (Assante and Lee, 2015).

This reflects decades of digital transformation that have lowered barriers for adversaries while increasing the impact of cyber-attacks (UcedaVelez and Morana, 2015). Aerial, ground, and underwater vehicles (UAV, UGV, UUV), are cyber-physical threat vectors that empower adversaries (Khawaja et al., 2022; Rassler and Veilleux-Lepage, 2025). Comprehensive threat modeling for CPS must address not only cybersecurity vulnerabilities but also physical and cyber-physical vectors, CI interdependencies, and constraints of limited domain expertise and human resources (Huang, Poskitt and Shar, 2024; Yampolskiy et al., 2012).

Our approach addresses this 'physical gap' with CPSTRIDE, an extension of STRIDE. CPSTRIDE expands its definitions of security properties and threats, introducing a Cyber-Physical Flow Diagram (CPFD) specification (Figure 1) that offers advantages over STRIDE's DFD. We address human resource limitations with LLM-assisted

threat modeling grounded in context engineering principles. Specifically, we equip Anthropic's Claude Sonnet 4.5 with subject-matter expertise, enabling comprehensive enumeration of cyber-physical threats to a water treatment facility, including those posed by UAVs and UUVs.

Cyber-physical Flow Diagram (CPFD)

Cyber indicates information, data, or signal; *physical* indicates material, energy, or force; *cyber-physical* indicates the integration or combination of the two. Each CPFD element is categorized as either cyber, physical, or cyber-physical depending on whether the element can reasonably be considered vulnerable to cyber and/or physical threats.

Element name, abbreviation, and description	Graphical Symbol	Examples C: *Cyber* P: *Physical* CP: *Cyber-physical*
Interactor (I) An entity that exchanges data, energy, or material with the CPS but remains outside its design scope and/or control boundary.	CI/PI/CPI	**CI**: External APIs, the Internet and other networks. **PI**: Raw material sources; water, gas or electric mains. **CPI**: Humans (e.g., employees, contractors), external orgs (e.g., supply chain providers, partners, customers), technological entities (e.g., autonomous delivery and service robots).
Trust Boundary (TB) A virtual and/or physical zone of privileged access.	CTB/PTB/CPTB	**CTB**: Password-protected systems, encrypted files, or trusted computing environments. **PTB**: Physically secured areas with controlled access, locked rooms, fenced perimeters, analog safes, motor housings, machine casings. **CPTB**: Secured areas with both physical barriers (locks, fences) and cyber controls (authentication, surveillance monitoring).
Store (S) Data, energy, or material at rest, distinct from its storage medium or container. Nesting is allowed.	CS/PS/CPS NCS/NPS/NCPS	**CS**: Files, databases, registry keys. **PS**: Raw materials, simple manufactured objects, physical keys. **CPS**: Smart materials, physical key cards, 3D-printed objects. *Note: The 3D-printed object's transformation from cyber to physical makes it vulnerable to time-dependent cyber-physical threats.*
Flow (F) Data, energy, or material in motion, distinct from its enabling path, channel or medium. Flows interconnect Stores, Processes, Devices, and Interactors.	CF/PF/CPF	**CF**: Function calls, network communications, data transfers. **PF**: Material flows, energy transfers, mechanical forces. **CPF**: Sensor data streams, HVAC / IoT communications, transport of smart materials or devices, cyber-physical process I/Os.
Link (L) - New for CPFD A logical and/or physical path, channel, or medium that enables Flows, and interconnects Devices and Interactors.	CL/PL/CPL	**CL**: File formats / schema, data structures; communication ports, channels, & protocols. **PL**: Geographic routes, power lines, fluid pipes. **CPL**: RF spectrum, air (visible light / IR / acoustic transmission).
Process (P) Activity that transforms inputs into outputs.	CP/PP/CPP	**CP**: Only digital inputs and outputs, e.g. **any running code**. **PP:** Only physical inputs and outputs, e.g. manual manufacturing, simple raw material mixing / refining. **CPP**: Cyber-physical inputs and/or outputs, e.g. OT processes, smart manufacturing, automated logistics, robotic assembly, adaptive environmental control, etc.
Device (D) - New for CPFD An instantiation of computational capability and/or physical functionality that enables or abstracts Processes and Stores; a virtually- and/or physically-embodied enabler of Processes and/or Storage in a cyber-physical system. *Devices may be modeled as enablers of explicit Processes or as abstraction layers when internal process logic is not decomposed for threat analysis.*	CD/PD/CPD	**CD**: Abstracted virtual / digital resources, e.g. virtual sensors and machines, Docker containers, digital twins, cloud compute instances, remote database servers, content delivery networks (CDN), cloud storage instances, distributed blockchain ledgers. **PD**: Mechanical actuators, manual valves, analog gauges, hydraulic motors, physical key storage lockboxes, material storage tanks, pressure vessels, chemical reagent containers. **CPD**: Embedded systems, smart thermostats, autonomous vehicles, IoT-enabled medical implants, OT actuators, desktop computers, 3D printers, smart inventory management systems, RFID-enabled storage cabinets, IoT-connected storage tanks with sensors.

Figure 1: Cyber-Physical Flow Diagram (CPFD) specification for the CPSTRIDE threat modeling framework

The remainder of the paper is organized as follows: Section 2 presents background on STRIDE-based CPS threat modeling, W/WWS threat modeling, UAV/UUV capabilities, and LLM-assisted security analysis. Section 3 describes CPSTRIDE . Section 4 details our methodology for LLM-assisted analysis, system modeling, and UAV/UUV attack scenario integration. Section 5 presents our results. Section 6 reflects on the results and Section 7 concludes.

2. Background

2.1 STRIDE-based CPS Threat Modeling

Researchers frequently adapt Microsoft's STRIDE framework for CPS threat modeling (Saßnick et al., 2024). The STRIDE acronym represents six threat types and corresponding security properties (see Figure 3). STRIDE's systematic four-phase methodology constructs system data flow diagrams (DFDs), catalogues potential threats, analyzes vulnerabilities, and plans mitigations. The framework's five DFD symbols depict cyber-entities but lack established conventions for representing physical components (Shostack, 2025).

Applying STRIDE to CPS reveals the framework's promise and limitations. In the UAV domain, Yampolskiy et al. (2012) noted STRIDE's inability to adequately differentiate CPS interactions, e.g., cyber versus physical communications. Their extended DFD (xDFD) notation added four symbols to represent physical components, signals, communication media, and optional data flows, but did not expand STRIDE's core security properties or threat definitions. Similarly, Khan et al. (2017) demonstrated STRIDE's application to microgrid electrical systems, acknowledging DFD limitations for physical components and scoping their analysis exclusively to cyber-vulnerable components.

2.2 W/WWS Threat Modeling

W/WWS comprise water and wastewater treatment systems; this paper focuses on water treatment systems (WTS), which ensure safe drinking water. Water passes through five treatment processes to make it potable. The process begins with coagulation, where chemicals destabilize and bind suspended particles such as dirt and organic matter. This is followed by flocculation, where gentle mixing and sometimes chemicals encourage the formation of larger aggregates (flocs). Next, sedimentation, gravity forces flocs to settle in the basin, allowing

separation from the water. The clarified water then undergoes filtration, passing through layers of media removing impurities. Activated carbon accompanies filtration eliminating unpleasant tastes and odors. The final step is disinfection, in which disinfectants inactivate any remaining pathogens. Storage tanks hold the disinfected water for distribution to businesses and communities (CDC, 2025).

Historically, security for W/WWS has relied heavily on isolation and access restriction (Tuptuk et al., 2021). Threat modeling focused on contamination from hazardous chemical and biological agents (Zechman Berglund et al., 2020). However, IoT introduces new threats. Previous researchers evaluated the effectiveness of STRIDE/DREAD threat modeling for W/WWS and found limitations in its ability to assess the likelihood of a threat (Davis and Keskin, 2024).

2.2.1 Environment and architecture

The complex environment and architecture of W/WWS creates unique security challenges. Legacy systems, resource constraints, interdependencies, and cascading effects are confounding factors. Moreover, W/WWS routinely exhibit poor basic cyber hygiene, lacking secure default credentials, adequate network segmentation, and incident response planning (Joint Cybersecurity Advisory, 2021).

Its technology foundation, Industrial Control Systems (ICS), supports monitoring and control of physical components (Figure 2). These include Supervisory Control and Data Acquisition (SCADA) systems, programmable logic controllers (PLCs), human-machine interfaces (HMIs), master/remote terminal units (M/RTUs), distributed control systems (DCS), and OT networks. Physical components, which are geographically dispersed, often in remote areas, include tanks, pipes, pumps, filters, clarifiers, chemical dosing equipment, reservoirs, and treatment basins. These are susceptible to sabotage, theft, and vandalism, which can lead to contamination or service disruptions. Cyber-physical interfaces converge ICS and physical components using sensors, actuators, Internet of Things (IoT) devices, and networked controllers (Hassanzadeh et al., 2020; Parsons, 2024).

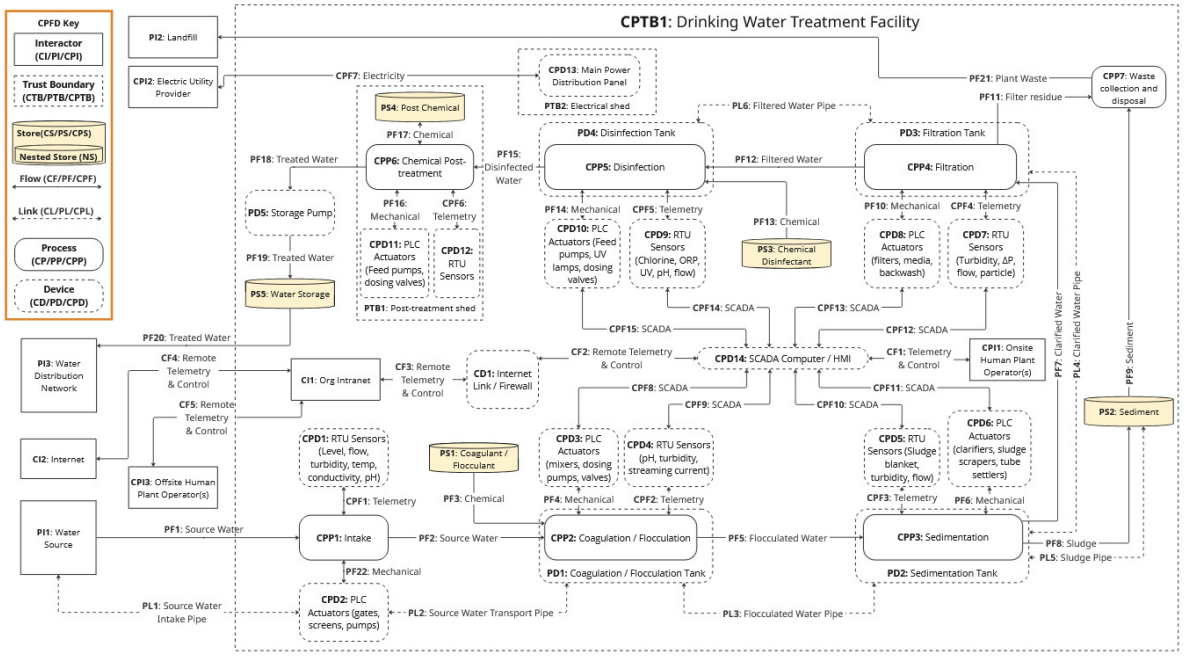

Figure 2: Cyber-Physical Flow Diagram (CPFD) for a representative drinking water treatment facility

2.2.2 Emerging threats and security issues

While OT environments incorporate security measures in highly regulated industries such as oil and gas, W/WWS remain comparatively vulnerable due to less regulation. As these systems become more connected, their attack surfaces expand, increasing the risk of compromise (Muncaster, 2025; Western Water, 2025). Recent data highlight the threat to W/WWS. Across the US and UK, approximately 62% of water and electricity utilities reported experiencing at least one cyberattack in the past year, with 80% of those affected suffering multiple incidents. Among these, 54% resulted in permanent data or system corruption, and 59% caused operational disruptions (Muncaster, 2025). An EPA assessment identified 97 drinking water systems serving 26.6 million people with critical or high-risk cybersecurity vulnerabilities (Joint Cybersecurity Advisory, 2021).

Several incidents illustrate the impact of these attacks. In November 2023, the hacktivist group CyberAv3ngers compromised PLCs in water utilities across North America, Europe, and Australia, causing at least one community to experience a two-day service disruption (Dragos, 2024). In the US, a Pennsylvania water authority was targeted by a state-backed actor, necessitating a switch from automated to manual operations to maintain water pressure (AP News, 2024). In the UK, Southern Water experienced a ransomware attack, compromising the data of 4.6 million customers (Coker, 2024).

W/WWS are also susceptible to physical attacks. In April 2013, approximately 400 residents were affected when an intruder tampered with chemical settings at a treatment plant in Chatsworth, Georgia (Barbour, 2013). This incident illustrates how physical threats bypass cybersecurity controls are not designed to monitor physical infrastructure. A threat may originate from physical intrusion, making digital protections a secondary line of defense (Barbour, 2013). And while STRIDE helps identify threats to IT and OT, it may overlook physical or hybrid attack vectors.

2.3 Emerging UAV/UUV Capabilities

Unmanned aerial and underwater vehicles (UAVs and UUVs) are remotely piloted or autonomous platforms that present significant security concerns (Sathyamoorthy, 2015). The Russia-Ukraine war and other recent conflicts have seen commercial-off-the-shelf (COTS) UAVs operationally weaponized by nation-states and violent extremist organizations (VEOs) against critical infrastructure as performance advances and integration with AI and additive manufacturing have elevated their relevance as low-cost, asymmetric weapons (Rassler and Veilleux-Lepage, 2025; World Bank, 2025). Increased UAV-enabled terrorism is anticipated as these battlefield developments create opportunities for lone-wolves and VEOs to amplify impact and surprise. COTS UUV deployments have grown from hundreds to thousands of units since the early 2000's (Race and Piskura, 2009; Campagnaro et al., 2023). Modern platforms feature modular payload architectures and operate at depths of 300 meters for only a few thousand dollars; acoustic modems affordably enable remote and coordinated operations (Blue Robotics, 2020; Water Linked, n.d.).

In 2023, a COTS UAV dropped dye packets into swimming pools, causing thousands of dollars in damage (NBC Philadelphia, 2024); in 2025, the EPA and WaterISAC reported a major utility burglary preceded by UAV surveillance (EPA and WaterISAC, 2025). These and other incidents underscore likely COTS UAV threats to critical water infrastructure, including reconnaissance and delivery of contaminants, explosives, or cyber-intrusion devices. COTS UUVs, in contrast, may kinetically threaten submerged components such as intake structures and pipelines where detection and defense are challenging, or deliver chemical or biological payloads directly into water intake streams. Although fewer cases of COTS UUV terror attacks on water infrastructure have been documented, security experts emphasize that water utilities must address these evolving threats from unmanned and autonomous systems (Edwards, 2024).

2.4 LLM-assisted Security Analysis

Large Language Models (LLMs) have rapidly transformed cybersecurity and catalyzed research spanning defensive applications and offensive capabilities in software, hardware, networks, systems, penetration testing, malware detection, threat intelligence, forensics, and many other areas (Yao et al, 2024; Jaffal, Alkhanafseh, and Mohaisen, 2025; Zhang et al. 2025). Although barriers to efficacy such as hallucinations and inaccuracy persist, advanced models continue to surpass benchmarks, and research demonstrates substantial performance gains through model pre-training, domain-specific fine-tuning, and inference-time *context engineering* techniques. Context engineering is the emerging formal discipline of designing and managing the informational payloads that LLMs receive, and which are primary factors steering LLM behavior. As LLM capabilities have evolved from basic instruction-following to complex reasoning, methods for managing their context have correspondingly matured. Mei et al. (2025) survey over 1400 papers and formulate a taxonomy that includes context retrieval, generation, processing, and management as foundational components used to implement dynamic knowledge injection, persistent memory, tool-integrated reasoning, agent-environment interaction, and multi-agent coordination and communication protocols.

LLM-assisted threat modeling offers one promising approach to address the labor-intensive, expert-dependent nature of traditional security analysis. Tools are proliferating rapidly in the space; the ThreatModeling-LLM framework demonstrates successful combined application of LoRA fine-tuning, Chain of Thought (CoT) reasoning, and Optimization by PROmpting (OPRO) for STRIDE-based threat modeling on banking system datasets, with demonstrated improved accuracy from 0.36 to 0.69 on NIST 800-53 control codes by the open Llama-3.1-8B model (Wu, et al. 2025). Similarly, Elsharef et al.'s NDSS 2024 workshop study found RAG-enhanced

Llama 2 model outputs to prove more concise and contain higher-density meaningful information than base LLMs. Deep-STRIDE enables automated STRIDE threat analysis directly from visual Data Flow Diagrams through a consortium of fine-tuned open-source vision-language models integrated with the DeepSeek-R1 reasoning LLM, representing the first application of a VLM-LLM hybrid reasoning pipeline for threat modeling (Bandara et al., 2025). STRIDE GPT offers multi-modal threat modeling across frontier and local models via Ollama (Adams, 2025).

Despite these advances, significant challenges remain that require human oversight. TM-Bench, the first benchmark specifically designed to evaluate LLM capabilities in security threat modeling, reveals substantial variation in how different models approach threat identification (*TM-Bench,* n.d.). The research consistently shows that base LLMs perform well on generic threats but struggle with business-logic vulnerabilities, organization-specific operational risks, and novel attack patterns, manifesting contextual errors and logical failures. The appropriate deployment model therefore currently positions LLMs as initial draft generators and safety nets against oversight rather than autonomous decision-makers, with fine-tuned smaller models and custom datasets outperforming prompt engineered large models in specialized domains.

3. CPSTRIDE

CPSTRIDE extends Microsoft's STRIDE framework to address security challenges in CPS (Elleman & Hale, 2025). While STRIDE's focus on data flows, data stores, and software processes is effective for IT systems, it creates significant gaps when modeling systems with material and energetic components. CPSTRIDE addresses these limitations by expanding STRIDE's scope to consider cyber-physical threats while preserving its four-step methodology: system modeling, threat identification, vulnerability investigation, and mitigation planning.

CPSTRIDE's primary innovation is the Cyber-Physical Flow Diagram (CPFD) (Figure 1), which extends DFDs. CPFD specifies seven element types using C/P/CP notation to indicate cyber, physical, or cyber-physical characteristics: Interactors (entities beyond system boundaries), Trust Boundaries (zones of privileged access), Stores (data/materials/energy at rest), Processes (transformation activities), Flows (data/materials/energy in motion), and two new elements—Links (pathways enabling flows, such as pipes or RF spectrum) and Devices (instantiations enabling processes and storage, such as machines, sensors, and actuators). This enables comprehensive modeling of CPS architectures while maintaining visual clarity.

CPSTRIDE redefines STRIDE's six security properties and threat categories (Figure 3) to encompass physical contexts. Authentication becomes Authenticity (verifying data/material/energy components), while Confidentiality becomes Containment (preventing unauthorized data/material/energy extraction). The six threat categories - Spoofing, Tampering, Repudiation, Interception (replacing Information Disclosure), Denial of Service, and Elevation of Privilege - receive expanded definitions that include physical attack vectors such as counterfeiting, sabotage, and material/energy diversion. Unlike STRIDE's restrictive susceptibility assumptions, CPSTRIDE recognizes that all CPFD elements are potentially vulnerable to all threat categories, enabling more comprehensive threat analysis for systems where cyber-attacks can manifest as physical harms.

CPSTRIDE Cyber-Physical Threats

Each Threat potentially violates a corresponding Security Property. In the Examples column, *Cyber* indicates threats to information, data, control signal, etc.; *Physical* indicates threats to material, energy, force, etc.; *Cyber-physical* indicates the integration or combination of the two. Highlighted rows contain new CPSTRIDE Threats.

Threat (Security Property)	Definition	Examples **C:** Cyber Only **P:** Physical Only **CP:** Cyber-physical
Spoofing	Falsification of identity, source, or authenticity of system elements, including users, processes, signals, or physical/cyber-physical stores, undermining trust mechanisms and authentication controls within the CPS. *Violates Authenticity.*	**C:** Phishing, smishing, social engineering, malicious broadcast of trusted WiFi network SSID, typosquatting, deepfaking. **P:** Faking physical credentials, passing off counterfeit parts and materials as genuine, forging signatures on physical documents. **CP:** Broadcasting fake GPS to misguide autonomous vehicles or drones, injection of counterfeit OT sensor readings.
Tampering	Unauthorized modification, corruption, or alteration of legitimate cyber-physical entities including data, structures, energy flows, material compositions, or control signals, that compromises system integrity. *Violates Integrity.*	**C:** Modifying control logic in industrial automation software. **P:** Physically adjusting valve settings or equipment calibration screws. **CP:** Altering sensor readings through electromagnetic interference, causing the system to respond to fabricated conditions.
Repudiation	Denial of responsibility for actions within the system, either through passive rejection of accountability or active measures to destroy, corrupt, or disable auditing mechanisms or evidence trails that would establish proof of activities, legitimate or malicious. *Violates Non-repudiation.*	**C:** Disabling logging mechanisms to hide evidence of digital access. **P:** Destroying physical access records or tampering with surveillance footage. **CP:** Cross-domain log corruption.
Interception *(previously Information Disclosure)*	Unauthorized acquisition or monitoring of system resources, including data, energy flows, or physical materials, violating containment. Interception replaces and includes the traditional Information Disclosure threat, and *Violates Containment.*	**C:** Capturing sensitive control data through network sniffing. **P:** Physically extracting / diverting material from manufacturing processes. **CP:** Harvesting energy from wireless power transmission systems through unauthorized coupling.
Denial of Service	Impairment or prevention of system availability through any means that renders services, functions, or resources inaccessible or unreliable for legitimate users. *Violates Availability / Reliability.*	**C:** Network flooding, resource exhaustion, communication jamming. **P:** Blockage of moving parts; permanent damage by physical destruction, component sabotage, or irreversible physical alterations; energy disruption through power supply manipulation or battery depletion; environmental manipulation to introduce adverse conditions. **CP:** Creating electromagnetic interference to disrupt wireless communications and/or electronic sensors, physical obstruction of sensors/actuators.
Elevation of Privilege	Exploitation of system vulnerabilities to gain unauthorized higher-level access rights beyond assigned permissions. *Violates Authorization.*	**C:** Traditional privilege elevation cyber-techniques such as exploiting software vulnerabilities to gain administrative access to control systems. **P:** Obtaining master keys or accessing restricted physical areas without authorization. **CP:** Using physical access to maintenance ports to install privileged software that bypasses normal authorization controls.

Figure 3: CPSTRIDE Cyber-physical threats

4. Methodology

Our methodology is informed by context engineering principles described by Mei et al. (2025), who highlight the fundamental asymmetry between LLM proficiency in understanding complex contexts and limitations in generating long-form outputs, and find LLM performance to be strongly determined by inference-time context quality, length, coherence, and self-consistency. We engage Claude 4.5 Sonnet LLM in 10 rounds of conversation via Anthropic's Desktop and Code applications, systematically and iteratively refining context and progressively mitigating hallucinations and inaccuracies in output by identifying and resolving contradictory and incoherent inputs (Figure 4). Successive rounds of conversation produce more detailed, concise, and coherent outputs, which are fed into a clean context window during subsequent rounds. Early conversations (1-4) establish a contextual foundation by refining the CPSTRIDE specification and visual CPFD, enumerating elements, and developing JSON schemas; conversations 5-6 apply these schemas but reveal critical specification violations; conversation 7 documents root cause analysis that leads to further specification disambiguation; conversations 8-10 successfully apply the refined specifications to produce accurate system models and comprehensive threat matrices. All conversations can be found in full in our project repository.

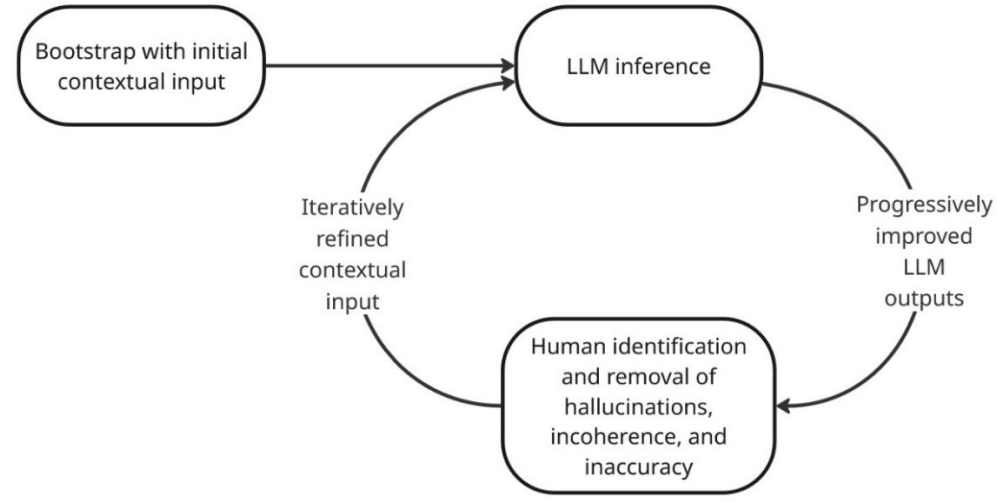

Figure 4: Iterative context-refinement loop for LLM-assisted threat modeling

These rounds of conversation with Claude 4.5 Sonnet demonstrate the value of iterative context refinement and specification-aware validation, and produce: (1) updated CPSTRIDE and CPFD specification documents (v3.0) that provide a more coherent and realistic threat modeling framework for CPS involving SCADA/ICS architectures; (2) progressively refined drinking water treatment facility CPFDs (v1 through v4) that accurately capture cyber-physical interactions; (3) specification-compliant JSON system representations that enable machine-readable threat modeling and automated analysis; and (4) a comprehensive CPSTRIDE threat matrix documenting threats from adversarial UAVs and UUVs to 53 drinking water treatment facility elements (58% of total) that would likely remain invisible using traditional STRIDE analysis. Our process reveals that LLM hallucinations and errors often propagate from inconsistencies in source materials and shows the need for specification-aware validation tools as well as the value of LLMs in identifying contradictions between formal specifications and practical system implementations.

5. Results

We applied our iterative LLM-assisted CPSTRIDE methodology to a representative drinking water treatment facility CPFD (Figure 2) and produced a comprehensive threat matrix that identifies attack scenarios across all threat categories and system elements, with emphasis on emerging UAV and UUV vectors. The matrix includes physical and cyber-physical interactors (water source, landfill, distribution network, electric utility), trust boundaries (post-treatment and electrical sheds), stores (coagulant, sediment, disinfectant, post-chemicals, water storage), processes (intake through waste disposal), devices (treatment tanks and pumps), flows (water, chemicals, mechanical forces, electricity), and links (power distribution infrastructure and treatment stage connections). The matrix identifies distinct UAV-enabled attack scenarios spanning reconnaissance (thermal imaging of facilities, mapping distribution networks, identifying chemical storage), physical attack (explosive/incendiary payloads targeting infrastructure), contamination (aerial delivery of chemical/biological agents into open storage or chemical supplies), and infrastructure destruction (targeting power systems, treatment tanks, pipelines). UUV threats prove particularly dangerous in the identified scenarios due to detection challenges and direct access to intake structures: contamination injection bypassing or confounding treatment stages, physical destruction of underwater infrastructure, and blockage of intake systems. High-impact threat vectors identified include UUV-enabled contamination injection at source intake, UAV-delivered contaminant payloads; UAV/physical attacks on main power distribution causing facility-wide operational disruption; UAV-enabled breaches of physical trust boundaries providing surveillance opportunities and access to chemicals and electrical infrastructure; and multi-stage contamination injection via UAV/UUV at various physical flows throughout the treatment process.

The threat analysis reveals systematic patterns across element types. Physical stores prove vulnerable to authenticity violations (counterfeit chemicals, contaminated materials) and availability attacks (UAV-delivered explosives, supply chain disruption). Cyber-physical processes demonstrate the highest threat density, with distinct scenarios across all water treatment stages, particularly in tampering (PLC manipulation, mechanical sabotage) and denial of service (infrastructure destruction, chemical supply disruption). Disinfection-stage threats carry severe public health implications: under-chlorination or UV lamp sabotage could allow pathogen survival, while post-filtration pipeline tampering before disinfection creates further contamination pathways to distribution networks.

6. Discussion

Our results demonstrate that CPSTRIDE bridges an urgent gap in CPS threat modeling. The emergence of affordable COTS UAVs and UUVs as attack vectors fundamentally alters the threat calculus for critical infrastructure by introducing asymmetries that favor adversaries. UAV reconnaissance capabilities can provide attackers with detailed intelligence while remaining difficult to detect. UUV threats to underwater intake infrastructure prove especially dangerous due to detection challenges and the ability to inject contaminants that bypass downstream treatment processes.

Our LLM-assisted methodology demonstrates significant benefits and noteworthy limitations. The primary advantage is scalability: Claude 4.5 Sonnet systematically enumerated CPSTRIDE threats across 91 system elements and 6 threat categories, producing a comprehensive threat matrix that would otherwise require substantial human expert hours. The iterative context refinement process enabled progressive accuracy and revealed that initial LLM errors originating from specification inconsistencies could be mitigated, and highlights LLMs' value in identifying contradictions between formal specifications and implementations.

However, significant challenges remain. Our methodology required 10 rounds of carefully structured conversations with human oversight, specification refinement, and validation. Conversational error rates demonstrate that base LLM outputs cannot be trusted without verification, consistent with TM-Bench findings showing substantial variation in model threat identification capabilities (TM-Bench, n.d.). The most significant challenge emerged during conversations 5-7, when two independent attempts to generate JSON representations of the drinking water treatment facility CPFD both produced errors. Initial analysis suggested these were LLM generation failures, but deeper investigation revealed that Claude 4.5 Sonnet had reproduced specification violations present in the source materials themselves. The root cause was a mismatch between the v2.0 CPSTRIDE specification and the v3.0 CPFD's representation of several SCADA devices. This insight led to a refinement of the CPSTRIDE specification to v3.0, with relaxed CPFD conventions that allow Devices to serve a dual role as both enablers and abstraction layers for Processes. So while LLMs have remarkable generative capacities, human domain experts are currently still required to provide output validation and guide specification coherence.

Security, privacy, and reliability warrant consideration when deploying LLM-assisted threat modeling for critical infrastructure. Using commercial LLM services like Anthropic's Claude involves transmitting information to third-party servers, creating information disclosure risks unacceptable for classified or highly sensitive systems. Organizations must evaluate whether: (1) sanitized/abstracted system models provide sufficient threat coverage while protecting sensitive details, (2) on-premises deployment of open-source models meets security requirements, or (3) the sensitivity level necessitates human-expert-only approaches. Reliability concerns extend beyond hallucinations to include adversarial risks: as Zhang et al. (2025) document, LLMs themselves face security vulnerabilities including prompt injection, data poisoning, and model extraction attacks. Threat modeling workflows must therefore implement verification protocols, maintain human oversight for critical decisions, and avoid over-reliance on LLM outputs for security decisions.

7. Conclusion

This research demonstrates that cyber-physical critical infrastructure requires frameworks capable of representing physical and hybrid attack vectors that conventional cyber-centric approaches overlook. CPSTRIDE addresses this need by extending STRIDE's flow diagram paradigm, security properties, and threat categories to encompass material, energy, and physical domains. LLM-assisted threat modeling offers a promising approach to address the labor-intensive, expert-dependent nature of traditional security analysis, but presents hallucination and inaccuracy risks that must be mitigated through careful human oversight. Our methodology demonstrates iterative human-in-the-loop refinement of contextual inputs to achieve successively more detailed, concise, and coherent outputs. We leveraged Claude 4.5 Sonnet as a subject matter expert and performed CPSTRIDE analysis to produce a comprehensive cyber-physical threat matrix for a drinking water treatment facility and modeled emerging UAV and UUV hybrid attack vectors, providing an example of practical, scalable threat modeling for vital public infrastructure.

Ethics Declaration: Ethical clearance was not required for this research.

AI Declaration: The researchers employed Anthropic's Claude Sonnet 4.5 for LLM-assisted threat modeling, as described in Section 4. During the preparation of this work the authors used Claude to assist in literature review synthesis and initial drafting. The authors have reviewed and edited the content as needed and take full responsibility for the content of the publication.

Project Repository: The repository for this project can be found at: https://github.com/DallasElleman/CPSTRIDE

References

Adams, M. (2025) "mrwadams/stride-gpt." Available at: https://github.com/mrwadams/stride-gpt (Accessed: 01 November 2025).

Alexander, O., Belisle, M. and Steele, J. (2020) *MITRE ATT&CK® for Industrial Control Systems: Design and Philosophy*. McLean, VA: The MITRE Corporation. Available at: https://attack.mitre.org/docs/ATTACK_for_ICS_Philosophy_March_2020.pdf (Accessed: 1 September 2025)

AP News (2024) 'States and Congress wrestle with cybersecurity after Iran attacks small town water utilities', 2 January. Available at: https://apnews.com/article/water-utilities-hackers-cybersecurity-1c475f5d2ef3b5d52410c93bdeab3aad (Accessed: 10 September 2025).

Assante, M.J. and Lee, R.M. (2015) 'The Industrial Control System Cyber Kill Chain', SANS Institute, White Paper, October. Available at: https://www.sans.org/white-papers/36297

Bandara, E. et al. (2025) "Deep-Stride: Automated Security Threat Modeling with Vision-Language Models," in 2025 International Conference on Software, Telecommunications and Computer Networks (SoftCOM). 2025 International

Conference on Software, Telecommunications and Computer Networks (SoftCOM), pp. 1–7. Available at: https://doi.org/10.23919/SoftCOM66362.2025.11197424

Barbour, K. (2013) 'Do not drink order for hundreds in North GA; FBI investigating', April. Available at: https://fluoridealert.org/news/do-not-drink-order-for-hundreds-in-north-ga-fbi-investigating/ (Accessed: 9 August 2025).

Bello, A., Jahan, S., Farid, F. and Ahamed, F. (2023) 'A Systemic Review of the Cybersecurity Challenges in Australian Water Infrastructure Management', Water, 168(15).

Blue Robotics (2020) 'BlueROV2 (300m)', *RobotShop*. Available at: https://www.robotshop.com/products/bluerobotics-bluerov2-aluminum-300m (Accessed: 1 September 2025).

Campagnaro, F., Steinmetz, F., Renner, B.-C. and Zorzi, M. (2023) 'Affordable underwater acoustic modems and their application in everyday life: a complete overview', in *The 17th International Conference on Underwater Networks & Systems (WUWNet'23)*, Shenzhen, Guangdong, China, 24-26 November. New York: ACM, pp. 1-8. doi: 10.1145/3567600.3568156.

CDC (2025) 'How Water Treatment Works', Drinking Water. Available at: https://www.cdc.gov/drinking-water/about/how-water-treatment-works.html (Accessed: 10 July 2025).

Coker, J. (2024) 'Southern Water Confirms Data Breach Following Black Basta Claims', Infosecurity Magazine, 25 January. Available at: https://www.infosecurity-magazine.com/news/southern-water-data-breach-black/ (Accessed: 10 September 2025).

Davis, R. and Keskin, O.F. (2024) 'Cyber Threat Modeling for Water and Wastewater Systems: Contextualizing STRIDE and DREAD with the Current Cyber Threat Landscape', presented at the 2024 Systems and Information Engineering Design Symposium, May. doi: 10.1109/SIEDS61124.2024.10534706.

Dragos (2024) 'The Rising Tide of Water Utility Cyber Threats: How Dragos Shields Water Systems', 2 May. Available at: https://www.dragos.com/blog/water-utility-cyber-threats (Accessed: 10 September 2025).

Duo, W., Zhou, M. and Abusorrah, A. (2022) 'A Survey of Cyber Attacks on Cyber Physical Systems: Recent Advances and Challenges', IEEE/CAA Journal of Automatica Sinica, 9(5), pp. 784–800. doi: 10.1109/JAS.2022.105548.

Edwards, W. (2024) 'Protecting Water Utilities from Drone Threats: Understanding the Steps of a Drone Security Methodology that Support the J100 framework', *Utility Security*. Available at: https://utilitysecurity.com/blog/protecting-water-utilities-from-drone-threats-understanding-the-steps-of-a-drone-security-methodology-that-support-the-j100-framework/ (Accessed: 16 September 2025).

Elleman, D. and Hale, J. (2025) 'CPSTRIDE: A threat Modeling Framework for Cyber-Physical Systems', The 20th International Conference on Critical Information Infrastructures Security, Jönköping, Sweden, 21-23 October 2025.

Elsharef, I., Zeng, Z. and Gu, Z. (2024) "Facilitating Threat Modeling by Leveraging Large Language Models," in *Proceedings 2024 Workshop on AI Systems with Confidential Cmputing*. *Workshop on AI Systems with Confidential Computing*, San Diego, CA, USA: Internet Society. Available at: https://doi.org/10.14722/aiscc.2024.23016.

EPA and WaterISAC (2025) 'National Security Information Sharing Bulletin, Q3 2025', *WaterISAC*. Available at: https://www.waterisac.org/system/files/articles/WaterISAC_EPA%20Bulletin_Q3-25_Final.pdf (Accessed: 16 September 2025).

Hassanzadeh, A., Rasekh, A., Galelli, S., Aghashahi, M., Taormina, R., Ostfeld, A. and Banks, M.K. (2020) 'A Review of Cybersecurity Incidents in the Water Sector', Journal of Environmental Engineering, 146(5). doi: 10.1061/(asce)ee.1943-7870.0001686.

Huang, S., Poskitt, C.M. and Shar, L.K. (2024) 'Security Modelling for Cyber-Physical Systems: A Systematic Literature Review', 19 September, arXiv: arXiv:2404.07527. doi: 10.48550/arXiv.2404.07527.

Jaffal, N.O., Alkhanafseh, M. and Mohaisen, D. (2025) "Large Language Models in Cybersecurity: A Survey of Applications, Vulnerabilities, and Defense Techniques," *AI*, 6(9), p. 216. Available at: https://doi.org/10.3390/ai6090216.

Joint Cybersecurity Advisory (2021) 'Ongoing Cyber Threats to U.S. Water and Wastewater Systems', October.

Khan, R., McLaughlin, K., Laverty, D. and Sezer, S. (2017) 'STRIDE-based threat modeling for cyber-physical systems', in 2017 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe), September, pp. 1–6. doi: 10.1109/ISGTEurope.2017.8260283.

Khawaja, W., Semkin, V., Ratyal, N.I., Yaqoob, Q., Gul, J. and Guvenc, I. (2022) 'Threats from and Countermeasures for Unmanned Aerial and Underwater Vehicles', Sensors, 22(10), p. 3896. doi: 10.3390/s22103896.Muncaster, P. (2025) 'Over Half of Attacks on Electricity and Water Firms Are Destructive', Infosecurity Magazine, 3 April. Available at: https://www.infosecurity-magazine.com/news/half-attacks-electricity-water (Accessed: 10 September 2025).

Mei, L. et al. (2025) "A Survey of Context Engineering for Large Language Models." arXiv. Available at: https://doi.org/10.48550/arXiv.2507.13334.

NBC Philadelphia (2024) 'Sky-high vandal: NJ man used drone to drop dye in AC pools, police say', 18 December. Available at: https://www.nbcphiladelphia.com/news/local/sky-high-vandal-nj-man-used-drone-to-drop-dye-in-ac-pools-police-say/4055348/ (Accessed: 16 September 2025).

Parsons, D. (2024) 'Protecting Critical Water Systems with the Five ICS Cybersecurity Critical Controls', March.

Piekarski, M., Wolbach, M. and Okuniewska, M. (2025) 'Employment of uncrewed systems in attacks on critical infrastructure: a hybrid threat perspective. Security challenges related to recent developments in technology', Open Research Europe, 4, p. 129. doi: 10.12688/openreseurope.17797.2 (Accessed: 10 September 2025).

Race, R. and Piskura, J. (2009) *Tethered Antennas for Unmanned Underwater Vehicles: Phase I Final Report*. SBIR Report N08-194, Contract N00014-09-M0038. New Bedford, MA: Brooke Ocean Technology (USA) Inc.

Rassler, D. and Veilleux-Lepage, Y. (2025) 'On the Horizon: The Ukraine War and the Evolving Threat of Drone Terrorism', CTC Sentinel, 18(3), pp. 1-24, March. Available at: https://ctc.westpoint.edu/on-the-horizon-the-ukraine-war-and-the-evolving-threat-of-drone-terrorism/.

Saßnick, O., Rosenstatter, T., Schäfer, C. and Huber, S. (2024) 'STRIDE-based Methodologies for Threat Modeling of Industrial Control Systems: A Review', in 2024 IEEE 7th International Conference on Industrial Cyber-Physical Systems (ICPS), May, pp. 1–8. doi: 10.1109/ICPS59941.2024.10639949.

Sathyamoorthy, D., 2015. A review of security threats of unmanned aerial vehicles and mitigation steps. *J. Def. Secur*, *6*(1), pp.81-97.

Shostack, A. (2025) adamshostack/DFD3, 10 April. Available at: https://github.com/adamshostack/DFD3 (Accessed: 3 May 2025).

TM-Bench: Threat Modeling Benchmark (n.d.). Available at: https://www.tmbench.com/about (Accessed: December 14, 2025).

Tuptuk, N., Hazell, P., Watson, J. and Hailes, S. (2021) 'A Systematic Review of the State of Cyber-Security in Water Systems', Water, 81(13), pp. 1–20.

UcedaVelez, T. and Morana, M.M. (2015) Risk Centric Threat Modeling: Process for Attack Simulation and Threat Analysis. Wiley. Available at: https://books.google.com/books?id=pHtXCQAAQBAJ.

Water Linked (n.d.) 'Underwater Acoustic Modem'. Available at: https://waterlinked.com/modem (Accessed: 16 September 2025).

Western Water (2025) 'Water utilities face rising cybersecurity threats', 15 April. Available at: https://www.western-water.com/2025/04/15/water-utilities-face-rising-cybersecurity-threats/ (Accessed: 10 September 2025).

World Bank (2025) Ukraine Fourth Rapid Damage and Needs Assessment (RDNA4). Available at: https://documents.worldbank.org/en/publication/documents-reports/documentdetail/099022025114040022 (Accessed: 12 September 2025).

Wu, T. *et al.* (2025) "ThreatModeling-LLM: Automating Threat Modeling using Large Language Models for Banking System." arXiv. Available at: https://doi.org/10.48550/arXiv.2411.17058.

Yampolskiy, M., Horvath, P., Koutsoukos, X.D., Xue, Y. and Sztipanovits, J. (2012) 'Systematic analysis of cyber-attacks on CPS-evaluating applicability of DFD-based approach', in 2012 5th International Symposium on Resilient Control Systems, August, pp. 55–62. doi: 10.1109/ISRCS.2012.6309293.

Yao, Y. *et al.* (2024) "A survey on large language model (LLM) security and privacy: The Good, The Bad, and The Ugly," *High-Confidence Computing*, 4(2), p. 100211. Available at: https://doi.org/10.1016/j.hcc.2024.100211.

Zechman Berglund, E., Pesantez, J.E., Rasekh, A., Shaflee, E., Sela, L. and Haxton, T. (2020) 'Review of Modeling Methodologies for Managing Water Distribution Security', Water Resource Plan Management, 146(8), pp. 1–23. doi: 10.1061/(ASCE)WR.1943-5452.0001265.

Zhang, J. *et al.* (2025) "When LLMs meet cybersecurity: a systematic literature review," *Cybersecurity*, 8(1), p. 55. Available at: https://doi.org/10.1186/s42400-025-00361-w.