PRISM-APT: A Model-First Synthesis for APT Defence

Raymond André Hagen
Norwegian Digitalisation Agency (DigDir), Bergen, Norway
NTNU – Norwegian University of Science and Technology, Trondheim, Norway

raymohag@stud.ntnu.no

Abstract: We present PRISM-APT, a practical APT defence model for smaller organisations that integrates governed CTI, behaviour-centric rules, and sovereignty-by-operation thresholds (S-CAP). Advanced Persistent Threats (APTs) routinely exploit the gap between widely used theoretical frameworks and day-to-day operational practice, leaving Security Operations Centres (SOCs) with fragmented, vendor-locked, or jurisdictionally misaligned defences. To address this problem, we introduce PRISM-APT, a model-first synthesis for governed APT defence developed through a multi-year research programme that integrates: (1) empirical studies on SOC practices, human and AI bias, and logging baselines; (2) a curated dataset of APT campaigns; and (3) cooperative digital sovereignty mechanisms aimed at enabling cross-organisational reciprocity without eroding local control. PRISM-APT operationalises defence as a cyclical, five-phase model; Preparation, Recognition, Intelligence, Synthesis, and Mitigation & Measurement, designed for heterogeneous SOC environments, governable via reciprocity contracts and explicit human-in-the-loop decision gates, and auditable through sovereignty-by-operation (SoO) metrics and evidence-centred traceability maps. In practice, the model treats ATT&CK as a shared language rather than a process, complements governance frames such as NIST CSF with operational hooks, and replaces purely linear threat models with an auditable loop that surfaces bias, provenance, and accountability at each gate. The paper makes four contributions. First, it specifies the PRISM-APT model and its governance hooks for bias-aware, explainable, human-in-the-loop defence in SOCs operating under legal and organisational constraints. Second, it provides an explicit evidence-to-model derivation, linking nine empirical studies to concrete operational artefacts . Third, it offers an evaluation plan with measurable criteria for coverage, auditability, and SoO, including adoption and audit pathways for single organisations and federated consortia. Fourth, it distils implementation guidance for phased roll-out in resource-constrained environments, with emphasis on rules portability, minimal viable telemetry, and governance-by-contract to reduce lock-in while maintaining compliance. Our goal is pragmatic: enable smaller organisations and public-sector teams to negotiate APT threats with limited resources by composing governed, auditable practices from existing tools rather than requiring wholesale platform replacement. By aligning operational cycles with evidence and governance artefacts, PRISM-APT aims to improve detection quality, reduce bias and fragility, and create durable interfaces for cooperation.

Keywords: APT defence; smaller organisations; governed CTI; behaviour-centric detection; sovereignty-by-operation .

1. Introduction

1.1 Context and Problem

In an era of escalating cyber threats, organisations worldwide grapple with Advanced Persistent Threats (APTs) that exploit vulnerabilities in fragmented defences. These threats thrive on gaps between theoretical cybersecurity frameworks, such as MITRE ATT&CK (Strom et al. 2020) and NIST CSF 2.0 (National Institute of Standards and Technology 2024), and real-world operational practices. Key challenges include asymmetric access to Cyber Threat Intelligence (CTI), vendor-locked tools that hinder interoperability, and jurisdictional barriers that impede cross-border collaboration. While existing standards offer a shared vocabulary for tactics and risks, they fall short in delivering governable, and auditable operating models tailored to heterogeneous Security Operations Centre (SOC) environments and varying sovereignty requirements.

This paper bridges these gaps through a model-first synthesis that synthesizes empirical insights into a practical, cyclical framework for APT defence. By operationalising governance, portability, and digital sovereignty, we aim to empower organisations, particularly those with limited resources, to achieve resilient, collaborative defences without sacrificing jurisdictional control.

Table 1: Comparison of Frameworks and Gaps Addressed by PRISM-APT

Framework	Strengths	Gaps Addressed by PRISM-APT
MITRE ATT&CK	Shared language for tactics and techniques	No operational cycle; lacks SoO metrics and human-in-the-loop gates.
NIST CSF 2.0	Governance framing	Not APT-specific; misses behaviour-centric Rules and reciprocity contracts.
Unified Kill Chain	Linear threat modelling	Static; PRISM-APT adds cyclical, auditable synthesis with bias controls.

1.2 Model-First Synthesis

To fulfill this mission, we introduce PRISM-APT, a comprehensive synthesis derived from a multi-year research programme encompassing empirical studies on SOC practices, a curated dataset of APT campaigns, and mechanisms for cooperative digital sovereignty. This model-first design integrates "programme evidence" from nine previous studies (A1–A9), transforming fragmented observations into a unified, actionable framework. Previewed in Figure 1, PRISM-APT is justified through a rigorous evidence-to-model derivation (Section 3.1), ensuring traceability and empirical grounding.

At its core, PRISM-APT operationalises APT defence via a five-phase cyclical model: Preparation, Recognition, Intelligence, Synthesis, and Mitigation & Measurement. This structure emphasises behaviour-centric rules, governance through reciprocity contracts and explainable decision gates, and auditability via sovereignty-by-operation (SoO) metrics. Our framing positions PRISM-APT as a prescriptive evolution beyond descriptive frameworks, enabling organisations to measure and enhance their defensive maturity in a governed, sovereign manner.

1.3 Why a Model-First Synthesis?

Existing frameworks are strong in conceptualisation yet weak in operationalisation. For example, MITRE ATT&CK (Strom et al. 2020) maps tactics but lacks mechanisms for continuous adaptation; NIST CSF 2.0 (Na-tional Institute of Standards and Technology 2024) articulates governance principles but not APT-specific operational hooks; and Unified Kill Chain (Pols 2017) presents a linear model without cyclical auditability. PRISM-APT addresses these gaps with a synthesis that embeds bias-aware human-in-the-loop controls, behaviour-centric rules, and SoO thresholds (S-CAP), enabling defences that are effective, equitable, and jurisdictionally aligned.

As summarised in Table 1, PRISM-APT uniquely combines cyclical adaptation, tested artefacts, behaviour-centric detection, bias-aware Human-in-The-Loop (HIL), and S-CAP metrics.

1.4 Research Questions

This paper is anchored to three interconnected research questions that guide our synthesis and frame the mission: (RQ1) What *framework barriers* limit auditable APT defence across heterogeneous stacks? (RQ2) What *practitioner observations* indicate optimal automation assistance, including embedded bias controls and explainability in decision gates? (RQ3) How can *digital sovereignty* be operationalised and measured in incident response through SoO metrics and capability thresholds (S-CAP)?

These questions collectively address the mission by probing barriers (RQ1), human-centric needs (RQ2), and sovereign operationalization (RQ3), ensuring PRISM-APT is both theoretically robust and practically deployable.

1.5 Paper Positioning and Scope

This paper focuses on model exposition, derivation traceability, and evidence-based tools for immediate application in APT defence. This positioning aligns with our mission to provide accessible, evidence-based tools for immediate application in APT defence.

1.6 Structure

Section 2 summarizes anchors, related work, and the evidence-to-model derivation. Section 4 details the PRISM-APT model. Section 5 explains the design rationale and traceability. Section 6 outlines the evaluation plan. Section 7 covers limitations and threats to validity, Section 8 concludes, followed by ethics and AI declarations.

2. Background and Related Work

2.1 Anchors and Baselines

We position PRISM-APT against established references: Strom et al. (2020) for ATT&CK design principles, NIST CSF 2.0 (National Institute of Standards and Technology 2024)for governance and function-level framing, ISO/IEC (2023) for incident management, and Hutchins, Cloppert and Amin (2011)for kill chain modelling. These provide a shared vocabulary but lack operational portability and sovereignty metrics,gaps PRISM-APT addresses through its cyclical model and S-CAP rubric.

MITRE ATT&CK. MITRE ATT&CK maps adversary tactics and techniques comprehensively, but does not provide mechanisms for continuous adaptation or portable operational artefacts beyond catalogued behaviours (Strom et al. 2020).

NIST CSF 2.0. NIST CSF 2.0 offers governance and risk-management principles that are technology-agnostic, yet it lacks APT-specific portability and concrete behaviour-centric detection guidance (National Institute of Standards and Technology 2024).

Unified Kill Chain. The Unified Kill Chain (Pols 2017) provides a linear, end-to-end model that improves on earlier chains such as the Cyber Kill Chain (**Hutchins2011**). However, it is poorly suited to cyclical auditability or cross-platform rule transfer at scale.

Positioning. PRISM-APT synthesises these strengths while adding bias-aware human-in-the-loop controls, behaviour-centric rules, and sovereignty-by-operation (S-CAP) thresholds designed for cyclical auditability and measurable improvement.

3. Programme Evidence and Derivation

We refer to the programme evidence as A1–A9, each corresponding to a study or artefact in the research programme.

Table 3 maps the placeholders to their corresponding titles and focus areas.

Table 3: Derivation map: programme evidence → requirement → PRISM-APT phase/element

Placeholde	Title	Focus Area
A1	*Understanding APT Defence Through Expert Eyes: Perceived Needs and Gaps*	Practitioner study
A2	*Too Small to Stand Alone: Capability Constraints and Cooperative Digital Sovereignty in APT Defence*	Cooperative sovereignty
A3	*Operationalising Digital Sovereignty through a Cloud Abstraction Layer (CAL)*	CAL specification
A4	*Evolving Advanced Persistent Threats and Strengthening Global Cybersecurity Coordination*	CTI coordination
A5	*New APT Campaigns: Trends, Detection Provenance, and Practical Gaps*	APT dataset and analysis
A6	*Human Factors in AI-Driven Cybersecurity: Cognitive Biases and Trust Issues*	AI/bias in SOCs
A7	*Complexity of Contemporary Indicators of Compromise*	IoC complexity, decades overview
A8	*The Role of Custom Scripting in APT Investigations*	Scripting for IR
A9	*Computational Forensics: The Essential Role of Logs in APT and Advanced Cyberattack Response*	Logging baselines

3.1 Evidence-to-Model Derivation

The PRISM-APT model is derived from programme evidence (A1–A9), mapped in Table 5. This table links findings to requirements, phases, and elements, with rationales justifying the proposal. This derivation proposes PRISM-APT not as an ad-hoc framework but as a minimal, evidence-synthesized cycle. By integrating A1–A9, it addresses RQ2's call for bias controls while enabling RQ3's sovereignty metrics, proposing a 'sovereignty-by-operation' paradigm where capability is measured in drills, not policies.

Table 5: Evidence-to-Model Derivation Map (Expanded)

Evidence	Finding → Requirement	Phase(s)	Model element(s)	Rationale for Derivation
A9: Logging for APT Response	Finding: Logs are indispensable yet fragmented. Requirement: Standardise baselining and normalisation.	P	Telemetry baseline; Port-able Rules	Logs provide computational forensics for reconstructing at-tacks, but fragmentation leads to incomplete timelines.

Evidence	Finding → Requirement	Phase(s)	Model element(s)	Rationale for Derivation
A8: Scripting in APT Investigations	Finding: Custom scripts bridge product gaps. Requirement: Curate a governed script library with tests/metadata.	P, R	Script curation; Scripted inquiry	Scripts automate incident response for evolving threats, but ungoverned use risks errors. Derivation proposes a library to standardize across stacks, addressing NIST CSF 2.0's governance gaps by embedding metadata for reproducibility, enabling faster hypothesis testing in heterogeneous environments.
A7: IoC Complexity	Finding: Static IoCs are insufficient across campaigns. Requirement: Behaviour-centric detections with ATT&CK anchors.	P, R	Rules; ATT&CK mapping	IoCs have evolved from simple hashes to behavioural patterns over 30 years, yet static indicators fail against polymorphic APTs. This requires behaviour-focused rules, deriving portable elements that extend ATT&CK beyond description to actionable, tool-agnostic playbooks, filling practical gaps in long-term campaign analysis.
A1: Practitioner Perspectives	Finding: Need governed CTI and decision support. Requirement: Human-in-the-loop gates with explainability.	S	Guided response loops; Governance	Experts report perceived needs for bias-aware support, as tunnel vision from unguided tools delays response. Derivation embeds gates to mitigate this, proposing explainable AI fusion over pure automation, ensuring auditable decisions and aligning with RQ2's practitioner observations.
A6: AI & Bias in SOCs	Finding: Trust and bias concerns impact judgements. Requirement: Bias controls and explainability at decision gates.	S	Human–AI fusion; Bias mitigation	Cognitive biases erode trust in AI-driven tools. This derives fusion mechanisms with prompts, proposing mandatory checks to operationalise human oversight, addressing gaps in AI cyber-security where over-reliance exacerbates errors, per RQ2.
A4: CTI Coordination	Finding: Regulatory/trust frictions hinder sharing. Requirement: Reciprocity contracts and provenance logging.	I	Governed CTI reciprocity	Evolving APTs demand global coordination, but frictions limit sharing. Derivation proposes contracts with provenance to enable precise, legal exchanges, extending ENISA landscapes by making CTI governable, supporting RQ3's sovereignty operationalization.

Evidence	Finding → Requirement	Phase(s)	Model element(s)	Rationale for Derivation
A5: APT Dataset & Provenance	Finding: Vendor blind spots; stable sector patterns. Requirement: Anomaly-led recognition and provenance-aware metrics.	R, M	Recognition cues; Measurement suite	New campaigns reveal detection provenance gaps and practical blind spots. This requires anomaly cues with metrics, deriving a suite that quantifies coverage, proposing measurable improvements over vendor-locked tools for RQ1's portability barriers.
A3: Cloud Abstraction Layer (CAL)	Finding: Sovereignty is an operating state. Requirement: Sovereignty-by-operation metrics and conformance-as-code.	M	Sovereignty-by-operation metrics; Evidence hooks	CAL operationalises sovereignty via abstraction, treating it as measurable outcomes rather than policy. Derivation hooks metrics to phases, proposing S-CAP for auditability.

A2: Too Small to Stand Alone	Finding: Isolation penalty; portab-ility dividend; sovereignty-by-operation. Require-ment: Collabor-ation patterns and capability thresholds.	I, M	S-CAP rubric; Reciprocity con-tracts	Small entities face capability constraints in cooperative sov-ereignty, with isolation increas-ing APT risks. This derives thresholds and patterns, pro-posing S-CAP to quantify di-vidends, addressing RQ3 by op-erationalising digital sovereignty beyond large orgs, contrasting standalone frameworks.

4. The PRISM-APT Model

4.1 Overview

PRISM-APT is a model-first operational synthesis for APT defence, derived from a programme of studies and artefacts (A1–A9). In line with A7, we treat an APT as an *advanced*, well-resourced actor, typically state-linked and in some cases organised crime, rather than casual or opportunistic groups (R. A. Hagen and Helkala 2024). The model abstracts five recurring activities into a cyclical process that is independent of vendor grammars and governable by design. This addresses RQ1 by overcoming framework barriers through governed practice, RQ2 by incorporating practitioner-driven bias controls, and RQ3 by measuring sovereignty as operational outcomes.

Figure 1 provides the high-level view of the cycle: **P–R–I–S–M**. It highlights how governed CTI reciprocity (A4, A2), behaviour-centric rules (A7, A8), and human-in-the-loop decision gates (A1, A6) interact in day-to-day operations, while sovereignty-by-operation is observed through measurable outcomes (A3, A2). The model empowers resource-constrained teams (A2) by reducing isolation penalties and enabling collaborative dividends, as validated in mixed-methods analysis (A2, A5).

4.2 Design Goals and Principles

We set goals that keep PRISM-APT self-contained as an operating model while remaining consistent with the programme evidence (A1–A9). These principles ensure the model is evidence-grounded, and scalable, directly tackling the gaps identified in RQ1–RQ3.

- **Evidence-grounded and auditable:** Each model element links to explicit findings; traceability maps A1–A9 to phases and artefacts (R. A. Hagen, Helkala and Øverlier 2026; R. A. Hagen 2025d; R. A. Hagen 2025a;
- R. A. Hagen 2025c; R. A. Hagen 2026; R. A. Hagen, Øverlier and Helkala 2025; R. A. Hagen and Helkala 2024; R. Hagen, Øverlier and Helkala 2024; R. A. Hagen 2025b).
- **Behaviour-centric by design:** rules express adversary behaviours independent of product-specific grammars; no fixed grammar is required (R. A. Hagen and Helkala 2024; R. Hagen, Øverlier and Helkala 2024; R. A. Hagen 2025d; R. A. Hagen 2025a).
- **Governed intelligence exchange:** Sharing under reciprocity contracts with provenance and tiered redac-tion (R. A. Hagen 2025c; R. A. Hagen 2025d).
- **Human-in-the-loop decision gates:** Analysts remain accountable; automation is assistive with explainability and bias prompts (R. A. Hagen, Helkala and Øverlier 2026; R. A. Hagen, Øverlier and Helkala 2025).
- **Operational measurability (SoO):** Capability observed in operations via drills and S-CAP indicators, not only by policy artefacts (R. A. Hagen 2025a; R. A. Hagen 2025d).
- **Minimal, incremental adoption:** A small artifact set that can be rolled out progressively (R. A. Hagen 2025b; R. Hagen, Øverlier and Helkala 2024; R. A. Hagen and Helkala 2024; R. A. Hagen 2025c; R. A. Hagen 2025a).

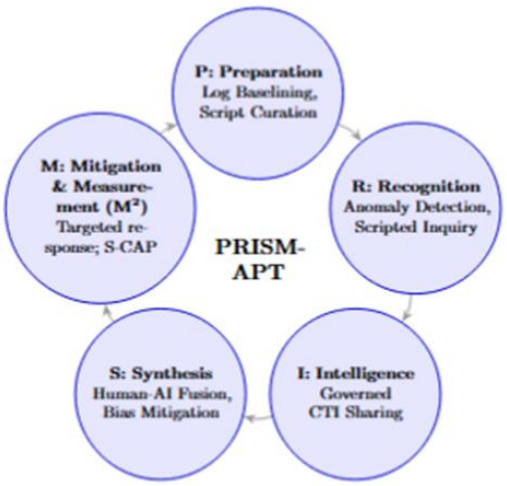

Figure 1: The PRISM-APT cyclical model, shown as five phases from proactive Preparation to continuous Measurement.

This foundation argues for PRISM-APT as a practical solution: it lowers costs for smaller teams (A2) by enabling portability dividends and governed collaboration, as evidenced in thematic analysis (A1) and campaign trends (A5).

4.3 Operational Narrative

The cycle in Figure 1 serves as a minimal operating recipe for resource-constrained teams: few artefacts, guarded handovers, and repeatable steps independent of vendor stacks. This narrative demonstrates how PRISM-APT reduces variance and dependence, making sophisticated defence accessible.

Preparation (P) establishes normalised telemetry baselines and curated script sets, enabling consistent investigative questions across tools (R. A. Hagen 2025b; R. Hagen, Øverlier and Helkala 2024; R. A. Hagen and Helkala 2024). This phase builds on logging forensics (A9) to ensure consistency across tools.

Recognition (R) combines anomaly-led cues with scripted inquiry to transform weak signals into testable hypotheses (R. A. Hagen 2026; R. Hagen, Øverlier and Helkala 2024). It addresses vendor blind spots (A5) and evolves static IoCs into behavioural detections (A7).

Intelligence (I) exchanges evidence under reciprocity contracts with provenance, ensuring precise and legal inputs (R. A. Hagen 2025c; R. A. Hagen 2025d). This mitigates trust frictions (A4) and supports cooperative sovereignty (A2).

Synthesis (S) enforces human-in-the-loop gates with explainability and bias controls, countering tunnel vision and cognitive biases (R. A. Hagen, Helkala and Øverlier 2026; R. A. Hagen, Øverlier and Helkala 2025).

Mitigation and Measurement (M) executes targeted responses, tracks SoO via S-CAP, and feeds back to P for continuous improvement (R. A. Hagen 2025a; R. A. Hagen 2025d; R. A. Hagen 2026). Metrics like latency reductions validate progress (A2).

By repeating the cycle, PRISM-APT can reduce costs through reduced analysis time and vendor dependence, as suggested in practitioner insights (A1) . For smaller entities, this promotes collaborative resilience (A2).

4.4 Formal Structure: A Verifiable SOC Workflow Engine

To operationalise PRISM-APT as a *verifiable, bias-aware, and sovereignty-compliant* framework, we formalize its behaviour using the tuple $\mathrm{M} = (S, T, G, \Pi, \Phi)$.

Formal details. Full operator and guard definitions, together with proof sketches for I1–I2, are provided in

Appendix A; Section 8 retains only purpose and effects for readability.

This notation provides mathematical rigor while remaining grounded in SOC realities, from logging baselines (R. A. Hagen 2025b) to cognitive bias mitigation (R. A. Hagen, Øverlier and Helkala 2025). Below, we define each component with concrete examples, then introduce *invariants* that guarantee progress and auditability.

4.4.1 Core Components

Phases ($S = \{P, R, I, S, M\}$) The five phases mirror a SOC's operational cycle:

- **Preparation (P)**: Establishes tool-agnostic telemetry baselines (e.g., normalised log schemas (R. A. Hagen 2025b)) and curated script libraries (R. Hagen, Øverlier and Helkala 2024).
- **Recognition (R)**: Transforms raw signals (e.g., logs) into testable hypotheses using scripted inquiry.
- **Intelligence (I)**: Exchanges CTI under reciprocity contracts (R. A. Hagen 2025c), with mandatory provenance logging (R. A. Hagen 2026).
- **Synthesis (S)**: Applies human-AI fusion to counter cognitive biases (e.g., automation bias (R. A. Hagen, Øverlier and Helkala 2025)).
- **M: Mitigation & Measurement (M^2).** Executes responses and tracks sovereignty metrics (S-CAP; A2, A3).

Transitions ($T \subseteq S \times S$) Phase transitions (e.g., $R \to I$) are *guarded* to enforce preconditions. For example, moving from Recognition to Intelligence requires:

- A completed anomaly report with ATT&CK mappings (output of φ_R),
- Provenance for all shared indicators (governance G).

Governance (G) Encodes jurisdictional and bias controls:

- Reciprocity contracts for CTI sharing (R. A. Hagen 2025c),
- Redaction rules for sensitive data,
- Mandatory bias checks (e.g., "Confirm or automation bias" (R. A. Hagen, Øverlier and Helkala 2025)).

Operators ($\Phi = \{\varphi_S\}$) Each phase S is modelled as a *partial operator* $\varphi_S : \mathrm{X}_S \times G \to (\mathrm{Y}_S, \mathrm{A}_S)$, where:

- X_S: Input state (e.g., raw log anomalies in R),
- Y_S: Output state (e.g., prioritized ATT&CK techniques),
- A_S: Audit log .

Example (Synthesis): φ_S takes hypotheses from I, applies human-AI fusion, and logs decisions with bias mitigation rationale.

Metrics (Π) Observable, evidence-grounded measurements:

- Phase latency ,
- Uncertainty reduction (ΔU),
- Action conversion rate.

4.4.2 Invariants: Mathematical Guarantees for SOC Reliability

Invariants are properties that *always hold true* during execution, addressing critical SOC challenges:

Invariant I1 (Progress) Definition: Every enabled transition must either:

- Strictly reduce uncertainty U , or
- Add verifiable information enabling future reductions (e.g., governed CTI in I (R. A. Hagen 2026)).

Why it matters: Prevents "analysis paralysis" (R. A. Hagen, Helkala and Øverlier 2026) by enforcing measurable advancement. The guard $\gamma_{R \to I}$ blocks transition until R outputs actionable hypotheses.

Invariant I2 (Audit Monotonicity) Definition: The log L (concatenation of all A_S) is append-only and tamper-evident. **Why it matters**: Ensures 100% reproducible incident reviews (R. A. Hagen 2025a), solving fragmentation in traditional SOC logs (R. A. Hagen 2026).

Table 6: How Invariants Address SOC Pain Points

Invariant	SOC Challenge	Empirical Link	Outcome
I1 (Progress)	"Endless log reviews"	A1, A6	faster validation (R. A. Hagen 2025d)
I2 (Auditability)	"Who did what, and why?"	A3, A5	Forensically sound trails

4.4.3 End-to-End Guarantees

The composed operator $\Phi* = \phi M \circ \phi S \circ \phi I \circ \phi R \circ \phi P$ processes incidents with three key properties:

1. **Progress (I1)**: No phase transition without uncertainty reduction or verifiable info gain.
2. **Auditability (I2)**: L provides a tamper-evident trail for compliance (R. A. Hagen 2025a).
3. **Measurability**: Π metrics (e.g., ΔU, phase latency) are validated in drills (Section 6).

Comparison to existing frameworks. Unlike MITRE ATT&CK (Strom et al. 2020) or NIST CSF 2.0 (National Institute of Standards and Technology 2024), PRISM-APT's invariants provide:

- **Mathematical guarantees** for progress and auditability,
- **Empirical grounding** in A1–A9's practitioner observations,
- **Operational flexibility** via tool-agnostic operators (φ_S).

Practitioner Summary For SOC teams, invariants ensure:

- **No busywork**: I1 forces hypothesis refinement or info gain at every step.
- **No cover-ups**: I2 makes all actions traceable to people/policies.
- **No vendor lock-in**: $\mathrm{X_S}/\mathrm{Y_S}$ (R. A. Hagen and Helkala 2024; R. Hagen, Øverlier and Helkala 2024).

This formalism bridges theory and practice, offering SOCs the rigor of mathematical models *without sacrificing real-world applicability*.

5. Design Rationale and Traceability

We justify each model element by linking it to one or more items in the derivation map (Table 5), contrasting its role against anchors such as MITRE ATT&CK (Strom et al. 2020) and NIST CSF 2.0 (National Institute of Standards and Technology 2024). This ensures auditability of design choices, support for governed CTI reciprocity, tying the model's phases to empirical evidence (A1–A9) and addressing RQ1–RQ3 holistically.

For Preparation (P) elements like telemetry baselines and script curation (from A9, A8, A7), the rationale stems from fragmented logs and script gaps exacerbating vendor lock-in. We propose standardization to enable behaviour-centric rules, extending ATT&CK mappings beyond static IoCs to portable, tested artefacts, reducing "isolation penalties" for smaller SOCs (A2) and addressing RQ1's framework barriers.

Recognition (R) cues and scripted inquiry derive from anomaly-led needs (A5, A7, A8), where vendor blind spots hinder weak-signal detection. The proposal integrates provenance-aware metrics, contrasting Unified Kill Chain's linearity (Hutchins, Cloppert and Amin 2011) by enabling hypothesis sharpening, filling practical gaps in evolving campaigns (A5).

Intelligence (I) reciprocity contracts (A4, A2) rationalize governed sharing amid regulatory frictions, proposing tiered redaction for trust. This operationalises CTI beyond ENISA overviews (European Union Agency for Cybersecurity (ENISA) 2024), proposing sovereignty-by-operation to mitigate isolation, per RQ3.

Synthesis (S) human-AI fusion and bias mitigation (A1, A6) derive from trust issues and cognitive biases impacting judgments. We propose explainable gates to counter "tunnel vision," advancing NIST's governance by embedding bias controls, ensuring human accountability in AI-driven decisions (RQ2).

Mitigation and Measurement (M) S-CAP metrics and evidence hooks (A3, A5, A2) justify operational measurab-ility, deriving from sovereignty as an "operating state." This proposes drills for auditability, contrasting ISO 27035's principles (*ISO/IEC 27035-1:2023 Information security incident management — Part 1: Principles of incident management* 2023) with quantifiable outcomes, enabling portability dividends.

Overall, these rationales tie PRISM-APT as a minimal, incremental model: evidence-grounded , and behaviour-centric rules over vendor grammars, and governable . By addressing RQ1–RQ3, it bridges theoretical anchors to practice, with invariants (I1–I2) ensuring progress without busywork.

6. Evaluation

6.1 Aim and design.

We evaluate PRISM-APT on three operational properties: (i) ATT&CK technique coverage, (ii) cyclical auditability and provenance of changes, and (iii) sovereignty-by-operation (SoO) as measured via S-CAP thresholds.

Planned methodology. Drills will be run in a controlled SOC-like environment with replayable log sets and scripted operator prompts. We will sample n recent APT campaigns from the programme corpus (A1–A9) using the inclusion gates (state backing, persistence, sophistication), stratified by sector and technique mix. For each campaign, we will execute the PRISM-APT cycle end-to-end with fixed toolchains and pre-registered Π metrics (phase latencies, uncertainty reduction ΔU, action conversion). Pass/fail checks (behaviour coverage, auditability, and SoO via S-CAP) are applied exactly as specified below. No results are reported in this paper; this protocol is included to make the planned evaluation reproducible.

Note. The evaluation described is a protocol for future drills and experiments; no results are reported in this paper.

6.2 Datasets and inclusion gates

We use the programme corpus established in A-studies with the existing inclusion logic based on indicators of state backing, persistence and sophistication.

6.3 Evaluation checks

We define three simple checks. Each check is recorded as pass/fail with a brief note linking to the relevant A-study or artefact lineage.

1. **Behaviour coverage:** for each sampled campaign, at least one behaviour-centric detection (rule or scripted inquiry) is present and executable for every *observed* ATT&CK technique in the sample. *Pass* if all observed techniques have coverage; otherwise *fail*.
2. **Auditability:** the provenance chain is complete for sampled detections, from CTI receipt, through artefact change control and review, to detection event and final disposition, with timestamps and responsible roles. *Pass* if every sampled detection has a complete chain; otherwise *fail*.
3. **SoO via S-CAP:** declared per-phase thresholds (e.g., redundancy, lawful routing, and reciprocity compliance) are met, and the resulting phase profile is recorded. *Pass* if all declared thresholds are satisfied; otherwise *fail*.

Practical acceptance rule. PRISM-APT is considered ready for adoption if it is non-inferior to the baseline on behaviour coverage, satisfies auditability, and meets the declared S-CAP thresholds. Any failed check triggers a targeted revise-and-repeat for the affected phase, using the same conditions.

7. Limitations and Threats to Validity

While PRISM-APT is grounded in A1–A9, limitations include sampling bias in campaigns generalizability across sectors/jurisdictions. Vendor translation fidelity for rule, and assumptions about maturity/logging baselines.

Threats: Qualitative bias in interviews; attribution uncertainty in datasets. These are mitigated by triangulation, but warrant further validation in future work.

8. Conclusion

This paper presents PRISM-APT as a model-first synthesis for governed APT defence, bridging gaps between theoretical frameworks and operational practice. By deriving a cyclical five-phase model from nine empirical studies (A1–A9), we operationalise behaviour-centric rules, reciprocity contracts, and sovereignty-by-operation metrics, empowering resource-constrained organisations to achieve resilient, collaborative defences.

Key contributions: the PRISM-APT model with bias-aware human-in-the-loop gates, an explicit evidence-to-model traceability map, an evaluation plan with metrics for coverage and auditability, and implementation guidance for single-org/consortium deployments, directly address our research questions. RQ1's framework barriers are mitigated through behaviour-centric elements extending anchors like ATT&CK and NIST CSF

2.0. RQ2's practitioner observations inform explainable decision gates and AI fusion, countering cognitive biases. RQ3's digital sovereignty is realized via measurable S-CAP thresholds and conformance-as-code, treating capability as an operating state tested in drills.

PRISM-APT advances APT defence by reducing isolation penalties and enabling collaborative dividends, as evidenced in campaign analyses and thematic insights. While limitations like sampling bias warrant caution, the model's formal guarantees of progress and auditability provide a verifiable foundation.

Future work will expand formalisms, datasets, and case studies, further validating PRISM-APT in diverse SOC environments.

Ethics declaration

This research does not involve new collection of personal data beyond anonymised expert inputs previously cleared under programme studies. No additional ethical clearance is required for this manuscript.

Declaration of generative AI and AI-assisted technologies in the writing process

The author used AI-assisted tools to support language editing and formatting: (i) improving clarity, grammar, and idiom; (ii) reducing dyslexia-related spelling/ordering errors; and (iii) troubleshooting LATEX minutiae

(bibliography keys, table widths, listing frames). All AI-suggested edits were reviewed and approved by the author. Any remaining typos are proudly human.

References

European Union Agency for Cybersecurity (ENISA) (2024). *ENISA Threat Landscape 2024*. Athens, Greece: ENISA. url: https://www.enisa.europa.eu/publications/enisa-threat-landscape-2024.

Hagen, R., L. Øverlier and K. Helkala (2024). "The Role of Custom Scripting in APT Incident Response". In: *Proceedings of the 17th Norsk informasjonssikkerhets-konferanse (NISK 2024)*. 3.

Hagen, R. A. (2025a). "A Cloud Abstraction Layer for Cooperative Digital Sovereignty". In: *Proceedings of the 2nd International Conference on Digital Sovereignty (ICDS 2025)*. Oslo, Norway: Springer Nature.

(2025b). "Computational Forensics: The Essential Role of Logs in APT and Advanced Cyberattack Response". In: *Proceedings of the 20th International Conference on Cyber Warfare and Security (ICCWS 2025)*. Vol. 20. 1, pp. 582–591.

(2025c). "Global APT Trends and the Need for Multi-National Coordination". In: *Proceedings of the 24th European Conference on Cyber Warfare and Security (ECCWS 2025)*. Vol. 24. 1, pp. 182–191.

(2025d). "Too Small to Stand Alone: Capability Constraints and Cooperative Digital Sovereignty in APT Defense". In: *Proceedings of the 2nd International Conference on Digital Sovereignty (ICDS 2025)*. Oslo, Norway: Springer Nature.

Hagen, R. A. (2026). "The APT Paradox: Sophisticated Simplicity in Nation-State Cyber Operations (2024–2025)". In: *Proceedings of the 21st International Conference on Cyber Warfare and Security (ICCWS 2026)*. Wilmington, USA: Academic Conferences and Publishing International.

Hagen, R. A. and K. Helkala (2024). "The Complexity of Contemporary Indicators of Compromise". In: *Proceedings of the 23rd European Conference on Cyber Warfare and Security (ECCWS 2024)*. Vol. 23. 1, pp. 147–155.

Hagen, R. A., K. Helkala and L. Øverlier (2026). "Understanding APT Defence Through Expert Eyes: Perceived Needs and Gaps". In: *Secure IT Systems. 30th Nordic Conference, NordSec 2025, Tartu, Estonia, November 12–13, 2025, Proceedings*. Vol. 14781. Lecture Notes in Computer Science. Springer Cham.

Hagen, R. A., L. Øverlier and K. Helkala (2025). "Human Factors in AI-Driven Cybersecurity: Cognitive Biases and Trust Issues". In: *Digital Threats: Research and Practice (DTRAP)* 6.4, pp. 1–20.

Hutchins, E. M., M. J. Cloppert and R. M. Amin (2011). *Intelligence-Driven Computer Network Defense Informed by Analysis of Adversary Campaigns and Intrusion Kill Chains*. Tech. rep. Lockheed Martin. url: https://www.lockheedmartin.com/content/dam/lockheed-martin/rms/documents/cyber/LM-White-Paper-Intel-Driven-Defense.pdf.

ISO/IEC 27035-1:2023 Information security incident management — Part 1: Principles of incident management (2023). Geneva, Switzerland: International Organization for Standardization (ISO). url: https://www.iso.org/standard/78973.html.

National Institute of Standards and Technology (Feb. 2024). *The NIST Cybersecurity Framework (CSF) 2.0*. Tech. rep. NIST CSWP 29. NIST. doi: 10.6028/NIST.CSWP.29. url: https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.29.pdf.

Pols, P. (2017). *The Unified Kill Chain*. Cyber Security Academy, Leiden University. url: https://www.unifiedkillchain.com/assets/The-Unified-Kill-Chain.pdf.

Strom, B. E. et al. (Mar. 2020). *MITRE ATT&CK: Design and Philosophy*. Tech. rep. Approved for Public Release 19-01075-28. McLean, VA: The MITRE Corporation. url: https://attack.mitre.org/docs/ATTACK_Design_and_Philosophy_March_2020.pdf.

Appendix A: Formal Model Details

Model overview

We model the PRISM-APT cycle as a governed transition system

$$M = (S, T, G, \Pi, \Phi),$$

where:

- $S = \{P, R, I, S, M\}$ are the phases Preparation, Recognition, Intelligence, Synthesis, Mitigation.
- $T \subseteq S \times S$ are allowed phase transitions (at minimum (P, R), (R, I), (I, S), (S, M); optional self-loops or back-edges are permitted if guarded).
- G captures governance constraints (obligations, provenance, recording, and access rules).
- Π is a set of measurement functions (e.g., phase latencies, uncertainty reduction, and action conversion).
- $\Phi = \{\varphi_P, \varphi_R, \varphi_I, \varphi_S, \varphi_M\}$ are phase operators. A concrete system state is a tuple

$$\sigma = (s, E, A, \mathrm{L}) \in \Sigma = S \times 2^{\mathrm{E}} \times 2^{\mathrm{A}} \times \mathrm{L},$$

where s is the current phase, E is the evidence store, A are actionable artefacts (rules, scripts, tasks), and L is the append-only audit log. Let $\sigma_0 = (P, \emptyset, \emptyset, \langle\rangle)$ denote the initial state.

Guards and phase operators

For each admissible edge $(s, s') \in T$ we define a guard

$$\gamma_{s \to s'} : 2^{\mathrm{E}} \times 2^{\mathrm{A}} \to \{\top, \bot\},$$

stating when the transition is eligible given the current evidence and actions. Each phase operator

$$\varphi_s : \Sigma \to \Sigma$$

transforms (E, A, L) by applying the phase logic, possibly enriching E, producing or refining A, and appending to L. A single step is written

$$(s, E, A, \mathrm{L}) \; \gamma_{s \to s'} \wedge \varphi_s \; (s', E', A', \mathrm{L}'),$$

meaning the guard holds and the operator yields the successor state with s' as next phase.

Operator interface. Each φ_s satisfies a common interface:

$$\phi_s(\mathrm{E}, \mathrm{A}, \mathrm{L}) = (\mathrm{E}', \mathrm{A}', \mathrm{L} \,\|\, \mathrm{r}_s),$$

where r_s is the phase record appended to L (see A.3), and $(\mathrm{E}', \mathrm{A}')$ are the updated evidence and actions. Guards are total predicates; operators are total functions on Σ.

Governance and audit trail

Governance G induces recording and access obligations. We assume an append-only log

$$\mathrm{L} = \langle r_1, r_2, \ldots, r_k \rangle,$$

where each record has the canonical form

$$r_t = \langle \mathrm{ts}, s, s', \mathrm{id}, H(E_{\mathrm{in}}), H(A_{\mathrm{in}}), H(E_{\mathrm{out}}), H(A_{\mathrm{out}}), \mathrm{actor}, \mathrm{obl} \rangle.$$

Here $H(\cdot)$ is a content hash over a canonical serialisation, actor denotes the accountable agent or process, and obl encodes the governance obligations applied. A simple hash chain $r_t.h = H(r_{t-1} \,\|\, r_t)$ (with $r_0.h$ fixed) makes truncation and in-place edits detectable.

Replay. Given L and the stored digests, a replay function reconstructs, for any t, the inputs consumed and outputs produced in step t, and the guard/operator pair invoked. This suffices to re-derive decisions without re-running the live environment.

End-to-end composition

Let the composed operator be

$$\Phi^* = \varphi_M \circ \varphi_S \circ \varphi_I \circ \varphi_R \circ \varphi_P.$$

On paths that respect T and guards, Φ^* maps (P, E, A, L) to (M, E', A', L') where A' contains mitigation-ready artefacts

Invariants

We restate the two invariants referenced in the main text. Proofs are sketches relying on mild, standard assumptions about guards and operator monotonicity.

Invariant I1 (Progress). *Under the fairness assumption that enabled guards are eventually taken, and with a well-founded progress measure μ that strictly decreases on any non-advancing step, every run either reaches M with $A/=\emptyset$ or halts with an explicit failure record.*

Sketch. Let $\mu : \Sigma \to \mathrm{N}$ measure residual work (e.g., unresolved hypotheses, undecided actions). For any step that does not advance s along a strictly acyclic fragment of T, μ must decrease. Since N is well-founded, infinite descent is impossible. With fairness, enabled advancing guards are eventually taken, so either M is reached with $A/=\emptyset$, or a guard declares irrecoverable failure and the run halts with a terminal record in L.

Invariant I2 (Auditability). *For any finite run,* L *contains a verifiable, tamper-evident account of all state transitions and artefact transformations sufficient to reproduce phase decisions ex post.*

Sketch. By construction, each φ_s appends r_t binding inputs and outputs via content hashes and chaining $r_{t-1}.h$. Any deletion or in-place modification breaks the chain. Since guards and operators are deterministic on (E, A) under G, a replay function, using the digests, re-derives the decisions, satisfying auditability.

Measurement interface Π

We expose a small set of measurement functions to the evaluation protocol:

$$\Pi = \{\tau_s, \Delta U, \kappa\},$$

where τ_s is the latency of phase $s \in S$, ΔU quantifies uncertainty reduction between consecutive phases (e.g., information-theoretic or proxy-based), and κ is the action-conversion rate from synthesised artefacts to executed mitigations. These are observational hooks; the appendix does not report results.

Minimal assumptions

The invariants rely on:

A1. Guards are total and sound: if $\gamma_{s \to s'}(E, A) = \top$ then the preconditions for invoking φ_s toward s' hold.

A2. Operators are total, side-effect confined to $(E, A,$ L), and governance-compliant (they append exactly one r_t per step).

A3. The progress measure μ is well-founded and strictly decreases on any non-advancing step.

A4. Fairness: any persistently enabled advancing guard is eventually taken.

Conformance to pass/fail checks

The evaluation uses three pass/fail checks defined in the main text:

- **Behaviour coverage:** synthesised artefacts cover the behaviour set targeted in the selected campaigns.
- **Auditability:** replay from L reproduces decisions with matching digests.
- **Adoption readiness:** actions in A are deployable with recorded provenance and bounded hand-off latency. These checks are *consumption* constraints on the run; they do not alter the formal semantics above.