Comprehensive Threat Model Evaluation Report (Combined)

Executive Summary & Comparative Analysis

1. Threats & Mitigations Maturity Ranking (Across Models)

Rank	Model Name	Threats & Mitigations Score	Maturity	Reasoning
1	infosecotb-model-openai-gpt-5	95	🏆 Excellent	Broad, precise STRIDE coverage across all critical elements; strong supply-chain and logging/repudiation treatments; mitigations are actionable (e.g., SLSA, signed artifacts, mTLS, CSP, DLP) and proportionate to exposure.
2	infosecotb-model-anthropic-claude-sonnet-4-5-20250929	92	🏆 Excellent	Comprehensive and balanced threats with meaningful mitigations (prompt filtering, RASP, mTLS, strict CSP, audit logging). Good inclusion of repudiation and development-environment risks; minor gaps in consistency and prioritization.
3	infosecotb-model-anthropic-claude-opus-4-1-20250805	88	🌟 Good	Strong coverage across processes, stores, external APIs, and admin flows; detailed mitigations (SRI, cert pinning, OAuth, rate limiting). Slightly uneven category balance and fewer repudiation items.
4	infosecotb-model-novita-deepseek-deepseek-v3.1-terminus	62	✅ Adequate	Broad set of threats across flows and admin operations with workable mitigations; some duplication and overgeneralization; a few misaligned internal-vs-public network assumptions reduce contextual accuracy.
5	infosecotb-model-novita-qwen-qwen3-coder-480b-a35b-instruct	52	⚙️ Fair	Reasonable STRIDE spread, but recurring assertions of “unencrypted” on already-HTTPS flows; mitigations are often generic and miss root causes; inconsistent treatment of in-zone vs cross-zone threats.
6	infosecotb-model-xai-grok-4-latest	47	⚙️ Fair	Coverage exists but is uneven and repetitive; several mitigations are high-level (e.g., “validate,” “sign”) without placement specificity; some STRIDE categories are underrepresented.
7	infosecotb-model-gemini-gemini-2.5-pro	55	⚙️ Fair	Partial and scattered coverage; many elements have no threats; mitigations are plausible but not comprehensive; limited treatment of repudiation and development pipeline risks.

8	infosecotb-model-novita-deepseek-deepseek-r1,	40	⚠️ Poor	Sparse, generic threats; limited link to modeled flows/zones; mitigations do not target root causes; several assumptions about encryption/public networks misapplied.

2. Overall Model Maturity

2.1 Evaluation Summary

The shared DFD across all models is materially consistent and gives a workable view of trust zones (BlueHost, Hugging Face, Dev/Admin), key processes (WordPress, vMeNext App, Dev tools), and major external dependencies (OpenAI API, SMTP2GO). Visual separation of zones and principal flows (iFrame, admin and API paths) is clear enough for attack surface analysis. However, several flows are inconsistently flagged (e.g., HTTPS yet marked as unencrypted or non-public), and element naming is occasionally inconsistent (e.g., "WrodPress"). Data classifications, data types within flows, and decomposition of key services (e.g., WordPress stack) are limited, which constrains precision of risk inference and control mapping.

2.2 Scoring Table

Dimension	Weight	Score	Reasoning
Clarity and Readability	25%	70	Trust zones and principal nodes are consistently present; some naming typos and inconsistent flow labeling reduce polish; protocols and encryption states are not always coherent with labels.
Completeness and Coverage	30%	72	Core actors, processes, stores, and external APIs are modeled; missing data classifications, data types, and deeper decomposition of WordPress and CI/CD limit full coverage.
Accuracy and Logical Consistency	25%	60	Multiple flows marked as non-public or unencrypted while using HTTPS across the internet; a few cross-zone assumptions are internally contradictory.
Usability for Security Analysis	20%	75	Attack surface is visible (iFrame boundary, admin paths, external APIs); sufficient for meaningful analysis and prioritization despite missing data sensitivity annotations.

Overall Model Maturity Total Score (0–100): 69 **Overall Model Maturity:** ✅ Adequate

3. Individual Model Evaluations (Threats & Mitigations Only)

##

infosecotb-model-anthropic-claude-opus-4-1-20250805

This section provides the dedicated Threats & Mitigations analysis for this specific model.

##

Threat Landscape Snapshot "(Per Model)"

STRIDE Category	High	Medium	Low	Observations

Spoofing	4	1	0	Strong coverage for admin takeovers (BlueHost, Hugging Face), session hijack, and unauthorized access to the app.
Tampering	4	3	0	Prompt injection, RAG poisoning, iFrame/XSS, API response manipulation; mitigations are concrete (CSP, SRI, integrity checks).
Repudiation	0	1	0	Logging/auditing present mainly around WordPress; could add logs for CI/CD and iFrame events.
Information Disclosure	4	2	0	Key exposure, MITM on iFrame, RAG context extraction; thoughtful origin/CORS and key management advice.
Denial of Service	1	2	0	API/usage cost abuse and site DDoS; mitigations are practical (rate limits, CAPTCHA, usage caps).
Elevation of Privilege	4	1	0	WordPress/plugin, BlueHost/SMTP admin, and developer account risks handled with MFA, RBAC, PAM guidance.

Balanced and plausible; minor under-representation of repudiation vs. other categories.

Mitigation Quality & Alignment "(Per Model)"

infosecotb-model-anthropic-claude-opus-4-1-20250805

Control Area	Adequacy	Observations
Relevance & Specificity	✅	Controls map well to flows (e.g., cert pinning, SRI on iFrame, OAuth/JWT for embedding).
Practicality	✅	Most mitigations are implementable in standard stacks (CSP, rate limiting, MFA, WAF).
Completeness & Coverage	✅	Covers app, store, external APIs, admin flows; minor logging gaps.
Effectiveness	✅	Ties mitigations to root causes (e.g., secrets in code, cross-origin messaging).
Standards Alignment	✅	Aligns with common industry practices (CSP, TLS hygiene, RBAC, logging).
Traceability & Justification	⚠️	Thorough, but could add mapping between specific assets and log events, and reference where controls live (infra vs. app).

Summary Rating: ✅ Adequate

Gaps, Blind Spots & Prioritized Fixes "(Per Model)"

#	Finding	Impact	Effort	Recommendation
1	Thin repudiation coverage outside WordPress	Medium	Low	Add audit trails for iFrame message handling, CI/CD deployments, and API admin actions (immutable logs).
2	Limited data classification for RAG	Medium	Medium	Tag documents by sensitivity; enforce rule-based exclusion/redaction before prompts.
3	Inconsistent flow flags (encryption/public)	Low	Low	Normalize model flags to reflect HTTPS/public paths to avoid control gaps or false assumptions.

Threats & Mitigations Maturity Assessment "(Per Model)"

Evaluation Focus

Assess the model on five dimensions as described in table.

Threats & Mitigations Maturity Levels

- 90–100: 🏆 Excellent — Broad, balanced, and contextually accurate coverage of both DFD elements and methodology categories.
- 75–89: 🌟 Good — Strong overall coverage with minor category or contextual gaps.
- 60–74: ✅ Adequate — Reasonable coverage but missing some methodology areas or underdeveloped mitigations.
- 45–59: ⚙️ Fair — Partial or uneven coverage; key threats or categories underrepresented.
- 30–44: ⚠️ Poor — Major omissions in coverage or unrealistic threats/mitigations.
- 0–29: ❌ Inadequate — Very limited or irrelevant threats and mitigations.

Weighted Scoring Table

Dimension	Weight	Score	Reasoning
DFD Element Coverage	30%	90	Addresses processes, flows, stores, and external actors with meaningful threats and mitigations.
Methodology Coverage & Balance	30%	88	Good spread across STRIDE; minor underweighting of repudiation.
Contextual Accuracy	20%	88	Mostly accurate per trust zone and flow; a few flags contradict HTTPS/public reality.
Mitigation Validity	10%	85	Controls are effective and implementable; could add more log integrity specifics.
Proportionality & Realism	10%	85	Severities and controls are proportionate; cost/abuse threats well handled.

Threats & Mitigations Total Score (0–100): 88 **Threats & Mitigations Maturity:** 🌟 Good

Strategic Recommendations "(Per Model)"

infosecotb-model-anthropic-claude-opus-4-1-20250805

1. Add immutable, centralized audit logs for iFrame postMessage handlers and CI/CD events (deploy, rollback, permission changes).
2. Introduce a data classification and DLP/redaction layer before RAG ingestion and before responses leave the app.
3. Normalize data-flow attributes (isEncrypted/isPublicNetwork) to reflect real transport characteristics.
4. Expand repudiation coverage for SMTP2GO administrative changes (signed webhooks, change receipts).
5. Add explicit incident response triggers tied to usage/cost anomalies for OpenAI API and SMTP2GO (e.g., auto key-rotate, traffic shaping).
6. Specify CSP directives (frame-ancestors, script-src with nonces) and sandbox attributes for the embedded chatbot.

7. Require deployment artifact signing and verification (e.g., Sigstore) for the Hugging Face pipeline.

##

infosecotb-model-anthropic-claude-sonnet-4-5-20250929

This section provides the dedicated Threats & Mitigations analysis for this specific model.

##

Threat Landscape Snapshot "(Per Model)"

STRIDE Category	High	Medium	Low	Observations
Spoofing	4	2	0	Covers WordPress admin compromise, session hijack, API key theft; good origin validation guidance.
Tampering	5	3	1	Strong treatment of iFrame MITM/tampering, RAG content manipulation, CI/CD injection.
Repudiation	2	2	0	Better than most: logging for WP and admin/API activities included; could extend to iFrame messages.
Information Disclosure	4	3	0	Solid handling of API/key leakage, dev logs, transport security specifics (mTLS, schema validation).
Denial of Service	2	2	1	Balanced acknowledgement of DoS via iFrame load and API rate limits; includes graceful degradation.
Elevation of Privilege	4	2	0	Consistent coverage for platform admin (BlueHost, Hugging Face) and dev credentials.

Well-balanced, realistic, and tied to specific flows and zones.

Mitigation Quality & Alignment "(Per Model)"

infosecotb-model-anthropic-claude-sonnet-4-5-20250929

Control Area	Adequacy	Observations
Relevance & Specificity	✅	mTLS, response signature verification, CSP/frame-ancestors, version control for RAG files.
Practicality	✅	Feasible for typical cloud/hosted WordPress/Spaces stacks.
Completeness & Coverage	✅	Thorough across app, store, pipelines, admin portals, and external APIs.
Effectiveness	✅	Tied well to root causes (e.g., prompt/filtering, dev telemetry risk management).
Standards Alignment	✅	Consistent with strong TLS posture, secure logging, and access governance.
Traceability & Justification	✅	Rationales are included; suggests runtime/app self-protection in places.

Summary Rating: ✅ Adequate

Gaps, Blind Spots & Prioritized Fixes "(Per Model)"

#	Finding	Impact	Effort	Recommendation

1	Limited treatment of iFrame event logging	Medium	Low	Add structured logging for cross-origin postMessage events and failures; include origin and message type.
2	Data sensitivity not codified	Medium	Medium	Add classification tags to RAG corpus; enforce filtering rules per tag pre-prompt and pre-response.
3	Dev environment isolation	Medium	Medium	Codify isolation (containers/VMs), require signed extensions, and enforce EDR policy baselines.

Threats & Mitigations Maturity Assessment "(Per Model)"

Weighted Scoring Table

Dimension	Weight	Score	Reasoning
DFD Element Coverage	30%	95	Nearly all modeled elements are covered with specific threats+mitigations.
Methodology Coverage & Balance	30%	90	Very balanced STRIDE representation; minor room for more detailed repudiation on iFrame interactions.
Contextual Accuracy	20%	92	Strong trust-zone awareness and transport security detail; minimal inconsistencies.
Mitigation Validity	10%	90	Appropriate and actionable controls; minor opportunities to deepen log integrity.
Proportionality & Realism	10%	92	Prioritization matches exposure and business context.

Threats & Mitigations Total Score (0–100): 92 **Threats & Mitigations Maturity:** 🏆 Excellent

Strategic Recommendations "(Per Model)"

infosecotb-model-anthropic-claude-sonnet-4-5-20250929

8. Add structured iFrame telemetry (message schemas, origin, decision outcomes).
9. Formalize RAG classification/DLP rules (deny-lists of sensitive entities; content policies).
10. Specify immutable logging controls (e.g., WORM storage) for admin and CI/CD events.
11. Add automated prompts/outputs testing for injection patterns prior to production pushes.
12. Expand change-control guidance for SMTP2GO administrative actions (dual control/workflows).
13. Include detailed rollback verification steps post-deploy (hash comparison against attested release).
14. Add explicit secrets handling for developer tooling (session isolation, short-lived tokens, scoped access).

##

infosecotb-model-gemini-gemini-2.5-pro

This section provides the dedicated Threats & Mitigations analysis for this specific model.

##

Threat Landscape Snapshot "(Per Model)"

STRIDE Category	High	Medium	Low	Observations
Spoofing	2	1	0	API key compromise and admin spoofing mentioned but limited depth.
Tampering	2	1	0	iFrame clickjacking noted; lacks coverings for CI/CD and RAG integrity.
Repudiation	0	0	0	Not meaningfully addressed.
Information Disclosure	2	1	0	API traffic interception flagged; little on dev logs/secrets.
Denial of Service	1	1	0	High-level DDoS mention; limited throttling and fallback detail.
Elevation of Privilege	1	0	0	Underdeveloped across admin and plugin surfaces.

Partial, skewed coverage; several DFD elements lack threats.

Mitigation Quality & Alignment "(Per Model)"

infosecotb-model-gemini-gemini-2.5-pro

Control Area	Adequacy	Observations
Relevance & Specificity	⚠️	Generally relevant but broad; lacks placement (infra vs. app) and control depth.
Practicality	✅	Suggested controls (TLS, CSP, WAF) are implementable.
Completeness & Coverage	⚠️	Major omissions (repudiation, CI/CD, dev telemetry, RAG store integrity).
Effectiveness	⚠️	Controls mitigate symptoms more than root causes (e.g., no secrets governance).
Standards Alignment	✅	Aligns with common baseline web controls.
Traceability & Justification	⚠️	Limited linkage between threats and specific components or data.

Summary Rating: ⚠️ Partially adequate

Gaps, Blind Spots & Prioritized Fixes "(Per Model)"

#	Finding	Impact	Effort	Recommendation
1	Missing repudiation	Medium	Low	Add immutable logs for admin/API/CI events with correlation to user and device.
2	Sparse dev/secrets	High	Medium	Introduce secret scanning, managers, scoped keys, and pre-commit hooks.

	handling			
3	RAG integrity not addressed	Medium	Medium	Add signing, checksums, and write controls to about_me corpus with alerts.

Threats & Mitigations Maturity Assessment "(Per Model)"

Weighted Scoring Table

Dimension	Weight	Score	Reasoning
DFD Element Coverage	30%	55	Many elements unaddressed (stores, several flows).
Methodology Coverage & Balance	30%	50	Missing or under-represented repudiation and EoP.
Contextual Accuracy	20%	60	Mostly correct but lacks zone-aware nuance.
Mitigation Validity	10%	50	High-level mitigations without root-cause targeting.
Proportionality & Realism	10%	55	Prioritization acceptable but shallow.

Threats & Mitigations Total Score (0–100): 55 **Threats & Mitigations Maturity:** ⚙ Fair

Strategic Recommendations "(Per Model)"

infosecotb-model-gemini-gemini-2.5-pro

15. Add dev/CI/CD threat set (supply chain, signed artifacts, SCA/SAST policy gates).
16. Introduce secrets lifecycle controls (rotation, scanning, environment separation).
17. Expand RAG store threats (poisoning, exfiltration) and precise mitigations (FIM, signatures).
18. Add repudiation with log integrity and change receipts for admin actions.
19. Specify traffic-level protections (mTLS/cert pinning) for API requests.
20. Include cost-abuse and quota exhaustion scenarios with circuit breakers and alerts.

##

infosecotb-model-novita-deepseek-deepseek-r1,

This section provides the dedicated Threats & Mitigations analysis for this specific model.

##

Threat Landscape Snapshot "(Per Model)"

STRIDE Category	High	Medium	Low	Observations
Spoofing	1	0	0	Minimal admin/developer spoofing attention.
Tampering	1	1	0	Generic notes; little mapping to specific flows.
Repudiation	0	0	0	Absent.
Information	1	1	0	Broad statements; little tie-in to data stores or logs.

Disclosure				
Denial of Service	1	0	0	Not elaborated on throttling or fallback.
Elevation of Privilege	1	0	0	Limited surface treatment.

Sparse and generic; weak contextual alignment to DFD elements.

Mitigation Quality & Alignment "(Per Model)"

infosecotb-model-novita-deepseek-deepseek-r1,

Control Area	Adequacy	Observations
Relevance & Specificity	❌	Lacks placement and detail; not targeted at root causes.
Practicality	⚠️	Broad guidance feasible, but incomplete.
Completeness & Coverage	❌	Major omissions across flows, stores, admin, and CI/CD.
Effectiveness	⚠️	Generic measures unlikely to reduce real risk materially.
Standards Alignment	⚠️	Mentions encryption and access controls; no depth.
Traceability & Justification	❌	No clear mapping of threats to controls or affected elements.

Summary Rating: ❌ Inadequate

Gaps, Blind Spots & Prioritized Fixes "(Per Model)"

#	Finding	Impact	Effort	Recommendation
1	Lack of STRIDE breadth	High	Medium	Add threats across all STRIDE categories for app, flows, store, and admin endpoints.
2	Missing CI/CD and secrets governance	High	Medium	Introduce detailed supply-chain threats and concrete mitigations.
3	No repudiation	Medium	Low	Add immutable logging across admin and API flows with correlation IDs.

Threats & Mitigations Maturity Assessment "(Per Model)"

Weighted Scoring Table

Dimension	Weight	Score	Reasoning
DFD Element Coverage	30%	40	Large portions of the DFD lack threats/mitigations.
Methodology Coverage & Balance	30%	35	STRIDE categories are thin and unbalanced.
Contextual Accuracy	20%	45	Some claims contradict HTTPS/public flows; low zone-awareness.

Mitigation Validity	10%	35	Generic mitigations; not rooted in specific failure modes.
Proportionality & Realism	10%	40	Prioritization unclear; missing emphasis on real hotspots.

Threats & Mitigations Total Score (0–100): 40 **Threats & Mitigations Maturity:** ⚠️ Poor

Strategic Recommendations "(Per Model)"

infosecotb-model-novita-deepseek-deepseek-r1,

21. Enumerate threats for each critical flow (iFrame, admin portals, OpenAI/SMTP2GO APIs) and about_me store.
22. Add dev/CI/CD supply-chain threats and strong mitigations (signing, attestation, gated deployments).
23. Establish secrets lifecycle and scanning; separate dev/prod keys with minimal scopes.
24. Introduce event-level repudiation with immutable logs and alerting for admin changes.
25. Clarify flow attributes (public/encrypted) to align mitigations with actual transport.
26. Add cost/abuse and quota protections for external APIs.

##

infosecotb-model-novita-deepseek-deepseek-v3.1-terminus

This section provides the dedicated Threats & Mitigations analysis for this specific model.

##

Threat Landscape Snapshot "(Per Model)"

STRIDE Category	High	Medium	Low	Observations
Spoofing	3	2	0	Admin and API endpoint spoofing broadly covered.
Tampering	4	2	0	Good attention to flows; some internal-vs-public inconsistencies.
Repudiation	2	1	0	Better than average; recommends audit logs and receipts.
Information Disclosure	3	3	0	Transport/auth leakage and RAG responses recognized.
Denial of Service	1	1	0	Limited treatment of quotas/fallbacks.
Elevation of Privilege	3	2	0	Addresses admin channels; could add plugin/WordPress role misuse detail.

Broad and generally plausible; some misapplied assumptions about encryption and public networks.

Mitigation Quality & Alignment "(Per Model)"

infosecotb-model-novita-deepseek-deepseek-v3.1-terminus

Control Area	Adequacy	Observations
Relevance &	✅	Often accurate, placed on the right flow/components.

Specificity		
Practicality	✅	Controls are implementable with common tooling.
Completeness & Coverage	⚠️	RAG corpus integrity addressed, but CI/CD and secrets could be deeper.
Effectiveness	⚠️	Some mitigations high-level or duplicated.
Standards Alignment	✅	TLS, MFA, audit logging, access control align with standard practice.
Traceability & Justification	⚠️	Mapping between threats, data assets, and control locations could be clearer.

Summary Rating: ✅ Adequate

Gaps, Blind Spots & Prioritized Fixes "(Per Model)"

#	Finding	Impact	Effort	Recommendation
1	Inconsistent flow attributes	Low	Low	Align isEncrypted/isPublicNetwork with actual HTTPS/public usage to prevent misprioritization.
2	Limited dev/secrets governance	High	Medium	Add secret management, rotation, and scanning gates in CI/CD.
3	Quota and fallback posture	Medium	Low	Add circuit breakers, retries, and failover for OpenAI/SMTP2GO.

Threats & Mitigations Maturity Assessment "(Per Model)"

Weighted Scoring Table

Dimension	Weight	Score	Reasoning
DFD Element Coverage	30%	65	Most areas addressed; a few elements still thin.
Methodology Coverage & Balance	30%	60	Reasonable spread; some categories light compared to others.
Contextual Accuracy	20%	60	Occasional contradictions on transport/public network flags.
Mitigation Validity	10%	55	Effective but sometimes generic; lacks control placement precision.
Proportionality & Realism	10%	70	Prioritization roughly matches exposure.

Threats & Mitigations Total Score (0–100): 62 **Threats & Mitigations Maturity:** ✅ Adequate

Strategic Recommendations "(Per Model)"

infosecotb-model-novita-deepseek-deepseek-v3.1-terminus

27. Normalize transport attributes for all cross-zone flows (HTTPS and public networks).
28. Introduce secret managers, rotation policies, and pre-commit/CI scanners.

29. Add SLSA/attestation and artifact signing in deployments.
30. Strengthen RAG ingestion gate with signed manifests and integrity checks.
31. Define hard limits and circuit breakers for external API usage with automatic alerting and fallbacks.
32. Clarify control locations (app, host, WAF, provider) per mitigation.

##

infosecotb-model-novita-qwen-qwen3-coder-480b-a35b-instruct

This section provides the dedicated Threats & Mitigations analysis for this specific model.

##

Threat Landscape Snapshot "(Per Model)"

STRIDE Category	High	Medium	Low	Observations
Spoofing	2	1	0	Admin/API spoofing acknowledged.
Tampering	3	1	0	iFrame and API tampering noted; some mislabels of encryption.
Repudiation	1	1	0	Present but minimal; needs immutable logs and receipts.
Information Disclosure	3	2	0	Emphasis on “unencrypted” even where HTTPS exists; refine assumptions.
Denial of Service	1	1	0	High-level; throttling/fallback specifics absent.
Elevation of Privilege	2	1	0	Recognizes admin/EoP paths; limited specificity for plugin misuse.

Moderate coverage with several inconsistencies on transport attributes.

Mitigation Quality & Alignment "(Per Model)"

infosecotb-model-novita-qwen-qwen3-coder-480b-a35b-instruct

Control Area	Adequacy	Observations
Relevance & Specificity	⚠️	Generally relevant; transport mislabels erode precision.
Practicality	✅	Controls feasible to implement.
Completeness & Coverage	⚠️	Some flows and store integrity underdeveloped.
Effectiveness	⚠️	Targets symptoms; needs stronger root cause focus (secrets, supply chain, attestation).
Standards Alignment	✅	TLS, CSP, MFA, logging are standard.
Traceability & Justification	⚠️	Missing mapping to data sensitivities and asset-level logs.

Summary Rating: ⚠️ Partially adequate

Gaps, Blind Spots & Prioritized Fixes "(Per Model)"

#	Finding	Impact	Effort	Recommendation
1	Misinterpretation of encryption/public networks	Medium	Low	Correct flags to align with HTTPS/public flows and adjust controls accordingly.
2	Weak supply-chain/CI coverage	High	Medium	Add signed artifacts, SCA/SAST, and attestation in pipeline.
3	RAG store controls	Medium	Medium	Enforce signed content and change monitoring; add access segregation.

Threats & Mitigations Maturity Assessment "(Per Model)"

Weighted Scoring Table

Dimension	Weight	Score	Reasoning
DFD Element Coverage	30%	55	Some coverage, yet notable gaps in store and pipeline.
Methodology Coverage & Balance	30%	50	STRIDE skewed; repudiation underplayed.
Contextual Accuracy	20%	50	Repeated confusion about encryption/public network.
Mitigation Validity	10%	50	Largely generic; lacking root cause measures.
Proportionality & Realism	10%	55	Prioritization is acceptable but shallow.

Threats & Mitigations Total Score (0–100): 52 **Threats & Mitigations Maturity:** ⚙ Fair

Strategic Recommendations "(Per Model)"

infosecotb-model-novita-qwen-qwen3-coder-480b-a35b-instruct

33. Correct transport attributes and recalibrate mitigations (e.g., add cert pinning/mTLS vs. “unencrypted” claims).
34. Add end-to-end RAG integrity (signing, checksums, read-only mounts, FIM).
35. Introduce secrets governance/rotation and pre-commit scanners.
36. Expand repudiation with immutable logging for admin/API/CI and message-level provenance.
37. Add resilience measures (circuit breakers, retries, fallback models/providers).

##

infosecotb-model-openai-gpt-5

This section provides the dedicated Threats & Mitigations analysis for this specific model.

##

Threat Landscape Snapshot "(Per Model)"

STRIDE Category	High	Medium	Low	Observations
Spoofing	4	2	0	Endpoint impersonation and admin identity compromise well-treated (pinning, mTLS, device posture).
Tampering	5	3	1	CI/CD tamper (Sigstore, SLSA), API response manipulation, iFrame integrity robustly covered.
Repudiation	2	2	0	Solid: immutable logs, change receipts, correlation with IdP.
Information Disclosure	4	3	0	Sensitive-data handling for RAG and prompts with DLP/masking and zero-retention options.
Denial of Service	2	2	1	Cost/usage abuse, circuit breakers, throttling, caching, graceful degradation.
Elevation of Privilege	4	2	0	Least privilege RBAC, just-in-time access, short-lived tokens, approval gates.

Comprehensive, precise, and proportionate across all categories.

Mitigation Quality & Alignment "(Per Model)"

infosecotb-model-openai-gpt-5

Control Area	Adequacy	Observations
Relevance & Specificity	✅	Highly targeted to flows/zones with clear placement.
Practicality	✅	Uses commonly available capabilities (e.g., Sigstore, mTLS, CSP, WAF).
Completeness & Coverage	✅	Covers app, store, dev tooling, CI/CD, and admin surfaces extensively.
Effectiveness	✅	Root-cause oriented (supply chain, secrets lifecycle, provenance).
Standards Alignment	✅	Aligns with modern best practices (SLSA, signed artifacts, zero-retention, FIDO2).
Traceability & Justification	✅	Strong mapping from threat → control → outcome; includes prioritization levers.

Summary Rating: ✅ Adequate

Gaps, Blind Spots & Prioritized Fixes "(Per Model)"

#	Finding	Impact	Effort	Recommendation
1	Flow attribute inconsistencies remain in shared DFD	Low	Low	Ensure encryption/public flags are corrected and consistently applied.
2	Formalize data	Medium	Medium	Codify sensitivity tiers and automated

	taxonomy for RAG			exclusion/redaction before API calls.
3	Expand quantitative guardrails	Medium	Low	Add cost/latency SLOs, auto-rate limits, and budget-driven key rotation triggers.

Threats & Mitigations Maturity Assessment "(Per Model)"

Weighted Scoring Table

Dimension	Weight	Score	Reasoning
DFD Element Coverage	30%	95	Nearly all elements covered with high fidelity.
Methodology Coverage & Balance	30%	95	Balanced STRIDE treatment with no major gaps.
Contextual Accuracy	20%	95	Trust-zone and transport-aware, precise control placement.
Mitigation Validity	10%	90	Strong and implementable; few places could add even more specifics.
Proportionality & Realism	10%	95	Controls tied to exposure and business constraints.

Threats & Mitigations Total Score (0–100): 95 **Threats & Mitigations Maturity:** 🏆 Excellent

Strategic Recommendations "(Per Model)"

infosecotb-model-openai-gpt-5

38. Standardize flow flags across the shared DFD; publish a mapping table per flow.
39. Institutionalize content classification with policy enforcement in both ingestion and response paths.
40. Add threat-informed, auto-remediation runbooks (e.g., key rotation on anomaly, deployment freeze on failed attestations).
41. Extend automated tests for prompt injection and output policy violations in pre-deploy CI.
42. Include periodic tabletop exercises focused on supply-chain and admin account compromise scenarios.

##

infosecotb-model-xai-grok-4-latest

This section provides the dedicated Threats & Mitigations analysis for this specific model.

##

Threat Landscape Snapshot "(Per Model)"

STRIDE Category	High	Medium	Low	Observations
Spoofing	2	1	0	Admin and API spoofing mentioned but lightly treated.

Tampering	2	2	0	API/iFrame tampering addressed; control placement not always clear.
Repudiation	1	0	0	Minimal; auditing guidance lacks integrity details.
Information Disclosure	2	1	0	General transport leakage; missing store/log specifics.
Denial of Service	1	1	0	High-level; lacks concrete throttling/circuit breaker actions.
Elevation of Privilege	2	1	0	Present but generic for hosted control panels.

Coverage exists but tends to be repetitive and generic.

Mitigation Quality & Alignment "(Per Model)"

infosecotb-model-xai-grok-4-latest

Control Area	Adequacy	Observations
Relevance & Specificity	⚠️	Generally relevant but lacks actionable detail and control locations.
Practicality	✅	Feasible when specified; many are baseline controls.
Completeness & Coverage	⚠️	Repudiation and RAG integrity need more depth.
Effectiveness	⚠️	Less root-cause focus (secrets, supply chain, attestation).
Standards Alignment	✅	Aligns with common security headers and TLS practices.
Traceability & Justification	⚠️	Thin linkage between threats, assets, and mitigations.

Summary Rating: ⚙️ Fair

Gaps, Blind Spots & Prioritized Fixes "(Per Model)"

#	Finding	Impact	Effort	Recommendation
1	Limited repudiation	Medium	Low	Add WORM/immutable logging and change receipts for admin and deployment actions.
2	Weak RAG integrity	Medium	Medium	Enforce signed content and read-only mounts with FIM and alerts.
3	Generic guidance for supply chain	High	Medium	Introduce signed artifacts, SLSA levels, and attestation verification.

Threats & Mitigations Maturity Assessment "(Per Model)"

Weighted Scoring Table

Dimension	Weight	Score	Reasoning
DFD Element Coverage	30%	50	Not all critical elements receive threats/mitigations.
Methodology Coverage &	30%	45	STRIDE not balanced; repudiation/EoP under-emphasized.

Balance			
Contextual Accuracy	20%	45	Several mitigations not clearly tied to specific flows/zones.
Mitigation Validity	10%	45	Generic and repetitive, lacking root-cause measures.
Proportionality & Realism	10%	50	Relative priorities broadly correct but shallow.

Threats & Mitigations Total Score (0–100): 47 **Threats & Mitigations Maturity:** ⚙️ Fair

Strategic Recommendations "(Per Model)"

infosecotb-model-xai-grok-4-latest

43. Add immutable audit logs and administrative change receipts with signed events.
44. Introduce supply-chain protections (artifact signing, provenance, dependency policies).
45. Strengthen RAG integrity and sensitivity controls (classification, exclusion lists, redaction).
46. Specify precise control locations (WAF vs. app vs. provider) for each mitigation.
47. Add quotas, circuit breakers, and fallback paths for external APIs.

4. Conclusion

- Comparative strengths/weaknesses (threats & mitigations):
- The OpenAI GPT-5 and Anthropic Sonnet models provide the strongest, most balanced STRIDE coverage with precise, actionable mitigations. Anthropic Opus follows closely, with only minor gaps in repudiation scope and control traceability.
- The DeepSeek v3.1 model reaches adequacy but is held back by inconsistent flow attributes and lighter CI/CD and secrets governance.
- Gemini 2.5, xAI Grok, Qwen, and DeepSeek r1 are comparatively weaker—coverage is patchy, mitigations are often generic, and several controls are not tied to root causes, particularly around supply chain, secrets, and log integrity.
- Common DFD-only maturity outcome and implications:
- The shared architecture earns an Overall Model Maturity of 69 (✅ Adequate). Trust zones and main flows are consistently modeled, enabling solid attack surface identification. However, inconsistent encryption/public flags, naming glitches, and missing data classifications reduce accuracy and analytical precision. Addressing these will materially improve the usefulness and credibility of all derived threat models.
- Next steps to elevate shared architecture and per-model threat quality:

48. Standardize flow attributes (encryption/public) across all diagrams to match real transport paths; correct naming inconsistencies.
49. Add data classification to the RAG corpus and enforce redaction/exclusion at ingestion and egress.
50. Decompose WordPress and the CI/CD pipeline one level deeper (plugins/themes, build/signing/attestation steps) to anchor supply-chain and plugin risks.
51. Institutionalize immutable logging and change receipts for admin, API, and CI/CD operations; correlate with IdP events.

52. Raise the baseline for dev/secrets governance (scoped keys, rotation, pre-commit scanning, secret vaults).
53. Implement quantitative guardrails for external APIs (circuit breakers, quotas, cost alerts, failover providers).

By addressing these DFD common gaps and adopting the higher-maturity practices exemplified by the top-ranked models, the organization can achieve consistent, defensible, and action-ready threat models across the full portfolio.