Comprehensive Threat Model Evaluation Report (Combined)

Executive Summary & Comparative Analysis

This report evaluates nine AI-generated threat models for an online payments processing platform based on the OWASP Threat Dragon schema. All models share a common data flow diagram (DFD) structure depicting customer interactions, merchant processing, and Stripe integration across trust boundaries (Customer/Internet, Merchant/Web, Stripe/Web). The DFD includes actors (Customer), processes (Customer Client, Merchant Web Server, Stripe API service, Stripe Payment Service), and flows for login, order intent, payment confirmation, and responses. Key differences emerge in the depth, specificity, and coverage of threats and mitigations, with some models providing detailed, context-aware threats while others are superficial or repetitive. Overall, the models demonstrate varying maturity in applying STRIDE, with stronger ones offering balanced coverage and practical mitigations, while weaker ones focus narrowly on encryption gaps or lack nuance. The shared DFD is logically sound but shows opportunities for enhancement in labeling and boundary details. Recommendations prioritize improving threat realism, mitigation specificity, and cross-model consistency to strengthen security posture.

1. Threats & Mitigations Maturity Ranking (Across Models)

Rank	Model Name	Threats & Mitigations Score	Maturity	Reasoning
1	payments-processing-platform-openai-gpt-5	90	🏆 Excellent	Comprehensive coverage across all STRIDE categories with highly contextual threats tied to specific flows and boundaries; mitigations are detailed, standards-aligned (e.g., TLS 1.3, mTLS), and address root causes like replay attacks and credential leaks.
2	payments-processing-platform-anthropic-claude-sonnet-4-5-20250929	85	🌟 Good	Strong, detailed threats with in-depth descriptions referencing trust zones and flows; mitigations are practical and PCI DSS-aligned, though slightly repetitive in encryption emphasis.
3	payments-processing-platform-xai-grok-4-latest	82	🌟 Good	Balanced STRIDE coverage with realistic threats focused on boundary crossings; mitigations are comprehensive but occasionally generic (e.g., "encrypt all data").
4	payments-processing-platform-anthropic-claude-opus-4-1-20250805	78	🌟 Good	Good coverage of tampering and disclosure threats with flow-specific details; mitigations are actionable but less emphasis on repudiation and DoS compared to top models.
5	payments-processing-	75	🌟 Good	Solid threats tied to positions and flows, with practical mitigations; reasoning is efficient but

	platform-xai-grok-4-fast-reasoning-latest			some threats overlap without deeper chaining.
6	payments-processing-platform-gemini-gemini-2.5-pro	70	✅ Adequate	Adequate coverage but uneven; focuses on client-side and API threats with relevant mitigations, but misses some internal flows and has fewer threats overall.
7	payments-processing-platform-novita-qwenqwen3-coder-480b-a35b-instruct	65	✅ Adequate	Repetitive threats centered on encryption gaps; mitigations are consistent but lack depth in addressing specific STRIDE categories like elevation of privilege.
8	payments-processing-platform-novita-deepseek-deepseek-v3.1-terminus	60	✅ Adequate	Basic threats with good coverage of tampering and disclosure; mitigations are straightforward but underdeveloped for complex scenarios like repudiation.
9	payments-processing-platform-ollama-qwen330b	50	⚙️ Fair	

2. Overall Model Maturity

2.1 Evaluation Summary

The shared DFD across all models effectively captures the core payment flow with clear trust boundaries (Customer/Internet, Merchant/Web, Stripe/Web) and key elements like actors, processes, and bidirectional flows for login, order intent, and payment processing. Strengths include logical sequencing of steps (1-11) and identification of boundary crossings, which aids in spotting exposure points. Key gaps are inconsistent encryption flags (often false despite HTTPS protocol) and lack of data stores, limiting visibility into persistence risks; the layout is readable but could benefit from explicit data labels on flows for better traceability.

2.2 Scoring Table

Dimension	Weight	Score	Reasoning
Clarity and Readability	25%	85	Flows and elements are well-labeled with step numbers and descriptions; trust boundaries are clearly delineated, though some flows lack protocol details for quick visual assessment.
Completeness and Coverage	30%	80	Covers essential actors, processes, and flows across boundaries; includes bidirectional elements and positions, but omits data

			stores and external dependencies (e.g., banks), reducing full-system coverage.
Accuracy and Logical Consistency	25%	90	Flows follow a coherent sequence (login → intent → confirmation → response) without contradictions; boundary crossings are accurately implied, and elements align with payment orchestration logic.
Usability for Security Analysis	20%	85	Enables identification of high-risk boundary flows; positions and connections highlight exposure, but missing data flows and encryption flags make risk inference somewhat manual.

Overall Model Maturity Total Score (0–100): 85 **Overall Model Maturity:** 🌟 Good

3. Individual Model Evaluations (Threats & Mitigations Only)

##

payments-processing-platform-anthropic-claude-opus-4-1-20250805.json

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

##

Threat Landscape Snapshot

STRIDE Category	High	Medium	Low	Observations
Spoofing	3	2	0	Balanced focus on identity threats in actors and flows.
Tampering	8	3	0	Dominant category, covering data manipulation in most flows.
Repudiation	2	2	0	Adequate but limited to basic logging needs.
Information Disclosure	7	2	0	Strong emphasis on data exposure in client and boundary flows.
Denial of Service	3	2	0	Covers resource exhaustion but not deeply.
Elevation of Privilege	3	1	0	Focuses on privilege escalation in processes.

The model provides balanced STRIDE coverage with a focus on tampering and disclosure, plausible for payment flows, though repudiation threats are underdeveloped.

Mitigation Quality & Alignment

Control Area	Adequacy	Observations
Relevance & Specificity	✅	Mitigations directly address flow-specific risks like token hijacking.
Practicality	✅	Actionable steps like MFA and server validation.
Completeness & Coverage	⚠️	Covers encryption and validation but gaps in DoS mitigations.
Effectiveness	✅	Targets root causes, e.g., PKCE for OAuth.
Standards Alignment	✅	References TLS 1.3, PCI DSS.

Traceability & Justification	✅	Clear links to threats with practical examples.

Summary Rating: ✅ Adequate

Gaps, Blind Spots & Prioritized Fixes

#	Finding	Impact	Effort	Recommendation
1	Underdeveloped repudiation threats for flows like (10).	Medium	Low	Add threats for log tampering and mitigations like immutable logs.
2	No threats for internal Stripe flows like (9)/(10).	High	Medium	Include intra-service threats with mTLS mitigations.
3	Generic DoS mitigations across models.	Medium	Low	Specify WAF rules for payment endpoints.

Threats & Mitigations Maturity Assessment

This section evaluates the **completeness, contextual quality, and methodological balance** of threats and mitigations within the model. It focuses on whether the threat model demonstrates a *credible and comprehensive application of the selected methodology* (e.g., STRIDE) across all relevant elements of the DFD.

Dimension	Weight	Score	Reasoning
DFD Element Coverage	30%	85	All major processes and flows have threats; actors and boundaries are addressed, but no data stores covered.
Methodology Coverage & Balance	30%	80	All STRIDE categories present, but DoS and repudiation underrepresented relative to tampering.
Contextual Accuracy	20%	85	Threats align with boundary crossings and payment sensitivity; plausible for e-commerce.
Mitigation Validity	10%	80	Mitigations effective but some generic (e.g., "encrypt channel" without specifics).
Proportionality & Realism	10%	75	High-severity threats prioritized correctly, but lacks prioritization for low-effort fixes.

Threats & Mitigations Total Score (0–100): 82 **Threats & Mitigations Maturity:** 🌟 Good

Strategic Recommendations

1. Expand repudiation threats to include blockchain-based proofs for high-value transactions to enhance non-repudiation.
2. Add threats for supply chain attacks on client libraries, with mitigations like SRI for stripe.js.
3. Include DoS threats for all boundary-crossing flows, recommending API gateways with behavioral analysis.

4. Balance coverage by adding low-severity threats like session fixation, with mitigations tied to OAuth.
5. Remove redundant encryption mitigations; consolidate into a single, comprehensive TLS policy recommendation.

##

payments-processing-platform-anthropic-claude-sonnet-4-5-20250929.json

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

##

Threat Landscape Snapshot

STRIDE Category	High	Medium	Low	Observations
Spoofing	4	3	0	Detailed on endpoint spoofing and phishing.
Tampering	9	4	0	Extensive coverage of data manipulation in flows.
Repudiation	3	3	0	Improved with audit trail focus.
Information Disclosure	8	3	0	Strong on card data and token leakage.
Denial of Service	4	3	0	Covers client and server exhaustion.
Elevation of Privilege	4	2	0	Addresses client manipulation and API bypass.

Strong balance with deeper chaining of threats across flows.

Mitigation Quality & Alignment

Control Area	Adequacy	Observations
Relevance & Specificity	✅	Highly specific to flows, e.g., 3D Secure for cardholder auth.
Practicality	✅	Detailed, e.g., nonce validation for replays.
Completeness & Coverage	✅	Covers all categories with multi-layered defenses.
Effectiveness	✅	Root-cause focused, like token binding.
Standards Alignment	✅	PCI DSS, PSD2, TLS 1.3 explicit.
Traceability & Justification	✅	Threats linked to specific boundary risks.

Summary Rating: ✅ Adequate

Gaps, Blind Spots & Prioritized Fixes

#	Finding	Impact	Effort	Recommendation
1	Limited internal Stripe threats.	High	Medium	Add threats for service-to-service auth bypass.

2	Repetition in encryption mitigations.	Low	Low	Consolidate into reusable security controls section.
3	No threats for mobile client variants.	Medium	Low	Include device-specific threats like jailbreak detection.

Threats & Mitigations Maturity Assessment

Dimension	Weight	Score	Reasoning
DFD Element Coverage	30%	90	Comprehensive threats on all processes and flows; boundaries well-integrated.
Methodology Coverage & Balance	30%	85	Even distribution, with good depth in tampering/disclosure.
Contextual Accuracy	20%	90	Threats precisely tied to positions, flows, and zones (e.g., client-side encryption gaps).
Mitigation Validity	10%	85	Effective and compliant, but some overlap reduces uniqueness.
Proportionality & Realism	10%	85	Realistic high-impact threats prioritized; mitigations feasible for payments.

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

Strategic Recommendations

6. Enhance repudiation with blockchain audit trails for immutable transaction proofs.
7. Add supply chain threats for third-party SDKs, mitigating with vendor attestations.
8. Include threats for quantum-resistant crypto in long-term mitigations.
9. Balance by adding low-severity threats like session hijacking with token binding.
10. Remove generic "encrypt" repeats; focus on implementation details like key rotation.

##

payments-processing-platform-gemini-gemini-2.5-pro,.json

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

##

Threat Landscape Snapshot

STRIDE Category	High	Medium	Low	Observations
Spoofing	2	1	0	Basic identity threats.
Tampering	3	1	0	Focused on client and API tampering.
Repudiation	1	1	0	Limited to transaction disputes.
Information Disclosure	2	1	0	Covers data leakage in flows.
Denial of Service	2	1	0	Addresses server DoS.

Elevation of Privilege	1	0	0	Minimal coverage.

Uneven, with tampering dominant but overall sparse.

Mitigation Quality & Alignment

Control Area	Adequacy	Observations
Relevance & Specificity	⚠️	Tied to elements but lacks flow details.
Practicality	✅	Practical like CSP for XSS.
Completeness & Coverage	⚠️	Gaps in repudiation and elevation.
Effectiveness	✅	Addresses threats directly.
Standards Alignment	⚠️	Basic TLS mentions.
Traceability & Justification	⚠️	Some links to flows but generic.

Summary Rating: ⚠️ Partially adequate

Gaps, Blind Spots & Prioritized Fixes

#	Finding	Impact	Effort	Recommendation
1	Sparse threats; misses internal flows.	High	Medium	Add threats for all 11 steps with STRIDE balance.
2	No elevation threats in processes.	High	Low	Include privilege escalation in Merchant Server.
3	Generic mitigations for DoS.	Medium	Low	Specify DDoS mitigation tools.

Threats & Mitigations Maturity Assessment

Dimension	Weight	Score	Reasoning
DFD Element Coverage	30%	70	Covers main elements but threats are few and not all flows addressed.
Methodology Coverage & Balance	30%	65	All categories touched but shallow; elevation underrepresented.
Contextual Accuracy	20%	75	Threats plausible for client-server interactions.
Mitigation Validity	10%	70	Effective but not comprehensive.
Proportionality	10%	70	Realistic but lacks depth in high-impact areas.

& Realism			

Threats & Mitigations Total Score (0–100): 70 **Threats & Mitigations Maturity:** ✅ Adequate

Strategic Recommendations

11. Expand to cover all STRIDE categories equally, adding 2-3 threats per category.
12. Include threats for data stores if added to DFD.
13. Enhance mitigations with specific tools like Fail2Ban for DoS.
14. Add cross-flow threats like race conditions in (3)/(6).
15. Balance by reducing tampering focus; add repudiation threats.

##

payments-processing-platform-novita-deepseek-deepseek-v3.1-terminus.json

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

##

Threat Landscape Snapshot

STRIDE Category	High	Medium	Low	Observations
Spoofing	1	1	0	Basic identity spoofing.
Tampering	2	1	0	Focus on data flows.
Repudiation	1	1	0	Logging-focused.
Information Disclosure	2	1	0	Encryption emphasis.
Denial of Service	1	1	0	Service availability.
Elevation of Privilege	2	1	0	Privilege bypass.

Basic coverage, heavy on tampering/disclosure.

Mitigation Quality & Alignment

Control Area	Adequacy	Observations
Relevance & Specificity	⚠️	Tied to flows but brief.
Practicality	✅	Straightforward like MFA.
Completeness & Coverage	⚠️	Gaps in DoS and elevation.
Effectiveness	⚠️	Basic encryption focus.
Standards Alignment	⚠️	Mentions PCI DSS vaguely.
Traceability & Justification	⚠️	Limited justification.

Summary Rating: ⚠️ Partially adequate

Gaps, Blind Spots & Prioritized Fixes

#	Finding	Impact	Effort	Recommendation
1	Few threats overall.	High	Medium	Add 5+ threats per major flow.
2	No internal boundary threats.	Medium	Low	Include service-to-service risks.
3	Repetitive encryption mitigations.	Low	Low	Diversify with access controls.

Threats & Mitigations Maturity Assessment

Dimension	Weight	Score	Reasoning
DFD Element Coverage	30%	65	Threats on key processes but misses some flows.
Methodology Coverage & Balance	30%	60	All categories but shallow depth.
Contextual Accuracy	20%	70	Plausible for basic payment risks.
Mitigation Validity	10%	65	Practical but not layered.
Proportionality & Realism	10%	60	Realistic but unbalanced.

Threats & Mitigations Total Score (0–100): 64 **Threats & Mitigations Maturity:** ⚙️ Fair

Strategic Recommendations

16. Add detailed threats for all boundary crossings.
17. Include elevation threats with RBAC mitigations.
18. Enhance DoS coverage with redundancy recommendations.
19. Balance categories by adding spoofing variants.
20. Remove duplicates; focus on unique risks per flow.

##

payments-processing-platform-novita-qwenqwen3-coder-480b-a35b-instruct.json

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

##

Threat Landscape Snapshot

STRIDE Category	High	Medium	Low	Observations

Spoofing	3	2	0	Repetitive encryption focus.
Tampering	3	2	0	Encryption-heavy.
Repudiation	3	2	0	Logging emphasis.
Information Disclosure	3	3	0	Dominant category.
Denial of Service	2	1	0	Basic coverage.
Elevation of Privilege	3	2	0	Client/server focus.

Repetitive but covers all.

Mitigation Quality & Alignment

Control Area	Adequacy	Observations
Relevance & Specificity	⚠️	Encryption repeated across threats.
Practicality	✅	Basic but implementable.
Completeness & Coverage	⚠️	Over-relies on encryption.
Effectiveness	⚠️	Generic for complex threats.
Standards Alignment	⚠️	TLS mentions but no PCI.
Traceability & Justification	⚠️	Poor links to specific flows.

Summary Rating: ⚠️ Partially adequate

Gaps, Blind Spots & Prioritized Fixes

#	Finding	Impact	Effort	Recommendation
1	Repetitive threats lack variety.	Medium	Low	Diversify with flow-specific variants.
2	No unique mitigations per threat.	High	Medium	Tailor mitigations to categories.
3	Misses internal Stripe threats.	Medium	Low	Add intra-service risks.

Threats & Mitigations Maturity Assessment

Dimension	Weight	Score	Reasoning
DFD Element Coverage	30%	70	Covers processes but repetitive.
Methodology Coverage &	30%	70	All categories but unbalanced.

Balance			
Contextual Accuracy	20%	65	Encryption focus is relevant but narrow.
Mitigation Validity	10%	60	Effective basics but lacks depth.
Proportionality & Realism	10%	65	Realistic but not prioritized.

Threats & Mitigations Total Score (0–100): 68 **Threats & Mitigations Maturity:** ✅ Adequate

Strategic Recommendations

21. Vary threats to avoid repetition, adding behavioral analysis.
22. Include PCI-specific mitigations for card flows.
23. Add threats for repudiation in (3)/(6).
24. Balance with more DoS and elevation threats.
25. Enhance traceability by referencing exact flow steps.

##

payments-processing-platform-ollama-qwen330b.json

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

##

Threat Landscape Snapshot

STRIDE Category	High	Medium	Low	Observations
Spoofing	1	0	0	Minimal.
Tampering	1	0	0	Encryption-focused.
Repudiation	0	0	0	Absent.
Information Disclosure	1	0	0	Single threat on unencrypted flow.
Denial of Service	0	0	0	Absent.
Elevation of Privilege	0	0	0	Absent.

Severely limited coverage.

Mitigation Quality & Alignment

Control Area	Adequacy	Observations
Relevance & Specificity	❌	Generic encryption advice.
Practicality	⚠️	Basic but incomplete.
Completeness & Coverage	❌	Only addresses one flow.
Effectiveness	⚠️	TLS but no depth.

Standards Alignment	⚠️	Mentions TLS vaguely.
Traceability & Justification	❌	No threat links.

Summary Rating: ❌ Inadequate

Gaps, Blind Spots & Prioritized Fixes

#	Finding	Impact	Effort	Recommendation
1	Only one threat total.	High	High	Generate full STRIDE threats for all elements.
2	No coverage for most categories.	High	Medium	Add balanced threats per flow.
3	Ignores internal and boundary risks.	High	Low	Include all 11 flows.

Threats & Mitigations Maturity Assessment

Dimension	Weight	Score	Reasoning
DFD Element Coverage	30%	20	Only one flow threatened; vast gaps.
Methodology Coverage & Balance	30%	10	Almost no STRIDE balance.
Contextual Accuracy	20%	50	Single threat is plausible but isolated.
Mitigation Validity	10%	40	Basic but irrelevant to most risks.
Proportionality & Realism	10%	30	Unrealistic minimalism.

Threats & Mitigations Total Score (0–100): 25 **Threats & Mitigations Maturity:** ❌ Inadequate

Strategic Recommendations

26. Massively expand threats to cover all STRIDE and flows.
27. Add mitigations specific to payment sensitivity.
28. Include threats for actors and stores.
29. Balance with equal category distribution.
30. Focus on boundary-specific risks.

##

payments-processing-platform-openai-gpt-5.json

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

##

Threat Landscape Snapshot

STRIDE Category	High	Medium	Low	Observations
Spoofing	4	2	0	Detailed on impersonation.
Tampering	5	3	0	Comprehensive flow coverage.
Repudiation	2	2	0	Audit-focused.
Information Disclosure	5	3	0	Strong on data leaks.
Denial of Service	3	2	0	Covers exhaustion.
Elevation of Privilege	3	2	0	Privilege bypass emphasis.

Excellent balance and depth.

Mitigation Quality & Alignment

Control Area	Adequacy	Observations
Relevance & Specificity	🏆	Highly contextual to flows/zones.
Practicality	🏆	Detailed, e.g., certificate pinning.
Completeness & Coverage	🏆	Layered defenses per threat.
Effectiveness	🏆	Root-cause addressed.
Standards Alignment	🏆	TLS 1.3, mTLS, PCI explicit.
Traceability & Justification	🏆	Threats directly mapped.

Summary Rating: 🏆 Excellent

Gaps, Blind Spots & Prioritized Fixes

#	Finding	Impact	Effort	Recommendation
1	Minor gaps in quantum threats.	Low	Low	Add future-proof crypto notes.
2	No mobile-specific threats.	Medium	Medium	Include app-specific risks.
3	Overlap in TLS mitigations.	Low	Low	Consolidate security controls.

Threats & Mitigations Maturity Assessment

Dimension	Weight	Score	Reasoning
DFD Element Coverage	30%	95	Near-complete; all flows and boundaries threatened.
Methodology Coverage & Balance	30%	95	Even, deep coverage across STRIDE.
Contextual Accuracy	20%	95	Precisely tied to positions and risks.
Mitigation Validity	10%	95	Highly effective and realistic.
Proportionality & Realism	10%	95	Prioritized correctly for payments.

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

Strategic Recommendations

31. Add advanced threats like supply chain compromises in SDKs.
32. Include zero-trust mitigations for internal flows.
33. Enhance with threat modeling for webhooks.
34. Balance by adding low-severity threats like enumeration.
35. Recommend automated threat scanning tools.

##

payments-processing-platform-xai-grok-4-fast-reasoning-latest.json

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

##

Threat Landscape Snapshot

STRIDE Category	High	Medium	Low	Observations
Spoofing	2	1	0	Boundary-focused.
Tampering	3	1	0	Flow-specific.
Repudiation	1	1	0	Basic.
Information Disclosure	2	1	0	Encryption emphasis.
Denial of Service	2	1	0	Server/client DoS.
Elevation of Privilege	1	0	0	Limited.

Moderate coverage, efficient.

Mitigation Quality & Alignment

Control Area	Adequacy	Observations
Relevance & Specificity	✅	Tied to positions/flows.
Practicality	✅	Actionable like CSP.
Completeness & Coverage	⚠️	Gaps in elevation/repudiation.
Effectiveness	✅	Good for identified risks.
Standards Alignment	✅	TLS, CSP mentioned.
Traceability & Justification	✅	Clear element links.

Summary Rating: ✅ Adequate

Gaps, Blind Spots & Prioritized Fixes

#	Finding	Impact	Effort	Recommendation
1	Limited elevation threats.	High	Low	Add privilege escalation details.
2	No repudiation depth.	Medium	Low	Include audit trails.
3	Sparse internal threats.	Medium	Medium	Cover Stripe intra-flows.

Threats & Mitigations Maturity Assessment

Dimension	Weight	Score	Reasoning
DFD Element Coverage	30%	75	Covers key elements but not all flows.
Methodology Coverage & Balance	30%	70	Balanced but shallow in some areas.
Contextual Accuracy	20%	80	Plausible for fast-reasoning model.
Mitigation Validity	10%	75	Effective basics.
Proportionality & Realism	10%	80	Realistic prioritization.

Threats & Mitigations Total Score (0–100): 76 **Threats & Mitigations Maturity:** 🌟 Good

Strategic Recommendations

36. Deepen elevation threats with code examples.
37. Add repudiation via digital receipts.
38. Include threats for all 11 steps.

39. Balance DoS with client-side mitigations.
40. Enhance with standards like OWASP.

##

payments-processing-platform-xai-grok-4-latest.json

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

##

Threat Landscape Snapshot

STRIDE Category	High	Medium	Low	Observations
Spoofing	3	2	0	Comprehensive.
Tampering	4	2	0	Flow-detailed.
Repudiation	2	1	0	Logging focus.
Information Disclosure	4	2	0	Boundary risks.
Denial of Service	3	2	0	Multi-level.
Elevation of Privilege	3	2	0	Privilege focus.

Well-balanced.

Mitigation Quality & Alignment

Control Area	Adequacy	Observations
Relevance & Specificity	✅	Specific to elements.
Practicality	✅	Detailed logging.
Completeness & Coverage	✅	Covers categories well.
Effectiveness	✅	Root-cause oriented.
Standards Alignment	✅	TLS, RBAC.
Traceability & Justification	✅	Good links.

Summary Rating: ✅ Adequate

Gaps, Blind Spots & Prioritized Fixes

#	Finding	Impact	Effort	Recommendation
1	Some generic mitigations.	Low	Low	Add specifics like tools.
2	Limited internal threats.	Medium	Medium	Include service risks.

3	No advanced threats.	Medium	Low	Add supply chain.

Threats & Mitigations Maturity Assessment

Dimension	Weight	Score	Reasoning
DFD Element Coverage	30%	85	Strong on processes/flows.
Methodology Coverage & Balance	30%	85	Even across STRIDE.
Contextual Accuracy	20%	90	Highly plausible.
Mitigation Validity	10%	85	Effective and practical.
Proportionality & Realism	10%	85	Well-prioritized.

Threats & Mitigations Total Score (0–100): 86 **Threats & Mitigations Maturity:** 🌞 Good

Strategic Recommendations

41. Add advanced threats like API abuse.
42. Include mobile and API-specific mitigations.
43. Enhance repudiation with blockchain.
44. Balance with low-severity threats.
45. Recommend continuous monitoring.

4. Conclusion

The top models (GPT-5, Claude Sonnet) excel in detailed, balanced STRIDE application with contextual threats and robust mitigations, making them suitable for high-stakes payments, while lower-ranked ones (e.g., Ollama-Qwen) suffer from sparsity and repetition, limiting their utility. The common DFD is 🌞 Good in maturity, providing a solid foundation for analysis but needing better encryption flags and data store inclusion to fully support security reviews. To elevate, standardize threat templates across models for consistency, prioritize boundary-crossing risks in future iterations, and integrate automated tools for ongoing validation; this would transform adequate models into excellent ones, enhancing overall platform resilience against evolving threats like API abuse and supply chain attacks.