**Phase 0 – Platform Architecture & Infrastructure Setup**

* Collaborate with Data Engineer & Backend Engineer to set up **ML/LLM serving infrastructure** (GPU servers, cloud ML endpoints, on-prem inference).
* Select **model hosting strategy** (on-prem, cloud API, hybrid).
* Set up **CI/CD for AI models** (training → validation → deployment pipelines).

**Phase 1 – Data Acquisition & Ingestion**

* Advise Data Engineer on **which raw data sources** are most valuable for AI model training and inference (e.g., text-heavy intel reports, historical IOCs).
* Define **metadata requirements** for training future models (source reliability, campaign tags).
* Ensure ingestion captures **rich context**, not just raw IOCs.

**Phase 2 – Data Normalization & Storage**

* Define **data formats and structures** required for ML/LLM inputs.
* Collaborate with Data Engineer to implement **data preprocessing pipelines** for AI readiness (tokenization, text cleaning, vectorization).
* Maintain a **feature store** for ML models (e.g., enriched IOC features, network activity embeddings).

**Phase 3 – Threat Data Management**

* Design **automatic IOC classification** models (malware, phishing, ransomware, etc.).
* Build **threat attribution models** (linking IOCs to known APT groups or campaigns).
* Create **priority scoring models** to help SOC teams focus on high-impact threats.

**Phase 4 – Analysis & Correlation**

* Develop **ML-based correlation algorithms** (e.g., link campaigns using shared infrastructure or TTPs).
* Build **graph-based AI models** for entity relationship mapping (Neo4j + GNNs).
* Implement **predictive models** for early warning of future attacks.
* Integrate **LLMs** to:
  + Summarize raw intel feeds into human-readable reports.
  + Perform Q&A over historical threat data.
  + Suggest mitigation strategies based on past incidents.

**Phase 5 – Visualization & Reporting**

* Build **AI-assisted dashboards** that surface anomaly alerts or predicted campaigns.
* Integrate **natural language query capability** so analysts can type “Show me phishing campaigns targeting finance sector in last 30 days” and get results.
* Enable **automated executive summaries** with LLMs for reports generated in this phase.

**Phase 6 – Security Operations Integration**

* Integrate AI-driven alerts into **SIEM/SOAR systems**.
* Create **ML/LLM-powered enrichment microservices** that SOC playbooks can call automatically.
* Collaborate with Security Engineers to ensure **model outputs are explainable** for SOC compliance.

**Phase 7 – Governance, Compliance & Security**

* Implement **bias checks** and model fairness audits (important if models score threats differently by geography or actor attribution).
* Ensure **LLM data handling policies** align with GDPR/CCPA (no sensitive customer data leaks into training).
* Document **model decisions** for forensic and compliance needs.

**Phase 8 – AI & Automation (Dedicated AI Phase)**

* Maintain a **model training loop** using feedback from analysts (human-in-the-loop refinement).
* Experiment with **multi-modal AI** (text, network logs, images of phishing websites, malware binary features).
* Fine-tune LLMs for domain-specific cybersecurity vocabulary.
* Develop **continuous learning pipelines** so models adapt to evolving threats.

**Expected Outcomes from AI Engineer’s Work**

* **Faster threat triage** via automated classification, correlation, and scoring.
* **Higher analyst productivity** by offloading repetitive analysis tasks to ML/LLM systems.
* **Smarter threat predictions** that warn SOC teams before large-scale incidents occur.
* **Explainable AI outputs** that can be trusted in compliance-heavy environments.
* **Seamless integration** of AI features into backend APIs, dashboards, and the Flutter mobile app.

**UPDATED PLAN:**

**AI Engineer – Job Responsibilities in the TIP Pipeline**

**1. Threat Intelligence Data Modeling & Processing**

* Design and implement **NLP/NLU models** to process unstructured threat intelligence data (e.g., CVEs, CTI reports, dark web chatter).
* Build entity recognition pipelines for **Indicators of Compromise (IoCs)**, tactics/techniques (MITRE ATT&CK), and threat actor profiles.
* Develop automated methods for **data normalization** across multiple intelligence formats (STIX/TAXII, JSON, CSV, etc.).

**2. Threat Pattern Recognition & Scoring**

* Train ML models to **detect patterns** and anomalies in incoming threat feeds (e.g., phishing campaigns, malware C2 servers).
* Create **risk scoring algorithms** based on severity, confidence, and potential impact of threats.
* Implement **automated correlation** between internal logs and external threat feeds.

**3. AI-driven Threat Prediction**

* Build models for **threat trend forecasting** using time-series and historical attack data.
* Use **graph neural networks** or clustering to detect related threat campaigns across data sources.
* Integrate contextual enrichment from **OSINT and dark web sources**.

**4. Integration with Backend APIs**

* Work closely with **Backend Engineers** to:
  + Deploy trained models as REST/GraphQL APIs.
  + Optimize inference performance for **real-time** threat detection.
  + Ensure model outputs align with backend database schema and indexing.

**5. Model Lifecycle Management**

* Establish **MLOps pipelines** for training, testing, and deploying models.
* Implement **continuous learning** with new threat data.
* Ensure **version control** for models, datasets, and inference pipelines.

**6. Collaboration with Security/Network Engineers**

* Use security engineers’ expertise to **validate model detections** and reduce false positives.
* Incorporate **real-world attack patterns** and SOC feedback into training datasets.
* Support security automation by **integrating AI outputs into SOC playbooks**.

**7. Visualization & Reporting Support**

* Provide **structured AI outputs** that can be consumed by the Flutter front-end for:
  + Threat heatmaps
  + Confidence scores
  + Attack chain visualizations
* Collaborate with data engineers to ensure data is **aggregated, cleaned, and labeled** before modeling.

✅ **Expected Outcomes**:

* Production-ready AI models integrated into the TIP backend.
* Automated detection, classification, and prioritization of threat intelligence.
* Predictive analytics improving proactive threat hunting.
* Measurable reduction in analyst workload via AI-assisted triage.