Top 10 Models and Technologies for Cyber Threat Visualization

---

AI & Machine Learning Models

1. Isolation Forest

- **Description:** An unsupervised machine learning algorithm that identifies anomalies by isolating observations in a data forest. Unlike other methods that profile "normal" data, it focuses specifically on the "few and different" points.
- **Why it helps your project:** In **Module 2**, this is essential for **Anomaly Detection**. It can automatically flag unusual spikes in traffic or non-standard attack patterns in your logs that might otherwise be missed by static thresholds, helping analysts catch zero-day threats.

2. LSTM (Long Short-Term Memory) Networks

- **Description:** A type of Recurrent Neural Network (RNN) capable of learning long-term dependencies, specifically designed for processing and predicting sequences of data over time.
- **Why it helps your project:** It powers the **Trend Detection** outcome. By analyzing historical incident data, an LSTM can forecast future attack volumes or identify "seasonality" in cyber threats (e.g., increased attacks during holiday weekends), allowing for proactive resource allocation.

3. XGBoost (Extreme Gradient Boosting)

- **Description:** A highly efficient and scalable implementation of gradient-boosted decision trees designed for speed and performance with structured data.
- **Why it helps your project:** Useful in **Module 1** for **Threat Classification**. It can be trained to automatically categorize raw security logs into specific **MITRE ATT&CK** tactics with high accuracy, automating the "Data Structuring" phase and reducing manual effort for the analyst.

4. Autoencoders (Deep Learning)

- **Description:** A neural network that learns a compressed representation of input data and then tries to reconstruct it. High reconstruction error indicates an anomaly.
- **Why it helps your project:** It enhances **Systemic Hotspot Identification**. By training on "healthy" system behavior, the dashboard can visually highlight systems that are behaving erratically, providing a "heat map" of potentially compromised internal assets.

Visualization & Frontend Technologies

5. Plotly Dash

- **Description:** A productive Python framework for building web analytic applications. It is built on top of Flask, Plotly.js, and React.js.
- **Why it helps your project:** It is the primary engine for **Module 4 (Integration)**. It allows you to create a **responsive, professional-grade dashboard** using only Python, ensuring that your interactive filters (like date pickers or dropdowns) update all charts across the dashboard in real-time.

6. Kepler.gl (via Deck.gl)

- **Description:** A powerful open-source geospatial analysis tool for large-scale data sets, capable of rendering millions of points in a web browser using GPU acceleration.
- **Why it helps your project:** Essential for **Geospatial Risk Mapping**. Standard maps often lag when showing global attack data. Kepler.gl provides high-performance 3D "arc" visualizations that can show the flow of attacks from origin to target across the globe without performance drops.

7. D3.js (Data-Driven Documents)

- **Description:** A JavaScript library for producing dynamic, interactive data visualizations in web browsers using SVG, HTML5, and CSS.
- **Why it helps your project:** Perfect for **Module 3's Hierarchical Visualization**. While Plotly handles standard charts, D3.js allows for highly customized **Sunburst or Treemap** charts that are necessary to represent the deep, nested layers of the MITRE ATT&CK framework (Tactics → Techniques → Sub-techniques).

Data Engineering & Backend Stack

8. Apache Kafka

- **Description:** A distributed streaming platform that can publish, subscribe to, store, and process streams of records in real-time.
- **Why it helps your project:** Critical for **Module 1 (Data Acquisition)**. If your dashboard aims for "immediate understanding," Kafka acts as the ingestion layer that feeds live attack logs into your dashboard pipeline without losing data during high-traffic "spikes."

9. Polars (DataFrame Library)

- **Description:** A blazingly fast DataFrame library for Python and Rust, optimized for multi-threaded processing of massive datasets.
- **Why it helps your project:** It improves the **User Experience (UX)**. By using Polars instead of traditional Pandas, your "Interactive Filtering" (e.g., filtering 5 million logs by a specific CVE) becomes near-instant, ensuring the dashboard remains snappy and responsive for the analyst.

10. Elasticsearch

- **Description:** A distributed, RESTful search and analytics engine capable of addressing a growing number of use cases, particularly log and event data.
- **Why it helps your project:** It serves as the **Database Backend**. It is designed specifically for searching through billions of lines of text (like security logs or CVE descriptions). It enables your dashboard to provide "Executive-ready reports" by aggregating complex data queries in milliseconds.
