SAFE

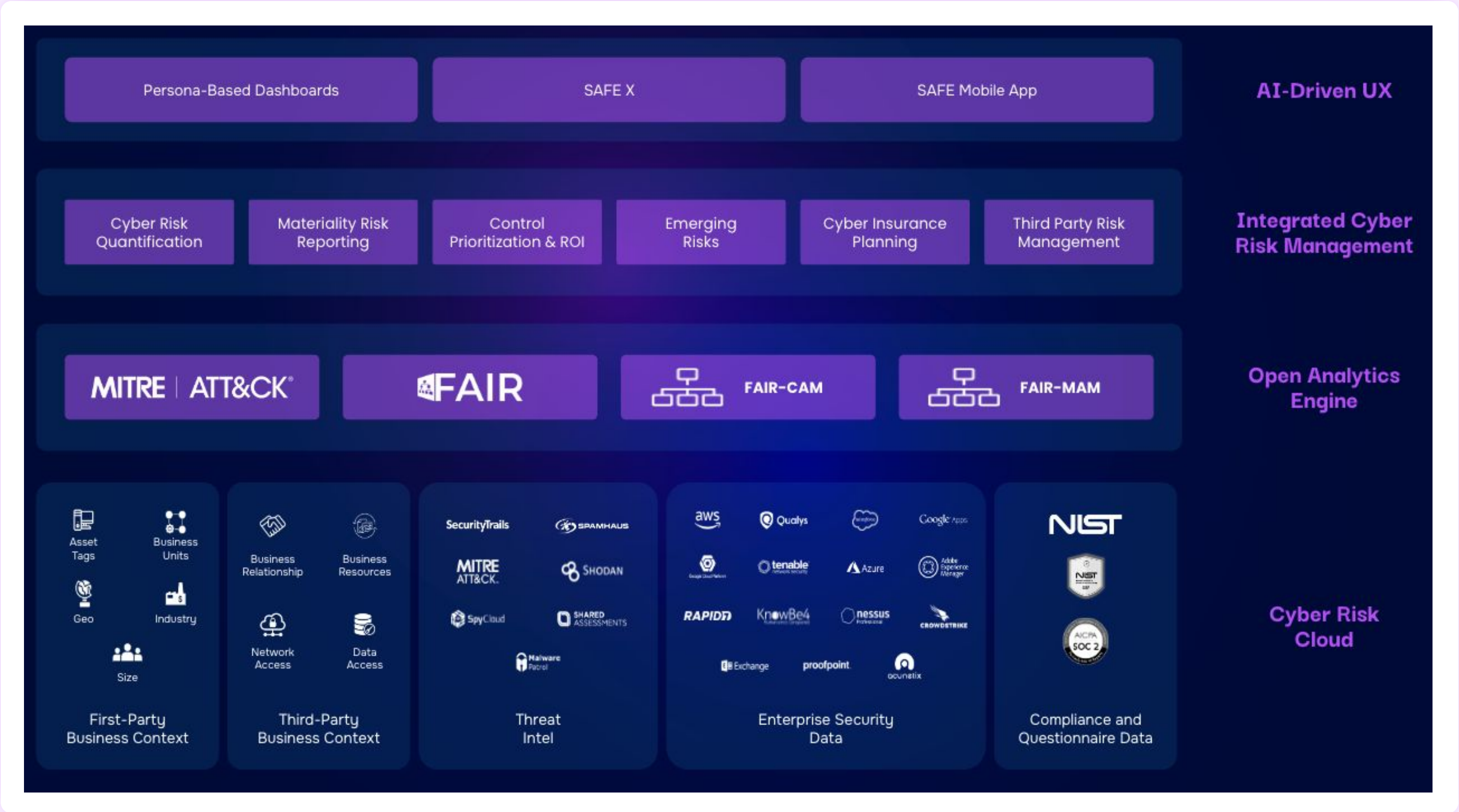
Use of AI in SAFE Third-Party Risk Management

AI Integration and Data Privacy

- **Use of OpenAI Models**: SAFE utilizes OpenAI's Turbo models for preprocessing user queries. This preprocessing step involves structuring the input data to ensure that the questions posed are understood accurately by our AI systems. Importantly, these models do not have access to any customer sensitive information, even in encrypted form.
- **Primary AI Models**: The main AI engine model handles more complex data processing and decision-making tasks. It has the capability to access customer-specific data, but only within a tightly controlled environment on AWS. This ensures that sensitive data remains secure and is never exposed to external threats.
- **Data Handling Protocols**: No customer sensitive data is sent to OpenAI. All customer-specific data is processed and stored within AWS infrastructure, leveraging AWS's robust security measures including data encryption.

AI Model Training and Operations

- **Training Data:** The training of our AI models, particularly the smaller models, utilizes generic data that is not customer-specific. These models are fine-tuned on public documentation and safe terminology, ensuring that the training process does not compromise customer privacy.
- **Operational Use of AI Models:** The primary model interacts with customer data. It performs data extraction and summarization directly within the customer's environment. This model's interaction with data is confined to providing contextually relevant answers based on the prompts generated from the immediate customer query.



AI Enhanced TPRM: Clarifying Your Most Common Queries

☑ **Do you use OPEN AI?**

Yes, OPEN AI is used only for preprocessing user questions without access to any customer sensitive information.

☑ **Which model in OPEN AI do you use?**

The GPT Turbo models are used for internal preprocessing, without access to customer data.

☑ **Do you send customer sensitive data to OPEN AI?**

No, customer sensitive data is not sent to OPEN AI.

☑ **Are you training models?**

Yes, models are being trained, but not with customer data.

☑ **How do you use the trained models?**

Trained models are used for internal processing and do not access customer specific data.

☑ **What corpus of data is the model trained on?**

The main model is used with prompts for answer generation without pre-training on specific data.

☑ **For generating peer insights, do you encrypt and anonymize the data?**

Yes, the data used for generating peer insights are encrypted and anonymized. We don’t aggregate based on customer-specific entities, but rather on general risks, known hacks, and CAM controls.

☑ **Which model has access to customer specific data?**

The AWS model has access to customer-specific data, which is extracted from a specific tenant then summarized by the LLM.

☑ **Can the AI use data of customer x to answer question asked by customer y?**

No, based on the question we first extract the relevant data from a specific tenant then LLM helps in summarization. Therefore, LLM can never access data from tenant X to answer a question asked by tenant Y.

☑ **Have you reviewed the terms of service (including open-source terms) that govern the model?**

Yes, we have reviewed the terms of service for all our models. There is no data retention in any of the models except for "abuse" analysis which any model/service provider would ask for and it is an industry standard.

Security and Compliance

- **Data Isolation**: Data from one customer is never used to inform or answer queries from another. This isolation ensures that each customer's data is handled securely and independently.
- **Prompts and Data Storage:** Prompts used for AI interactions are stored dynamically within SAFE’s code base and services. They are tailored to the specific context of each query, ensuring that only relevant and secure prompts are used. While these prompts are retained for analysis and to prevent abuse, customers can request the removal of specific prompts on a case-by-case basis.
- **Peer Insights and Anonymization**: When generating insights that may involve peer comparisons, the data is encrypted and anonymized. This process ensures that insights are derived from generalized data sets rather than any specific customer data.

Monitoring and Feedback

- **Model Oversight**: SAFE conducts regular health checks and monitors the latency and performance of the AI models. This proactive management helps maintain high standards of operational efficiency and accuracy.
- **Customer Interaction**: Customers can provide feedback on the AI’s performance via SAFE’s support portal. This feedback is crucial for refining the AI’s responses and ensuring that the model remains aligned with user needs and expectations.

Ensuring Trust and Transparency

SAFE is committed to maintaining the highest standards of data privacy and AI security. By utilizing robust AI models and strict data governance protocols, we ensure that our AI-driven TPRM solutions are both effective and secure. Our approach to AI integration respects and protects customer data, while delivering the insights necessary to manage third-party risks effectively.

SAFE ONE

RESEARCH SPONSOR
MITRE ATT&CK, TOP CONTRIBUTOR

PUBLISHER’S CHOICE AWARD RISK MANAGEMENT
GLOBAL INFOSEC AWARDS - RSAC 2023

BEST NEXT GEN CYBER INSURANCE PLATFORM
GLOBAL INFOSEC AWARDS - RSAC 2023

BEST RISK MANAGEMENT SOLUTION
CISO CHOICE AWARDS 2022™

www.safe.security | getintouch@safe.security