FRENOS

CYBERSECURITY VENDOR AI & LLM CLAIMS CHECKLIST

In the rapidly evolving world of cybersecurity, Artificial Intelligence (AI) and Large Language Models (LLMs) have become buzzwords that many vendors are eager to tout. However, not all AI and LLM solutions are created equal, and it's crucial to critically assess the claims made by vendors before making any commitments. We designed this checklist to help you ask the right questions about AI and LLM technologies to make sure they are robust, secure, and truly beneficial to your organization.

This resource will guide you through key concepts like understanding LLM architecture to evaluating its deployment and security practices. For each question, we provide examples of both good and bad answers, along with reasoning to help you identify whether a vendor's claim holds up under scrutiny. This will empower you to make informed decisions, ensuring that the AI and LLM solutions you invest in align with your organization's needs and security standards.

HOW TO USE THIS CHECKLIST

This checklist is your tool for cutting through the hype and making sure that the AI and LLM solutions you choose will deliver on their promises.

01 PREPARE AHEAD

Before meeting with vendors, review this checklist to familiarize yourself with the key topics and questions. It will help you lead the conversation and ensure that all critical areas are covered.

02 ASK THE RIGHT QUESTIONS

During discussions with vendors, use these questions to probe into the details of AI and LLM offerings. Pay attention to how vendors respond—specific, detailed answers indicate a well-thought-out, reliable solution.

03 EVALUATE RESPONSES

Compare the vendor's response to the good and bad examples provided. This will help you assess the quality of the vendor's claims and determine if their AI and LLM solutions meet your standards.

04 DOCUMENT YOUR FINDINGS

Take notes on vendor's responses and how they compare to the this checklist. This document will be invaluable when it comes time to make a final decision, allowing you to compare vendors.

05 CONSULT WITH STAKEHOLDERS

Share your findings with relevant stakeholders in your organization. Use this checklist to guide discussions and ensure that all concerns are addressed before moving forward with a vendor.

06 MAKE INFORMED DECISIONS

Choose a vendor with AI and LLM solutions that are secure, robust, and aligned with your goals. This ensures you're investing in technology that will effectively strengthen your cybersecurity efforts.

FRENOS

THE CHECKLIST

LLM ARCHITECTURE

What architecture does your LLM use? (Transformer, RNN, CNN, etc.)

GOOD ANSWER

"We use a Transformer-based architecture, specifically GPT-like models, which are state-of-the-art for NLP tasks. Transformers excel at handling long-range dependencies in text and are the foundation of most modern LLMs."

Reasoning: Transformers are widely recognized for their effectiveness in NLP tasks, indicating the vendor is using a robust, modern architecture.

BAD ANSWER

"We use a basic RNN architecture because it's easy to implement."

Reasoning: RNNs are less capable of handling complex language tasks compared to Transformers, suggesting the vendor may not be utilizing the best technology available.

LLM AUGMENTATION

How is your LLM augmented? (MoE, Quantization, Retrieval Augmented Generation, Agents, etc.)

GOOD ANSWER

"Our model is augmented with Retrieval Augmented Generation (RAG) and Mixture of Experts (MoE). RAG allows the model to retrieve relevant information from a knowledge base, improving accuracy, while MoE enables efficient scaling without compromising performance."

Reasoning: These augmentations are advanced techniques that enhance the model's capabilities, indicating that the vendor is serious about optimizing their LLM.

BAD ANSWER

"We don't use any augmentations because the base model is sufficient."

Reasoning: This response may suggest a lack of innovation or an unwillingness to invest in enhancing the model's capabilities, which could limit the LLM's effectiveness.

LLM TRAINING

How many parameters does the model have?

GOOD ANSWER

"Our model has 13 billion parameters, striking a balance between performance and resource efficiency. This size is optimal for our target use cases."

Reasoning: The vendor provides a specific number and justification, showing they have thoughtfully designed their model.

BAD ANSWER

"The model size doesn't matter; what matter is how it works."

Reasoning: While true to some extent, this vague answer avoids the specifics, possibly indicating a lack of transparency or understanding.

How many days did it take to train the model?

GOOD ANSWER

"It took 30 days to train our model on 8 A100 GPUs, ensuring that the model was thoroughly trained with high-quality data."

Reasoning: A specific timeframe and resource details indicate careful training, contributing to the model's reliability.

BAD ANSWER

"We trained it quickly to get it to market faster."

Reasoning: Rushing training can lead to poor model performance and overlooked issues, suggesting a focus on speed over quality.

How do you fine-tune your LLM?

GOOD ANSWER

"We fine-tune the model using domain-specific data to ensure it is tailored to the unique needs of our clients."

Reasoning: This shows that the vendor customizes the model for different applications, increasing its relevance and effectiveness.

BAD ANSWER

"We don't do much fine-tuning; the base model should be good enough."

Reasoning: Lack of fine-tuning may lead to a less accurate or relevant model for specific tasks, indicating a lower commitment to quality.

FRENOS

LLM TRAINING CONT'D

Did you need to do anything special to train the model?

GOOD ANSWER

"Yes, we implemented gradient accumulation and mixed precision training to efficiently train large batches without overwhelming the hardware."

Reasoning: Special training techniques often indicate a sophisticated approach to model development, reflecting positively on the vendor's expertise.

BAD ANSWER

"No, we just used the standard settings."

Reasoning: A lack of special measures may suggest a lack of innovation or optimization, which could impact the model's performance.

ML AUGMENTATION

What machine learning techniques do you use to augment your LLM?

(Regression, Clustering, MCTS, MuZero, AlphaMu, Ensemble, etc.)

GOOD ANSWER

"We use a combination of Monte Carlo Tree Search (MCTS) and Ensemble methods to enhance decision-making and prediction accuracy."

Reasoning: The use of advanced techniques indicates a well-rounded and robust AI solution, reflecting the vendor's commitment to delivering high-quality outcomes.

BAD ANSWER

"We rely solely on the LLM with no additional ML techniques."

Reasoning: While LLMs are powerful, augmentation with other ML techniques often leads to better performance, so this response may indicate a less comprehensive approach.

DEPLOYMENT

Is the model deployed on-premise or in the cloud?

GOOD ANSWER

"We offer both on-premise and cloud deployment options to meet different client needs. On-premise ensures data control, while cloud deployment provides scalability."

Reasoning: Offering both options demonstrates flexibility and an understanding of different client requirements.

BAD ANSWER

"We only offer cloud deployment because it's easier for us."

Reasoning: A single deployment option may not meet all clients' needs and suggests a lack of flexibility.

If on-premise, how many resources do I need to run this?

GOOD ANSWER

"You would need at least 4 GPUs with 64GB of RAM each to run our model effectively on-premise."

Reasoning: A specific and detailed resource requirement helps the client understand the feasibility of running the model on-premise.

BAD ANSWER

"We're not sure; it depends on your setup."

Reasoning: This vague answer shows a lack of preparation or understanding of the model's deployment requirements, which could lead to operational challenges.

If in the cloud, how do you gate the model against unauthorized access?

GOOD ANSWER

"We implement multi-factor authentication and role-based access control, along with continuous monitoring to protect against unauthorized access."

Reasoning: Strong security measures indicate a commitment to protecting the client's data and maintaining the integrity of the AI model.

BAD ANSWER

"We use basic password protection."

Reasoning: Basic security measures may be insufficient for protecting sensitive data, indicating a potential risk for the client.

FRENOS

SECURITY AND PRIVACY

Can the LLM be accessed directly?

GOOD ANSWER

"No, we have implemented a gated API that controls access to the LLM, ensuring that only authorized users can interact with it."

Reasoning: Controlled access is crucial for security, showing that the vendor prioritizes safeguarding their AI.

BAD ANSWER

"Yes, anyone can access it if they have the link."

Reasoning: Unrestricted access poses significant security risks, making this a poor practice.

Do you have safeguards to filter out harmful questions and responses?

GOOD ANSWER

"Yes, we have implemented content filtering and moderation tools to ensure that harmful or inappropriate queries are handled appropriately."

Reasoning: Safeguards are essential to prevent misuse of the AI, indicating responsible AI development practices.

BAD ANSWER

"We trust the users to ask appropriate questions."

Reasoning: This hands-off approach can lead to the generation of harmful or biased content, posing ethical

Do you need access to my data for any LLM or ML Augmentation?

GOOD ANSWER

"We only require access to anonymized data to fine-tune the model specifically for your use case."

Reasoning: Limited data access, especially anonymized, shows respect for client privacy while ensuring the model's effectiveness.

BAD ANSWER

"We need full access to all your data for training."

Reasoning: This could lead to unnecessary exposure of sensitive information, posing significant privacy risks.

What do you store?

GOOD ANSWER

"We store only the metadata necessary for model performance optimization, with no personal data retained."

Reasoning: Storing minimal data reduces privacy risks and shows the vendor's commitment to protecting client information.

BAD ANSWER

"We store everything indefinitely, just in case."

Reasoning: Indiscriminate data storage increases privacy risks and regulatory compliance concerns, making this a bad practice.

What's the retention period on that data?

GOOD ANSWER

"We retain data for 30 days, after which it is securely deleted, unless otherwise required by regulation."

Reasoning: A clear retention policy ensures data is not kept longer than necessary, reducing privacy risks.

BAD ANSWER

"We haven't set a specific retention period."

Reasoning: Lack of a clear retention policy can lead to unnecessary data retention, increasing the risk of data breaches or misuse.

What data of mine do you use to fine-tune your model?

GOOD ANSWER

"We use only anonymized and aggregated data for fine-tuning, ensuring that individual data points cannot be traced back to specific users."

Reasoning: Using anonymized data respects user privacy while still enabling model improvement.

BAD ANSWER

"We use all available data to get the best results."

Reasoning: This broad approach could lead to privacy violations, particularly if personal