Artificial Intelligence (AI) Review

ABOUT THE CHECKLIST

This checklist was created to help you review vendors that offer artificial intelligence (AI) systems or that integrate AI functions into their products and services. Some vendors may make these kinds of information available on their websites or in due diligence packets. Other vendors may need to be contacted directly. To learn more about third-party risk management, visit our website: Tandem.App.

GENERAL INFORMATION

What purpose does the AI serve as part of this vendor service?

AI systems have a broad range of top-level functions. It is important to document what this system does for the organization to help identify and mitigate the risks associated with it.

- [ ] Content generation (e.g., text, images, audio, video, etc.)
- [ ] Data analysis (e.g., statistical analysis, anomaly detection, fraud detection, etc.)
- [ ] Insight generation (e.g., automated valuation models (AVM), predictive modeling, forecasting, etc.)
- [ ] Business automation (e.g., robotic process automation (RPA), task automation, data entry, etc.)
- [ ] Natural language process (e.g., document summarization, translation, sentiment analysis, etc.)
- [ ] Security functions (e.g., system monitoring, anti-malware, endpoint detection and response (EDR), etc.)
- [ ] Software development (e.g., code authoring, review, optimization, debugging, etc.)
- [ ] Customer support (e.g., chatbots, ticket classification, etc.)
- [ ] It does something else (see comments)

COMMENTS

Tandem

Have roles and responsibilities been defined to avoid potential conflicts of interest?

There are several potential points of conflict which may exist when implementing a new AI system (e.g., if the person implementing the system is also responsible for validating it, if the vendor stands to benefit from the organization's use of the system, etc.). Because of the potential for conflicts of interest, it is important to clearly define roles and responsibilities with fairness, accuracy, and accountability in mind.

- ○ Yes, roles and responsibilities have been clearly defined
- ○ No, roles and responsibilities have not yet been defined

COMMENTS

MODEL RISK MANAGEMENT

Which of the following best describes the vendor's AI model?

Assessing the vendor's AI model is important for understanding things like control over the system, resource demands, security functions, and strategic alignment. While no one model is inherently better than another, each does present unique risks and opportunities that need to be identified and managed appropriately.

- ☐ Proprietary, internally-developed model
- ☐ Proprietary, third-party-developed model
- ☐ Open-source model
- ☐ Hybrid model
- ☐ Other (see comments)
- ☐ I don't know

COMMENTS

How is the AI model trained?

Understanding how the vendor's AI model is trained is essential for ensuring accuracy, fairness, transparency, and overall effectiveness. When a vendor is willing to be transparent about how the model is trained, this helps build trust, improves security, and ensures compliance with legal and regulatory requirements (e.g., security regulations, privacy laws, fair lending laws, etc.).

- ☐ It is pre-trained on large datasets
- ☐ It is trained and/or fine-tuned based on specific datasets
- ☐ It is continuously trained based on user input
- ☐ It is trained on data that has been anonymized
- ☐ It is designed to learn from its own outputs
- ☐ It is trained in another way (see comments)
- ☐ I don't know

COMMENTS

Tandem

Can the organization opt out of participating in the AI model's training?

As mentioned in the previous topic, some AI models learn from user input. Sometimes, the vendor offers an "opt out" option. Other times, data is collected automatically and no opt out option exists. How the vendor plans to collect and use the organization's data should be disclosed in a privacy policy.

- ○ Yes, we can fully opt out
- ○ Yes, we can opt out of sharing certain types of data
- ○ No, we cannot opt out
- ○ I don't know

COMMENTS

Which of the following controls has the vendor implemented to protect the AI model from biased, malicious, and/or unauthorized input?

AI models are susceptible to a variety of threats (e.g., scripting, prompt injection, training data manipulation, targeted poisoning, backdooring, etc.). If the vendor has not implemented controls to prevent, detect, and respond to these threats, this can lead to inaccuracies, exploited vulnerabilities, and regulatory noncompliance.

- ☐ Data sanitization to exclude biased, harmful, irrelevant, and/or proprietary data from training datasets
- ☐ Input validation to confirm data is consistent and accurate
- ☐ Anomaly detection to detect unusual or suspicious input patterns
- ☐ Quality assurance and testing to validate the model's preparedness for malicious inputs
- ☐ Authentication and access controls to limit who can modify or input data into the model
- ☐ Ongoing monitoring to continuously watch for signs of unexpected behavior, drift, or degradation
- ☐ Human oversight to exercise independent judgement and handle exceptions
- ☐ Staff training to promote awareness of risks and best practices for maintaining AI model integrity
- ☐ Other (see comments)
- ☐ None of the above

COMMENTS

Which of the following methods are used to validate the AI model?

There is no one right way to validate an AI model. Like most risk management techniques, a layered approach is often best. AI model validation can be performed by the vendor and/or the organization. While the vendor may handle the more complex technical validation and can provide certifications for that, ultimate responsibility for output accuracy ultimately remains with the organization.

- [ ] Professional review (e.g., an expert assesses the results for accuracy)
- [ ] Historical comparison (e.g., comparing current results with previous results)
- [ ] Model benchmarking (e.g., comparing outputs from multiple models)
- [ ] Confidence scores (e.g., ensuring outputs fall into an approved confidence range)
- [ ] Outlier detection (e.g., alerting of outputs which fall outside an approved range)
- [ ] Random sampling (e.g., selecting a subset of data from a larger dataset to validate effectiveness)
- [ ] Other (see comments)
- [ ] None of the above

COMMENTS

How often is AI model validation performed?

For many applications, performing model validation on a regular schedule (e.g., monthly or quarterly) is sufficient. If the underlying data changes frequently (e.g., real-time data streams, rapidly evolving user behavior, etc.), the model may need more frequent validation (e.g., weekly or even daily). For critical systems, more frequent validation is crucial to ensure the model remains reliable and accurate. Additionally, it is important to validate models after significant updates to ensure changes have not adversely affected performance.

- ○ Annually
- ○ Semiannually
- ○ Quarterly
- ○ Monthly
- ○ Weekly
- ○ Daily
- ○ Other (see comments)
- ○ I don't know

COMMENTS

SECURITY & TECHNOLOGY

Is proprietary organization and/or customer information allowed to be input into the AI system?

While it is important to be aware of the risks associated with putting proprietary information into any system, there may be scenarios in which it is appropriate and approved (e.g., if the system is hosted internally, if appropriate controls are in place, etc.). List the allowed types of proprietary information in the comments.

- ◯ Yes (see comments)
- ◯ No

COMMENTS

If proprietary information is not allowed to be input into the system, how is this communicated to employees?

If the organization has prohibited proprietary information from being put into the AI system, this must be communicated to employees in a manner that is clear and understood. Like most training techniques, a layered approach is often best.

- ☐ Initial and ongoing system training
- ☐ Acceptable use policies (AUP)
- ☐ Security awareness training (SAT)
- ☐ Nondisclosure agreements (NDA)
- ☐ System instructions / manuals
- ☐ Other (see comments)
- ☐ N/A – Proprietary information is allowed

COMMENTS

How is the AI system and its data hosted?

Understanding the hosting infrastructure of an AI system is important for evaluating things such as control over the system, required resources, and security configurations. Each type of hosting infrastructure has distinct risks that need managed.

- ◯ It is hosted locally (a.k.a., on premises)
- ◯ It is hosted in a private cloud
- ◯ It is hosted in a community cloud
- ◯ It is hosted in a public cloud
- ◯ Other (see comments)
- ◯ I don't know

COMMENTS

Tandem

Which of the following security controls are in place to protect the AI system?

It is important to protect AI systems both physically and technically. The following is not a comprehensive list of controls, but rather is intended to provide a starting point for you to consider when evaluating the vendor's security practices and available controls.

- [ ] Authentication and access controls to limit who can administer, use, and view data generated by the AI
- [ ] Anti-malware controls to protect against malicious attacks
- [ ] Monitoring and log management to oversee modifications and use of the system
- [ ] Encryption of data at-rest and in-transit to protect information from unauthorized access
- [ ] Data backups of critical data and configurations to protect against data loss
- [ ] Vulnerability and patch management to keep systems up to date and secure
- [ ] Change and configuration management to prevent unauthorized changes to systems
- [ ] Data loss prevention (DLP) controls to monitor for and prevent leakage of proprietary data
- [ ] Environmental controls to ensure business continuity and resilience
- [ ] Other (see comments)
- [ ] None of the above

COMMENTS