Data Governance in the Age of AI

Addressing Vendor Misuse and Protecting Your Assets

11 Apr 2024

Coverbase

By the end of this session, you'll get...

Contract clauses

To protect yourself against errant AI vendors

A free tool

To keep track of which vendors are *AI vendors*

But more importantly, become the AI security expert on your team

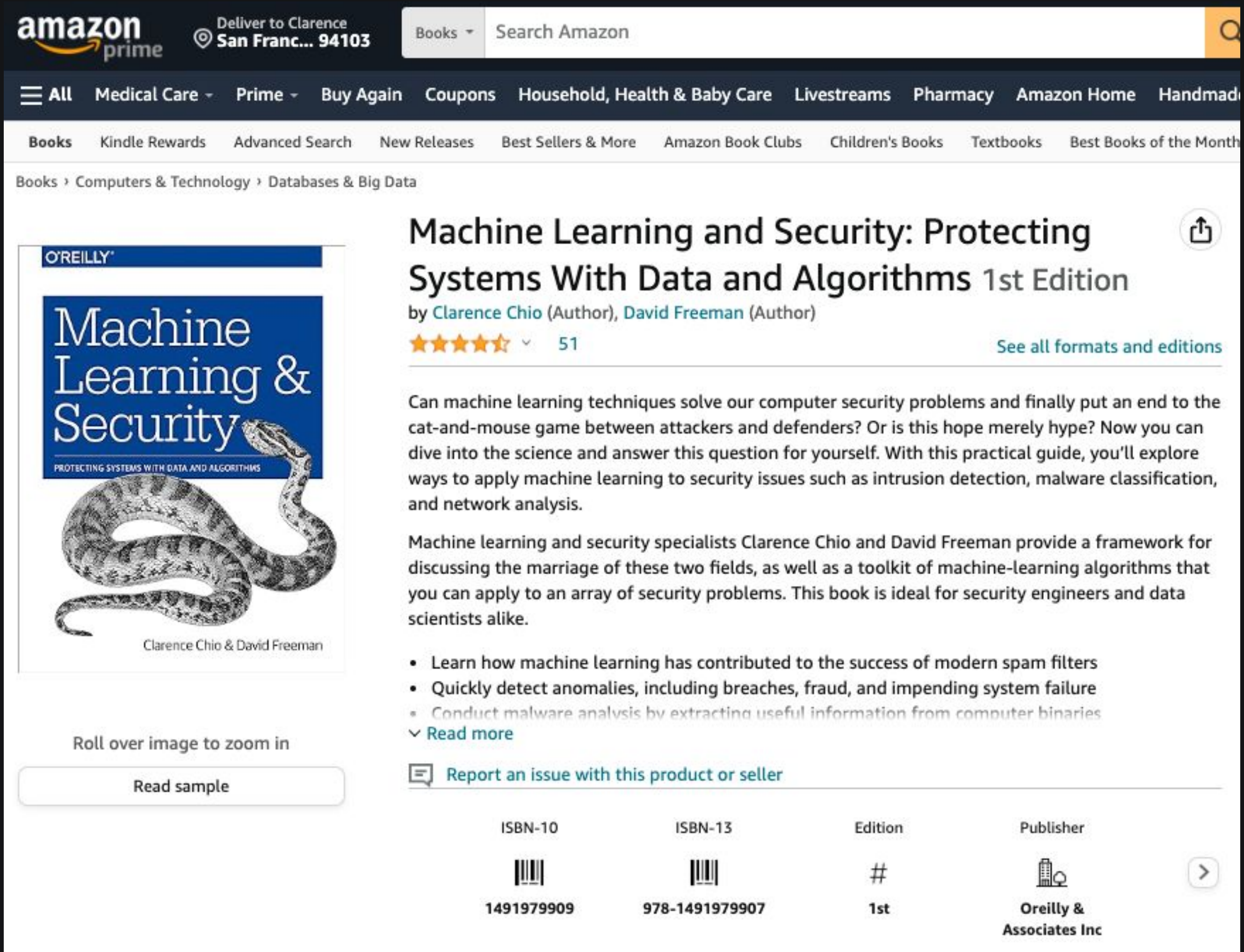
Who are these guys??

Clarence Chio

- Co-founder & CTO, Unit21
  Risk / compliance, raised $92m, $700m val
- Applied ML Lecturer at UC Berkeley
- Author, “Machine Learning and Security” (O’Reilly, 2018)

Kao Zi Chong

- Early engineer, Stripe Issuing
  Grew the product from $30m → $1.1b GMV
- Eng manager, Stripe BaaS Onboarding & Compliance
- Tech lead, Quora topic labeling & entity recognition



Books › Computers & Technology › Databases & Big Data

Machine Learning and Security: Protecting Systems With Data and Algorithms 1st Edition

by Clarence Chio (Author), David Freeman (Author)

★★★★☆ 51

Can machine learning techniques solve our computer security problems and finally put an end to the cat-and-mouse game between attackers and defenders? Or is this hope merely hype? Now you can dive into the science and answer this question for yourself. With this practical guide, you'll explore ways to apply machine learning to security issues such as intrusion detection, malware classification, and network analysis.

Machine learning and security specialists Clarence Chio and David Freeman provide a framework for discussing the marriage of these two fields, as well as a toolkit of machine-learning algorithms that you can apply to an array of security problems. This book is ideal for security engineers and data scientists alike.

- Learn how machine learning has contributed to the success of modern spam filters
- Quickly detect anomalies, including breaches, fraud, and impending system failure
- Conduct malware analysis by extracting useful information from computer binaries

ISBN-10	ISBN-13	Edition	Publisher
1491979909	978-1491979907	1st	Oreilly & Associates Inc

Worries that you have

Is my AI vendor using my data to train their models?

Is my AI vendor co-mingling my data with my competitor's data?

Is my AI vendor encrypting my data before using it for inference or training?

Is my AI vendor's product's results reproducible and auditable?

Is my AI vendor using my data in a way that puts me out of compliance?

Is my AI vendor's product tested for resilience against vulnerabilities?

Is my AI vendor's product audited for bias?

Is my AI vendor going to expose me to reputation risk?

Is my AI vendor's product going to hallucinate?

Is my AI vendor sending my data to 4th parties and subprocessors?

Is my AI vendor's product adequately protected against insider threats?

Is my AI vendor protecting themselves against data leakage?

Is my AI vendor's models susceptible to undetected model drift?

Is my AI vendor's underlying AI infrastructure hosted on infrastructure they operate?

...

Worries that you have + addressed by current TPRM processes

Is my AI vendor using my data to train their models?

Is my AI vendor co-mingling my data with my competitor’s data?

Is my AI vendor encrypting my data before using it for inference or training?

Is my AI vendor’s product’s results reproducible and auditable?

Is my AI vendor using my data in a way that puts me out of compliance?

Is my AI vendor’s product tested for resilience against vulnerabilities?

Is my AI vendor’s product audited for bias?

Is my AI vendor going to expose me to reputation risk?

Is my AI vendor’s product going to hallucinate?

Is my AI vendor sending my data to 4th parties and subprocessors?

Is my AI vendor’s product adequately protected against insider threats?

Is my AI vendor protecting themselves against data leakage?

Is my AI vendor’s models susceptible to undetected model drift?

Is my AI vendor’s underlying AI infrastructure hosted on infrastructure they operate?

...

Hold on... who are these “AI Vendors” anyway?

We don’t have many AI vendors today

Yes you do, and you’ll have more, whether you know / like it or not

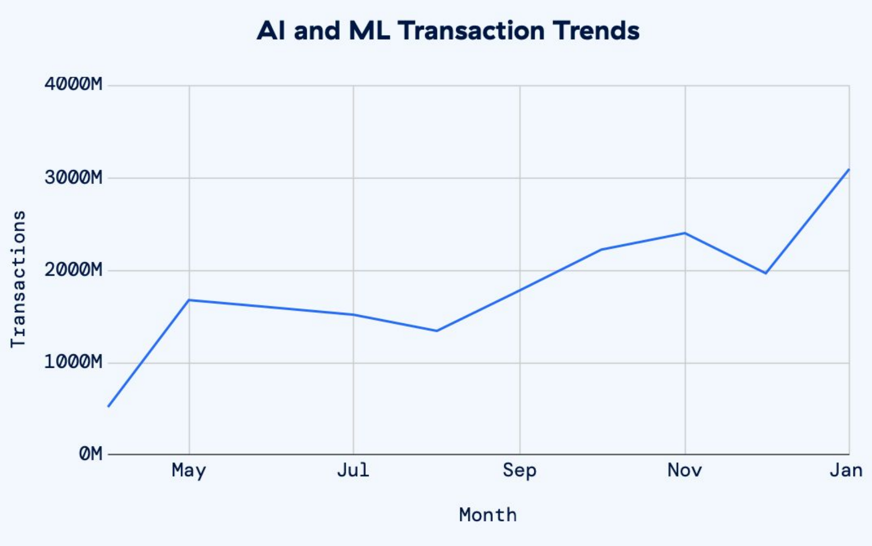


FIGURE 1 AI transactions from April 2023 to January 2024

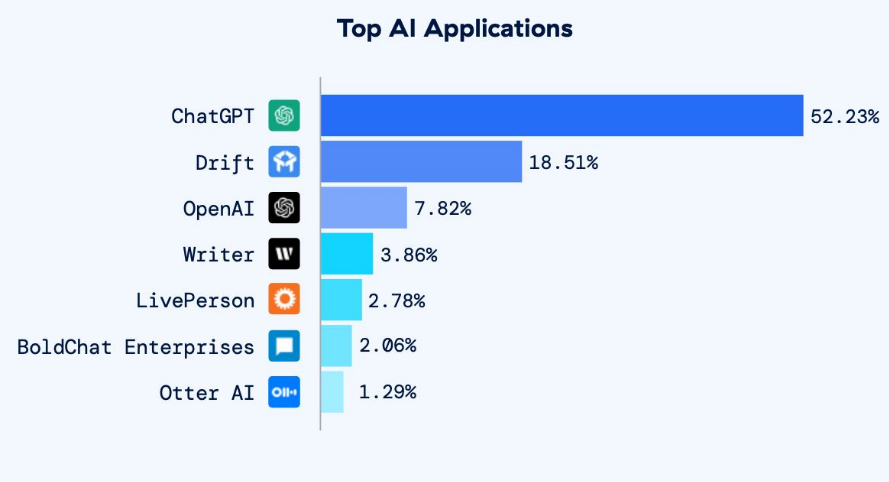


FIGURE 2 Top AI applications by transaction volume

ZSCALER THREATLABZ REPORT 2024

How do you know which vendors are “AI vendors”?

Do all AI vendors misuse data?

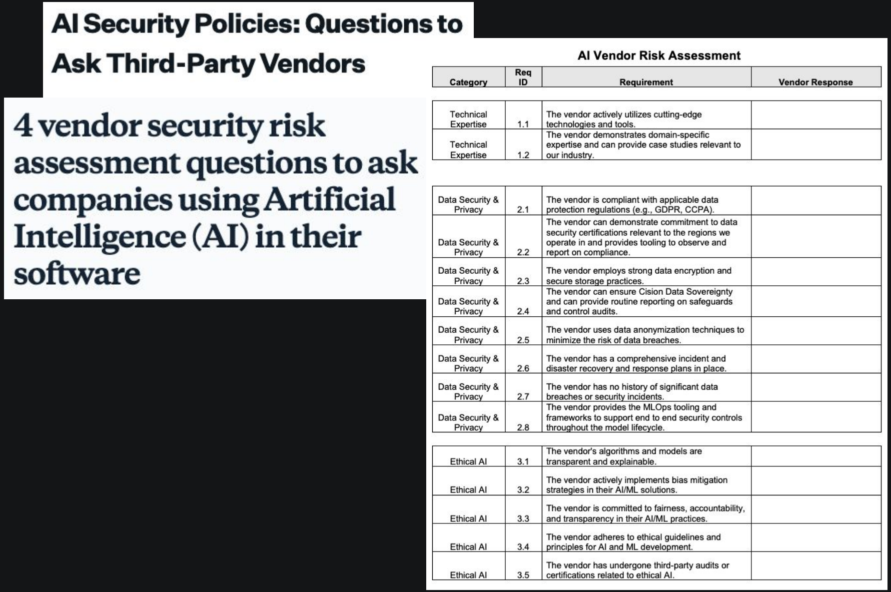
Risks & implications

What have others done about it?

1. Come up with AI vendor policies

Why Your Business Needs a Generative AI Policy
1 Regulation of AI use
Provides guidelines on AI
2 Legal compliance
Helps prevent breaching … related to data …, discrimination, and more
3 Bias reduction
Addresses potential AI biases, which could lead to discriminatory outcomes
4 Data protection
Reduces risks of exposing or mishandling confidential information
5 Human-AI balance
Promotes the importance of human judgment complemented with AI
…ons
…r …AI
7 Quality assurance
Encourages employee evaluations of AI outputs to maintain quality
8 Accountability issues
Defines responsibilities for AI-related mistakes and disputes
9 Employee trust
Communicates the purpose of AI, mitigating potential fears of replacement
AIHR ACADEMY TO INNOVATE HR
AI
CORPORATE
POLICY

2. Issue a new questionnaire



AI Security Policies: Questions to Ask Third-Party Vendors

4 vendor security risk assessment questions to ask companies using Artificial Intelligence (AI) in their software

AI Vendor Risk Assessment

Category	Req ID	Requirement	Vendor Response
Technical Expertise	1.1	The vendor actively utilizes cutting-edge technologies and tools.	
Technical Expertise	1.2	The vendor demonstrates domain-specific expertise and can provide case studies relevant to our industry.	
Data Security & Privacy	2.1	The vendor is compliant with applicable data protection regulations (e.g., GDPR, CCPA).	
Data Security & Privacy	2.2	The vendor can demonstrate commitment to data security certifications relevant to the regions we operate in and provides tooling to observe and report on compliance.	
Data Security & Privacy	2.3	The vendor employs strong data encryption and secure storage practices.	
Data Security & Privacy	2.4	The vendor can ensure Cision Data Sovereignty and can provide routine reporting on safeguards and control audits.	
Data Security & Privacy	2.5	The vendor uses data anonymization techniques to minimize the risk of data breaches.	
Data Security & Privacy	2.6	The vendor has a comprehensive incident and disaster recovery and response plans in place.	
Data Security & Privacy	2.7	The vendor has no history of significant data breaches or security incidents.	
Data Security & Privacy	2.8	The vendor provides the MLOps tooling and frameworks to support end to end security controls throughout the model lifecycle.	
Ethical AI	3.1	The vendor's algorithms and models are transparent and explainable.	
Ethical AI	3.2	The vendor actively implements bias mitigation strategies in their AI/ML solutions.	
Ethical AI	3.3	The vendor is committed to fairness, accountability, and transparency in their AI/ML practices.	
Ethical AI	3.4	The vendor adheres to ethical guidelines and principles for AI and ML development.	
Ethical AI	3.5	The vendor has undergone third-party audits or certifications related to ethical AI.	

What have others done about it?

3. Block AI vendors

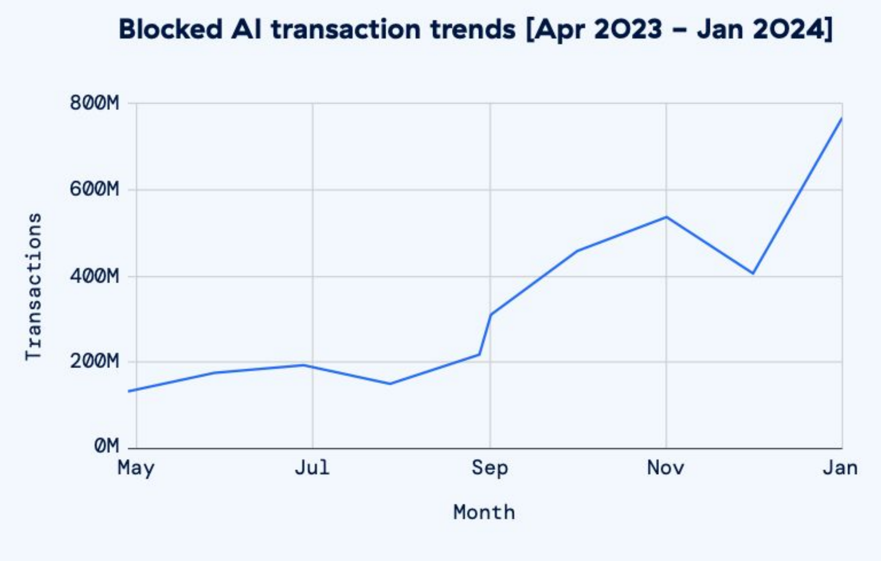
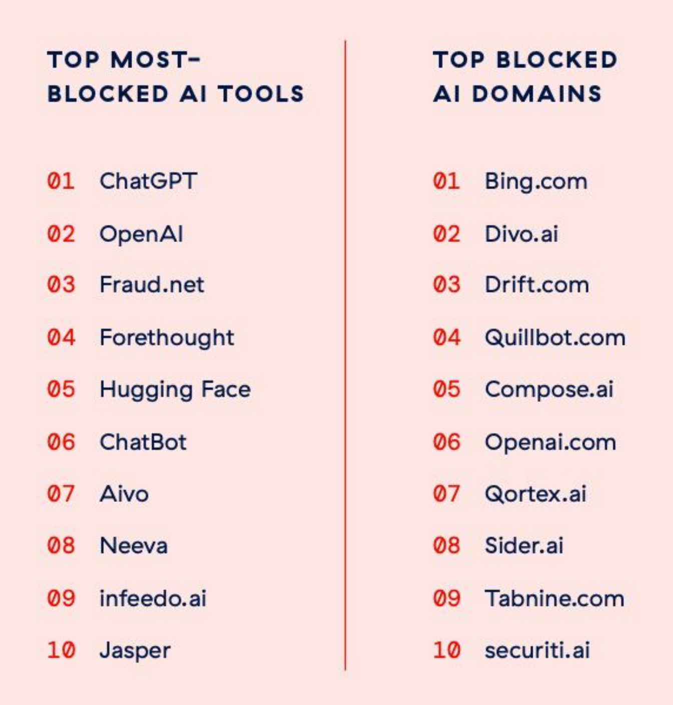


FIGURE 4 Number of AI/ML transactions blocked over time



TOP MOST-BLOCKED AI TOOLS		TOP BLOCKED AI DOMAINS	
01	ChatGPT	01	Bing.com
02	OpenAI	02	Divo.ai
03	Fraud.net	03	Drift.com
04	Forethought	04	Quillbot.com
05	Hugging Face	05	Compose.ai
06	ChatBot	06	Openai.com
07	Aivo	07	Qortex.ai
08	Neeva	08	Sider.ai
09	infeedo.ai	09	Tabnine.com
10	Jasper	10	securiti.ai

But really though... what should you do about it?

Balancing innovation & security

Coverbase

1

Know your stand

Develop your AI vendor policies,
assessment procedures & requirements

2

Develop Contractual Safeguards

Build in legal protections to show you mean it

AI vendor risk mitigation strategy

3

Scope your the problem

Keep track your AI vendors,
put them through appropriate assessments

4

Trust, but verify

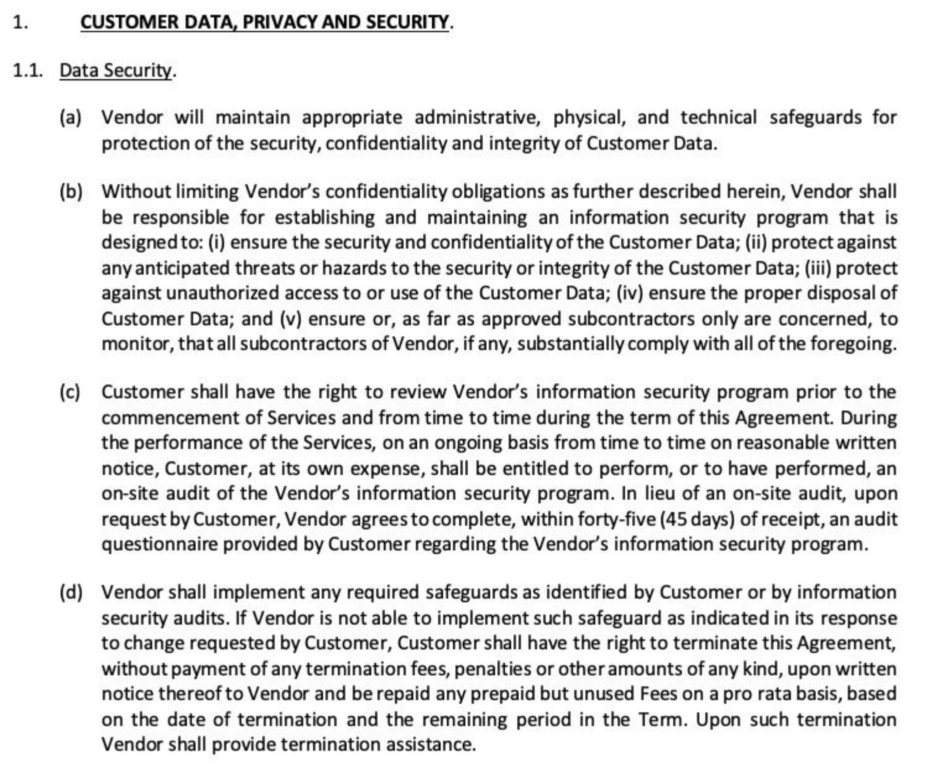
Don’t just take their word for it,
attest to their proper handling of your data.

1. Know your stand

Develop your AI vendor policies, assessment procedures & requirements

2. Develop Contractual Safeguards

Build in legal protections to show you mean it



1. **CUSTOMER DATA, PRIVACY AND SECURITY**.

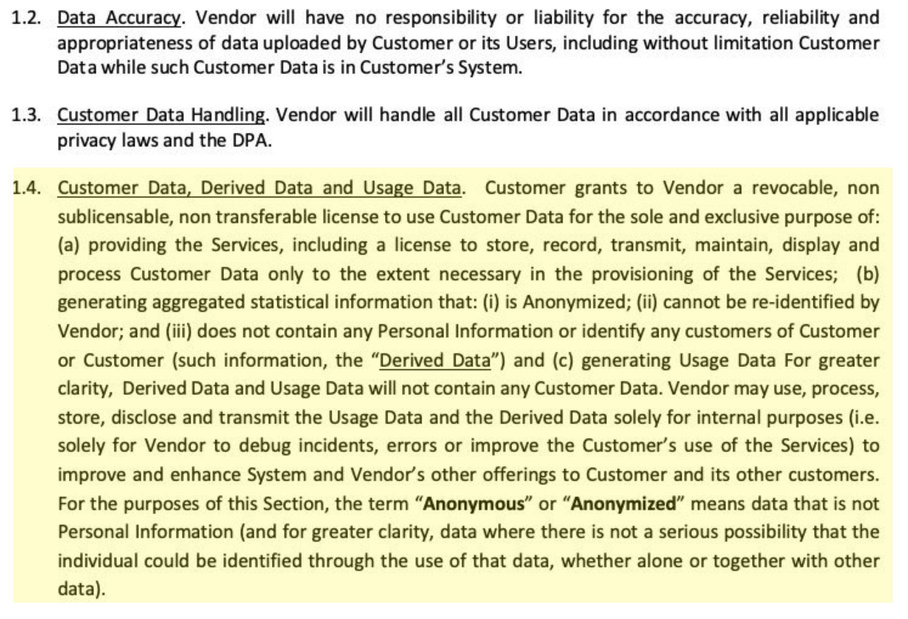
1.1. Data Security.

(a) Vendor will maintain appropriate administrative, physical, and technical safeguards for protection of the security, confidentiality and integrity of Customer Data.

(b) Without limiting Vendor's confidentiality obligations as further described herein, Vendor shall be responsible for establishing and maintaining an information security program that is designed to: (i) ensure the security and confidentiality of the Customer Data; (ii) protect against any anticipated threats or hazards to the security or integrity of the Customer Data; (iii) protect against unauthorized access to or use of the Customer Data; (iv) ensure the proper disposal of Customer Data; and (v) ensure or, as far as approved subcontractors only are concerned, to monitor, that all subcontractors of Vendor, if any, substantially comply with all of the foregoing.

(c) Customer shall have the right to review Vendor's information security program prior to the commencement of Services and from time to time during the term of this Agreement. During the performance of the Services, on an ongoing basis from time to time on reasonable written notice, Customer, at its own expense, shall be entitled to perform, or to have performed, an on-site audit of the Vendor's information security program. In lieu of an on-site audit, upon request by Customer, Vendor agrees to complete, within forty-five (45 days) of receipt, an audit questionnaire provided by Customer regarding the Vendor's information security program.

(d) Vendor shall implement any required safeguards as identified by Customer or by information security audits. If Vendor is not able to implement such safeguard as indicated in its response to change requested by Customer, Customer shall have the right to terminate this Agreement, without payment of any termination fees, penalties or other amounts of any kind, upon written notice thereof to Vendor and be repaid any prepaid but unused Fees on a pro rata basis, based on the date of termination and the remaining period in the Term. Upon such termination Vendor shall provide termination assistance.



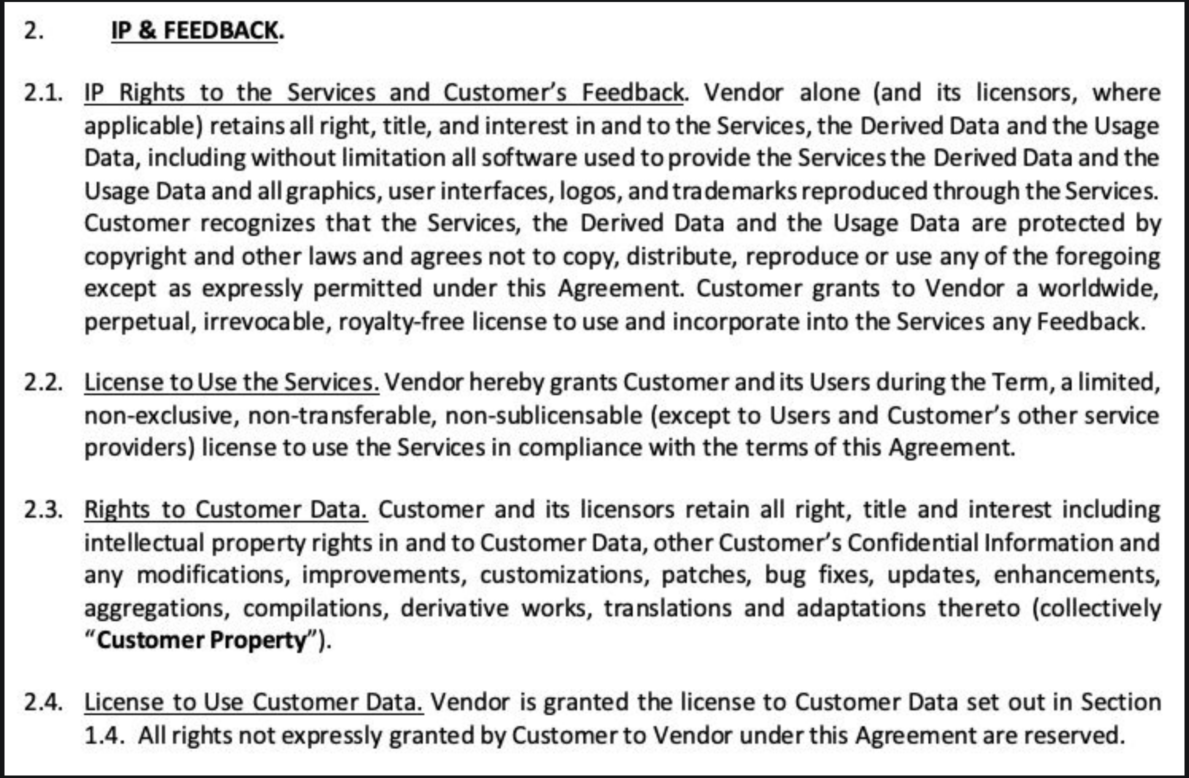
1.2. Data Accuracy. Vendor will have no responsibility or liability for the accuracy, reliability and appropriateness of data uploaded by Customer or its Users, including without limitation Customer Data while such Customer Data is in Customer's System.

1.3. Customer Data Handling. Vendor will handle all Customer Data in accordance with all applicable privacy laws and the DPA.

1.4. Customer Data, Derived Data and Usage Data. Customer grants to Vendor a revocable, non sublicensable, non transferable license to use Customer Data for the sole and exclusive purpose of: (a) providing the Services, including a license to store, record, transmit, maintain, display and process Customer Data only to the extent necessary in the provisioning of the Services; (b) generating aggregated statistical information that: (i) is Anonymized; (ii) cannot be re-identified by Vendor; and (iii) does not contain any Personal Information or identify any customers of Customer or Customer (such information, the "Derived Data") and (c) generating Usage Data For greater clarity, Derived Data and Usage Data will not contain any Customer Data. Vendor may use, process, store, disclose and transmit the Usage Data and the Derived Data solely for internal purposes (i.e. solely for Vendor to debug incidents, errors or improve the Customer's use of the Services) to improve and enhance System and Vendor's other offerings to Customer and its other customers. For the purposes of this Section, the term "**Anonymous**" or "**Anonymized**" means data that is not Personal Information (and for greater clarity, data where there is not a serious possibility that the individual could be identified through the use of that data, whether alone or together with other data).

2 Develop Contractual Safeguards

Build in legal protections to show you mean it



2. **IP & FEEDBACK.**

2.1. IP Rights to the Services and Customer's Feedback. Vendor alone (and its licensors, where applicable) retains all right, title, and interest in and to the Services, the Derived Data and the Usage Data, including without limitation all software used to provide the Services the Derived Data and the Usage Data and all graphics, user interfaces, logos, and trademarks reproduced through the Services. Customer recognizes that the Services, the Derived Data and the Usage Data are protected by copyright and other laws and agrees not to copy, distribute, reproduce or use any of the foregoing except as expressly permitted under this Agreement. Customer grants to Vendor a worldwide, perpetual, irrevocable, royalty-free license to use and incorporate into the Services any Feedback.

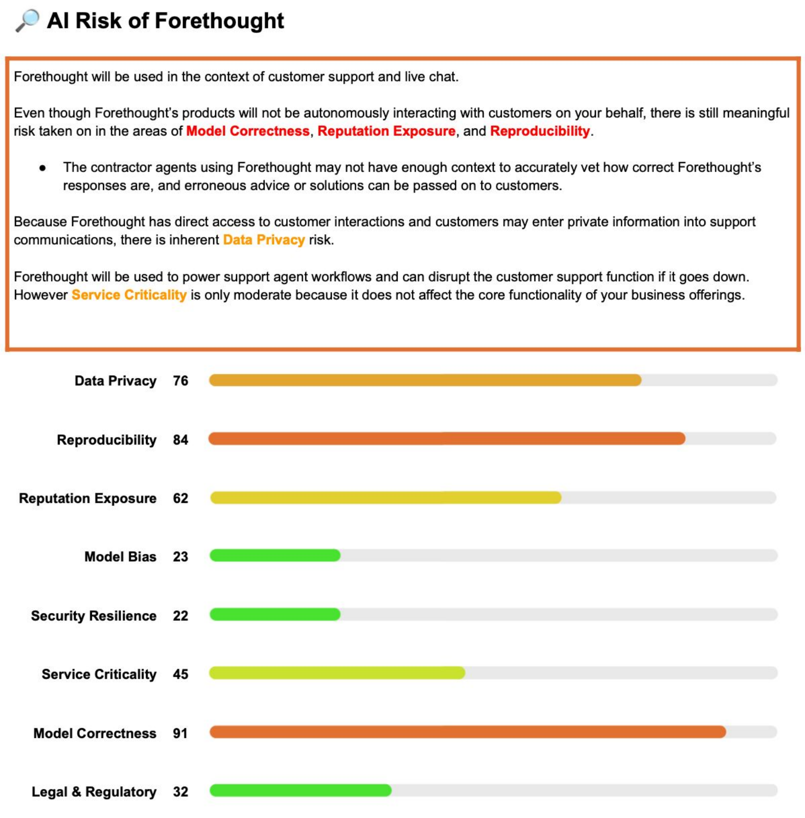
2.2. License to Use the Services. Vendor hereby grants Customer and its Users during the Term, a limited, non-exclusive, non-transferable, non-sublicensable (except to Users and Customer's other service providers) license to use the Services in compliance with the terms of this Agreement.

2.3. Rights to Customer Data. Customer and its licensors retain all right, title and interest including intellectual property rights in and to Customer Data, other Customer's Confidential Information and any modifications, improvements, customizations, patches, bug fixes, updates, enhancements, aggregations, compilations, derivative works, translations and adaptations thereto (collectively "**Customer Property**").

2.4. License to Use Customer Data. Vendor is granted the license to Customer Data set out in Section 1.4. All rights not expressly granted by Customer to Vendor under this Agreement are reserved.

3 Scope your the problem

Keep track your AI vendors, put them through appropriate assessments



🔎 AI Risk of Forethought

Forethought will be used in the context of customer support and live chat.

Even though Forethought's products will not be autonomously interacting with customers on your behalf, there is still meaningful risk taken on in the areas of **Model Correctness**, **Reputation Exposure**, and **Reproducibility**.

- The contractor agents using Forethought may not have enough context to accurately vet how correct Forethought's responses are, and erroneous advice or solutions can be passed on to customers.

Because Forethought has direct access to customer interactions and customers may enter private information into support communications, there is inherent **Data Privacy** risk.

Forethought will be used to power support agent workflows and can disrupt the customer support function if it goes down. However **Service Criticality** is only moderate because it does not affect the core functionality of your business offerings.

Data Privacy	**76**
Reproducibility	**84**
Reputation Exposure	**62**
Model Bias	**23**
Security Resilience	**22**
Service Criticality	**45**
Model Correctness	**91**
Legal & Regulatory	**32**

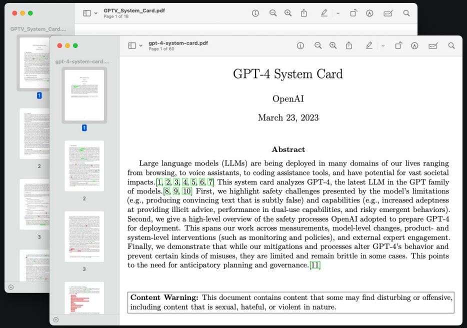
4 Trust, but verify

Don’t just take their word for it, attest to their proper handling of your data

How does Coverbase solve this?

1

A way for enterprises to understand the safety, reliability, and risks of AI products



GPT-4 System Card

OpenAI

March 23, 2023

Abstract

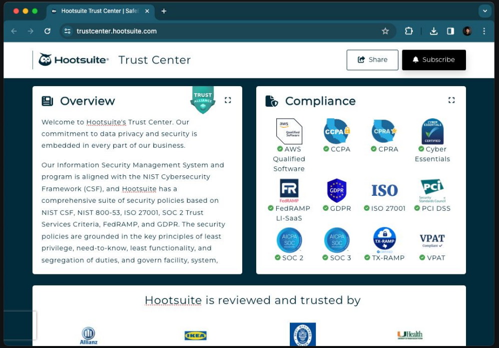
Large language models (LLMs) are being deployed in many domains of our lives ranging from browsing, to voice assistants, to coding assistance tools, and have potential for vast societal impacts.[1, 2, 3, 4, 5, 6, 7] This system card analyzes GPT-4, the latest LLM in the GPT family of models.[8, 9, 10] First, we highlight safety challenges presented by the model's limitations (e.g., producing convincing text that is subtly false) and capabilities (e.g., increased adeptness at providing illicit advice, performance in dual-use capabilities, and risky emergent behaviors). Second, we give a high-level overview of the safety processes OpenAI adopted to prepare GPT-4 for deployment. This spans our work across measurements, model-level changes, product- and system-level interventions (such as monitoring and policies), and external expert engagement. Finally, we demonstrate that while our mitigations and processes alter GPT-4's behavior and prevent certain kinds of misuses, they are limited and remain brittle in some cases. This points to the need for anticipatory planning and governance.[11]

Content Warning: This document contains content that some may find disturbing or offensive, including content that is sexual, hateful, or violent in nature.

Without reading 60-page papers published by vendors

2

A way for AI vendors to communicate their product's safety and reliability to customers



Hootsuite Trust Center

Overview

Welcome to Hootsuite's Trust Center. Our commitment to data privacy and security is embedded in every part of our business.

Our Information Security Management System and program is aligned with the NIST Cybersecurity Framework (CSF), and Hootsuite has a comprehensive suite of security policies based on NIST CSF, NIST 800-53, ISO 27001, SOC 2 Trust Services Criteria, FedRAMP, and GDPR. The security policies are grounded in the key principles of least privilege, need-to-know, least functionality, and segregation of duties, and govern facility, system,

Compliance

✓ AWS Qualified Software
✓ CCPA
✓ CPRA
✓ Cyber Essentials
✓ FedRAMP LI-SaaS
✓ GDPR
✓ ISO 27001
✓ PCI DSS
✓ SOC 2
✓ SOC 3
✓ TX-RAMP
✓ VPAT

Hootsuite is reviewed and trusted by

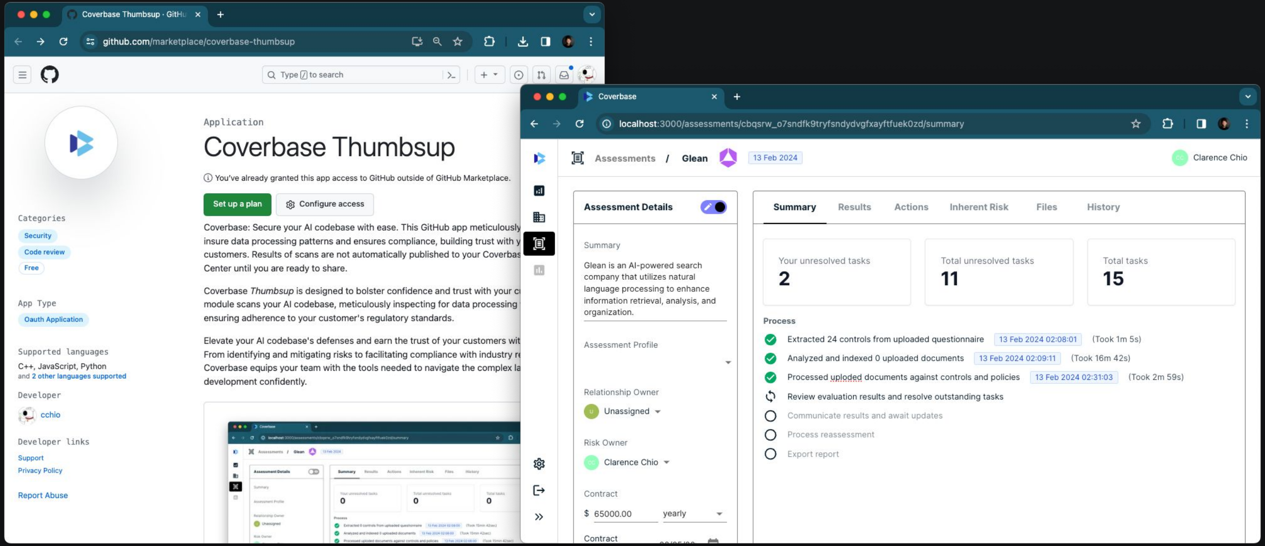
Without independently convincing each new customer of why it's safe to work with them

3

A way for both buyers and sellers to get independent attestation of the vendor's AI safety and compliance

So enterprises don't have to just take the vendor's word for it

Product



Application

Coverbase Thumbsup

ⓘ You've already granted this app access to GitHub outside of GitHub Marketplace.

Coverbase: Secure your AI codebase with ease. This GitHub app meticulously
insure data processing patterns and ensures compliance, building trust with y
customers. Results of scans are not automatically published to your Coverbas
Center until you are ready to share.

Coverbase *Thumbsup* is designed to bolster confidence and trust with your c
module scans your AI codebase, meticulously inspecting for data processing
ensuring adherence to your customer's regulatory standards.

Elevate your AI codebase's defenses and earn the trust of your customers wit
From identifying and mitigating risks to facilitating compliance with industry r
Coverbase equips your team with the tools needed to navigate the complex la
development confidently.

Categories: Security, Code review, Free

App Type: Oauth Application

Supported languages: C++, JavaScript, Python
and 2 other languages supported

Developer: cchio

Developer links: Support, Privacy Policy

Report Abuse

Assessment Details

Summary

Glean is an AI-powered search company that utilizes natural language processing to enhance information retrieval, analysis, and organization.

Assessment Profile

Relationship Owner: Unassigned

Risk Owner: Clarence Chio

Contract: $ 65000.00 yearly

Contract

Your unresolved tasks	Total unresolved tasks	Total tasks
2	11	15

Process

- ✓ Extracted 24 controls from uploaded questionnaire 13 Feb 2024 02:08:01 (Took 1m 5s)
- ✓ Analyzed and indexed 0 uploaded documents 13 Feb 2024 02:09:11 (Took 16m 42s)
- ✓ Processed uploded documents against controls and policies 13 Feb 2024 02:31:03 (Took 2m 59s)
- ↻ Review evaluation results and resolve outstanding tasks
- ○ Communicate results and await updates
- ○ Process reassessment
- ○ Export report

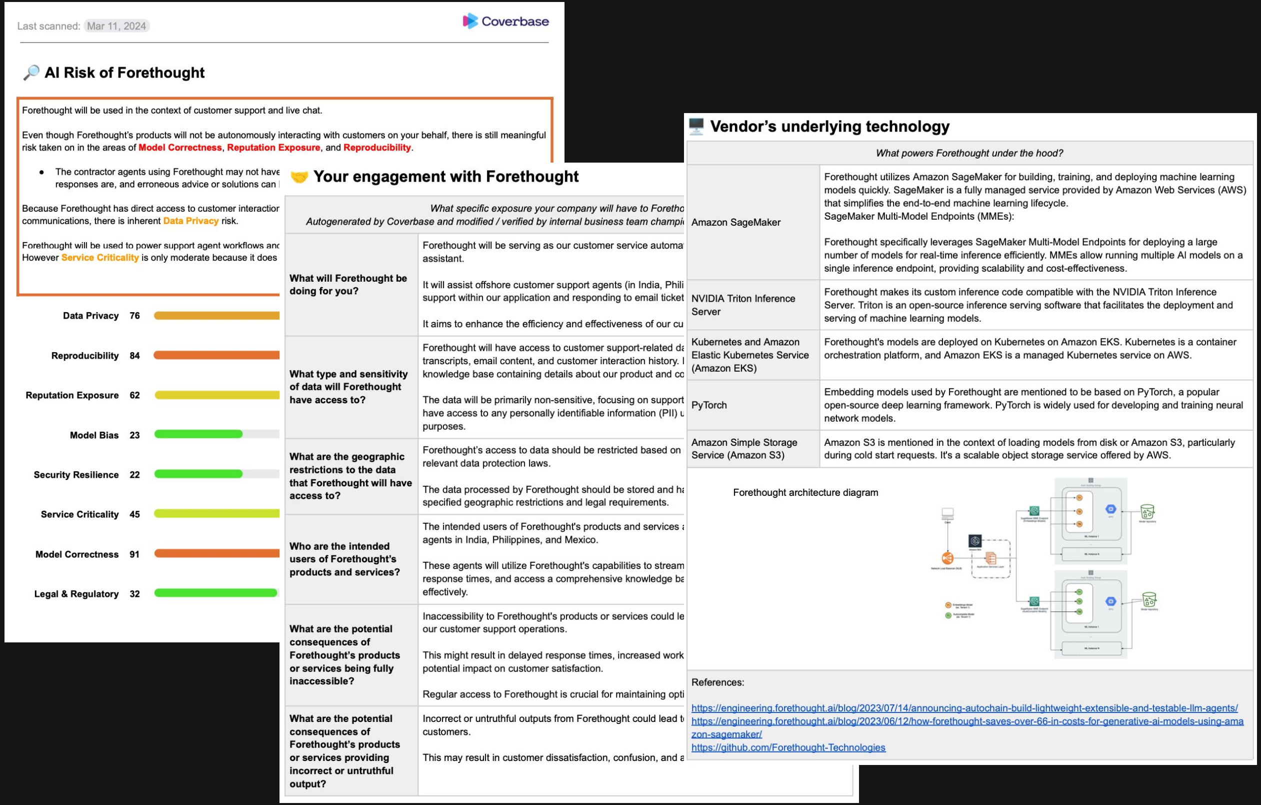
Software Code Analysis tool

free to use, whitebox code

Enterprise vendor management tool

enterprise saas license

Product



Last scanned: Mar 11, 2024

Coverbase

🔎 AI Risk of Forethought

Forethought will be used in the context of customer support and live chat.

Even though Forethought's products will not be autonomously interacting with customers on your behalf, there is still meaningful risk taken on in the areas of **Model Correctness**, **Reputation Exposure**, and **Reproducibility**.

- The contractor agents using Forethought may not have responses are, and erroneous advice or solutions can

Because Forethought has direct access to customer interactior communications, there is inherent **Data Privacy** risk.

Forethought will be used to power support agent workflows anc However **Service Criticality** is only moderate because it does

Data Privacy	76
Reproducibility	84
Reputation Exposure	62
Model Bias	23
Security Resilience	22
Service Criticality	45
Model Correctness	91
Legal & Regulatory	32

🤝 Your engagement with Forethought

What specific exposure your company will have to Foretho Autogenerated by Coverbase and modified / verified by internal business team champi	
What will Forethought be doing for you?	Forethought will be serving as our customer service automa assistant. It will assist offshore customer support agents (in India, Phili support within our application and responding to email ticket It aims to enhance the efficiency and effectiveness of our cu
What type and sensitivity of data will Forethought have access to?	Forethought will have access to customer support-related da transcripts, email content, and customer interaction history. I knowledge base containing details about our product and cc The data will be primarily non-sensitive, focusing on support have access to any personally identifiable information (PII) u purposes.
What are the geographic restrictions to the data that Forethought will have access to?	Forethought's access to data should be restricted based on relevant data protection laws. The data processed by Forethought should be stored and ha specified geographic restrictions and legal requirements.
Who are the intended users of Forethought's products and services?	The intended users of Forethought's products and services a agents in India, Philippines, and Mexico. These agents will utilize Forethought's capabilities to stream response times, and access a comprehensive knowledge ba effectively.
What are the potential consequences of Forethought's products or services being fully inaccessible?	Inaccessibility to Forethought's products or services could le our customer support operations. This might result in delayed response times, increased work potential impact on customer satisfaction. Regular access to Forethought is crucial for maintaining opti
What are the potential consequences of Forethought's products or services providing incorrect or untruthful output?	Incorrect or untruthful outputs from Forethought could lead t customers. This may result in customer dissatisfaction, confusion, and a

🖥️ Vendor's underlying technology

What powers Forethought under the hood?	
Amazon SageMaker	Forethought utilizes Amazon SageMaker for building, training, and deploying machine learning models quickly. SageMaker is a fully managed service provided by Amazon Web Services (AWS) that simplifies the end-to-end machine learning lifecycle. SageMaker Multi-Model Endpoints (MMEs): Forethought specifically leverages SageMaker Multi-Model Endpoints for deploying a large number of models for real-time inference efficiently. MMEs allow running multiple AI models on a single inference endpoint, providing scalability and cost-effectiveness.
NVIDIA Triton Inference Server	Forethought makes its custom inference code compatible with the NVIDIA Triton Inference Server. Triton is an open-source inference serving software that facilitates the deployment and serving of machine learning models.
Kubernetes and Amazon Elastic Kubernetes Service (Amazon EKS)	Forethought's models are deployed on Kubernetes on Amazon EKS. Kubernetes is a container orchestration platform, and Amazon EKS is a managed Kubernetes service on AWS.
PyTorch	Embedding models used by Forethought are mentioned to be based on PyTorch, a popular open-source deep learning framework. PyTorch is widely used for developing and training neural network models.
Amazon Simple Storage Service (Amazon S3)	Amazon S3 is mentioned in the context of loading models from disk or Amazon S3, particularly during cold start requests. It's a scalable object storage service offered by AWS.
Forethought architecture diagram	

References:

https://engineering.forethought.ai/blog/2023/07/14/announcing-autochain-build-lightweight-extensible-and-testable-llm-agents/
https://engineering.forethought.ai/blog/2023/06/12/how-forethought-saves-over-66-in-costs-for-generative-ai-models-using-amazon-sagemaker/
https://github.com/Forethought-Technologies

Our team is made up AI engineers & researchers that have built products used by top security, risk, and finance companies.

Google Quora stripe Unit21

Talk to us

contact@coverbase.ai