CLOUDFLARE

EBOOK

Ensuring safe AI practices

A CISO’s guide on how to create a scalable AI strategy

Table of Contents

Executive summary

Welcome, CISO!

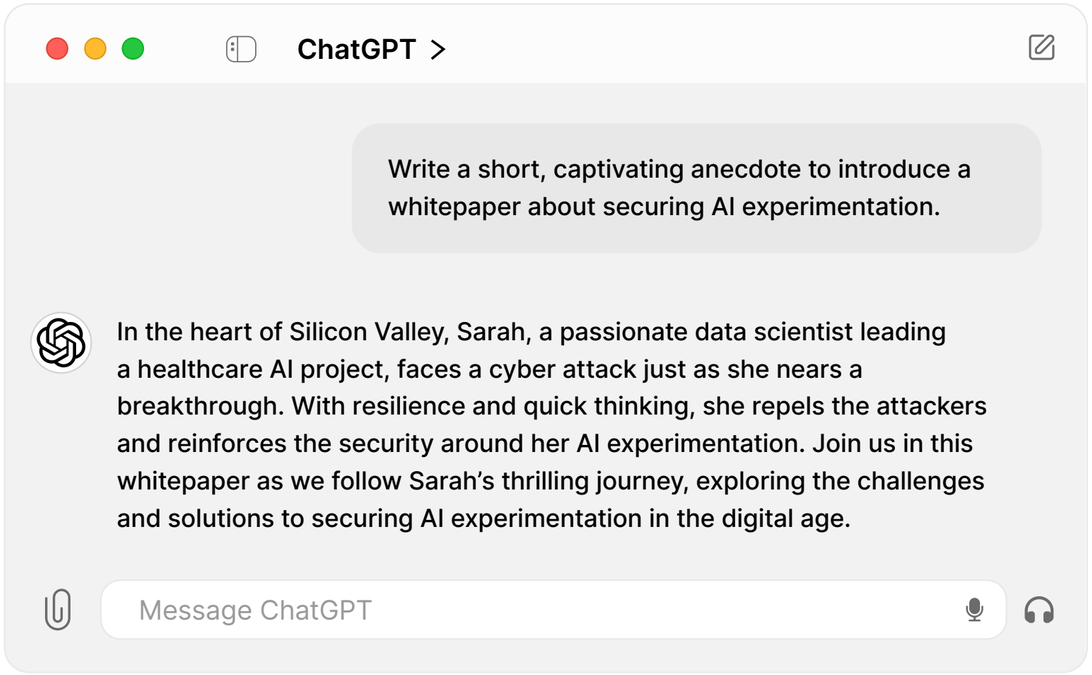
AI is possibly the buzziest word these days, and it is also one of the most pressing issues for the security community. Its influence demands our attention, which is why we at Cloudflare wrote this guide to help you think through secure Generative Artificial Intelligence (GenAI) experimentation in your organization.

AI tools are rapidly becoming more powerful and accessible, unlocking opportunities for innovations across industries. However, as with other paradigm-shifts, GenAI comes with unique security, privacy, and compliance challenges. Widespread GenAI adoption can trigger unforeseen usage spikes, instances of user abuse, malicious behaviors, and hazardous shadow IT practices, all increasing the risk of data breaches and sensitive information leaks.

As adoption expands in your workplace, you need to prepare with a GenAI blueprint that informs how to use, build, and secure at scale. Let's discuss the risks and review tips your team can use for securing GenAI based on maturity levels and usage. With these strategies, your organization can create a GenAI strategy that fits the needs of the business while protecting your data and ensuring compliance.

- Dawn Parzych, Director of Product Marketing, Cloudflare

Securing GenAI experimentation



Sorry to tell you, but Sarah's story ends there. While we're saying goodbye to our fictional character, as predictive and GenAI expands, there will be countless "Sarahs" in real life — each acting as a hero on IT and developer teams, as business technologists, and individual employees.

AI has enchanted technologists and everyday users alike, sparking curiosity and tinkering. This experimentation is necessary as we work to unlock the full potential of AI. But without caution and guardrails, it may also lead to compromising security or falling out of compliance.

To achieve the balance, and understand and manage AI initiatives more effectively, organizations must consider three key areas:

1 Using AI

Using AI technologies (e.g. ChatGPT, Bard, and GitHub Copilot) offered by third-party vendors while safeguarding assets (e.g. sensitive data, intellectual property, source code, etc.) and mitigating potential risks based on the use case

2 Building AI

Developing custom AI solutions tailored to an organization's specific needs (e.g. proprietary algorithms for predictive analytics, customer-facing co-pilots or chatbots, and AI-driven threat detection system)

3 Securing AI

Protecting AI applications and AI systems from bad actors manipulating it to behave unpredictably

Securing GenAI experimentation

GenAI transformation: today and in the future

GenAI's appeal to consumers and organizations has set it on an unprecedented adoption trajectory. A small group of power users grew quickly thanks in part to an active open source community and the consumer-driven experimentation of applications like ChatGPT and Stable Diffusion.

What users have found through it all is that robots will not, in fact, "replace us."

GenAI puts humans in the position of refining and augmenting, rather than creating everything from scratch, and can help businesses amplify their workforce efficiency. Predictive AI offers similar benefits by making it easier to tap into data to improve decision making, build smarter products, and personalize customer experiences, among a range of initiatives.

Today, **59% of developers** are currently using AI in their development workflows[1]

By 2026, >**80% of enterprises** will be using GenAI-enabled API, models, and/or apps deployed in production environments (up from 5% today)[2]

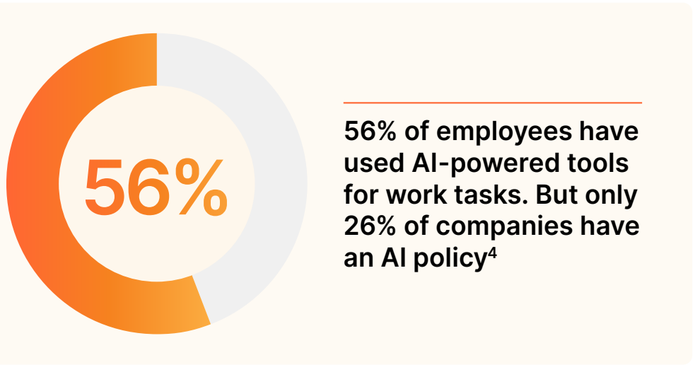
By 2030, GenAI will augment **50% of knowledge worker**s' tasks to boost productivity or raise average quality of work (up from <1% today)[3]

1. SlashData, "How developers interact with AI technologies," May 2024
2. Gartner, "A CTO's Guide to the Generative AI Technology Landscape", Sept 2023
3. Gartner, "Emerging Tech: The Key Technology Approaches That Define Generative AI", Sept 2023

Using GenAI securely

AI experimentation spans a spectrum from using pre-built AI tools and services to building custom AI solutions from scratch. While some organizations may advance towards creating their own AI models and applications, many will stick to consuming third-party AI tools.

In these instances, third-party AI tools create new risks because organizations only have limited direct controls on their security and privacy configurations.



56% of employees have used AI-powered tools for work tasks. But only 26% of companies have an AI policy[4]

Employees are likely using off-the-shelf AI tools for work right now via SaaS suites like Microsoft 365, chatbots built into search engines or public apps, and even APIs.

Organizations must do their due diligence to minimize risks, including:

- **Evaluating** the security risk of third-party tools
- **Addressing** data privacy concerns
- **Managing** reliance (or overreliance) on external APIs
- **Monitoring** potential vulnerabilities

An example of this would be when employees use public web apps like ChatGPT. Every input fed into a prompt becomes a piece of data leaving an organization's control. Users may overshare sensitive, confidential, or regulated information — like Personally Identifiable Information (PII), financial data, intellectual property, and source code. And even if they don't share explicit sensitive information, it is possible to piece together context from inputs to infer sensitive data.

To safeguard, employees can toggle a setting to prevent their inputs from training the model further, but they must do it manually. To ensure security, organizations need ways to keep people from entering private data.

Prepare for the security implications of AI

Data Exposure

To what extent are users improperly sharing sensitive data with external AI services? Are anonymization/pseudonymization techniques sufficient?

API Risks

How will you address vulnerabilities within third-party APIs that could be potential gateways for attackers?

Black-Box Systems

What are the decision-making processes of external AI models that could introduce unexpected risks?

Vendor Risk Management

What do you know about the security practices of your third-party AI providers? More importantly, what do you not know?

4. The Conference Board, September 2023

Steps to defend AI consumption

1 Manage governance and risk

- Develop policies on how and when to use AI, including what information the organization allows users to share with GenAI, access control guidelines, compliance requirements, and how to report violations
- Conduct an impact assessment to gather information, identify, and quantify benefits and risks of AI usage

2 Increase visibility and controls for security and privacy

- Log all connections, including to AI apps, to continuously monitor user activities, AI tool usage, and data access patterns to detect anomalies
- Discover what shadow IT exists (including AI tools)—and make decisions to approve, block, or layer additional controls
- Scan SaaS app configurations for potential security risks (e.g. OAuth permissions granted from approved apps to unauthorized AI-enabled apps, risking exposure of data)

3 Examine what data goes in and out of AI tools and filter out anything that could compromise IP, impact confidentiality, or violate copyright restrictions

- Apply security controls for how users can interact with AI tools (e.g. stop uploads, prevent copy/paste, and scan for and block inputs of sensitive/proprietary data)
- Put safeguards in place to block AI bots from scraping your website
- Block AI tools outright only if no other controls are possible. As we know, users will find workarounds, which puts security out of your control

4 Control access to AI apps and infrastructure

- Ensure that every user and device accessing AI tools undergoes strict identity verification to scope who gets to use AI tools
- Implement identity-based Zero Trust access controls. Apply least privilege to limit potential damage from compromised accounts or insider threats

5 Streamline costs and operational efficiency

- Understand how people are using AI applications with analytics and logging so that you have control over rate limiting, caching, as well as request retries, and model fallback as usage scales

Secure what you build

Train your AI model

AI pipelines are broadening the vulnerability spectrum. But with experience securing at the start and throughout the development process, we have insights into what leads to success. For AI security, the natural place to begin is in your model.

As the basis of AI applications, everything used to train your AI model will flow through to its outputs. Consider how you will secure that data initially to avoid negative repercussions later. If left unprotected, you risk expanding your attack surface and creating application issues down the road.

Security that ensures data integrity is pivotal in mitigating deliberate and accidental data compromise. Security risks in the AI pipeline can include:

- **Data poisoning:** Malicious datasets influence outcomes and create biases
- **Hallucination abuse:** Threat actors legitimize AI hallucinations — the invention of information in order to generate responses — so that malicious and illegitimate datasets inform outputs

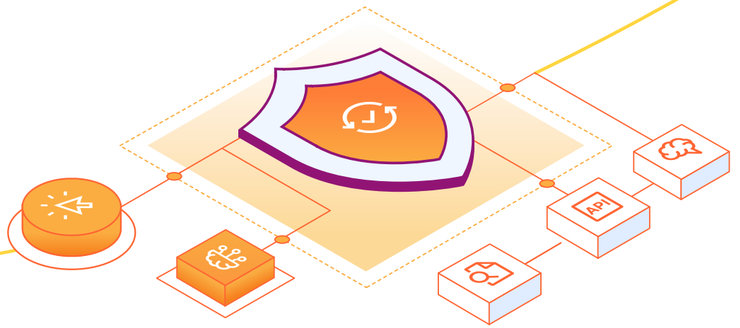
Alternatively, if you are not training models, your in-house AI would start with selecting a model to perform tasks. In these cases, you would want to explore how creators made and secured the model as it plays a role in inference.

Inference is the process that follows AI training. The better trained a model is, and the more fine-tuned it is, the better inferences will be — although they are never guaranteed to be perfect. Even highly-trained models can hallucinate.

Post-deployment security

Once you build and deploy your in-house AI, you will need to protect its private data and secure access to it. Along with recommendations we've already made in this paper, including enforcing tokens for each user and rate limiting, you should also consider:

- **Managing quotas:** Uses limits to protect against users' API keys from becoming compromised and shared
- **Blocking certain autonomous system numbers (ASNs):** Keeps attackers from sending overwhelming amounts of traffic to applications
- **Enabling waiting rooms or challenging users:** Makes requests more difficult or time-consuming, ruining the economics for attackers
- **Building and validating an API schema:** Outlines what intended usage is by identifying and cataloging all API endpoints, then lists all specific parameters and type limits
- **Analyzing the depth and complexity of queries:** Helps guard against outright DoS attacks and developer errors, keeping your origin healthy and serving requests to your users as expected
- **Building discipline around token-based access:** Protects from compromised access when tokens validate in the middleware layer or API Gateway



Robust threat protection across your GenAI experimentation

From adoption to implementation, every stage of the GenAI experimentation spectrum should progress with minimal or tolerated risk. With the knowledge gained from this paper, whether your organization is using, building, or planning on AI in some form in the future, you have the power to control your digital environment.

While it's natural to feel hesitant when adopting new capabilities, resources exist to give you the confidence to experiment with AI safely. Of those resources, what organizations need most today is a connective tissue for everything IT and security. One that acts as a common thread that reduces complexity by working with everything in the environment, is available everywhere, and performs necessary security, networking, and development functions.

With the connective tissue, you'll have confidence in a variety of use cases, including:

- Complying with regulations with the ability to detect and control movement of regulated data
- Regaining visibility and control for sensitive data across SaaS applications, shadow IT, and emerging AI tools
- Securing developer code by detecting and blocking source code in uploads and downloads. Plus, preventing, finding, and fixing misconfigurations in SaaS applications and cloud services, including code repositories

As AI continues to evolve, uncertainty is certain. That's why having a steadying force like Cloudflare is so beneficial.

Protect from AI risks across three types of LLMs

Depending on use, the level of risk exposure AI creates for an organization will vary. It is critical to understand the various risks associated with Large Language Models (LLM) usage and development, and then be actively involved in any LLM deployments.

Type of LLM	Key risk
Internal	Access to sensitive data and intellectual property
Product	Reputational risk
Public	Sensitive data leakage

Scale, ease of use, and seamless integration

Cloudflare's connectivity cloud puts control into your hands and improves visibility and security — making AI experimentation safe and scalable. Even better, our services fortify everything, ensuring that there's no trade-off between user experience and security.

Given that most organizations will either only use AI or will use and build, utilizing Cloudflare means never stalling on AI projects.

- Our **global network** enables you to scale and enforce controls with speed wherever you need them
- Our **ease of use** makes it simple to deploy and manage policies for how your users consume AI
- One **programmable architecture** enables you to layer security onto the applications you're building, without disrupting how your users consume AI

Cloudflare's connectivity cloud protects every facet of your AI experimentation, specifically:

- Our **Zero Trust & Secure Access Service Edge (SASE)** services help mitigate risk in how your workforce **uses** third-party AI tools
- Our **developer platform** helps your organization **build** your own AI tools and models safely and efficiently
- For **securing with AI**, our platform leverages AI and machine learning techniques to build threat intelligence that is then used to protect organizations across their AI experimentation

	How you use AI	How you build AI
Our global network scale	Scale and enforce controls everywhere with consistency	Accelerate inference, querying and caching
Our simplicity of management	One control plane with simple deployment and policies	Templates to get onboarded quickly
Our unified & programmable network architecture	Layer new security without disrupting how you use AI	Build-in privacy and compliance

Next steps

From protecting how your organization uses AI to defending the AI applications you build, Cloudflare for AI has you covered. With our services, you can adopt new capabilities in any order with limitless interoperability and flexible integrations.

Talk to an expert

For more information, visit cloudflare.com

1 888 99 FLARE | enterprise@cloudflare.com | Cloudflare.com

REV:BDES-6307.2024AUG19