CLOUDFLARE

EBOOK

Modernizing security for the AI era

The new playbook for securing people, data, and applications

Table of contents

Executive summary

If you could rebuild your security architecture, knowing what AI has unleashed, would it still resemble what you have now?

When I meet with security leaders around the world, the answer is almost always “no.”

Why? Because this era of hyper-accelerated AI adoption is forcing everyone to reevaluate their security stack. Even before generative AI (GenAI) became mainstream, IT and security teams struggled to manage dozens of disjointed point solutions. Layering more tools has only increased complexity, cost, and exposure.

Yesterday’s best practice of stitching together specialized, best-of-breed tools breaks down when AI itself becomes the new attack surface.

The workforce can use and misuse AI and GenAI apps in too many unpredictable ways. AI and machine learning (ML) models, frameworks, apps, agents, and compliance standards evolve too rapidly for siloed tools to adapt. And, with attackers now exploiting AI to move faster, relying on traditional, reactive security is dangerously inadequate.

This new reality makes rethinking security models not just prudent, but also essential to remain competitive and resilient. Organizations need a new kind of architecture that helps them harness the power of AI while safeguarding their systems and defending against emerging threats.

Read on to uncover the new security model required to succeed in this era, with strategies to harness AI confidently and to stay one step ahead of AI-powered threats.

Amit Chaudhry

Senior Director of Solutions Marketing, Cloudflare

Why AI demands a new security model

AI is reshaping the business landscape at a pace few security architectures can match. New models, frameworks, protocols like the Model Context Protocol (MCP), hardware advancements, and integration standards constantly emerge. Tools like Claude, ChatGPT, and Copilot empower business users outside of engineering teams, often bypassing IT approvals altogether. Organizations have also begun building with AI and launching new AI-powered capabilities.

Enterprises often find themselves struggling to keep pace with the latest developments, risking obsolescence, compliance failures, reputational damage, and competitive disadvantage if they fail to adapt quickly.

The rapid pace of AI adoption and creation require new security measures built on three pillars, which provide the foundation for secure and continuous innovation:

Protect people and data

Safeguarding employees and how they use data, devices, automated services, servers, and other systems that consume GenAI resources.

Secure AI-enabled applications

Embedding security throughout the AI app development lifecycle — and protecting apps themselves from data risks, model manipulation, and other malicious activity.

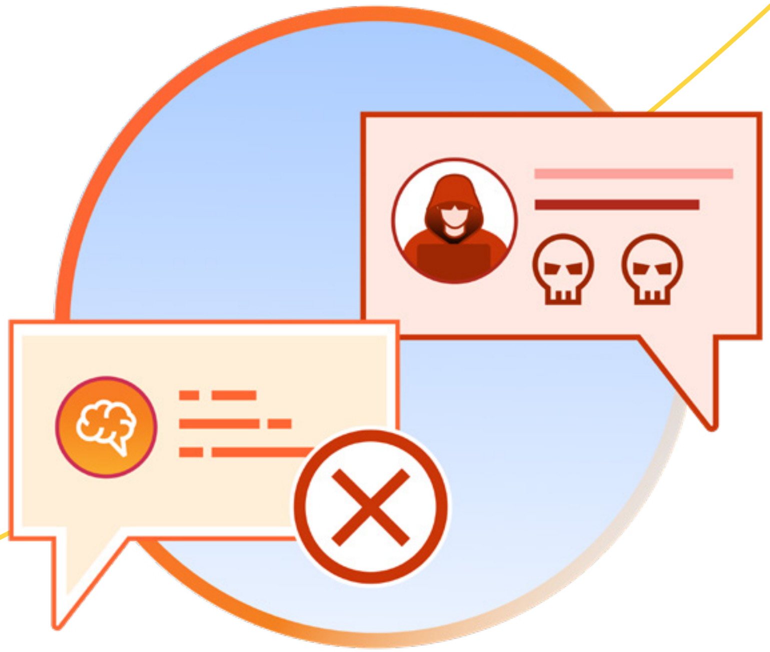
Defend with AI

Using AI to enhance organizations' security posture and reduce complexity by enabling advanced threat detection, automating responses, and more.

This playbook shares critical strategies for tackling these key security imperatives. It also shares how organizations can simplify their architecture by consolidating a range of tools and infrastructure into one modern security platform.

AI brings forward a broad set of security risks

Limited visibility into employee usage	85%	of IT leaders say employees are adopting AI tools before they can be assessed.[1]
Uncontrolled data exposure and compliance risks	90%	of employees trust unauthorized AI tools to protect their data; 50% believe there's little to no risk in using unapproved tools.[1]
Inconsistent governance	74%	of organizations have fragmented AI development toolchains — making governance harder.[2]
Unsafe app development	37%	of organizations have processes to assess the security of AI tools before deployment.[3]
Adversarial AI overwhelming traditional defenses	80%	of ransomware attacks use AI, from deepfakes to AI-generated phishing campaigns.[4]



USE CASE #1

Safeguard people and data as they consume AI

Today, the fastest growing risk with AI is not malicious hackers; it is employees unintentionally putting sensitive data into GenAI apps.

For instance, imagine a staff accountant prompting a GenAI tool to search and display a company's billable hours for the current month. They instruct the tool to send out invoices and payment reminders as customers' due dates approach. Now, AI initiates a transaction that needs to be tracked, documented, and — above all — kept private.

Employees are adopting AI well ahead of formal policy, often unaware of the risks. Shadow AI deployments sidestep traditional reviews, creating unseen attack surfaces and new compliance risks.

Defining acceptable use policies for AI apps is the critical first step to safeguarding AI consumption. **However, true enforcement will require having access controls and data restrictions built directly into the AI interaction surface.**

Five imperatives to safeguard employee AI use

1

Discover shadow AI usage

Organizations cannot protect what they cannot see, including shadow AI usage. Identify and filter all Internet-bound AI traffic; once GenAI app usage is discovered, implement the appropriate policies.

2

Monitor and control AI app access

Apply the zero trust security principle of least privilege to ensure that only authorized AI services, and authorized users on trusted devices, are allowed to connect with your network infrastructure.

3

Protect sensitive data

Data loss prevention (DLP) capabilities block attempts to share or upload proprietary code, personally identifiable information (PII), and other sensitive data with risky GenAI tools.

4

Block harmful or toxic prompts

Employees can inadvertently or intentionally submit inappropriate prompts or topics into an AI service. Block harmful interactions to prevent model poisoning and incorrect outputs, and enforce corporate policy.

5

Enhance posture management

An AI security posture management (AI-SPM) service featuring a cloud access security broker (CASB) scans for GenAI service misconfigurations and data exposure, giving you more control over the AI environment.

> “Despite its many legitimate users, AI presents major security and privacy concerns. Cloudflare helps us find what shadow AI risks exist and block unsanctioned AI apps and chatbots.”
>
> **Matthew Ortiz**
> Senior Manager Information Security, **Indeed**
>
> See how they do it >
>
> indeed

USE CASE #2

Protect AI-enabled apps and workloads

Whether an organization builds its own large language model (LLM) on internal data or or integrates third-party AI tools within their public-facing applications (such as through a website), the risks are significant.

For example, a customer support bot — if manipulated — could leak sensitive employee data or trade secrets. An attacker could also abuse a model by overloading it with requests, causing AI resource overconsumption or denial of service. AI-specific threat vectors such as these transform helpful AI tools into liabilities.

In other words: Risks from both homegrown and third-party AI applications are growing rapidly.

The Open Worldwide Application Security Project (OWASP) published a Top 10 risks for LLMs advisory warning organizations about sensitive information disclosure, prompt injection, model poisoning, and a number of other LLM threats that apply to any AI tool.

Despite these warnings, only 37% of organizations have processes in place to assess the security of AI tools before deployment.[3]

After an app is deployed, it's too late to bolt on security or try to claw back data from an external model.

Effective security must account for the model layer itself — across the entire ecosystem of AI tools an organization builds, buys, and uses.

Three requirements to protect AI-enabled apps

1

Extend visibility across apps and APIs

Specialized AI firewalls discover and label GenAI and application programming interface (API) endpoints, detect attempts to exfiltrate PII, and block malicious prompts before they impact AI model performance — or poison the model with toxic content or misinformation.

2

Block threats at the edge

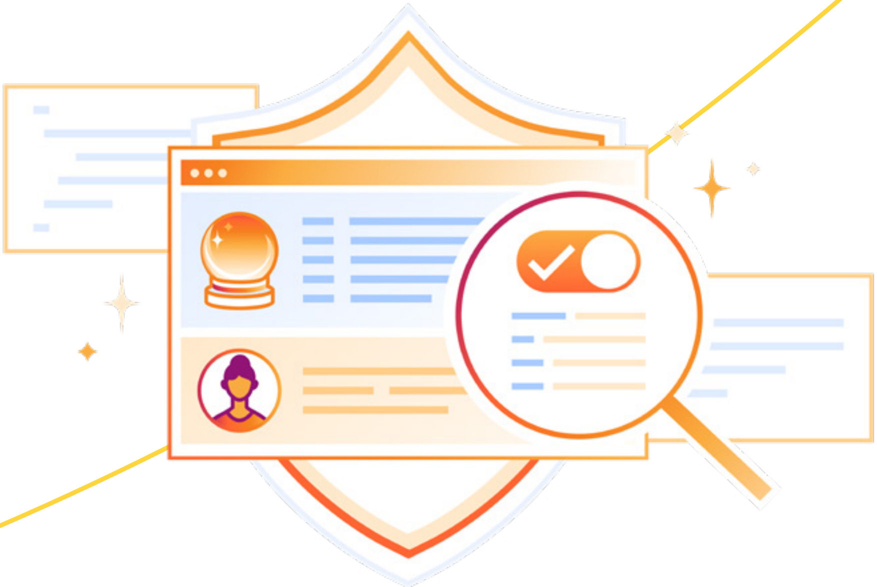
Real-time threat detection and mitigation comprehensively blocks AI-specific vulnerabilities, misconfigurations, and attack paths within AI pipelines — such as prompt injection, data poisoning, and model abuse.

3

Ensure compliant development

AI-aware data protection proactively manages data inputs, enforces strict access controls within AI models and pipelines, maintains audit trails, and helps comply with requirements like those set by the EU AI Act.

> “Cloudflare’s low cost and global presence offer us endless opportunities as we shift more toward infrastructure-as-code, programmability and automation, AI, and a greater understanding of how we can use Zero Trust to improve cloud security and visibility.”
>
> **Michael Lee**
> Manager of Network Engineering, **VistaPrint**
>
> See how they do it >
>
> vistaprint



USE CASE #3

Defend with AI

Because AI excels at pattern recognition, attackers can now more easily identify and exploit weaknesses in traditional defenses. For example, adversaries are leveraging AI to:

- Craft highly convincing phishing and social engineering scams that bypass legacy security measures
- Deploy automated bots to exploit network vulnerabilities at a scale and speed that makes it difficult for security teams to respond in a timely way
- Develop novel tactics, including more complex forms of identity fraud

As attackers continue to harness AI for harm, the shift from passive defense to proactive, AI-driven security is no longer optional — it is a necessity.

An organization's security architecture must constantly adapt and evolve to stay ahead of these and other threats.

AI-powered defense adapts to the unique behaviors and needs of individual organizations. This tailored approach leads to more accurate threat detection, faster automated incident response, and predictive risk management across the organization.

WHAT IS NEEDED

Five strategies to build proactive, AI-driven security

1

Drive comprehensive security and observability with AI

Unify logs, analytics, alerts, and forensics in a single interface to identify risks and their root causes.

2

Detect and neutralize threats in real time with AI-driven security

Leverage AI / ML-powered detection to analyze vast datasets, identify anomalies, and automate responses to emerging threats.

3

Defend against AI-powered phishing, deepfakes, and malware

Deploy AI-driven detection, phishing-resistant authentication, and adaptive security controls to counter evolving attacks.

4

Protect proprietary data from AI scrapers and automated threats

Utilize bot management, API authentication, and digital watermarking services to prevent data theft and exploitation.

5

Uncover insider threats in real time with AI-driven behavioral analytics

Continuously analyze user behavior, not just identity, to assess unusual access patterns, privilege escalations, and data exfiltration attempts.

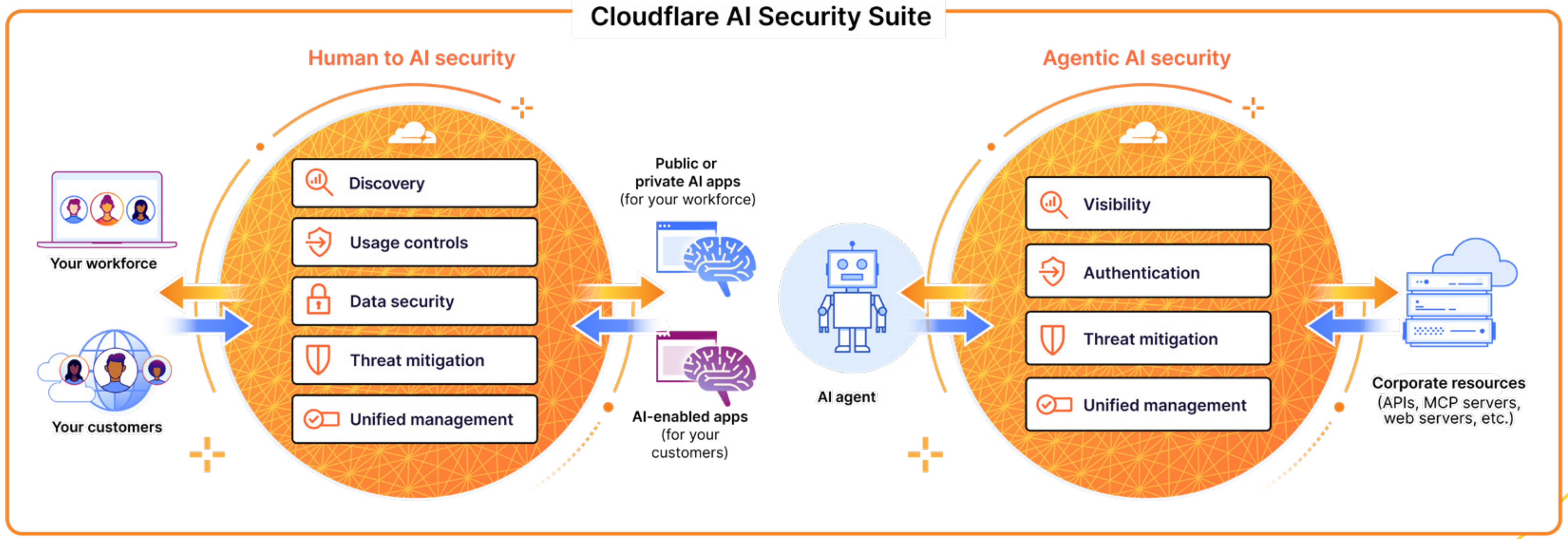
> **“After GPT4 was released, we analyzed our statistics in Cloudflare and found that attacks had jumped by 20% within one month. We rely on Cloudflare’s email security solution, which is handling the increased volume of these phishing attacks very well, and helping us protect our users and our company.”**
>
> **Roman Bugaev**
> CTO, **Flo Health**
>
> See how they do it >

One unified approach to securing people, data, and applications

To accelerate AI adoption without compromising security, organizations need a unified platform of security, connectivity, and development services powered by a programmable global network.

The Cloudflare connectivity cloud provides secure, low-latency, and infinitely scalable connectivity across applications, global user and customer bases, APIs, and hybrid networks. These benefits help organizations cut complexity, boost security and compliance, and speed up time-to-value from AI transformation.

Built on Cloudflare’s connectivity cloud, our AI security services work together to enhance the security, scalability, and efficiency of AI workloads. They represent a single end-to-end solution for safeguarding people, data, and apps from emerging threats while optimizing performance.



1. How Cloudflare safeguards people and data as they use AI

Secure access service edge (SASE), already a critical component for consolidating security and networking functions, is now even more foundational for effective AI security.

Within Cloudflare's SASE suite, Cloudflare One, protection for AI usage is multifaceted. Organizations gain real-time threat detection, governance, and protection tools that keep data safe, employees productive, and business operations resilient.

See with:

- **Shadow AI reporting** provides visibility into how employees are using AI tools.
- **Out-of-band API CASB integrations** detect misconfigurations and provide visibility into security posture inside popular public AI tools.

Measure with:

- **AI confidence scoring** assesses risk associated with third-party SaaS and AI applications, so organizations can write policies accordingly.

Protect with:

- **AI prompt protections**, a DLP capability, defends against malicious inputs and prevents sensitive data loss.

Secure with:

- **MCP security** untangles complex MCP deployments by centralizing policy enforcement and providing comprehensive visibility and logging.

2. How Cloudflare protects AI-enabled apps and workloads

Developer-friendly guardrails are required to mitigate the risk of harmful, inaccurate, or inappropriate content being incorporated into app development.

Cloudflare's AI solutions — AI Gateway, Firewall for AI, and Workers AI — work together to provide an end-to-end solution that protects AI-driven applications from emerging threats.

AI Gateway acts as a central point (proxy) between applications and the AI model provider, in order to:

- **Enforce** AI guardrails, which prevents harmful prompt and AI responses
- **Apply** advanced rate limiting to prevent model abuse
- **Secure** storage of API keys to model providers such as OpenAI, Anthropic, and others

Firewall for AI provides a layer of defense tailored for LLM vulnerabilities, in order to:

- **Analyze** incoming prompts using detection engines trained on AI threat intelligence
- **Identify** and block malicious patterns, anomalies, and policy violations
- **Ensure** only safe prompts reach the model

Workers AI (AI inference-as-a-service) ensures that the **AI inference itself runs in a secure, high-performance environment that is globally distributed.**

Workers AI does not employ user prompts to train models, and hosts 50+ open source models, ensuring transparency.

Organizations gain the AI inference workload they need for agents and AI applications, and can develop MCP servers with **built-in integrations (OAuth 2.1) for authentication and authorization.**

3. How Cloudflare defends with AI

Cloudflare's intelligent, adaptive security posture fights adversarial AI with AI.

On a daily basis, Cloudflare's vast global network analyzes trillions of signals across the Internet. This includes processing 84 million requests and 61 million DNS queries per second, and more. This unparalleled volume, velocity, and variety of traffic provides extensive threat telemetry. Consequently, our machine learning models offer advanced, proactive protection across an organization's entire digital estate.

The power of the Cloudflare network underpins our approach to safeguarding against sophisticated attacks — including those targeting AI systems, and those who maliciously exploit AI.

> **"...with the rapid development of generative AI, we expect to see new generations of complex online threats coming toward the industry. I believe Cloudflare has the visibility, access, and business intelligence to leverage the massive amounts of data on the global network and train defensive models to identify and mitigate these next-generation attacks."**
>
> **Mehdi Salour**
> Senior Vice President of Global Network and DevOps,
> **8×8**

Multi-vector AI

Protect against evolving cyber threats, including those that exploit AI technologies to increase attack volume and authenticity.

Customized solutions

Analyze traffic patterns specific to an organization and adapt strategies to fit the distinct behaviors and needs of their environments.

Advanced AI and real-time analytics

Enhance threat detection, mitigation, and protection across apps, APIs, networks, people, and AI workloads themselves.

Ready to modernize your security?

Cloudflare provides enterprises, developers, startups, and digital media and content owners with the tools and infrastructure needed to confidently navigate the complexities of AI adoption.

By leveraging Cloudflare's extensive global edge network (330+ cities in 125 countries), integrated security solutions, and developer-centric tools, Cloudflare empowers organizations to innovate faster, automate effectively, and build secure AI-driven applications while safeguarding their data and applications.

Learn more about building a secure foundation for AI innovation.

Request a custom enterprise demo

Further reading and resources

Preparing for the future of AI in cybersecurity

How do you start developing a strategy today that will help your organization maximize the value of AI in the future? Cloudflare's chief security officer, Grant Bourzikas, shares four key recommendations.

Read article

Cloudflare AI Security Suite

Adopt AI with confidence. Learn how Cloudflare AI Security Suite protects the entire AI lifecycle against attacks like prompt injection, data exfiltration, and model abuse.

Read solution brief

Security Signal: Adversarial AI

Khalid Kark, Field CIO at Cloudflare, Daniel Kendzior, Managing Director and AI Security Lead at Accenture, and Mike Hamilton, CIO at Cloudflare, dive into the real-world impact of AI in cyber. Watch to learn how to harness AI for defense — not just detection.

Watch interview

Best practices for securing generative AI with SASE

This guide dives into three key pillars for dealing with (human) employee access to AI — visibility, risk management, and data protection — as well as guidelines around deploying agentic AI in the enterprise using MCP.

Read blog post

References

1. ManageEngine. “97% of IT Decision Makers See Significant Risks of Shadow AI, While 91% of Employees See No Risk, Little Risk, or Risk That’s Outweighed by Reward.” Press release, 8 July 2025. https://www.manageengine.com/news/shadow-ai-report.html.

2. Pogorelec, Anamarija. “Bridging the AI model governance gap: Key findings for CISOs.” HelpNetSecurity, 18 Aug. 2025. https://www.helpnetsecurity.com/2025/08/18/ciso-ai-model-governance/.

3. Creese, Sadie and Joshi, Akshay. “A leader’s guide to managing cyber risks from AI adoption.” World Economic Forum, 21 Jan. 2025. https://www.weforum.org/stories/2025/01/a-leaders-guide-to-managing-cyber-risks-from-ai-adoption/.

4. Church, Zach. “80% of ransomware attacks now use artificial intelligence.” MIT Sloan, 8 Sept. 2025. https://mitsloan.mit.edu/ideas-made-to-matter/80-ransomware-attacks-now-use-artificial-intelligence.

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

REV:BDES-7856.2025OCT06

This document is for informational purposes only and is the property of Cloudflare. This document does not create any commitments or assurances from Cloudflare or its affiliates to you. You are responsible for making your own independent assessment of the information in this document. The information in this document is subject to change and does not purport to be all inclusive or to contain all the information that you may need. The responsibilities and liabilities of Cloudflare to its customers are controlled by separate agreements, and this document is not part of, nor does it modify, any agreement between Cloudflare and its customers. Cloudflare services are provided "as is" without warranties, representations, or conditions of any kind, whether express or implied.