

AI's role in cybersecurity

Using artificial intelligence and machine learning, we help people and organizations detect cyberthreats faster, mitigate risks, and confidently use digital technologies in their daily lives and work.

Why this matters

Attackers are increasingly using AI to automate attacks. Our experts regularly see AI used to [generate](#) phishing pages, ransomware, malware for advanced targeted attacks (such as [Bluenoroff](#)), and in campaigns targeting Russian organizations, such as Librarian Likho.

Attackers are automating more and more steps in the attack chain. However, it is important to note that AI does not radically change the threat landscape.

AI is also actively used in cyber defense: it can significantly improve threat detection performance, including detection of complex attack techniques such as DLL hijacking¹.

That said, reliable protection still relies on a multi-layered approach: endpoint and network protection, managed services (e.g. MDR), advanced end-to-end solutions such as XDR, and high-quality threat analytics (Threat Intelligence). AI plays a different role: it improves the response speed, scalability and accuracy of security technologies, helping to resist cyberattacks.

How we use AI in cyber defense

~20 years

Kaspersky has been using AI and ML technologies

135 AI-related patents

Kaspersky's intellectual property portfolio

Kaspersky has been [using](#) AI and ML technologies in its products and services for almost 20 years.

Each day, artificial intelligence helps analyze hundreds of thousands of suspicious and malicious files, identifying patterns and anomalies in a split second. At the same time, we view AI not as a replacement for human specialists, but as a support tool: algorithms take on routine signal processing, freeing up experts to analyze complex, targeted, and unconventional attacks.

Our [Kaspersky AI Technology Research Center](#) brings together data scientists, machine learning engineers, and experts on threat intelligence and infrastructure to solve the most ambitious challenges at the intersection of AI/ML and cybersecurity. This work includes developing and improving applied technologies, research into the safety of AI algorithms, raising awareness of AI risks, and much more.

We expand threat detection capabilities

Kaspersky has developed many AI/ML-based threat detection technologies, primarily for identifying malware, but also to detect attackers' suspicious activity. For example, in 2025, our experts [trained](#) an ML model to detect attempts to use DLL-hijacking techniques and used it to improve the KUMA SIEM system. Our solutions use not a monolithic algorithm that does everything, but a set of specialized models that each perform a particular task. This approach makes our protection more robust and accurate.

Key applications of AI/ML²

- Early file checks. Deep neural networks help identify malicious executable files based on static characteristics at early stages, even before the files are launched.
- Automatic creation of detection rules. Machine learning (ML) technologies based on decision trees help generate threat detection rules that can run

directly on the user's device. This is important when it comes to quickly translating threat knowledge into practical defenses.

- Behavioral analysis. Even if a file appears safe, malicious activity can occur while the program is running. Behavioral patterns help identify such threats through their atypical actions.
- Identification of malicious internet resources based on anonymous telemetry received from solutions installed on clients and other sources.
- Protection from phishing and spam. We reduce risks to users by applying specialized models, including an ML model for detecting fraudulent web pages and DeepQuarantine for quarantining emails suspected of spam.

Thanks to our cloud infrastructure, AI results are available to users almost instantly, and new threats are blocked immediately upon detection.

¹ An attack technique in which an attacker injects a malicious dynamic-link library (DLL), causing a legitimate program to load and execute malicious code.

² Read more about these technologies in our whitepaper [Machine Learning for Malware Detection](#).

We combat phishing and online fraud

Online fraud has become more sophisticated over the years: modern phishing sites look neat, their text is often error-free, and the visual elements mimic the interfaces of well-known services. To combat these threats, we use machine learning and content analysis.

In particular, Kaspersky protects against phishing through:

- optical character recognition (OCR) to identify malicious text hidden within images
- proprietary ML models that identify the telltale signs of fraud, having been trained on datasets of legitimate and counterfeit websites.

This is especially important as scammers increasingly use images and visual bait to try to bypass simple filters.

We help SOC specialists be more effective

In corporate environments, security professionals often encounter excessive "noise," i.e. a large number of alerts that do not represent real incidents. This wastes a lot of time.

Our Managed Detection and Response (MDR) services use AI algorithms to automatically analyze event streams and filter out false positives, allowing tens and hundreds of thousands of benign incidents to be resolved each year without human intervention. As a result, security operations center (SOC) specialists can focus on truly important attacks and respond more quickly to real threats.

Our enterprise monitoring and response solutions (Kaspersky SIEM and Kaspersky XDR) use machine learning for risk scoring when evaluating the behavior of devices and servers within the infrastructure. This helps identify hidden attacks and anomalies without sharing data outside the company, which is especially important for organizations with stringent confidentiality requirements.

We develop AI for industry and physical objects

Artificial intelligence is used for more than just protecting computers and networks. In industry, equipment failures and errors can lead to downtime, accidents and serious financial losses.

Machine learning-based solutions such as [Kaspersky MLAD](#) (Machine Learning for Anomaly Detection), which provides predictive analytics, are used to address such scenarios. These solutions analyze equipment telemetry and help identify early (hidden) signs of impending equipment failure, process disruptions, cyberattacks and human errors. By continuously training the neural network, MLAD analyzes the stream of "atomic" events from an object, structures the stream into patterns, and identifies abnormal behavior.

We prevent AI from being abused by malicious actors

Attackers actively use artificial intelligence to automate their work: creating phishing resources, accelerating the creation of malicious code, scaling up fraudulent schemes and creating audio and video deepfakes.

Our analysis of detected threats created using AI lets our specialists improve the effectiveness of protection against malware, phishing and scams, and then share recommendations with users. For example, Kaspersky experts recommend that users be suspicious of any unexpected messages, audio, and video, always double-checking information, and using reliable security solutions.

We research large language models

Generative AI and large language models have already become part of digital reality. We are developing infrastructure for researching and safely using their capabilities and for rapid prototyping. We deploy LLM tools, such as ChatGPT, in this environment, where they are available to employees across all departments for day-to-day tasks while also serving as a foundation for developing new solutions.

A key practical scenario is using language models to assist analysts. In particular, the [Kaspersky Threat Lookup](#) service (part of the [Kaspersky Threat Intelligence Portal](#)) now enables AI-enhanced open-source intelligence search, providing customers with summaries and article abstracts related to analyzed objects in the OSINT (Open-Source Intelligence)¹ tab, saving them time when searching for IoCs (Indicators of Compromise) or researching cybersecurity reports.

¹ Open-Source Intelligence - a branch of intelligence that analyzes information about people or organizations from sources available to the public

Examples of detected malware that was created using AI or that has an AI theme

- [FunkSec](#). Our analysis of the ransomware revealed signs of automated code generation. The targets include public sector, IT, financial, and educational organizations in Europe and Asia.
- [RevengeHotels](#). We detected a new wave of attacks on hotels that aims to steal bank card data. Samples created using AI have been identified in the campaign.
- Malware disguised as [DeepSeek and Grok](#). Campaigns with fake pages that distributed a stealer, a malicious PowerShell script¹, and a backdoor. Links to one of the malicious resources were posted online, including on the social network X (formerly Twitter).
- [BrowserVenom](#). A phishing resource imitating the DeepSeek website invited visitors to download a model for Windows, but in reality, it distributed a Trojan that intercepts traffic.
- [Jarka](#). Malicious packages distributed through the Python Package Index (PyPI) repository under the guise of tools for neural network-based chatbots infected devices with a stealer.
- [Gipy](#). The downloader was distributed under the guise of a neural network-based voice-changing app.

Our AI research

Loose-lipped neural networks and lazy scammers:

we analyzed the artifacts that LLMs may leave behind on phishing and scam pages (including characteristic phrases as well as traces in markup).

Why and how people manipulate neural networks:

we found out how and why people use indirect prompt injections² — for example, to point large language models at their resumes, etc.

Deepfake Services Are Now 400 Times Cheaper and More Accessible:

we analyzed multiple Russian- and English-language platforms and uncovered ads on darknet offering real-time video and audio deepfake services.



How we manage AI safety

We view AI safety as a complex issue that extends far beyond AI technologies themselves. It isn't just about protecting algorithms and models, but also about how AI is used at Kaspersky, what data it processes, what risks it creates, and how these risks are managed.

AI safety is multifaceted. It involves:

- legal and compliance issues—for example, what data is permissible to use and send to cloud-based AI services
- IT and information security processes, including access management, configuration control, and preventing the use of so-called "shadow AI," i.e. when employees use external AI tools without approval;
- processes used to develop and train models, where it is important to understand data sources, possible biases, model vulnerabilities, and scenarios where models could be abused.

Effective management of these risks requires specialists from various fields: information security, IT, and legal teams, as well as data science and machine learning experts. Ideally, dedicated AI security specialists would coordinate the work of these teams.

We also recognize AI security is a very rapidly evolving field. New approaches, vulnerabilities, attack methods and security tools are constantly emerging. Therefore, we believe AI risk management is an ongoing process that requires knowledge, practices and training [materials](#) to be regularly updated.

¹ A PowerShell script used by attackers to covertly execute commands, download malicious code or control an infected system.

² A method of manipulating AI in which instructions for the model are hidden within the data it is working with, rather than being fed to it directly.

We foster cooperation and share expertise

We grow our AI expertise systematically, performing fundamental research, building infrastructure for model training, and implementing practical applications in products and services. We believe that sustainable AI development is only possible with the active exchange of knowledge. Accordingly, Kaspersky participates in international initiatives and industry alliances, collaborates with the professional and academic communities, and appears in industry rankings.

In 2024, Kaspersky joined the Global Community on Artificial Intelligence for Industry and Manufacturing ([AIM Global](#)), which was created in 2023. AIM Global brings together governments, international organizations, commercial companies and industry leaders. By participating in AIM Global, we are able to exchange expertise, help develop unified approaches to applying AI, and support the development of technologies that reflect ethical, social, and technological considerations.

In 2025, Kaspersky joined the ranks of organizations supporting the [UN Global Digital Compact](#). The document sets out objectives and principles that will help achieve an inclusive, open and sustainable digital future.

We share our AI expertise with communities. Some research, such as our work on [monotonic machine learning algorithms](#) or [the application of neural networks to spam detection](#), is published as academic papers at leading machine learning conferences, and other research is published on specialized portals and at information security conferences.

For example, we publish research on the security of our own AI algorithms, including by writing about simulated attacks on [spam detection](#) and [malware detection](#) algorithms. We research the use of neural networks for [time series analysis](#).

We also release training materials and courses, including for information security professionals and developers, to help them safely implement AI solutions and consider potential risks.

In 2025, Kaspersky released a [course](#) to train developers and information security specialists in the fundamentals of protecting systems based on large language models. We also continually update our [expert training](#) portfolio to ensure our training materials meet the needs of businesses, government agencies and academic institutions.

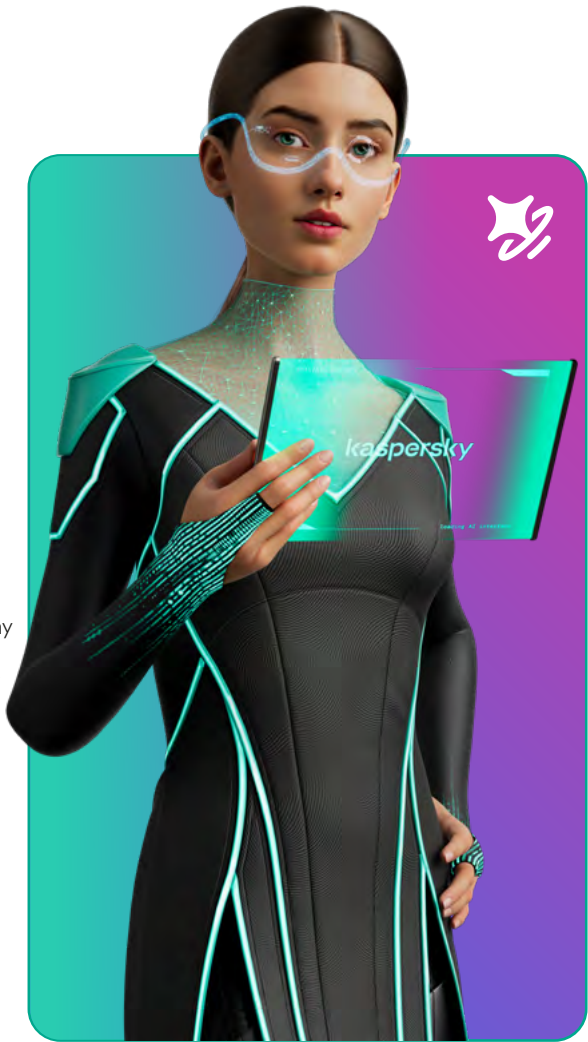


Kaspersky Unified Monitoring and Analysis Platform

Kaspersky's AI and ML expertise is also recognized externally. In 2025, Kaspersky won the IT Leader award in the Artificial Intelligence category. The award was given to our SIEM system: [Kaspersky Unified Monitoring and Analysis Platform](#) (KUMA) with the built-in AI-powered Kaspersky Investigation and Response Assistant (KIRA).



In 2025, Kaspersky also placed fifth in the Russian company Smart Ranking's ranking of AI firms, having demonstrated significant growth in revenue from sales of solutions using AI and machine learning. And in 2024, Kaspersky was recognized among key employers in Russia's AI sector based on the results of a TAdviser study.



We adhere to principles for responsible and ethical use of AI

We believe that artificial intelligence should be used responsibly and transparently, especially in such a sensitive area as cybersecurity. Accordingly, Kaspersky regularly participates in the development of legislation, policies, and other documents covering various aspects of security in working with new technologies, including AI.

At the 2024 UN Internet Governance Forum (IGF), Kaspersky [presented the Guidelines](#) for Secure Development and Deployment of AI Systems.

Earlier, in 2023, we formulated and then publicly presented at the IGF the first [principles for the ethical use of AI](#) in cybersecurity, which we adhere to in our work.

- **Transparency.** We explain how AI works: clients have the right to understand where and why machine learning technologies are used.
- **Security.** We make security a priority: all AI systems undergo validation, testing, and specialized audits. We are taking measures to minimize dependence on third-party datasets in the training of AI-powered solutions.
- **Human control.** Human experts can always intervene, check, and adjust the algorithms when analyzing complex threats.
- **Confidentiality.** We take steps to protect data and systems to ensure the digital privacy of our clients.
- **Commitment to cybersecurity goals.** By focusing exclusively on security technologies, we fulfill our mission to build a safer world and demonstrate our commitment to protecting users and their data.
- **Openness to dialogue.** We share best practices related to the ethical use of machine learning algorithms with all stakeholders.

Plans for 2026–2027

Kaspersky plans to expand its portfolio of machine learning-based solutions, complementing it with technologies that cover the full cycle of work with cybersecurity solutions. We seek a reasonable balance between resources used and results achieved: we use AI methods that are as reliable and fast as possible, while also being effective and understandable.

Our highest-priority plans include:

- develop classic statistical and ML models for detecting various types of malware, content attacks and anomalies, including searching for attacks involving lateral movement and unauthorized use of accounts.
- apply generative AI to new kinds of tasks inaccessible to traditional ML algorithms, primarily by expanding KIRA's skills and KIRA's use in various products—from Kaspersky SIEM to Kaspersky Container Security.
- develop a paradigm for agentic AI: create KIRA skills that combine existing decision functions and help in multi-step investigation and response scenarios without a pre-defined script. Such deep integration may require the creation of MCP (Model Context Protocol) interfaces to provide the AI model with access to our solutions' built-in tools.
- enter the market for vulnerability management solutions. A new AI-powered product will help organizations quickly identify and fix vulnerabilities and configuration errors in their IT infrastructure.