

GeeTest Adaptive CAPTCHA Technical White Paper

01 Introduction to GeeTest Adaptive CAPTCHA

02 The Human-or-Bot Dilemma

- Bad bots represent a serious threat to digital businesses
- Traditional CAPTCHAs and Their Limitations
- GeeTest's Breakthroughs

03 Bot Management Techniques of GeeTest Adaptive CAPTCHA

- Proactive CAPTCHA challenge updates
- Adaptive bot traffic detecting and labeling methods

04 Involve GeeTest Adaptive CAPTCHA in Your Business

- Tailor security schemes to your business
- First-rate service you can depend on
- Move closer and be responsive to your customers

05 Compatibility

06 Further Reading

07 Summing Up

Introduction to GeeTest Adaptive CAPTCHA

Preventing online resources from being abused by cybercriminals and fraudsters has long been an issue for website and app developers. The demand for bot protection and expectations of less user friction led to the creation and evolution of CAPTCHA technology.

But many website and app developers have discovered that bad bot traffic has become harder to detect and deter. That's where GeeTest Adaptive CAPTCHA comes in.

This white paper provides a current landscape of human/bot verification and the technology behind GeeTest Adaptive CAPTCHA.

GeeTest Adaptive CAPTCHA delivers a convenient and secure experience that protects websites, apps and average users from malicious bots. GeeTest's security-first dynamic approach helps online business take the initiative in bot detection and fraud prevention.

The Human-or-Bot Dilemma

Bad bots represent a serious threat to digital businesses

Bad bots now frequently become the first sign of online fraud in digital businesses, and pose a serious threat to customers of websites and mobile apps.

1. Bot status quo

Bots are computer programs that are automated to do certain tasks and interactions with web properties. Bots are not necessarily good or bad in the first place. It is the intent of bot operators that decides whether bots are good or bad. Good bots, such as search engine spiders, analytics tracking software, and chatbots, etc, help to improve efficiency and productivity for Internet users, while bots can also be manipulated and used for malicious goals, like hijacking website's bandwidth, posting spam, and stealing information through online platforms, etc.

Bad bot sophistication levels have been developing fast in recent years. Imperva classifies bad bots as the following four types:

Simple

- This group uses automated scripts to connect websites from a single IP address.

Moderate

- This group is more complex by using headless browser software to simulate browser technology.

Sophisticated

- This group can mimic human behavior, like mouse movements and clicks, to deceive bot detection methods.

Advanced

- This group is the mix of sophisticated and moderate bad bots.

Bad bot traffic accounted for 27.7% of all Internet traffic in 2021. It is the third year of continuous increase in bad bot traffic. And among all bad bot traffic, advanced bots are also on the rise. It increased from 16.7% in 2020 to 25.9% of all bad bot traffic in 2021.

2. How bots target online businesses?

The Open Web Application Security Project (OWASP) lists common types of bad bots in its Automated Threat Handbook. It classifies the frequently happening bot attacks on online businesses by industries and lists possible symptoms of each attack.

Based on GeeTest data studio, organizations in the e-commerce, blockchain, video games, and financial service are repeatedly targeted by cybercriminals. The four most common bad bot attacks are:

Account takeover (ATO) attacks

A form of online identity theft where a malicious actor illegally takes ownership of online accounts and gains access to confidential data (i.e. passwords, credit cards, email, bank accounts, SIM card and more).

Credit card fraud

It is a type of bot attack that involves multiple simultaneous requests to grant access to a stolen credit card and make transactions through it.

Web scraping

It is also known as screen scraping, web data extraction, or web harvesting, an automated bot operation aimed at acquiring, extracting, and stealing large amounts of content and data from websites.

Ticket scalping

Scalping, also known as bulk purchase, isn't limited to ticket selling. Bad actors obtain sought-after goods/services by unfair methods and then resell them at a profit.

Traditional CAPTCHAs and Their Limitations

Text-based CAPTCHA distinguishes human users from bots based on the premise that human beings can recognize specified text, but bots can't. The image below shows examples of the commonly seen traditional text-based CAPTCHAs.



1. Weakening security

Traditional text-based CAPTCHAs often use methods such as image noise, rotation, distortion, and adhesion to process texts in order to prevent computers from recognising CAPTCHA challenges. However, the security of text-based CAPTCHA has been gradually weakened with the development of computer technology, especially artificial intelligence (AI) and machine learning (ML). Bots can recognize distorted text more accurately than ever before. There are numerous mature and open-source CAPTCHA bypass techniques, and the solving rate of text-based CAPTCHA can be up to 95%.

2. Increasing user friction

CAPTCHA needs to deal with both human users and bots, which means it has to ensure a good user experience (UX) and reliable security at the same time. The horrible user experience of traditional text-based CAPTCHAs doesn't do any good to website owners and individual users. Vague text increases user friction, leading to customer dissatisfaction and loss.

Today's evolving and complex network environment offers an even bigger challenge for network security solutions. The security strategies of traditional text-based CAPTCHA can neither guarantee the security of website nor ensure a good user experience

GeeTest's Breakthroughs

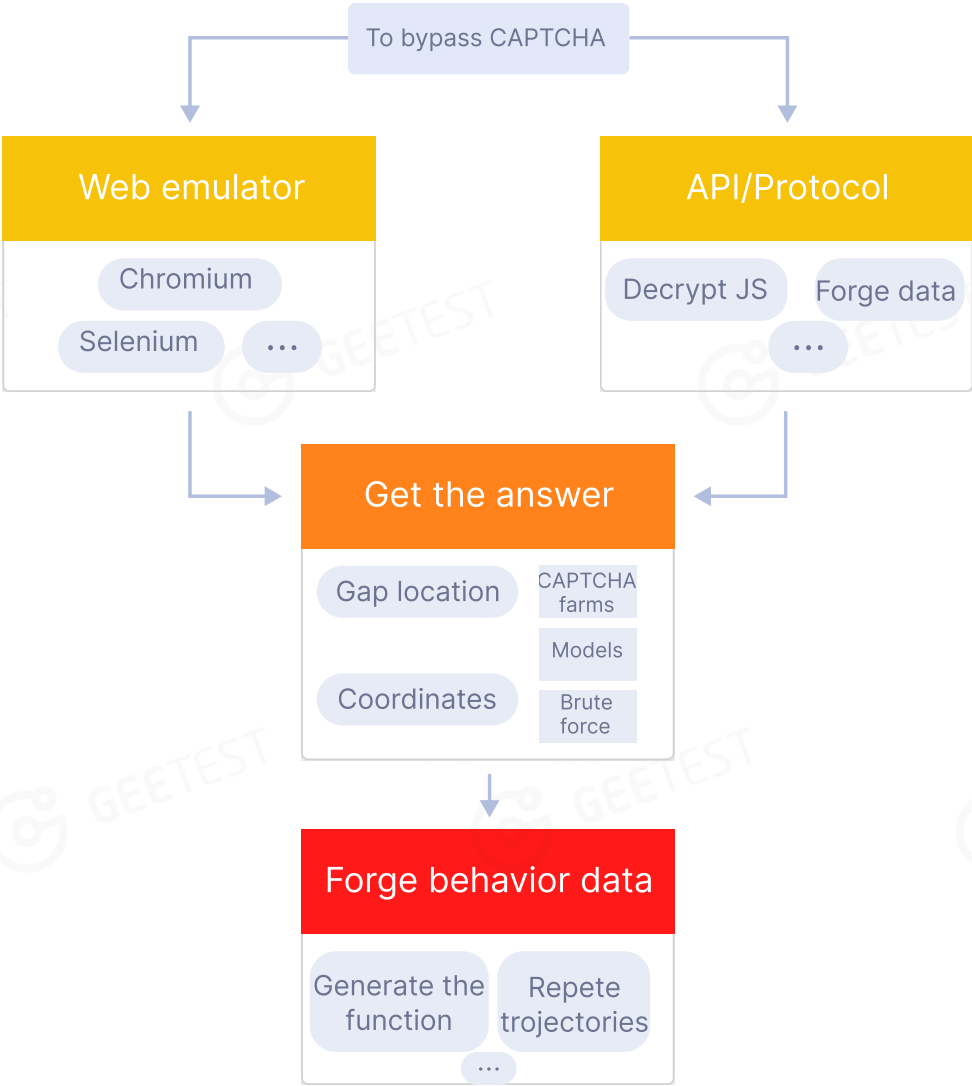
GeeTest has been developing AI-powered bot mitigation solution since 2012. It launched GeeTest CAPTCHA v1 (Behavior Verification) in 2013, v2 (Intelligent Verification) in 2015, and v3 (One-Click Verification) in 2017. In 2021, GeeTest released CAPTCHA v4 (Adaptive Verification), which adapts to bot attacks through dynamic and active security strategies in real-time with minimal user data while ensuring a smooth user experience for average users. It now protects over 320,000 websites, mobile apps, and APIs from automated bot-driven attacks, like ATO, credential stuffing, ticket scalping, etc.

The techniques details are explained in the following section.

Bot Management Techniques of GeeTest Adaptive CAPTCHA

GeeTest CAPTCHA v4 (Adaptive Verification) represents an innovation in terms of security and user experience of human/bot verification solutions.

GeeTest divides the ways of bypassing CAPTCHA into two categories. One is to bypass CAPTCHA with browser emulators that simulate human behaviors to solve CAPTCHA challenges, and the other is to directly send requests for completing verification through APIs or protocols.



No matter what methods that CAPTCHA solvers use, all roads lead to the same destination: getting the correct answer to a CAPTCHA challenge.

To get CAPTCHA answers, attackers need to

- collect as many CAPTCHA image samples as possible
- label the samples regarding different CAPTCHA types
- train a model based on the labels and samples
- connect the model to an API or protocol to solve CAPTCHA challenges

Therefore, preventing attackers from getting CAPTCHA answers becomes a primary ability of CAPTCHA solutions. As most CAPTCHA challenges are image-based, image processing and updating turn into significant CAPTCHA techniques.

Ways of getting CAPTCHA answers are roughly divided into two groups: method of exhaustion and mapping, and image recognition and labeling method.

- Method of exhaustion and mapping records the mapping relationship between all CAPTCHA challenges (images) and answers.
- Image recognition and labeling method recognizes image contents and then labels them as different groups.

Proactive CAPTCHA challenge updates

1. Anti-recognition model

GeeTest CAPTCHA v4's anti-recognition model is derived from a variety of image recognition models that GeeTest collected through years of bot defence. The anti-recognition model includes a pixeled-based confusion matrix to mislead attacker's image recognition model while leaving the images that individual users see unchanged, so that GeeTest CAPTCHA could prohibit attackers from collecting and labeling CAPTCHA image samples for browser emulator CAPTCHA solving by regularly updating images through the anti-recognition model.

Image rendering and stylization is also an important step in processing CAPTCHA images. Image processing methods will format different combinations to handle bot attacks on variable conditions, and the combination changes dynamically based on image recognition algorithms. As for API/protocol CAPTCHA solving, GeeTest adds new captcha tokens to JS, and every version of the token is unique. Each normal request should carry the newest token when it's returned to GeeTest server for validation, except for the requests sent directly from APIs/protocols. So GeeTest could label the requests without the newest tokens as risky ones and take further actions without alerting them.

2. Image database update

CAPTCHA image dataset is the key to getting CAPTCHA answers. Attackers target one dataset to crawl all the images and use the method of exhaustion and mapping to get the answers. GeeTest CAPTCHA v4 updates image dataset on an hourly basis, and can schedule updates for special occasions, like big deals, promotions, and campaigns, etc to adapt to client's business.

Adaptive bot traffic detecting and labeling methods

Security strategies and CAPTCHA challenges jointly form GeeTest CAPTCHA solution. CAPTCHA challenges build a scenario where human users can see and interact with as direct protection against bad bots and attackers. While security strategies that are imperceptible to human users serve as indirect protection.

GeeTest CAPTCHA v4 provides diverse security strategies to assist clients' security systems to accurately detect malicious bots and human-driven attacks more than ever before.

1. Dynamic Encryption

- JavaScript Obfuscation

GeeTest CAPTCHA v4 applies the ability of automated attacking analysis and anti-reverse engineering to JavaScript code, thus it enables JS code to defend itself. If JS code is detected to be tampered with or suspicious debugging activity is running, a combination of anti-tampering and anti-debugging techniques would undermine the execution of the program.

- Parameter Encryption

GeeTest encrypts front-end parameters of CAPTCHA v4 with encryption algorithms such as AES and RSA in order to hide the real trajectory and CAPTCHA answer, therefore attackers can not bypass CAPTCHA challenges by forging the parameters.

- Honeypot Method

Dynamic tokens are added to the front-end JS of GeeTest CAPTCHA v4 to confuse attackers. The dynamic token would be detected in the validation process, so that GeeTest could tell which are legit requests and which are not.

2. RAD Labeling

GeeTest CAPTCHA v4 provides a Real-time Anomalous Data Labeling (RAD Labeling), which labels features of suspicious requests during the verification process. RAD labeling operates based on GeeTest risk dataset, including the list of high risk IP addresses, the list of high risk device fingerprint, and the list of high risk trajectory data, etc. GeeTest risk dataset plays as a mesh engine to identify bot traffic across devices, networks, and user behaviors. Combining with honeypot method and other dynamic encryption, RAD labeling can pass the labeled data to website/app owners for further risk control decision making.

3. Behavior Analysis Models

GeeTest has been using Graph Convolutional Neural Network (GCN) since 2018 to upgrade defense models and improve data analysis ability. Before then, Convolutional Neural Networks (CNN) was the dominant model in GeeTest product.

GCN is an approach for semi-supervised learning on graph-structured data. A special feature of GCN model is to capture correlations from unstructured data. GeeTest uses GCN model for image composition training with the accumulated samples, and the trained model is deployed to the server to help image classification.

The GCN model helps to improve the accuracy of human/bot verification by quickly finding the patterns among the data.

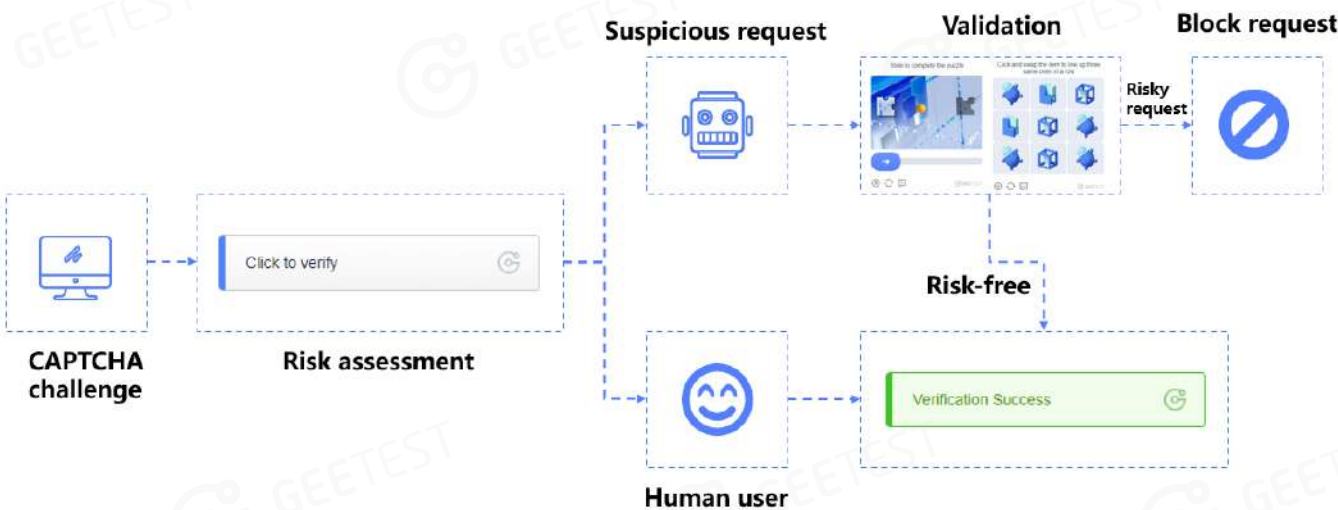
Involve GeeTest Adaptive CAPTCHA in Your Business

Tailor security schemes to your business

Three security schemes can be customized to meet various needs.

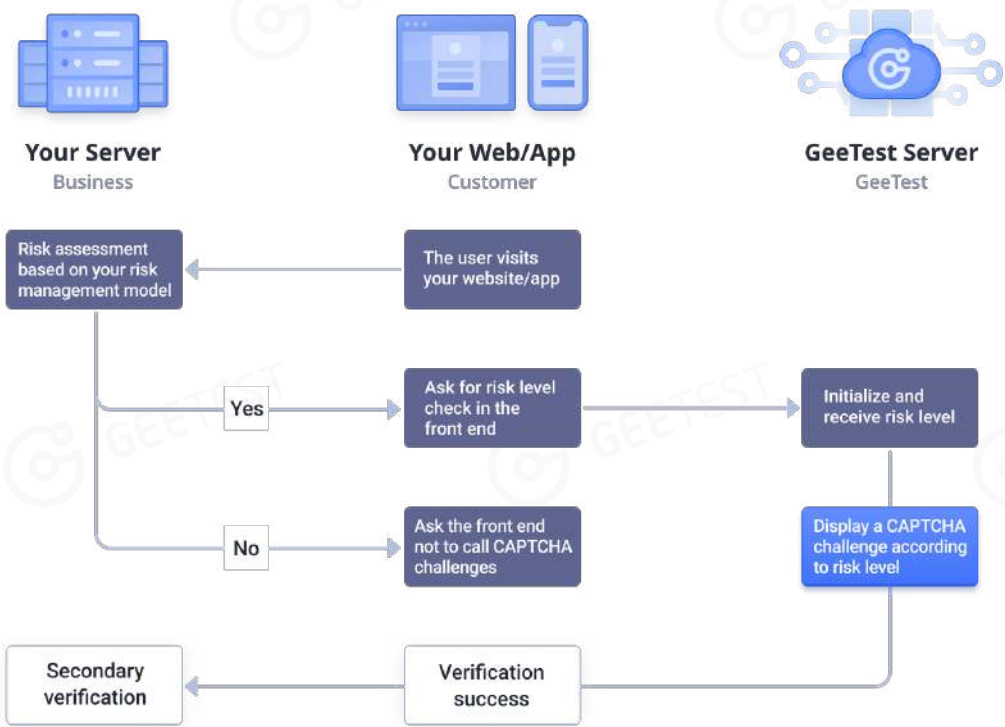
1. Intelligent Mode

In intelligent mode, GeeTest CAPTCHA v4 conducts a comprehensive risk assessment of whether the current request is risky or not by analysing the device environment and each request’s behavior data. CAPTCHA types and difficulty will change automatically based on the settings and the risk level of the request.



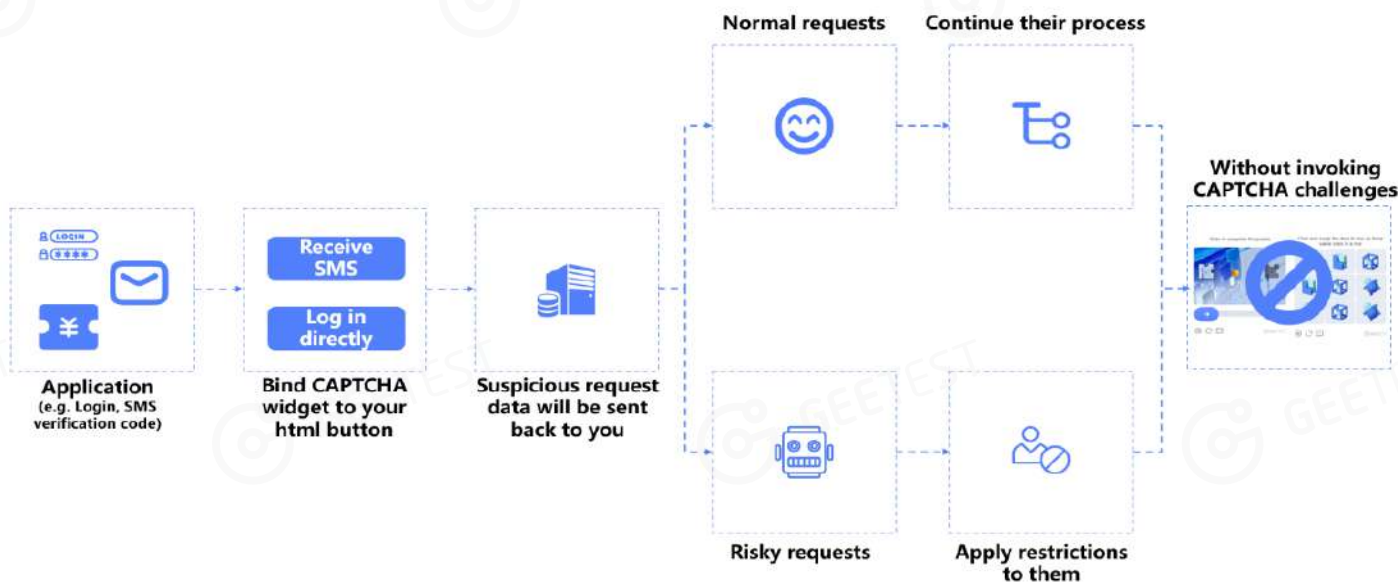
2. Direct Platform Integration

In this mode, GeeTest CAPTCHA works based on the risk assessment made by website/ app's security system and displays proper CAPTCHA challenges according to the risk level that website/app's security system assessed. It is designed for organizations with complex use cases to best suit their customer profile and business patterns



3. Invisible Mode

GeeTest Invisible Mode is a "No CAPTCHA Display" bot mitigation solution that identifies legitimate users and suspicious requests based on the network environment and user's behavior. It balances security and user experience to avoid unnecessary interruption to users. Each request will go through risk assessment and be labeled with risk level in this mode. The data will be returned to website/app owners via API in real-time and merge with their security system for further analysis and management.



First-rate service you can depend on

1. Visualized data analysis offers insights on bot mitigation

GeeTest CAPTCHA v4 offers an all-in-one bot traffic analysis dashboard to help website/app owners better understand attack trend on their business while they use GeeTest CAPTCHA, and allows them to make quick changes to system settings, such as security scheme, security strategy, CAPTCHA interface customization, etc to swiftly adapt to the ever-changing attack and fraud trend.

2. Customize settings for changing threats and commercial requirements

- Custom domains

A localized domain is essential for online businesses. GeeTest provides domain configuration solutions for clients who have such requirements. GeeTest splits domains into two types: dynamic and static. The solution supports http and https access, multi-domain disaster recovery, and will enable the parked domain once an error occurs.

- Multilingual Support

GeeTest CAPTCHA v4 supports up to 78 languages including English, Simplified Chinese, Traditional Chinese, Japanese, Indonesian, Korean, Russian, Arabic, Spanish, French, German, Uyghur, Brazilian Portuguese, European Portuguese, Latin American Spanish, etc.

- Customized image album

Customized CAPTCHA image albums assist website/app owners with commercial purposes, like brand building, better UX, etc. You can customize images for GeeTest Slide CAPTCHA and Icon CAPTCHA. Besides, updating CAPTCHA images on a regular basis helps to increase CAPTCHA security.

3. Audio challenge ensures accessibility for all

GeeTest team developed an “audio challenge” to make CAPTCHA v4 compliant with WCAG 2.0 and accessible for all users.

If you require any further information, please refer to the Further Reading section .

Move closer and be responsive to your customers

1. Global-scale service deployment

GeeTest service is deployed on servers around the world. User's CAPTCHA request and validation can be processed on the nearest server, respectively, to reduce the verification time for both websites/apps and individual users. GeeTest CAPTCHA v4 reduces the verification time to one third of its CAPTCHA v3, and the service cluster supports up to 70,000 QPS.

2. Prompt service recovery

GeeTest CAPTCHA v4's service recovery plan is designed from three aspects:

- Business Disaster Recovery: It ensures unblocked business processes (i.e., business continuity) of websites and apps in the downtime of CAPTCHA service.
- Database Disaster Recovery: By binding the metadata to the process and removing the central database, GeeTest improves service stability, and because of the loose coupling, GeeTest CAPTCHA v4 can rebuild the service in a few seconds.
- O&M Disaster Recovery: GeeTest invests a lot effort and resources in bandwidth construction to keep global services robust and stable.

Compatibility

GeeTest CAPTCHA v4 is compatible with most PC browsers, mobile browsers and various devices, and keeps updating its compatibility.

Device	System	Document type
Android	Android 4.4+	SDK
iOS	iOS 8+	SDK
web	IE9+, Chrome, Firefox, Safari, Opera, mainstream mobile browsers, and embedded WebView on iOS and Android	JS

Further Reading

For the latest details on GeeTest CAPTCHA v4 dashboard, security configuration, and deployment, please see:

[GeeTest Product Resources Library](#)

[GeeTest Developer Docs](#)

Summing Up

GeeTest CAPTCHA v4 (Adaptive Verification) provides you with best-in-class bot management and mitigation without compromising user experience. It's easily deployed, secure, user-friendly, and ensures privacy of both websites/apps and their customers.

Trusted by over 320,000 enterprises worldwide

Try Demo Now! →



Claim a free trial

I'm ready to talk

About GeeTest

A CAPTCHA and bot management provider founded in 2012 protects websites, mobile apps, and APIs from automated bot-driven attacks, like ATO, credential stuffing, ticket scalping, etc.

GeeTest has been developing human-bot verification technology since 2012 and has been studying bot traffic and its intent ever since. Now it processes 1.6 billion CAPTCHA requests daily and serves companies in sectors like blockchain, online games, e-commerce, etc. GeeTest CAPTCHA has been recognized by Gartner and Forrester in market reports for three consecutive years since 2020.

Awards and Recognition

- Gartner - Representative Vendors in OFD: Bot Mitigation 2020
- Forrester - Selected Vendor in Forrester Now Tech: Bot Management, Q4 2021
- Deloitte - 2021 Deloitte Optics Valley Technology Fast 20 & Rising Star program
- ...