Anti-Spam Report

Abstract

This report analyzes five anti-spam filters—Akismet, CleanTalk, OOPSpam, ModerationAPI, and ALTCHA Spam Filter—evaluating their GDPR compliance, accuracy, latency, and features. Most solutions struggle with data privacy compliance and show variable accuracy, often failing to detect spam reliably or generating false positives. Latency issues also affect their suitability for real-time applications.

ALTCHA Spam Filter excels with high accuracy, comprehensive language support, and robust privacy measures, making it the preferred choice for users needing a fast, reliable, and GDPR-compliant anti-spam solution. This report highlights the need for anti-spam filters that ensure both high detection accuracy and adherence to data privacy standards.

Preface

Server-side anti-spam solutions have been available for a significant period, including widely-known options such as Akismet and CleanTalk. However, their adoption outside of websites powered by WordPress has been limited, primarily due to their low classification accuracy and lack of advanced features, as illustrated in this report.

For many years, CAPTCHAs have been the dominant solution in the anti-spam market, providing reliable spam protection. However, with the emergence of new machine learning systems and Large Language Models (LLMs), it is now possible to achieve fast and highly accurate text classification, making server-side and real-time spam filters a viable option for websites and APIs.

Spam Filters Over CAPTCHAs

When protecting your website from spam, choosing the right solution is crucial. While CAPTCHAs are popular, server-side anti-spam solutions offer several compelling advantages that make them superior:

1. **Seamless User Experience**: CAPTCHAs can frustrate users with their often difficult and time-consuming challenges. Server-side solutions work silently in the background, providing a seamless experience without disrupting user engagement.
2. **Enhanced Security**: Server-side anti-spam solutions utilize advanced algorithms and machine learning to detect and block spam with higher accuracy. This approach can detect even sophisticated human and AI-generated spam, ensuring robust protection for your website.
3. **Accessibility**: CAPTCHAs can pose accessibility issues for users with disabilities, making it difficult for them to navigate your site. Server-side solutions eliminate this barrier, ensuring your website is inclusive and accessible to all users.
4. **Faster Performance**: CAPTCHAs can slow down page load times, negatively impacting user experience and SEO rankings. Server-side solutions operate efficiently without affecting the speed of your website.

In this report, we evaluate several commercially available spam filters—**Akismet, CleanTalk, OOPSpam, ModerationAPI, ChatGPT, and ALTCHA Spam Filter**—measuring their performance and accuracy of classification, and evaluating their privacy policies to highlight potential privacy and compliance issues.

Methodology

This section outlines the systematic approach employed to evaluate the performance, accuracy, and compliance of various spam filters. The methodology is designed to ensure fairness, transparency, and reproducibility of the results.

Objectives and Scope

The primary objective of this evaluation is to measure the performance and accuracy of selected spam filters in classifying user-submitted data across multiple languages. Additionally, the privacy policies of these filters are examined for compliance with GDPR and other relevant regulations.

Accuracy Tests

To evaluate the accuracy of the spam filters, we prepared a comprehensive set of 50 test samples, generated by ChatGPT. These samples are designed to reflect common messages submitted via contact forms, including both legitimate (non-spam) and spam content. The samples cover text in seven languages: English, German, Spanish, French, Italian, Portuguese, and Dutch. The evaluation criteria are as follows:

- **Legitimate Samples**: Messages that are genuine and non-spam.
- **Spam Samples**: Messages that contain spam or harmful content, including pills and drugs advertisements, profanity, and harmful code (e.g., SQL/HTML injection).

Each sample was submitted to the spam filters using their respective HTTP APIs. The classification results were compared with the expected outcomes to calculate accuracy scores. A test is considered *passed* if the platform correctly classifies the sample as either spam or non-spam.

Latency Tests

To measure the average latency of each spam filter, we conducted five consecutive classifications using the same content (text only). The latency, defined as the full HTTP round-trip time, was recorded. Tests were conducted from AWS data centers located in two regions:

- **EU Region**: AWS eu-central-1
- **US Region**: AWS us-west-1

By testing from these locations, we ensured a comprehensive assessment of the response times from different geographical points.

Statistical Analysis

The accuracy was determined by the proportion of correctly classified samples out of the total number of samples. Latency was averaged across the five tests for each region to provide a representative measure of performance.

Transparency and Reproducibility

The source code and samples used in this evaluation are available on GitHub [1]. This transparency allows for independent verification and reproduction of the tests, ensuring the credibility of our findings.

[1] https://github.com/altcha-org/altcha-vs-akismet

Akismet

Akismet [2] is a leading anti-spam solution widely deployed across WordPress websites.

Privacy and GDPR Compliance

Operated by Automattic Inc., a US-based company, Akismet's data practices raise concerns for GDPR compliance. According to their privacy policy [3] (as of April 22, 2024):

- Data is transferred to the USA and other non-EU countries.
- Submitted data, including text, email addresses, and IP addresses, may be stored for up to 90 days.
- User data is used for model training, with potential indefinite storage.

These practices present challenges for GDPR compliance, caution is advised for users seeking privacy-conscious solutions.

Features

Akismet's API integrates tightly with WordPress, offering basic features such as text classification, email and IP address verification. The API lacks customization options, operating with minimal user input.

Latency

Akismet demonstrates consistent server latency across regions:

Table 1: Akismet latency

Origin	Latency
EU (AWS eu-central-1)	~92ms
US (AWS us-west-1)	~95ms

Accuracy

Akismet exhibits significant limitations in accuracy:

- **Legitimate Samples:** Akismet correctly identified legitimate samples in non-English languages and two out of three English samples.
- **Spam Samples:** Akismet showed poor performance in detecting spam across different languages and types.

Table 2: Akismet accuracy - 46%

Sample Type	Accuracy
Legitimate Samples	20/21
Spam Samples	3/29

[2] https://akismet.com
[3] https://automattic.com/privacy/

CleanTalk

CleanTalk.org [4] is an anti-spam service widely used on platforms such as WordPress, Joomla, and Drupal, specializing in comment classification.

Privacy and GDPR Compliance

Operated by CleanTalk Inc., a US-based company, CleanTalk.org's data handling practices pose challenges for GDPR compliance. Key aspects from their privacy policy [5] (March 09, 2021) include:

- Data is transferred to the USA and other non-EU countries.
- Submitted data is retained for 31 days.
- Use of third-party cookies for promotional activities related to CleanTalk services.

CleanTalk's privacy policy raises concerns for GDPR compliance and may not be suitable for users prioritizing privacy-centric solutions.

Features

CleanTalk's API, akin to Akismet, provides basic functionalities for comment classification, requiring the user's IP address without offering detailed classification insights.

Latency

CleanTalk's API demonstrates notable latency fluctuations during testing, with peak latencies reaching up to 2.2 seconds, indicating potential reliability issues.

Table 3: CleanTalk latency

Origin	Latency
EU (AWS eu-central-1)	~165ms
US (AWS us-west-1)	~285ms

Accuracy

CleanTalk's API demonstrates notable limitations in accuracy:

- **Legitimate Samples:** CleanTalk accurately classified legitimate samples in English and several non-English languages (Spanish, French, Italian, Dutch, Portuguese).
- **Spam Samples:** CleanTalk performed poorly in identifying spam, particularly in non-English text, where it achieved a 0% accuracy rate. The system also reported three false positives in the "German Legitimate" samples.

Table 4: CleanTalk accuracy - 42%

Sample Type	Accuracy
Legitimate Samples	18/21
Spam Samples	3/29

CleanTalk's performance highlights challenges with non-English content and specific types of spam detection.

[4] https://cleantalk.org
[5] https://cleantalk.org/publicoffer#privacy

OOPSpam

OOPSpam [6] is an anti-spam service claiming to emphasize privacy-focused spam detection.

Privacy and GDPR Compliance

Operated by OOPSpam LLC, a US-based company, OOPSpam's data practices require caution for users concerned with GDPR compliance. Key points from their privacy policy [7] (February 07, 2024) include:

- Data transfers to the USA and other non-EU countries.
- Use of user data to improve the service, potentially involving AI model training.
- Utilization of cookies and similar tracking technologies, including third-party affiliate marketing cookies.

OOPSpam's privacy policy raises concerns for GDPR compliance and may not suit users seeking stringent privacy standards.

Features

OOPSpam's API offers advanced features, providing more details into the classification results and the ability to filter by language and country.

Latency

During testing, OOPSpam's API exhibited significant latency fluctuations, with peak latencies reaching up to 6.5 seconds, indicating potential reliability issues.

Table 5: OOPSpam latency

Origin	Latency
EU (AWS eu-central-1)	~1443ms
US (AWS us-west-1)	~1110ms

Accuracy

OOPSpam's API exhibits significant limitations in accuracy:

- **Legitimate Samples:** OOPSpam effectively classified legitimate samples in English and various non-English languages (Spanish, French, Italian, Dutch, Portuguese).
- **Spam Samples:** OOPSpam struggled with identifying spam, particularly in non-English text, achieving a 0% accuracy rate for languages other than English. Additionally, the system reported three false positives in the "German Legitimate" samples.

Table 6: OOPSpam accuracy - 46%

Sample Type	Accuracy
Legitimate Samples	18/21
Spam Samples	5/29

These results highlight OOPSpam's challenges in accurately detecting non-English spam and the occurrence of false positives in certain scenarios.

[6] https://www.oopspam.com
[7] https://www.oopspam.com/privacypolicy

ModerationAPI

ModerationAPI [8] is a service for content analysis and moderation, capable of classifying text and images for quality, toxicity, and spam.

Privacy and GDPR Compliance

ModerationAPI, operated by CUDE Development based in Denmark (EU), has data handling practices that may present challenges for GDPR compliance. Key points from their privacy policy [9] (February 10, 2021) include:

- Data sharing with third-party vendors, service providers, contractors, or agents, including Google.
- Data retention for "as long as necessary to fulfill the purposes."
- Use of cookies and other tracking technologies to collect and store information.
- Use of custom models powered by OpenAI's ChatGPT, raising privacy concerns.

ModerationAPI's privacy policy poses potential issues for GDPR compliance, making it less suitable for users seeking stringent privacy protections.

Features

ModerationAPI provides a modern API for data classification, allowing users to customize models for both text and images.

Latency

ModerationAPI demonstrated stable latency with minor fluctuations. However, the API appears to be hosted primarily in Europe, resulting in higher latencies for US users.

Table 7: ModerationAPI latency

Origin	Latency
EU (AWS eu-central-1)	~464ms
US (AWS us-west-1)	~802ms

Accuracy

ModerationAPI exhibits notable limitations in accuracy:

- **Legitimate Samples:** ModerationAPI struggled with legitimate samples, achieving lower accuracy across various languages, with several false positives.
- **Spam Samples:** ModerationAPI correctly detected spam across all tested languages, demonstrating a strong ability in spam identification.

Table 8: ModerationAPI accuracy - 66%

Sample Type	Accuracy
Legitimate Samples	6/21
Spam Samples	27/29

The service effectively detected spam in multiple languages but struggled with legitimate content, leading to a high rate of false positives.

[8] https://moderationapi.com
[9] https://moderationapi.com/privacy-policy

ChatGPT

OpenAI's ChatGPT [10] is a widely recognized and capable LLM system. For these tests, the chatgpt-3.5-turbo model was used as a customizable spam filter.

Privacy and GDPR Compliance

ChatGPT is operated by OpenAI OpCo, LLC, a US-based company, and caution is advised for customers seeking GDPR compliance. Key points from their privacy policy [11] (November 14, 2023) include:

- Data transfer outside your jurisdiction, to the USA.
- Use of data for "automated decision-making, profiling or AI training."
- Provision of personal information to vendors and service providers.

OpenAI's privacy policy may present challenges for GDPR compliance, making it less suitable for users prioritizing stringent privacy measures.

Features

ChatGPT offers advanced text classification capabilities, but it lacks specific features like email verification and IP address verification.

Latency

ChatGPT's latency is a notable concern, with fluctuations leading well over 1 second. Being hosted in the US, it results in higher latencies for EU users.

Table 9: ChatGPT latency

Origin	Latency
EU (AWS eu-central-1)	~828ms
US (AWS us-west-1)	~490ms

Accuracy

ChatGPT demonstrates exceptional accuracy but has some limitations:

- **Legitimate Samples:** ChatGPT accurately classified all legitimate samples across multiple languages (English, German, Spanish, French, Italian, Dutch, Portuguese).
- **Spam Samples:** ChatGPT effectively detected spam across all tested languages and scenarios. However, it failed in email verification and IP verification (TOR), which are not currently supported features.

Table 10: ChatGPT accuracy - 96%

Sample Type	Accuracy
Legitimate Samples	21/21
Spam Samples	27/29

Overall, ChatGPT achieved a 96% accuracy rate, correctly classifying 48 out of 50 samples, demonstrating high reliability in spam detection.

[10] https://openai.com
[11] https://openai.com/policies/privacy-policy/

ALTCHA Spam Filter

ALTCHA Spam Filter [12] is a new player in the anti-spam market, offering a privacy-respecting, GDPR-compliant solution with advanced features, such as language detection, security filter, and geo-fencing.

Privacy and GDPR Compliance

ALTCHA's Spam Filter is operated by BAU Software s.r.o., a private EU-based company in Czechia, committed to data privacy and regulatory compliance. Key points from the privacy policy [13] (June 21, 2024) include:

- Ensures local data residency and regulatory compliance, allowing customers to choose their region (EU or US).
- Does not store submitted data, use it to train AI models, or share data with any third parties.
- Does not log requests.
- Allows email verification using anonymized email addresses, avoiding the need to send PII.

For detailed GDPR compliance information, visit the documentation [14].

Features

ALTCHA's Spam Filter provides comprehensive spam detection and content moderation with advanced features. It includes automatic language detection, email verification with anonymization, IP address verification, and robust security filters for harmful content like SQL and HTML injections. Additionally, it supports geo-location and geo-fencing for enhanced customization and security.

Latency

ALTCHA demonstrated stable latency across both regions, with minimal fluctuations.

Table 11: ALTCHA latency

Origin	Latency
EU (AWS eu-central-1)	~35ms
US (AWS us-west-1)	~36ms

Accuracy

ALTCHA's Spam Filter exhibited exceptional accuracy:

- **Legitimate Samples:** ALTCHA accurately classified all legitimate samples across multiple languages (English, German, Spanish, French, Italian, Dutch, Portuguese).
- **Spam Samples:** ALTCHA effectively detected spam across all tested languages and scenarios, with no false positives. It also correctly identified harmful content (SQL and HTML injections), providing an additional security layer.

Table 12: ALTCHA accuracy - 100%

Sample Type	Accuracy
Legitimate Samples	21/21
Spam Samples	29/29

Overall, ALTCHA achieved a perfect 100% accuracy rate, correctly classifying all 50 samples, making it an ideal solution for both English and non-English speaking markets.

[12] https://altcha.org/anti-spam/
[13] https://altcha.org/privacy-policy/
[14] https://altcha.org/docs/api/gdpr/

Conclusion

The evaluation of various anti-spam filters reveals two critical areas of concern: compliance with data privacy regulations and the accuracy of spam classification.

Firstly, most anti-spam filters struggle with GDPR compliance. This is a significant issue as these filters often handle customer data that may include personal information. Using a non-compliant platform not only risks legal penalties but also undermines user trust and damages the company's reputation. Even in regions not bound by GDPR, respecting customer privacy has become a standard expectation and best practice.

Secondly, the accuracy of these filters varies significantly. Many of the tested solutions showed poor performance, either generating false positives or failing to detect actual spam. An unreliable spam filter defeats its own purpose, necessitating additional measures or manual reviews to ensure data integrity.

Latency is another important factor for real-time spam filtering. Filters need to operate swiftly to avoid impacting website performance. The tests revealed that while some platforms maintained stable latency, others exhibited fluctuations that could hinder real-time usability. However, these platforms might still be suitable for asynchronous data processing.

Among the tested solutions, ALTCHA's Spam Filter stood out with its exemplary performance and accuracy across all languages. Its commitment to GDPR compliance and robust privacy policies make it an ideal choice for those seeking a reliable, real-time spam filtering solution. ALTCHA not only meets the demands for high accuracy but also ensures data privacy, positioning it as the preferred option for modern, privacy-conscious applications.

FAQ

Where can I learn more about ALTCHA?

To learn more about ALTCHA's Spam Filter, visit the website or read the documentation.

Website: https://altcha.org/anti-spam
Documentation: https://altcha.org/docs/api/spam-filter-api/

What makes ALTCHA's Spam Filter fast and accurate?

Traditional anti-spam solutions like Akismet use Bayes Classifiers, which rely heavily on large amounts of high-quality training data. Despite Akismet's access to extensive data, its classification often suffers due to poor data quality.

ALTCHA's Spam Filter uses a different approach, similar to LLMs like ChatGPT, by methodically analyzing text and detecting common patterns. This method increases accuracy and reduces false positives. As a single-purpose program, its analytical components are optimized for speed.

What are the plans for supporting other languages?

ALTCHA's Spam Filter currently includes full support for English, German, Spanish, French, Italian, Portuguese, and Dutch. While it is possible to classify text in other languages, most text classification rules are language-agnostic, meaning the results may vary depending on the input. We are actively working on expanding full support to additional languages soon.

What types of content can ALTCHA's Spam Filter detect?

ALTCHA's Spam Filter can detect various types of content, including spam, harmful content like SQL and HTML injections, and profanities. It also supports email and IP address verification, as well as geo-location and geo-fencing for added security.

Is ALTCHA's Spam Filter suitable for real-time applications?

Yes, ALTCHA's Spam Filter is designed for real-time applications, offering minimal network latency and reliable performance to ensure it doesn't negatively impact website speed or user experience.

How easy is it to integrate ALTCHA's Spam Filter with my website or application?

ALTCHA's Spam Filter provides comprehensive API documentation and support, making it easy to integrate with your website or application using the HTTP API. Detailed guides and examples are available to assist developers with the integration process.

Does ALTCHA's Spam Filter support customization?

Yes, ALTCHA's Spam Filter offers a range of customization options, including geo-location and geo-fencing, and the ability to fine-tune text classification to meet specific needs.

Legal Notice

Attribution Requirement

In case of publication in printed media or websites, attribution to Altcha.org as the author of this report is required.

Disclaimer of Warranty

The information provided in this report is for informational purposes only. Altcha.org makes no representations or warranties of any kind, express or implied, about the completeness, accuracy, reliability, suitability, or availability of the information contained herein. Any reliance you place on such information is strictly at your own risk.

Limitation of Liability

Altcha.org shall not be liable for any loss or damage arising from the use of this report or its contents. This includes indirect, incidental, consequential, special, or punitive damages.

Intellectual Property Rights

All intellectual property rights in this report are owned by Altcha.org and it's legal operator BAU Software s.r.o.

Trademark

ALTCHA is a pending trademark filed with EUIPO, No. 019015396.