Report under Regulation (EU) 2022/2065 of the European Parliament and of the Council (Digital Services Act) on all moderation of content on the Allegro platform for 2025.

I. Introduction

Allegro sp. z o.o., as a marketplace subject to the provisions of the Digital Services Act, hereby submits this report on content moderation. The report has been prepared in accordance with the requirements of the Regulation and includes information on the measures taken as a result of the application and enforcement of the terms of service.

The key activity of Allegro sp. z o.o. is running an e-commerce platform (marketplace) that allows sellers to offer and sell products on the online market and reach their target buyers. On its main market, in Poland, Allegro manages an online platform connecting sellers and buyers (Allegro.pl). In addition, the platform also operates in Czechia, Slovakia, and Hungary. Sellers on Allegro offer a wide range of products in categories such as: Automotive, Home and Garden, Culture and Entertainment, Collections and Art, Fashion, Electronics, Kids, Health, Beauty, Sports and Travel, and Supermarket.

The rules for the sale of products and detailed information for sellers regarding allowed products, as well as the restrictions on the sale of products on specific markets, are presented in the Allegro Terms & Conditions.

Download the report for 2025 (.csv)

1. Summary of the content moderation engaged in at the providers' own initiative

Over the past year, Allegro has been actively implementing a comprehensive content moderation system, combining proactive and reactive measures to ensure transaction security and compliance of offers with the law and the Terms & Conditions. A key element of the strategy involves proactively searching for violations using advanced systems that analyze the database of offers to identify prohibited content. At the same time, the platform provides intuitive forms that allow users to report potentially illegal information.

For moderation of content on the platform, Allegro applies a diverse range of remedies:

- **In the area of sanctions on offers and users' accounts:** The most common restrictions include removal of the offer, and in the case of more serious violations — temporary suspension of the account. Moderation in this respect is both reactive and proactive.
- **In the area of reviews and ratings:** Moderation is mainly reactive. As a result of legitimate reports, the relevant content is hidden from public view. The platform

allows restoring a review if the user edits it in accordance with the Terms & Conditions.
- **For the Allegro Community**: Measures include removing content and the temporary or permanent suspension of the users' account if the Allegro Terms & Conditions have been violated.

In accordance with the policy of promoting good practices, Allegro applies "soft" restrictions. Instead of permanently removing content, the system may require the user to edit it, which allows correcting errors while maintaining sales activity. These measures are complemented by a broad educational program, including webinars, articles, and tutorial messages displayed directly while listing offers.

The detection systems are based on advanced algorithms analyzing text, images, and historical account data, which allows the creation of risk profiles. These processes are supported by:

- Specialist teams that analyze complex violation instances.
- Cooperation with law enforcement agencies and brand owners in the intellectual property rights protection program.
- Continuous real-time monitoring, which allows responding immediately to attempts of fraud and manipulation, such as false reviews

Summary: Own-initiative moderation measures are based on an integrated approach that uses modern technologies for proactive detection. As a result, the platform can effectively reduce the visibility of non-compliant content using a variety of measures – from the complete removal of offers and suspension of users' accounts, to flexible mechanisms allowing for self-correction of erroneous content.

2. Meaningful and comprehensible information regarding content moderation engaged in at the providers' own initiative

Allegro takes a multi-level approach to identifying potentially illegal content or content non-compliant with the Terms & Conditions, based on advanced technologies and expert knowledge. Key methods include:

- **Automated AI-based systems:** Using artificial intelligence algorithms to analyze text and images to detect patterns, keywords, and contexts that suggest violations.
- **Behavioral verification and risk profiling:** Creating risk profiles of users based on historical data, such as transaction history, communication patterns, and the type of products offered.
- **Real-time monitoring:** Continuous verification of activity on the platform, including automatic analyses of the content of messages between users and the content of new offers, which allows for immediate reaction in the event of fraud attempts (antifraud) and manipulation.

Depending on the channel of detection and the type of content, the platform takes various remedial steps:

- **Proactive measures:** The algorithms verify violations in offers and accounts, which may result in automatic prevention of offer listing, account suspension, or securing of funds.
- **Reactive measures:** In areas such as product reviews, seller ratings, and the Allegro Community forum, moderation is mainly based on user reports via intuitive forms. Once the violation has been confirmed, the content is immediately hidden from public view.
- **External cooperation:** As part of the Rights Protection Cooperation Program, Allegro allows brand owners to quickly report intellectual property violations through dedicated tools (fast-track procedure).

Automated systems are closely integrated with the work of specialized teams, in particular in the field of intellectual property protection. These experts conduct an in-depth analysis of cases flagged by the systems as complex, which ensures the high quality of moderation decisions and minimizes the risk of errors. In addition, regular trend analysis allows adapting detection methods to evolving threats, such as new fraud techniques.

Summary: Allegro's own-initiative moderation system is based on the synergy of advanced technological solutions and specialist human knowledge. By combining proactive AI-based monitoring with a reactive reporting system, the platform provides comprehensive protection for users against illegal content and fraud.

3. Qualitative description of the automated means

As part of its automated measures, Allegro uses advanced automated moderation mechanisms, the effectiveness of which is monitored for different types of interventions. In the current reporting period, the following accuracy indicators have been recorded:

- **Automatic moderation of offers and accounts:** Allegro algorithms show very high effectiveness in detecting content that directly violates the Terms & Conditions and is blocked automatically. This high effectiveness results from the use of a multi-layered control system based on the analysis of texts, images, and behavior patterns.
- **Identification for manual moderation**: Mechanisms classifying content for manual verification by moderators use broad filtering criteria (higher sensitivity indicator/recall) to ensure that any potentially risky content is flagged and is then ultimately evaluated by a human.

Accuracy indicators are shaped by a number of input criteria, such as analysis of keywords, regular expressions, historical user activity data, and risk profiles. The calculation methodology is based on:

- **Periodic audits**: Regular checks of monitored systems and tools allow verifying the correctness of content classification.
- **Analysis of internal signals:** Monitoring misclassification indicators and unexpected system behavior allows correcting algorithms on an ongoing basis.
- **Cooperation with experts:** Automated results are verified against the knowledge of specialists, which allows for precise determination of the level of hits (true positive rate).

To minimize the risk of incorrect decisions (false positives), Allegro uses strict security procedures:

- **Test environments:** Each new rule or algorithm undergoes tests before implementation to verify its effectiveness and exclude potential errors.
- **Continuous improvement:** Based on audit results, corrections are made that reduce the response time and increase detection accuracy.
- **Human verification:** In complex cases, the system automatically forwards the content to moderators, which prevents unjustified suspensions and allows for the context to be taken into account.

Summary: Automated moderation systems demonstrate very high effectiveness in directly detecting and blocking content that violates the rules in offers and users' accounts. In areas requiring deeper analysis, algorithms provide a selective filter, classifying potential breaches for final evaluation by expert teams. Thanks to regular audits and constant optimization of mechanisms, the platform gradually increases the precision of operations, which allows for effective minimization of errors and ensuring a high level of security without the need for excessive interference with allowed content.

4. Qualitative description of indicators of accuracy and possible rate of error of automated means

Automated content moderation measures on Allegro are based on a multi-layered surveillance system that is designed to provide the highest level of security. The key element of this solution is an automated mechanism for validating all offers, regardless of the channel in which they are created. This system works in a proactive manner, scanning new content in order to detect potential violations of the law and the Terms & Conditions.

Automated tools use advanced technical and algorithmic parameters, including:

- **Text and image analysis:** AI-based algorithms analyze descriptions and images for obscenity or threats, and recognize objects that violate rules.
- **Detection of behavior patterns:** Systems identify unusual activities, such as suspicious transactions between linked accounts or sudden sales spikes, allowing for a quick response to potential fraud.
- **Linguistic techniques:** Using keywords and regular expressions to identify prohibited phrases.
- **Substantive verification of offers:** The system automatically blocks attempts to misuse brand names, offer unauthorized software, and post contact details intending to drive traffic away from the platform.

Automated processes are closely integrated with the work of expert teams. These systems have two main functions:

- **Automatic suspension:** Content that clearly violates the law or Terms & Conditions is removed immediately without human intervention.
- **Selective flagging:** Content that requires in-depth contextual analysis is automatically sent for manual verification by moderators, which allows for precise resolution of complex cases. In addition, these mechanisms serve educational

purposes, generating messages for sellers explaining how to adapt the offer to the applicable requirements.

It should be noted that the described automated processes for detecting illegal or non-compliant content are used primarily in the area of offers and accounts.

Summary: Automated moderation measures are a key element of platform protection, proving highly effective in directly detecting and removing content that violates the rules in offers and users' accounts. In situations with a higher degree of complexity, algorithms provide a precise selection tool, forwarding selected cases to a thorough analysis by moderators.

5. Specification of the precise purposes to apply automated means

The main goal of the implementation of automated systems is to create a safe commercial environment that is fully compliant with the law and the Terms & Conditions. These systems are designed to continuously eliminate risks associated with the presence of illegal and undesirable content on the platform.

Key operational objectives:

- **Proactive prevention of violations:** Automation serves to scan all new offers to identify potential violations even before they are listed, which allows for the preventive blocking of content that violates the rules.
- **Elimination of dangerous and illegal products:** Another purpose of the systems is to prevent the listing and sale of products that may pose a risk to users or are prohibited by law.
- **Combating fraud and abuse:** The algorithms are designed to detect and block attempts at extortion, fraud, or any actions that violate the integrity of trading.
- **Protection against manipulation:** The systems prevent attempts to artificially influence the credibility of the platform, among others, by blocking the possibility of adding false reviews between linked accounts.

Automation allows for immediate blocking of payouts of funds to users against whom there is a reasonable suspicion of illegal activity.

The purpose of automated measures is not only to remove violations, but also to optimize the work of the moderator teams. With automatic processing of obvious violations, specialists can focus on analyzing the most complex and layered cases. These activities have a direct impact on increasing the trust of buyers and sellers in the platform as a reliable shopping site.

Summary Automated measures serve as the foundation of the platform’s security, with the aim of immediately responding to threats and proactively blocking illegal content. Their use effectively eliminates fraud, protects intellectual property rights and ensures the integrity of the rating system. As a result, the moderation process becomes more efficient, which allows the building of a professional and trustworthy environment for all service recipients.

6. Safeguards applied to the use of automated means

In order to ensure the highest quality and effectiveness of automated systems, Allegro has implemented a number of security mechanisms, which include technical, operational, and educational processes. The key element is to conduct operations in dedicated test environments, where each new rule or algorithm is thoroughly checked before they are implemented on the production platform. This allows their effectiveness to be verified while eliminating any potential errors that could negatively affect the user experience.

Automated systems are subject to continuous surveillance via:

- **Audits and analyses:** Regular internal audits and periodic analysis of operational data allow identifying gaps and irregularities in the operation of algorithms.
- **Monitoring internal signals:** The technical team tracks misclassification indicators and any unexpected behavior of the systems, allowing immediate response and correction.
- **Adapting to the environment:** The mechanisms are continuously optimized based on an analysis of the external environment, including new trends in the area of fraud and changing legal regulations.

Automated measures do not function in isolation, but are supported by extensive substantive background:

- **Team-to-team collaboration:** Internal teams work closely with developers to maintain and optimize the monitoring rules.
- **Documentation and standardization:** The moderation process is based on detailed internal procedures and clearly defined quality standards, which ensures the repeatability and objectivity of the decisions taken.
- **Employee education:** The staff responsible for supervising the systems have constant access to updated educational materials and participate in the processes of knowledge exchange within the organization.

Summary: The employed security mechanisms ensure that automated moderation measures operate in a predictable, secure, and highly effective manner. Thanks to the combination of rigorous testing, continuous monitoring of reports concerning irregularities, and constant development support, the platform effectively minimizes the risk of erroneous decisions. Systematic audits and educational processes allow for the quick adaptation of tools to new threats, which ensures stable protection of users and the high quality of services provided without relying solely on manual processes.