JOURNAL OF INFORMATION SYSTEM & ARTIFICIAL INTELLIGENCE

Design and Implementation of an Integrated Intelligent System for Online Content Moderation Using AI and Single Sign-On (SSO)

Syifa Ul Walidaien[1], Ari Muzakir[2]
[1,2] Informatics Engineering Study Program, Faculty of Science and Technology, Universitas Bina Darma
e-mail: [1]swalidaien@gmail.com, [2]arimuzakir@binadarma.ac.id

Submitted: December 17, 2025; Accepted: October 17, 2026

Abstract

The increase in user activity on digital platforms has driven significant growth in the volume of content published in real time. However, this development has also heightened the risk of spreading harmful content such as hate speech, spam, and misinformation, which, according to recent studies, remains one of the biggest threats in the global digital ecosystem. This research proposes the design and implementation of an integrated intelligent system for online content moderation using Artificial Intelligence (AI) and a Single Sign-On (SSO) mechanism, which has been proven to enhance authentication security and efficiency across various digital platforms since 2022. By applying the Agile methodology, system development is carried out iteratively, enabling rapid adaptation to user needs and behavioral dynamics. The evaluation results show that the integration of AI classification models and SSO can improve the speed of content moderation while strengthening user access security. These findings align with recent research trends emphasizing the importance of automated moderation and unified identity management in maintaining the security of digital spaces.

Keywords: *intelligent system, content moderation, artificial intelligence, SSO, user acceptance test*

1. Introduction

The development of digital technology in recent years has significantly increased user engagement across various social media platforms, discussion forums, and content-sharing services [1]. According to a 2023 study, the volume of digital content has grown by up to 32% annually, thereby heightening the risk of spreading harmful content such as hate speech, violence [2], pornography, and misinformation. Such content affects not only user safety but also the social stability and credibility of the digital platforms themselves [3].

Traditional moderation approaches based on manual review or rule-based methods have proven insufficient to cope with the speed of digital content distribution [4]. Research from 2022–2024 indicates that AI-based moderation systems capable of automatic classification have become the new standard for reducing the burden on human moderators while improving detection accuracy [5]. However, many moderation systems still operate independently and are not integrated, resulting in service fragmentation and inefficient authentication across digital platforms [6].

Meanwhile, Single Sign-On (SSO) technology has evolved rapidly since 2022 as a solution to simplify cross-platform authentication using centralized tokens [7]. SSO has been proven to enhance security, reduce the risk of credential leaks, and accelerate access to digital services. However, the integration of SSO into AI moderation systems has rarely been explored in depth, even though their combination has the potential to offer high efficiency in identity management and improved content detection performance.

Therefore, this study proposes the design and implementation of an intelligent AI-based content moderation system integrated with an SSO mechanism within a unified portal. The system is developed using the Agile methodology to enable rapid iteration, adaptive requirements, and continuous feature enhancement. This research contributes an adaptive

moderation model, a secure single-authentication architecture, and functional validation relevant to the needs of the digital era from 2022 onward.

2. Method

The research methodology is used to ensure that the system design and development process is carried out in a structured, systematic, and scientifically accountable manner. The selection of an appropriate method is crucial because the developed system involves the integration of Artificial Intelligence (AI) technology and a Single Sign-On (SSO) mechanism, both of which are complex and adaptive in nature. Therefore, the research approach is designed to support phased system development, enable iterative evaluation, and provide opportunities for improvement based on testing results.

2.1 Type of Research

This study uses the Research and Development (R&D) method because its main objective is to produce a new system in the form of an AI-based content moderation platform integrated with a Single Sign-On (SSO) mechanism. The R&D method was selected because it allows researchers to carry out a series of systematic processes, from needs identification, design, and development to comprehensive system evaluation. In this research context, R&D is used to produce a system prototype that can not only be tested functionally but also be developed further for large-scale implementation. This approach is highly relevant considering that digital content moderation is an evolving issue that requires adaptive systems capable of adjusting to the dynamics of user behavior and changes in harmful content types over time. In addition, the use of R&D provides flexibility for researchers to continuously improve the AI model and the SSO authentication module based on system testing results at each stage.

Furthermore, the R&D approach supports iterative refinement through evaluation and feedback cycles, which aligns with the dynamic nature of artificial intelligence development. Each development phase generates empirical data that can be analyzed to improve system accuracy, performance, and security. By applying R&D, this study does not merely focus on theoretical analysis but emphasizes the practical implementation and validation of the proposed system. This ensures that the resulting platform is not only technically feasible but also applicable in real-world environments where scalability, reliability, and adaptability are critical requirements.

2.2 Research Object

The object of this research is an integrated portal system designed to manage automated content moderation processes using AI and to regulate user authentication through an SSO service. The portal consists of several core modules, including a text analysis module for detecting harmful content, an admin dashboard for displaying moderation results, an API gateway that connects the modules, and an SSO authentication server that manages user access. This object was selected because of the growing need for reliable and secure moderation systems amid the increasing amount of digital content shared every day. The system also facilitates the integration of internal services into a single platform, allowing users to log in once to access various features. This has become increasingly important as many organizations have shifted to integrated service models to improve operational efficiency and data security.

In addition, the integrated portal system serves as a representative case of modern digital platforms that rely on centralized authentication and automated decision-making processes. By focusing on this object, the research can evaluate how AI-based moderation interacts with authentication mechanisms in a unified environment. The object also allows the analysis of system interoperability, data flow consistency, and access control effectiveness. These aspects are essential to ensure that content moderation decisions are not only accurate but also securely managed within a centralized system architecture.

2.3 Agile Methodology Approach

The Agile methodology is used as the primary approach because it provides flexibility, rapid adaptation, and continuous updates based on user feedback [8]. Agile emphasizes short iterations, continuous collaboration, and repeated improvements, which are highly suitable for developing complex AI systems and authentication services. In this study, Agile divides the development process into four main sprints, each of which produces an increment that can be tested immediately. This approach allows researchers to validate quickly, fix errors early, and add new features without disrupting the existing system structure.

Moreover, Agile supports continuous evaluation of both the functional and non-functional aspects of the system, such as performance, usability, and security. The iterative nature of Agile enables the research team to respond effectively to changes in requirements or technical challenges encountered during development. This is particularly important in AI-based systems, where model performance may vary depending on data quality and system integration. Through Agile, the development process becomes more transparent, structured, and responsive to research findings.

2.3.1 Sprint Planning

Sprint execution is the core stage of system development. In Sprint 1, the main focus is on building the system architecture, database structure, and data flow design. Sprint 2 continues with the implementation of the AI moderation model, including preprocessing, model training, performance evaluation, and the creation of APIs that connect the model with other services. Sprint 3 focuses on integrating the Single Sign-On (SSO) authentication module using OAuth 2.0 and JWT to ensure authentication security [9], [10]. Sprint 4 involves creating the admin dashboard interface and conducting comprehensive functional testing of all modules. In each sprint, the deliverables are tested and analyzed to ensure compliance with the functional and non-functional requirements.

Furthermore, sprint execution emphasizes continuous testing and validation throughout the development process. Any issues identified during implementation are documented and addressed within the same sprint or scheduled for subsequent sprints. This iterative execution approach ensures that system components evolve cohesively and reduces the likelihood of major integration problems at later stages. As a result, the overall quality and reliability of the system improve incrementally with each sprint.

2.3.2 Review and Retrospective

At the end of each sprint, a review session is conducted to assess the completed deliverables, including the testing of the developed features. Features are evaluated to determine whether they meet the specified requirements and user needs. After the review, the process continues with a retrospective session, which involves discussing the team's workflow during the sprint, including identifying aspects that worked well and elements that need improvement for the upcoming sprint. The retrospective phase plays a crucial role in improving team effectiveness and system development quality.

In addition, the review and retrospective stages provide valuable insights for refining both the technical and managerial aspects of the project. Feedback obtained during these sessions is used to adjust development strategies, improve communication, and optimize workflow efficiency. This reflective process ensures that lessons learned from each sprint contribute directly to the success of subsequent development cycles, leading to a more robust and well-engineered system.

2.4 Data Collection Techniques

This study uses several data collection techniques. A text moderation dataset is used as the primary data for training the AI model, comprising training and testing sets that contain safe and harmful text samples. Documentation of test results is used to analyze the system's performance in both AI detection and SSO authentication. User interaction observations are conducted to

understand how users interact with the system, particularly when using the dashboard and the login flow. Authentication activity logs are also examined to evaluate the validity of the SSO process, including login duration, token usage, and session security.

Combining multiple data collection techniques enables a more comprehensive evaluation of the system. Quantitative data from datasets and logs are used to measure accuracy, response time, and authentication reliability, while qualitative observations provide insights into usability and user experience. This mixed approach ensures that system evaluation is not limited to technical performance alone but also considers practical usability and security aspects, resulting in a more holistic assessment of the developed platform.

2.5 Research Flow

The research flow from the initial to the final stage is illustrated in Figure 1. The flowchart provides a visual representation of the R&D process conducted in this study, showing the logical sequence of and interconnections among the research stages. The research flow helps ensure that system development follows a structured, systematic approach, from problem identification to system deployment.

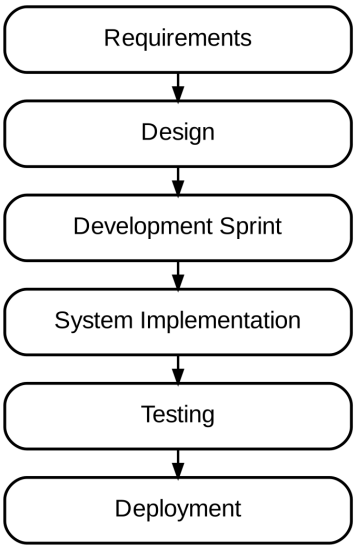


Figure 1. Research Flow

In addition, the research flow serves as a guide for researchers to control and evaluate progress at each stage. By visualizing the process, potential gaps or overlaps between stages can be identified early, allowing improvements to be made efficiently. The flowchart also supports transparency and clarity in explaining the research methodology to readers.

2.5.1 System Requirements

The system requirements stage is the initial step in the research and development process. At this stage, the researcher identifies the problems to be solved, namely the increasing spread of harmful content on digital platforms and the need for an AI-based automated moderation system. In addition, the functional requirements of the system are analyzed, such as content classification capability, user authentication through SSO, API integration, and the display of moderation results in the admin dashboard. These requirements are identified through literature studies, observations of digital platforms, and discussions with prospective system users. This stage serves as the main foundation for designing a targeted solution that meets security standards and user needs.

Furthermore, non-functional requirements are also identified at this stage, including system performance, scalability, security, and usability. These requirements ensure that the system can handle large volumes of content, provide secure authentication, and offer a user-friendly

interface. By clearly defining both functional and non-functional requirements, the development process can proceed in a more focused manner and reduce the risk of major design changes in later stages.

2.5.2 Design

The design stage focuses on preparing a comprehensive system blueprint before implementation begins. At this stage, the researcher designs the system architecture, including the relationships between modules such as the AI module, API Gateway, SSO server, and dashboard. The database structure is also designed to store moderation results, user data, and system activity logs. The researcher creates the user interface (UI) design to ensure that the admin dashboard is easy to use, informative, and supportive of the moderation workflow. The AI model is also designed by selecting algorithms, preprocessing techniques, and evaluation methods. This design stage ensures that system implementation aligns with technical standards and functional requirements.

In addition, design documentation is produced to serve as a reference during development and integration. This documentation helps ensure consistency between planned and implemented features. Proper system design also minimizes technical risks and improves maintainability, making it easier to update or expand the system in future development cycles.

2.5.3 Development Sprint

This stage involves building the system using the Agile methodology through four sprints. Each sprint has a specific objective and produces an increment or functional output. The first sprint focuses on creating the system architecture and database. The second sprint focuses on developing the AI model, from data collection, data cleaning (preprocessing), and model training to developing an API for accessing AI classification results. The third sprint covers the integration of the SSO module using OAuth 2.0 and JWT. The fourth sprint develops the admin dashboard interface and performs initial testing. This stage proceeds iteratively, with evaluations at the end of each sprint to ensure that all features meet the requirements.

Moreover, the sprint-based development approach allows continuous testing and refinement throughout the process. Feedback obtained from each sprint evaluation is used to improve subsequent development activities. This iterative mechanism reduces development risks and ensures that system components evolve in alignment with research objectives and user expectations.

2.5.4 System Implementation

The system implementation stage involves integrating all modules developed in each sprint. At this stage, the AI module is connected to the API Gateway to receive moderation requests from the dashboard. The SSO module is configured to authenticate users and issue valid tokens for accessing services. The dashboard is connected to all modules through the available API endpoints. Integration is conducted gradually to ensure that each module functions properly before full integration. The final output of this stage is a unified system ready for comprehensive testing.

Additionally, system configuration and environment setup are performed during this stage to ensure compatibility between components. Proper integration ensures seamless data exchange, stable authentication processes, and consistent moderation results. This stage is crucial for validating that all individual modules can operate cohesively within a single platform.

2.5.5 Testing

The testing stage is carried out to ensure that all system functions operate according to the requirements. Functional testing is conducted to examine all features, such as SSO login, content classification, dashboard display, and API usage. In addition, validation testing is performed to evaluate the accuracy of the AI model in detecting harmful content categories. System performance analysis is also conducted to assess API response speed, SSO token

security, and stability when handling high-volume requests. This testing is essential to ensure that the system is ready for use and free from technical issues that could disrupt operations.

Furthermore, testing results are documented and analyzed to identify system weaknesses and areas for improvement. Any issues found during testing are corrected before moving to the deployment stage. This thorough testing process helps improve system reliability, security, and overall performance in real-world usage scenarios.

2.5.6 Deployment

The deployment stage is the final phase of system development. At this stage, the tested system is uploaded to the server so that it can be accessed by users or administrators for further evaluation. Deployment includes server configuration, code deployment, creation of the production environment, and domain setup if needed. This stage ensures that the system can operate under real-world conditions and can be monitored for ongoing evaluation. After deployment, the system is ready to be used as a prototype or further developed into a full-scale product.

In addition, post-deployment monitoring is conducted to observe system behavior and performance under actual usage conditions. Feedback collected during this phase can be used as a reference for future system enhancements. This ensures that the deployed system remains adaptable, scalable, and aligned with evolving content moderation and security requirements.

3. Results and Discussion

The results describe the stages of development and testing of an online content moderation system based on Artificial Intelligence (AI) integrated with a Single Sign-On (SSO) mechanism. The discussion focuses on the results of the system architecture design, the implementation of the AI module in the content classification process, the integration of centralized authentication, and the presentation of information through an administrative dashboard. Each component is analyzed to assess the system's success in meeting the defined functional and non-functional requirements.

In addition, the results illustrate the effectiveness of inter-module integration, system performance in handling content moderation processes, and the reliability of the security and authentication mechanisms. This analysis provides a comprehensive overview of the system's capability to improve the efficiency of digital content management, strengthen access control, and support a more structured and centralized moderation process.

3.1 System Architecture Design Results

The system architecture design is a crucial stage that ensures that all major components, ranging from the Artificial Intelligence (AI) module, Single Sign-On (SSO) authentication mechanism, API Gateway, and admin dashboard to the database, can operate in an integrated manner. The architecture is designed using a modular approach, allowing each component to be updated or developed independently without disrupting the entire system. At this stage, the data flow between modules is analyzed thoroughly to ensure an efficient moderation process and secure handling of user data. The resulting architecture is shown in Figure 2.

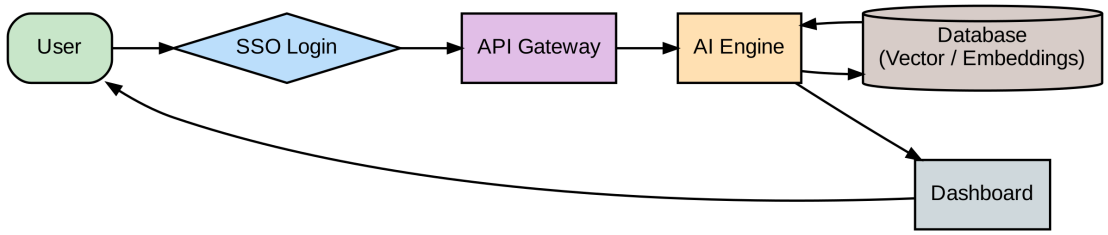


Figure 2. Architecture of the Integrated Content Moderation System

The system is designed such that every user attempting to access the service must go through the SSO authentication process. Once a user is authenticated, the SSO system generates an authorization token that serves as the identity for every request made to the API Gateway. This gateway functions as the centralized entry point for all user requests and ensures that each request is validated before being forwarded to other modules. Through this mechanism, the system prevents unauthorized access and ensures that only users with valid access rights can interact with the moderation module.

After passing through the API Gateway, the user's content is forwarded to the AI Engine module for analysis. The AI model classifies the content into categories such as safe, spam, hate speech, or other harmful content based on the data it was trained on. The AI analysis results are then sent back to the admin dashboard for visual presentation. This workflow not only ensures the system's security and processing speed but also provides administrators with the flexibility to take further action when needed.

3.1.1 User

The User component represents all individuals interacting with the system, including administrators, operators, and authorized users. Every system interaction begins with user input, whether submitting content for moderation or accessing the administrative dashboard. At this initial stage, the system does not yet recognize the user's identity, making authentication a mandatory prerequisite. This design ensures that all actions within the system can be traced back to an authenticated entity.

By positioning the user as the starting point of the workflow, the system emphasizes accountability and controlled access. Each user role determines the level of access and the available actions, which supports secure system operation. This role-based interaction model also simplifies activity tracking and auditing, especially when combined with centralized authentication and logging mechanisms.

3.1.2 SSO (Single Sign-On)

The Single Sign-On (SSO) module functions as a centralized authentication service that verifies user identities across the entire system. Upon successful login, the SSO module generates a JSON Web Token (JWT), which is used as proof of authentication for accessing other modules. This approach eliminates the need for repeated logins and significantly improves the user experience while maintaining security.

In addition to simplifying authentication, the SSO module plays a vital role in session management and access control. Token validation, expiration handling, and protection against unauthorized access attempts are managed centrally. As a result, the SSO module enhances both usability and security, making it a critical component of the integrated system architecture.

3.1.3 API Gateway

The API Gateway acts as the single entry point between authenticated users and the internal services of the system. Every incoming request must carry the JWT issued by the SSO module, and the gateway verifies the validity of this token before forwarding the request to the AI Engine, the database, or the dashboard services. Requests with missing, expired, or invalid tokens are rejected, preventing unauthorized access to the moderation module.

In addition to access control, the API Gateway routes each request to the appropriate module and returns the classification results from the AI Engine to the dashboard. Centralizing these functions in one component simplifies integration among modules, keeps the internal services decoupled from client applications, and makes it easier to monitor and log all traffic within the system.

3.1.4 AI Engine

The AI Engine serves as the core of the content moderation system. Once the API Gateway validates a request, the submitted content is processed by the AI module. The AI Engine

performs several stages, such as text preprocessing, tokenization, and classification using a pretrained model. The model is capable of detecting the following categories:

1. Safe
2. Spam
3. Hate speech
4. Other harmful content

The AI Engine generates an output containing the classification label and its probability score. These results are returned to the dashboard through the API Gateway. Additionally, the AI Engine is connected to the database to store classification results for archival and audit purposes.

3.1.5 Database

The database acts as the central repository for all system data, including user information, authentication tokens, moderation results, and activity logs. Each completed moderation process generates records that are stored for future reference, analysis, and validation. This centralized data management supports consistent information access across modules.

From a research and operational standpoint, the database enables traceability and performance evaluation. Administrators can generate reports, monitor moderation trends, and verify authentication activities. As such, the database is essential for maintaining data integrity, system transparency, and long-term scalability.

3.1.6 Dashboard

The dashboard functions as the primary interface for administrators and operators. It presents AI classification results, user authentication status, moderation statistics, and system activity logs in a structured and user-friendly manner. By translating complex AI outputs into visual information, the dashboard supports informed decision-making.

Moreover, the dashboard enables administrators to take follow-up actions, such as blocking content or reviewing flagged items. Its centralized role as a control panel enhances operational efficiency and ensures that moderation activities can be managed effectively within a single interface.

3.2 AI Implementation in Content Moderation

The implementation of the AI module begins with the data preprocessing stage, which includes text cleaning, character normalization, stopword removal, and tokenization. This stage forms a crucial foundation because the quality of preprocessing directly affects the model's accuracy in classifying content. The dataset consists of thousands of labeled text samples categorized into safe content, spam, hate speech, and potentially harmful content. This labeled dataset is then used to train the model to recognize harmful language patterns through statistical learning and feature extraction [11], [12].

During the model development stage, several text representation approaches were tested, such as TF-IDF and modern neural network embeddings. The classification algorithms were trained on these data and evaluated using metrics such as accuracy, precision, and recall. The best-performing model was selected based on its evaluation results, which are presented in Table 1. Testing showed that the AI model was able to identify harmful content with an accuracy above 90%, making it suitable for deployment within the system. This achievement demonstrates that manual moderation efforts can be significantly reduced and that potential human bias in the moderation process can be minimized.

Table 1. Performance Evaluation of the AI Classification Model

Metric	Value
Accuracy	93.8%
Precision	92.4%
Recall	94.1%

The results in Table 1 indicate that the model demonstrates excellent performance in detecting harmful content. The high precision reflects the model's ability to minimize false positives, while the high recall shows that the model is capable of effectively identifying nearly all harmful content.

3.3 Implementation and Validation of Single Sign-On (SSO)

The implementation of the Single Sign-On (SSO) module in this moderation system aims primarily to simplify the user authentication process. The SSO module is built using the OAuth 2.0 protocol, which generates a JSON Web Token (JWT). This token is then used to access all services within the system without requiring repeated logins. As a result, users only need to log in once to access all modules, including the admin dashboard, the API Gateway, and the AI moderation module.

To ensure that the SSO module functions properly, various authentication scenarios were tested. Login testing was conducted to verify that the system accepts user credentials and correctly generates authentication tokens. API testing was also performed to ensure that every user request is accompanied by a valid token. In addition, token security testing was carried out to confirm that the tokens cannot be forged or tampered with by unauthorized parties. The results of these tests show that the authentication process runs without errors and that the tokens are secure and accepted by all modules within the system.

The testing further demonstrated that authentication is significantly faster than with conventional login systems. SSO accelerates user access to various services within the integrated portal, improving both usability and efficiency. The integration of SSO also provides additional benefits, such as more structured and centralized user identity management, enabling administrators to monitor user activities more easily. The resulting login page and administrator dashboard are shown in Figure 3.



Figure 3. Implementation of the SSO Login Page and the Administrator Dashboard

3.3.1 Content Moderation Status Panel

The content moderation status panel is the primary component of the dashboard, functioning as a summary display of the AI's analysis results on incoming content. This panel presents information such as the number of items categorized as safe, spam, hate speech, and other harmful content. With this feature, administrators can immediately see the overall condition of content moderation without navigating to additional pages. The panel also displays the most recent analysis time, urgency level, and priority of content requiring review. This component helps administrators understand the content patterns circulating on the platform and determine further actions more quickly and accurately.

In addition, the status panel supports rapid situational awareness by highlighting critical trends and anomalies in the content flow. By presenting aggregated results in real time, it allows

administrators to respond proactively to sudden increases in harmful content categories. This feature not only improves reaction time but also enhances strategic decision-making in moderation management, particularly in high-volume digital environments.

3.3.2 Navigation Menu for Portal Services

The navigation menu serves as a gateway to the various features and services available within the portal. Through this menu, administrators can access pages such as content moderation, user management, SSO configuration, activity reports, and system settings. The navigation menu is designed to be concise and intuitive, allowing administrators to move between pages easily. With a structured navigation layout, admin workflows become more efficient and the risk of selecting incorrect features is reduced. This menu also accelerates monitoring and management processes, especially when handling large volumes of content.

Furthermore, a well-organized navigation structure reduces the cognitive load on administrators and supports usability principles. Clear labeling and logical grouping of menu items help ensure that critical system functions can be accessed quickly. This design choice contributes to improved operational efficiency and minimizes potential errors during system management.

3.3.3 User Activity Log

The user activity log records the history of user actions within the system, including login times, performed operations, data changes, and interactions with the moderation module. This log is essential for auditing, security monitoring, and operational analysis. By recording every activity, the system can identify suspicious actions, such as unauthorized login attempts or improper data modifications. The activity log also helps administrators and researchers understand system usage patterns and monitor the performance of the SSO module. Therefore, this element plays a crucial role in maintaining system security and data integrity.

Additionally, the activity log provides traceability and accountability for all system interactions. In the event of security incidents or system anomalies, administrators can analyze log records to identify root causes and take corrective action. This capability strengthens system governance and supports compliance with security and data management standards.

3.3.4 Moderation Summary Charts and Tables

Moderation summary charts and tables visualize AI analysis data in an easy-to-understand format. These visualizations may include bar charts, pie charts, or line graphs showing trends in content moderation over time. For example, a chart may illustrate the rise of harmful content within a specific period or compare the number of items in each category. Summary tables provide more detailed information, such as classification probabilities or a list of content that requires manual review. These visual components help administrators interpret data more accurately and make decisions more rapidly.

4. Conclusion

This research successfully designed and implemented an integrated intelligent system for online content moderation that leverages Artificial Intelligence (AI) and a Single Sign-On (SSO) mechanism. The developed system addresses key challenges in digital content moderation, particularly the high volume of user-generated content, the limitations of manual moderation, and the need for a secure, centralized authentication mechanism. Through an Agile-based development approach, each development stage was carried out iteratively and continuously, enabling systematic evaluation and refinement of each implemented module.

The system architecture was designed using a modular approach, with core components such as the AI Engine, API Gateway, SSO service, database, and administrative dashboard. This modular design provides flexibility in system development and maintenance while enabling efficient integration among modules without mutual interference. The implementation results demonstrate that the integration between the AI module and the centralized authentication

mechanism functions effectively, with every user request being validated before it is processed by the moderation system. This confirms that security and access control have been implemented optimally.

The AI model used in this research can classify content into several categories, including safe content, spam, hate speech, and other harmful content, with high accuracy. The testing results indicate that the application of AI significantly improves the efficiency of the moderation process compared with manual approaches. Moreover, the use of AI helps reduce potential subjectivity and human bias in moderation decisions, resulting in a more consistent and objective content evaluation process. Therefore, the developed system can serve as an effective solution for managing digital content at scale.

The implementation of Single Sign-On (SSO) contributes significantly to improving the user experience and system security. Through a single-login mechanism, users are required to authenticate only once to access all services within the integrated portal. The testing results show that the SSO module handles authentication processes in a stable, secure, and efficient manner, including access token and user session management. In addition to enhancing user convenience, SSO integration also simplifies centralized identity management and user activity monitoring for administrators.

Overall, the results of this research demonstrate that combining AI and SSO within a unified integrated system can enhance the effectiveness, efficiency, and security of online content moderation. The developed system functions not only as an automated moderation tool but also as a structured and accountable content management platform. In the future, the system can be further enhanced by incorporating multimodal detection capabilities such as image and video analysis, expanding integration with a wider range of digital platforms, and optimizing AI model performance to adapt to the continuously evolving patterns and types of harmful content in digital media.

References

[1] W. Tafesse and M. Dayan, “Content creators’ participation in the creator economy: Examining the effect of creators’ content sharing frequency on user engagement behavior on digital platforms,” *Journal of Retailing and Consumer Services*, vol. 73, p. 103357, 2023.

[2] D. Areri, “Social media and online hate speech: A content analysis study on Ethiopian diaspora-owned ethnically segregated social media platforms,” 2024.

[3] A. Gawer and N. Srnicek, “Online platforms: Economic and societal effects,” 2021.

[4] Q. He, Y. Hong, and T. S. Raghu, “Platform governance with algorithm-based content moderation: An empirical study on Reddit,” *Information Systems Research*, vol. 36, no. 2, pp. 1078–1095, 2025.

[5] S. Saharan, M. Sharma, M. Khiria, and A. Gupta, “Human-AI collaboration in moderating AI-generated harmful content,” in *Content Moderation in the Age of AI*, IGI Global Scientific Publishing, 2025, pp. 123–166.

[6] S. Y. Lim *et al.*, “Blockchain technology the identity management and authentication service disruptor: A survey,” *International Journal on Advanced Science, Engineering and Information Technology*, vol. 8, no. 4–2, pp. 1735–1745, 2018.

[7] L. T. T. Nguyen *et al.*, “BMDD: A novel approach for IoT platform (broker-less and microservice architecture, decentralized identity, and dynamic transmission messages),” *PeerJ Computer Science*, vol. 8, p. e950, 2022.

[8] I. Sommerville, *Software Engineering*, 10th ed. Pearson, 2016.

[9] Q. Chen and H. Zhang, “Single Sign-On (SSO) authentication mechanism based on OAuth 2.0,” *International Journal of Network Security*, vol. 22, no. 5, pp. 905–913, 2020.

[10] H. Sukmana and A. Pratama, “Implementasi OAuth 2.0 dalam sistem autentikasi terpusat pada aplikasi terintegrasi,” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 7, no. 3, pp. 435–442, 2020.

[11] P. Guzik and S. Wróbel, “Hate speech detection using machine learning techniques: A comparative study,” *Applied Sciences*, vol. 11, no. 8, p. 3414, 2021.

[12] G. Kaur and P. Kaur, “Moderation of harmful online content using AI-based classification techniques: A systematic review,” *Journal of Information Security and Applications*, vol. 67, p. 103176, 2022.