AMERICAN JOURNAL OF MANAGEMENT AND IOT MEDICAL COMPUTING
Peer Reviewed, Referred & Indexed Journal
E-ISSN: 3069-0110
Vol.5, No.2(2026)
www.ajmimc.com

REDACTIT-AI POWERED DATA REDACTION

[1] Mrs. K. Helini, [2] B. Hindu, [3] D. Sindhuja, [4] L. Chandana, [5] B. Shailaja, [6] M. Charitha
[1] Assistant Professor, [(2,3,4,5,6)] B. Tech 4th year Students,
Department of Information Technology,
Vignan's Institute of Management and Technology for Women, Hyderabad, India.
[1] helini@vmtw.in, [2] hindubejugam@gmail.com, [3] sindhujadodlapati@gmail.com, [4] chandanalodangi@gmail.com,
[5] shailajabandari01@gmail.com, [6] charitha7113@gmail.com

Abstract

Sensitive information like financial data, personal information and personal identities are more likely to be revealed due to the explosive growth of digital documents. The majority of traditional methods for redacting data are labor-intensive and manual, which often leads to errors and lack of data protection. An automated redaction system based on artificial intelligence that can effectively identify and redaction of sensitive information from digital documents and photos. To protect privacy and security, the system automatically redacts sensitive data after identifying it using machine learning techniques. An efficient and scalable solution is built by using technologies like Python, MongoDB, and web frameworks. This helps in increasing data protection, decreases manual work and improves accuracy in digital documents.

Keywords: Automated Redaction, Sensitive information detection, Data privacy, Document security, Optical character recognition.

1. Introduction

In the current digital world, a huge number of documents are generated, shared and stored electronically. Which increases the chances of leaking sensitive information like personal details, financial data and etc. Traditional redaction methods are mostly requiring a lot of time. This method is not efficient enough for redacting documents. Moreover, the traditional method involves a high chance of error due to human involvement. Therefore, to address this issue, a proposed automated redaction system based on artificial intelligence has been presented that will help in the efficient redaction of documents containing sensitive information like machine learning, image processing and optical character recognition. This proposed method will help in increasing the efficiency of the redaction method by automating the process. This will help in improving the security level of data in digital document management systems. The usage of digital technology has increased in organizations over the past few years. This has increased the need for maintaining the privacy of the data that is being shared on digital platforms. Personal identification numbers, addresses, financial data, etc., need to be maintained properly before being shared on digital platforms. An automated redaction system helps organizations maintain the privacy of their data by efficiently identifying the data that needs to be redacted. This system has integrated different technologies that help in

efficiently processing the data. This has reduced the chances of errors that might occur during the traditional method of redaction. Moreover, this method is efficient enough for redacting a number of documents at a given point in time.

2. Literature Survey

Google Cloud [1] developed a Sensitive Data Protection system that helps automatically find and hide sensitive information in large datasets. Their approach combines pattern matching with machine learning techniques to detect data like personal details and confidential records. It is especially useful for handling large-scale data efficiently in cloud environments.

Adobe Systems [2] provides redaction features in tools like Adobe Acrobat, where users can manually or semi-automatically remove sensitive content from documents. The system mainly works based on keyword searches and user selection. While it is useful for simple tasks, it still depends a lot on human effort and may not work well for complex or hidden information.

Li H., Kumar A., and Wang Y. [3] focused on improving privacy protection using Natural Language Processing (NLP). Their work tries to understand the meaning of the text rather than just looking for fixed patterns. By using techniques like entity recognition and context analysis, their system can identify sensitive information more accurately, even when it is not clearly structured.

Recent research using models like BERT [4] has further improved how systems understand text. These models can analyze the context of words in a sentence, which helps in detecting sensitive data more effectively. The process usually includes steps like text preprocessing, feature extraction, and classification to get better results.

Some studies [5] also focus on handling images and scanned documents by combining OCR (Optical Character Recognition) with machine learning. OCR helps in extracting text from images, and then NLP techniques are applied to detect sensitive information. Along with this, image processing methods like face detection and segmentation are used to hide sensitive areas in images.

3. Methodology

a. Data Input

The process starts when the user uploads a document through the system interface. The system supports multiple formats such as text files, PDFs, and images.

b. File Validation

The uploaded file is checked to ensure it is in a supported format and is safe to process. Invalid or unsupported files are rejected at this stage.

c. Data Classification

The system identifies whether the file is a text-based document or an image/scanned document. Based on this, it decides the next processing steps.

d. Text Extraction

For images or scanned documents, Optical Character Recognition (OCR) is used to extract text. This converts non-editable content into machine-readable text.

e. **Text Preprocessing**

The extracted or original text is cleaned and prepared using steps like removing noise, tokenization, and normalization. This helps improve accuracy in later stages.

f. **Sensitive Data Detection (NLP)**

Using Natural Language Processing (NLP) and Named Entity Recognition (NER), the system identifies sensitive information such as names, addresses, IDs, and other personal data.

g. **Image-Based Detection**

If the input contains images, the system applies computer vision techniques. Face detection (Haar Cascade) and segmentation models (like U-Net) are used to find sensitive visual content.

h. **Redaction Processing**

Once sensitive data is detected, the system applies redaction based on user choice:

- Black boxes
- Blurring
- Pixelation

i. **Output Generation**

Finally, the system generates the redacted document and provides it to the user in the same format as the input file.

4. Algorithm

1. **Named Entity Recognition (NER)**
   - Used to detect sensitive text like names, locations, IDs, etc.
   - Based on Natural Language Processing (NLP)
   - It uses GLiNER-based NER
   - It identifies entities from sentences using context, not just patterns
2. **OCR (Optical Character Recognition)**
   - Converts images or scanned documents into text
   - Common algorithm: Tesseract OCR
   - It converts visual text into actual text so that further processing (like NER) can be applied.
3. **Haar Cascade**
   - Used for face detection in images
   - Fast and efficient for real-time detection
4. **Redaction Algorithm (Custom Logic)**
   - Applies different methods:
     - Black boxes
     - Blurring
     - Pixelation

5. Results

AMERICAN JOURNAL OF MANAGEMENT AND IOT MEDICAL COMPUTING
Peer Reviewed, Referred & Indexed Journal
E-ISSN: 3069-0110
Vol.5, No.2(2026)
www.ajmimc.com

User Interface

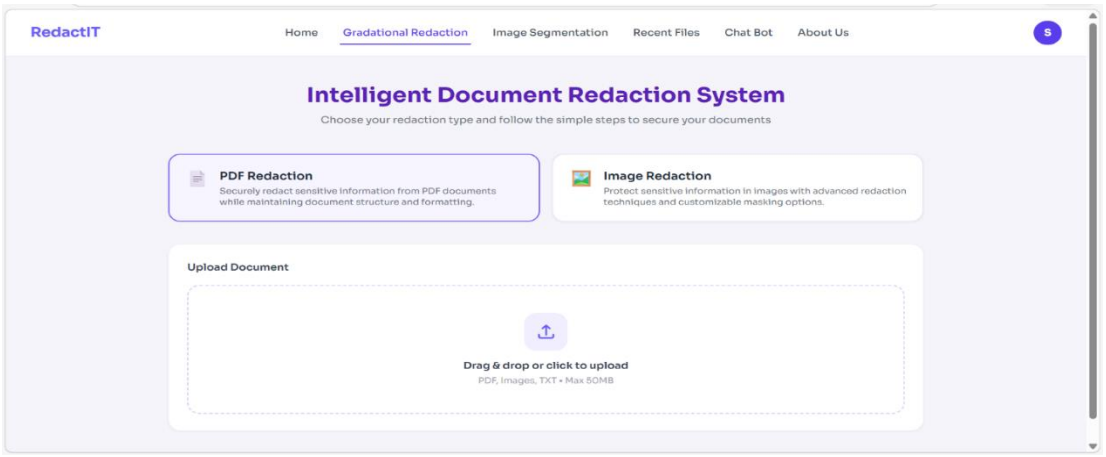


Fig 1: File Upload Process

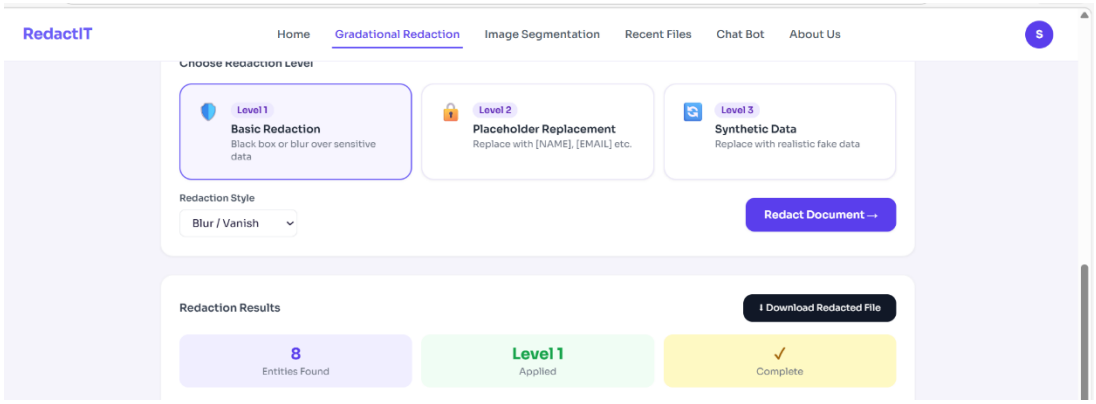
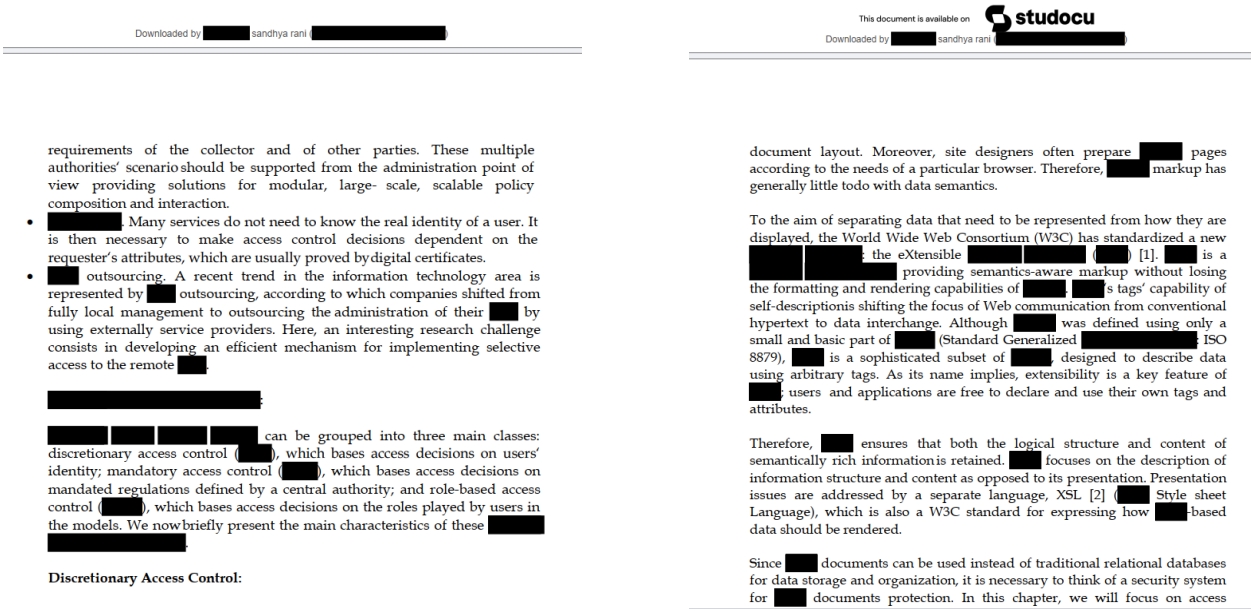


Fig 2: Redaction Process



requirements of the collector and of other parties. These multiple authorities' scenario should be supported from the administration point of view providing solutions for modular, large- scale, scalable policy composition and interaction.

- [redacted]. Many services do not need to know the real identity of a user. It is then necessary to make access control decisions dependent on the requester's attributes, which are usually proved by digital certificates.
- [redacted] outsourcing. A recent trend in the information technology area is represented by [redacted] outsourcing, according to which companies shifted from fully local management to outsourcing the administration of their [redacted] by using externally service providers. Here, an interesting research challenge consists in developing an efficient mechanism for implementing selective access to the remote [redacted].

[redacted]:

[redacted] can be grouped into three main classes: discretionary access control ([redacted]), which bases access decisions on users' identity; mandatory access control ([redacted]), which bases access decisions on mandated regulations defined by a central authority; and role-based access control ([redacted]), which bases access decisions on the roles played by users in the models. We now briefly present the main characteristics of these [redacted].

Discretionary Access Control:

document layout. Moreover, site designers often prepare [redacted] pages according to the needs of a particular browser. Therefore, [redacted] markup has generally little todo with data semantics.

To the aim of separating data that need to be represented from how they are displayed, the World Wide Web Consortium (W3C) has standardized a new [redacted]: the eXtensible [redacted] ([redacted]) [1]. [redacted] is a [redacted] providing semantics-aware markup without losing the formatting and rendering capabilities of [redacted]. [redacted]'s tags' capability of self-descriptionis shifting the focus of Web communication from conventional hypertext to data interchange. Although [redacted] was defined using only a small and basic part of [redacted] (Standard Generalized [redacted] ISO 8879), [redacted] is a sophisticated subset of [redacted], designed to describe data using arbitrary tags. As its name implies, extensibility is a key feature of [redacted]: users and applications are free to declare and use their own tags and attributes.

Therefore, [redacted] ensures that both the logical structure and content of semantically rich information is retained. [redacted] focuses on the description of information structure and content as opposed to its presentation. Presentation issues are addressed by a separate language, XSL [2] ([redacted] Style sheet Language), which is also a W3C standard for expressing how [redacted]-based data should be rendered.

Since [redacted] documents can be used instead of traditional relational databases for data storage and organization, it is necessary to think of a security system for [redacted] documents protection. In this chapter, we will focus on access

Fig 3: PDF Format

Fig 4: Image Format

After the user uploads PDF files or images into the system, the redactIT platform begins processing the input data to identify sensitive information. For image-based and scanned PDF documents, the system first extracts textual content using Tesseract OCR.

The processed text is then analyzed using Natural Language Processing (NLP) techniques. A Named Entity Recognition (NER) model is applied to identify sensitive entities such as names, addresses, identification numbers, and other confidential information.

After detecting all sensitive elements, the system applies appropriate redaction methods such as black boxes, blurring or pixelation based on user preferences. Finally, the redacted document is generated and returned to the user while preserving the original format and structure.

6. Conclusion

The automated redaction system is a way to keep sensitive information in documents safe. It uses tools to find and hide secret data like names, phone numbers and email addresses. Users can upload their documents to a website. Get them back with the secret parts hidden really fast. This system is helpful because it does the work for us. We do not have to spend a lot of time hiding information by ourselves. The system uses the internet and special text analysis tools to work with different kinds of documents.

7. Future Scope

The system can be improved to support more advanced detection of sensitive information using machine learning techniques. The system can also be improved to work with types of documents and bigger sets of data. Another thing that can be done is to make the system better, at finding and extracting text and sensitive data. The automated redaction system can also be connected to cloud storage and security systems to make it easier to access and keep data safe. The automated redaction system will be able to protect information in documents even better.

8. References

1. Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Proceedings of NAACL-HLT. Foundation transformer model widely used for entity detection in documents.
2. Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., & Dyer, C. (2016). Neural Architectures for Named Entity Recognition. Proceedings of ACL.
3. Akbik, A., Bergmann, T., & Vollgarf, R. (2019) Pooled Contextualized Embeddings for Named Entity Recognition. Proceedings of NAACL.
4. Katti, A. R., et al. (2018). Chargid: Towards Understanding 2D documents. Proceedings of EMNLP.
5. Brown, T., et al. (2020) Language Models are Few- Shot Learners. Advances in Neural Information Processing Systems (NeurIPS).
6. Dernoncourt, F., Lee, J., Szolovits, P., & Johnson, A. (2017). De-identification of Patient Notes with Recurrent Neural Networks. Journal of the American Medical Informatics Association.

7. Lison, P., & Meena, R. (2023). Named Entity Recognition for privacy Protection in Documents. ACL Conference.
8. Narayanan, A., & Shmatikov, V. (2008). Robust De-anonymization of large Sparse Datasets. IEEE Symposium on security and privacy.
9. Sweeney, L. (2002) k-Anonymity: A Model for protecting privacy. International Journal of Uncertainty, Fuzziness and knowledge-Based Systems.
10. Yang, Z., et al. (2019). XLNet: Generalized Autoregressive Pretraining for language Understanding. NeurIPS Conference.
11. Todupunuri, A. (2025). IMPROVING CUSTOMER EXPERIENCE WITH MODERN BANKING SOLUTIONS. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5120615
12. Babburi, S. (2024). Explainable AI Framework for Policy-Compliant Anomaly Detection in Data Pipelines.
13. Gaddam, S. Integrating Analytics into the Development Process: Bridging the Gap between Data Insights and Design Execution.
14. Reddy, S. K. R. Developing a Modular AI Framework to Enhance Scalability and Personalization in Next-Generation Reward Platforms.
15. Poojari, R. INTELLIGENT SYSTEMS+B108 AND APPLICATIONS IN ENGINEERING.
16. Vasagam, M. (2024, August 30). Ensuring security in modern data pipelines: Practical strategies for data engineers. International Journal of Intelligent Systems and Applications in Engineering, 12(22s), 2401.
17. Santthosh Saai Reddy Purmani. (2026). Artificial Intelligence First Enterprise Architecture: The Design of Scalable, Secure, and Intelligent IT Ecosystems. American Journal of AI Cyber Computing Management, 6(1(2)), 1–8. https://doi.org/10.64751/ajaccm.2026.v6.n1(2).pp1-8
18. Cyril, H. P., & Kumara, S. (2026, February). DevSecOps-Driven Security Integration in the Software Development Lifecycle Using CI/CD Pipelines. In 2026 IEEE 5th International Conference on AI in Cybersecurity (ICAIC) (pp. 1-6). IEEE.
19. Kotte, G. (2025). Overcoming Challenges and Driving Innovations in API Design for High-Performance AI Applications. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5283649
20. Mahtabi, M., Roshan, M., Muhit, M. M. I., Behvar, A., & Haghshenas, M. (2026). Cryogenic ultrasonic fatigue: Mechanisms, advancements, and insights. Cryogenics, 153, 104257. https://doi.org/10.1016/j.cryogenics.2025.104257
21. Viswanathan, V. (2024). Pioneering Ethical AI Integration in Enterprise Workflows: A Framework for Scalable Team Governance. Available at SSRN 5375619.

22. Akhilaiswarya, B., Sree, B. T., Lilly, K., Chowdary, K. H., & Sruthi, M. (2023). Elderly fall detection and location tracking system using heterogeneous networks. Journal of Engineering Sciences, 14(05).
23. Viswanathan, V. (2025). Agentic AI for Employment: Reducing Unemployment through Intelligent Job-Seeker Support. LEX LOCALIS–Journal of Local Self-Government.
24. Mudusu, S. K. (2026, February 9). AI-augmented data quality engineering. InfoWorld (Foundry Expert Contributor Network).
25. Viswanathan, V., Shah, A. K., Kubam, C. S., Dontu, S., Gandhi, A., & Singla, P. (2025, August). Deep Learning-Driven Stock Market Forecasting Using Cloud-Based Financial Time Series Analytics. In 2025 International Conference on Emerging Trends in Networks and Computer Communications (ETNCC) (pp. 1-6). IEEE.
26. Sruthi, M. V., Soundararajan, K., & Sree, V. U. (2012). Accurate Multimodality Registration of medical images. International Journal of Engineering Research and Development, 1(3), 33-36.
27. Viswanathan, V., Polagani, S. S., Agarwal, R., Akula, S., Dey, S., & Kashyap, R. (2025, September). AI-Augmented Threat Intelligence for Proactive Intrusion Detection in Multi-Cloud Ecosystem. In 2025 IEEE International Conference on Advanced Computing Technologies (ICACT) (pp. 567-572). IEEE.
28. Mudusu, S. K., & Gentyala, S. (2026). Zero-Trust Data Pipelines for AI Systems: A Framework for Secure, Verifiable, and Auditable Data Engineering. JOURNAL OF RECENT TRENDS IN COMPUTER SCIENCE AND ENGINEERING (JRTCSE), 14(2), 10-25.
29. DEVARASETTY, N. (2023). SCALABLE DATA ENGINEERING APPROACHES FOR AI-DRIVEN INDUSTRIAL IOT APPLICATIONS. INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH AND MANAGEMENT, 11(06), 954-968.
30. Agrawal, A. M., Gajula, S., Shinde, R. P., Shah, H., & Ghosh, H. (2025, July). Machine Translation for Long Sequences with Enhanced Attention Mechanisms. In 2025 5th International Conference on Electrical, Computer and Energy Technologies (ICECET) (pp. 1-6). IEEE.
31. Dayal, P. S., Chandra, B. R., Keerthi, M., Sruthi, M., Venkatesh, K., Appalaraju, G., & Eswari, G. (2013). Design of Pyramidal Horn Antenna at 10GHz Using WIPL-D Optimizer. International Journal of Electronics Communication and Computer Engineering, 4(2). Maturi, S. Y. (2023). Crowdsourced frontier: Unveiling autonomous adversarial cybercapabilities via open AI competition. International Journal of Intelligent Systems and Applications in Engineering, 11(1s), 275–284.
32. Hassan, T., Karim, M. F., Jeelani, H., Behnam, E., Green, R., & Syed, F. J. (2025). Optimizing Medical Question-Answering Systems: A Comparative Study of Fine-Tuned and Zero-Shot Large Language Models with RAG Framework. arXiv preprint arXiv:2512.05863.

33. Manoharan, D. (2026). Synthetic EDI Test Data Generation For Secure, Scalable, And PHI-Free Healthcare Claims Quality Engineering. Journal of International Crisis and Risk Communication Research, 9(1).

34. Ravishankara, M. (2026, February). CircuChain: Disentangling Competence and Compliance in LLM Circuit Analysis. In SoutheastCon 2026 (pp. 1-7). IEEE.

35. Sruthi, M. V., Sree, V. U., & Soundararajan, K. (2012). Specific removal of motion artifacts in medical image processing. IJECCE, 3(3), 227-229.