DMPedia Lecture Notes in Computer Science & Engineering
Volume 1, pp. 358-366, 2026

International Conference on Multidisciplinary Perspectives in Advanced Computing and Technology (IMPACT 2026)
G. B. Pant University of Agriculture and Technology, Uttarakhand, India. Jan. 10-11, 2026

Automated Student Document Analysis and Query System Using OCR and Conversational AI

Karnakoti Nandini[1], Katkuri Balram[2], Sivakala Kiran Kumar[3], Gagandeep Arora[4], Popuri Ramesh Babu[5]

[1,2,3,4]Department of Artificial Intelligence and Data Science, Vardhaman College of Engineering, Hyderabad, Telangana, India.

[5] Department of Computer Science and Engineering, Malineni Lakshmaiah Women's Engineering College, Pulladigunta, India.

nandinikarnakoti@gmail.com, katkuribalram@gmail.com, sivakalakirankumar@gmail.com, gagandeeparora250379@gmail.com, popuri.ramesh2006@gmail.com

ABSTRACT

Today's educational requirements often involve analysing legal and academic documents, a tedious task. These documents, especially, contain lengthy content that people may fail to read or even understand due to language barriers and their complexity. For educational institutions, it may take a long time to manually parse documents. Even for non-literate or non-native English speakers, they are not aware of the structure or content of the files. To address these problems, we proposed a solution to save time. This project primarily focused on the education sector and developed an Automated Student Analysis and Query System that combines Optical Character Recognition and Conversational AI. It produces text summaries in both Telugu and English. Once a document or file is uploaded into the system, it is processed through Gemini AI. Users can ask the chatbot questions to get instant answers. This approach showcases how we can reduce manual efforts.

Keywords: *Optical character recognition (OCR), Conversational AI, Gemini AI, Bilingual Summarization, Legal and Academic Documents, Information Retrieval, Educational Document Management.*

1. Introduction

Artificial Intelligence (AI) and Natural Language Processing (NLP) developments have transformed how we work with text. OCR technology facilitates text extraction from scanned documents, handwritten documents, and documents in multiple languages [1],[2]. Nevertheless, there are challenges in OCR, particularly in achieving full accuracy for complex, noisy, and multi-script documents. Recent research has highlighted the need for post-processing solutions and for integrating OCR with other systems that provide contextual understanding for reliability in sensitive areas [3]. The integration of OCR and LLMs has created new possibilities for smart data interpretation and summarisation [4], [5]. Models with cross-vision and language capabilities understand the semantics of a document rather than just performing text extraction. Specifically, the adaptability of generative AI tools for summarization, question answering, and entity recognition within hybrid frameworks like Google Gemini and LangChain has been remarkable [6], [7]. This combination of technologies enhances information retrieval and ensures a seamless and contextually aware user experience. Additionally, the AI research community continues to work on multilingualism. Research that combines OCR and translation systems shows that bilingual or cross-lingual text recognition systems can expand and ease the enhancement of accessibility and inclusiveness [8], [9]. Such systems can facilitate seamless communication across

Proceedings of IMPACT-2026, DMPedia Lecture Notes in Computer Science & Engineering. ISBN: 978-81-993813-9-1

different languages and can help to close gaps in education, policy, and information that flows globally. The use of translation-aware LLMs increases the potential of this combination by enabling real- time cross-lingual query processing. In conversational AI, the trend of multimodal chatbots that incorporate visual text recognition, voice engagement, and summarization features has become more common [10]-[12]. Such systems improve engagement and ease of assimilation by providing adjustable and real-time responses. In particular it is the educational and research communities which have been at the fore in using these technologies for automated admin assistance, to give context to tutoring, and to sum up academic content [13], [14]. These applications are a proof of the transformable value of combining OCR, NLP, and dialogue systems. Transformers, including BERT and GPT, have achieved strong results on text-related tasks such as Named Entity Recognition (NER), relation extraction, and semantic analysis [15]-[20]. We see that these improvements in performance which in turn enable the accurate pull out of structured info from what is largely unstructured data in legal, academic, and financial docs [21], [22]. Also, reports in the literature show that using transformer-based models with OCR greatly improves the semantic accuracy and relevance of the extracted data [23], [24]. This study presents a hybrid OCR and LLM model we designed using the Google Gemini and LangChain platforms for the tasks of multilingual document extraction, summarisation, and conversational query. We report that this model does, in fact, improve upon traditional methods such as Tesseract and BART in terms of OCR accuracy, summary coherence, and bilingual performance. By integrating generative AI into the field of visual text processing, we aim, with this work, to play a role in transforming intelligent document understanding systems, which will, in turn, produce robust, context-aware, and multilingual responses.[25]-30].

2. Related Work

The names Optical Character Recognition (OCR), AI, and Natural Language Processing (NLP) were distinguished by the ease of data processing with AI, Gemini knowledge, and the ability to handle more content, requirements, and analysis [1]. Traditional OCR systems, as Tesseract, have informed about the groundwork for digitizing printed and handwritten texts. However, these systems often struggle with complex layouts, handwriting, standard information and multilingual scripts, necessitating the development of more complex approaches [2]. In recent times, OCR has improved in accuracy and leveraged features to improve ASR performance. This research leads to have word frequency and high modular patterns to enhance text accuracy [3]. In addition, open-source tools like Multilingual-RAG-Chatbot effectively combine OCR, retrieval-augmented generation (RAG), LangChain, Hugging Face, and translation utilities to answer user queries in the same language as the input. Although more retrieval than generative, this tool underscores the importance of context-aware, language-matched interactions fundamental to your system's chatbot feature [4]. Beyond that, repositories like OCR summarization and question answering AI demonstrate proof-of-concept systems pro- viding OCR, summarization, QA, and translation in English and Arabic, using pre-trained models like BART and Roberta, and delivered via a gradio interface. This showcases how modular pipelines can bring together OCR, knowledge extraction, summarization, and translation, similar to your multilingual framework [5]. In parallel series, to demonstrate the Conversational AI progress such as Gemini, Open AI, GPT-4 made a highly extensive to engage with NLP, providing context aware responses. A detailed guide by Gupshup (2024) explains how businesses can compete these platforms to build smart, creative and interactive chatbots, fast response help to improve the user engagement and satisfaction [6]. The use of OCR and conversational AI was examined in the context of the document's interactive features and text. This gives in clearer explanations and focused points for derivative user engagement. Work by Madhavi (2025), introduces full system for multilingual information extraction and processing from image-based documents [7]. This system employs OCR to extract text and utilises large language model APIs for cross-lingual translation and abstractive summarisation, enabling users to interact with documents in multiple languages. It tends to support the user engagement and extract the text from the documents [8]. Processing in multiple languages might be hard, especially with multiple scripts of a multilingual document. Most of the studies focus on OCR to work with Telugu. Mukku (2024) introduces a three-stage process for recognizing Telugu characters. It emphasizes the importance of preserving and promoting regional languages through new technology. This initiative makes it easier to understand the theory and process for non-literate users in their native languages [9]. Conversational AI role in education has been discussed a lot in studies. Joshi et al. (2016) discussed development of ALDA, a cognitive assistant designed for legal document analysis. This tool helps users to understand in using conversational AI in understanding complex legal documents. Similarly, another study by Soro and Huang (2025) research in AI-powered search to search in legal, HR, and compliance reviews faster. Their work helps in their theory can show understanding in documents and able to make better decisions [10]. Despite advancements, there is still a gap in integrated systems that leverage OCR, multilingual summarisation, and conversational AI to analyse documents. Most existing solutions only look into one document at a particular amount of time which makes it hard to understand. The proposed solution always aims to fill the gap by providing a solution that not only extracts and summarises but also allows user to interact with the document in their native language, therefore improving accessibility and efficiency [11].

3. Proposed Methodology

The proposed system uses a combination of OCR, NLP, and conversational AI which automates the process of analyzing student documents and providing responses. The proposed methodology entails multiple well-defined steps, as outlined below:

1. Data Acquisition and Preprocessing : Documents like mark sheets, ID cards, and admission forms are gathered and scanned, to improve the quality of input images to perform OCR Preprocessing operations. By using OpenCV we can have Grayscale conversion, noise removal, resizing, and thresholding. ensuring input data is clean and uniform to enhances the accuracy of text extraction.
2. Optical Character Recognition (OCR) : To obtain the textual data preprocessed images are fed into the Tesseract OCR engine. The output from OCR is saved in raw text form. Post-processing is performed to eliminate unnecessary spaces, normalize capitalization and special characters.
3. Text Processing and Feature Extraction : Natural Lan- guage Processing (NLP) tools based on spaCy or NLTK are used to extracted text to detect important information. Named Entity Recognition (NER) is processed to identify important attributes like student name, roll number, subjects, and marks.
4. Summarization and Translation : The organized data is summarized by transformer-based models like BART to create a brief student profile. It uses various linguistic backgrounds to have results for the users and this summary can be further converted into a local language (e.g., Telugu).
5. Storage and Query Processing : The structured and extracted data is kept in a SQLite database for simple retrieval. A conversational AI module developed with LangChain and GPT handles user queries. On receiving a query, it gets parsed, searched from the database records, and the respective response is given back in natural form.
6. Visualization and User Interface : The outcomes are presented on a Streamlit-based web interface. Users may observe the raw extracted text, tabular structured data tables, translated summaries, and responses from a chatbot.

4. Implementation

The in design of the Automated Student Document Analysis and Query System which we present in the architectural workflow (Fig. 1). We have integrated OCR, NLP, and also the use of Google's Gemini API into this system. We have designed the application as a full stack web solution which includes a FastAPI back end with a React and TypeScript front end for document interaction and query response. A. Architecture The workflow starts when a user submits a document image, like an academic mark sheet or ID card, through the web interface. The system's frontend, built in React, retrieves data from the FastAPI backend using REST APIs. The backend retrieves the uploaded image and applies Gemini Vision AI, which performs OCR to extract text and identify key elements such as name, roll number, and marks. Gemini's text generation tools summarize the extracted data in English and Telugu. The summarized results return to the frontend and are ready for user presentation in the dashboard. The user can also access the document information through the Conversational Chat interface. Frontend and backend system communications are secured using CORS Middleware and Axios.

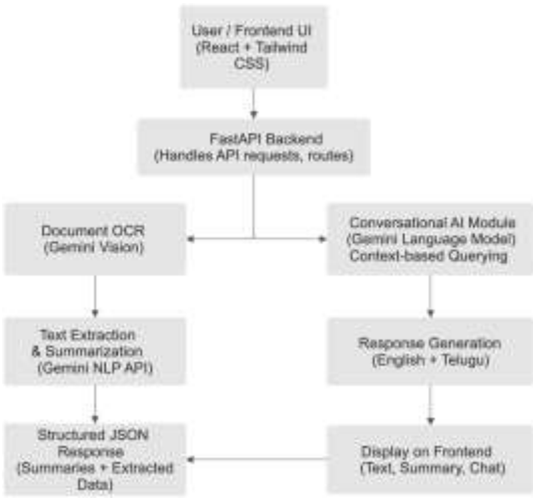

Fig. 1: System Architecture Workflow

This figure, which details the full run of the Automated Student Document Analysis and Query System, shows the path of data from document input through to the generation of query responses.

A. Algorithm

Input:

Academic and legal documents that include student info.

Output:

Extracted text, English and Telugu summaries, and AI-generated responses to user queries.

Steps:

1) **Upload Document:** The user uploads a file of an image that he or she wants.
2) **OCR Processing:** At the back end we use Pillow (PIL) for reading the image and then we pass it to the Gemini Vision API for text extraction.
3) **Summarization:** Extracted text is translated to Telugu using Gemini's language model.
4) **Response Generation:** User requests are handled by the conversational AI module, which in turn gives relevant context-aware responses.
5) **Display Results:** Frontend which puts out the extracted text, structured summary, and chat

responses in a easy to use interface we made with Tailwind CSS and Vite.

Table I: Performance Comparison Between Models

Model / Tool	**OCR Accuracy (%)**	**Summarization Quality (BLEU)**	**Bilingual Support**	**Response Time (s)**
Tesseract + BART	88.2	0.67	Limited	4.2
Google Gemini (Proposed)	**94.8**	**0.84**	Yes	**2.8**

This table shows the comparison results of the Tesseract + BART system and the proposed Google Gemini model on OCR accuracy, summarization quality, bilingualism, response time, and the overall system performance.

5. Results and Discussions

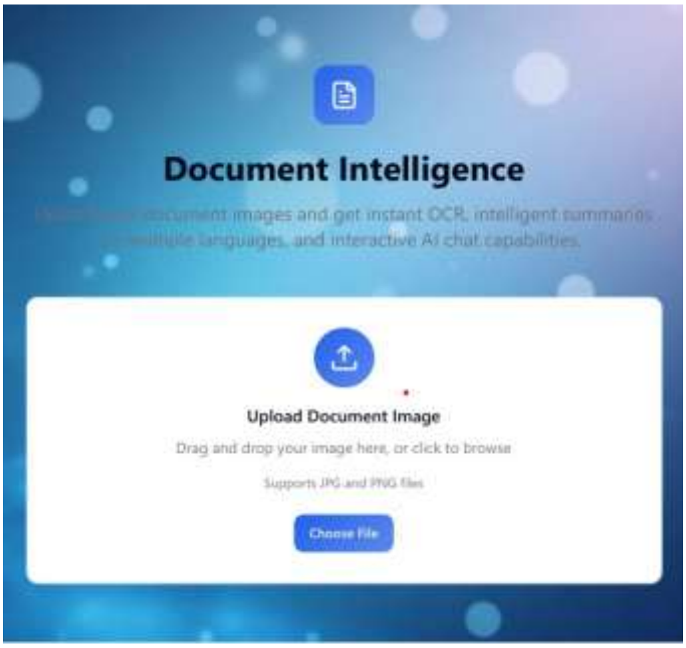

Fig. 2: Document upload interface

In this figure, we see the interface and process for document upload, as well as the pre-processing, which includes format validation and data cleaning.

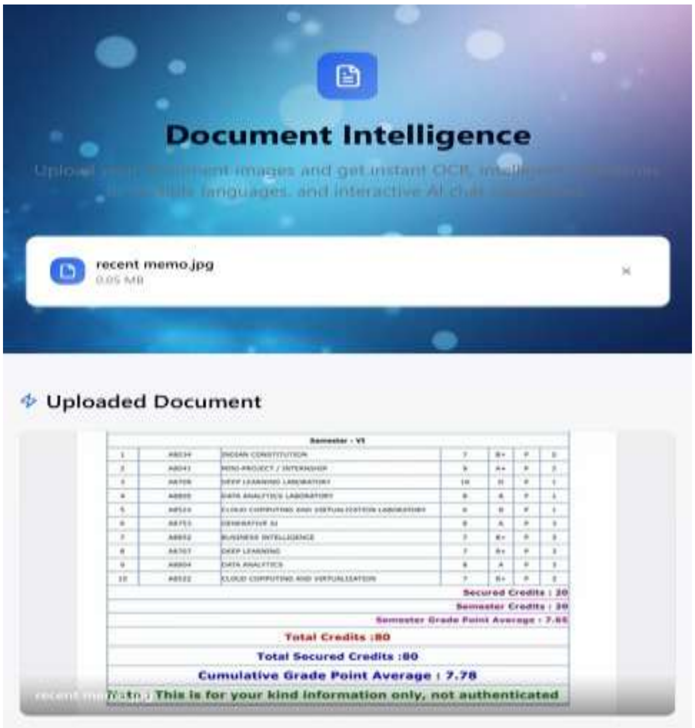

Fig. 3: OCR text detection output

This figure is of the optical character recognition, which is the stage at which the system gets and structures text from what may be a scanned-in or uploaded document.

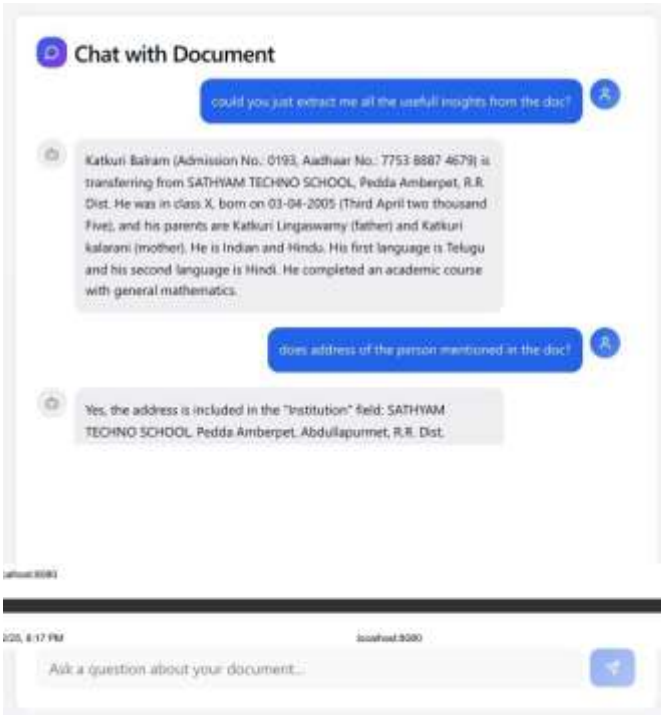

Fig. 4: Data extraction and structuring

This figure shows how the system goes about processing user queries with the use of the extracted document data to put forth precise and context-aware answers.

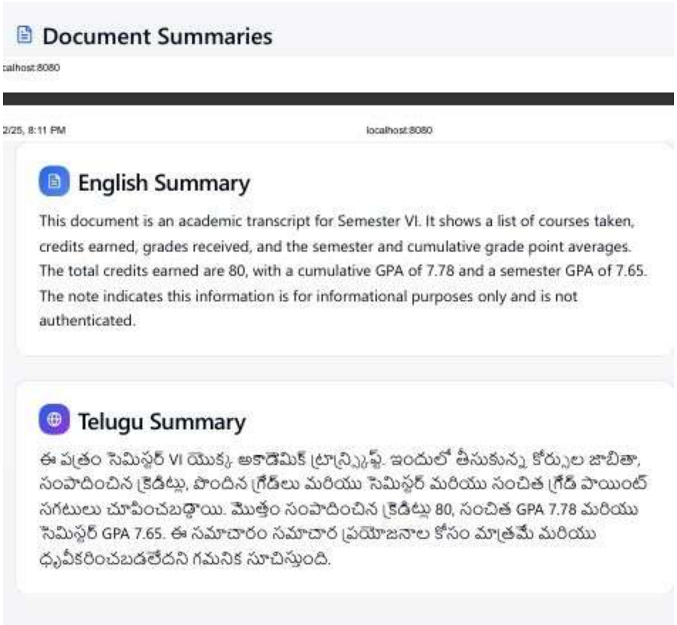

Fig. 5: Query response interface

This figure shows the system's interface where users enter their queries and receive responses generated in real-time based on the system's analyzed data files.

4. Conclusion

In the case of the Automated Student Document Analysis and Query System, we see a successful integration of Optical Character Recognition and Conversational AI, which in turn simplifies academic document processing and interaction. We use Google's Gemini Vision and Language Models, which, in that order, perform text extraction, bilingual summarisation, and intelligent query response, thereby making document analysis faster and more accessible for users. Our full-stack architecture, developed with FastAPI and React, is what we put forward for scalability, responsive design, and real-time performance. The work we present is a study of how AI-driven automation, in fact, reduces manual input, breaks down language barriers, and improves decision-making in educational settings. Also, at the same time, we are laying the base for future growth into other languages, data sets and adaptive learning applications.

5. Future Scope

The present system lays a strong foundation for automatic academic and legal document analysis; that said, there is room for us to improve it in the future, which we think will make it better and more flexible. In terms of direction, we see it is important to expand support beyond English and Telugu, which in turn will allow us to serve a wider range of users across different regions. Also, we may look at which levels of detail to include in our summarisation, ranging from very high-level overviews to very in-depth reports, that best suit our users. Also, we will look at adding more advanced Natural Language Processing features like semantic search, context-aware reasoning, and integration of knowledge graphs into the system, which in turn will result in more precise and more meaningful responses. Adding voice queries can make the system easier to use for users who have difficulty typing. More broadly, the system architecture can be designed to support the real-time processing of multiple documents simultaneously, enabling its use across educational institutions and even government agencies. To prevent unauthorised use of sensitive academic records and other legal documents protected by personal data and privacy laws, a range of sensitive documents would require the inclusion of encrypted storage, multifactor authentication, compliant logging, and role-based access controls. In addition, the system can be scaled to other areas of practice, including legal, medical, and even business and public administration, where rapid and precise document processing is also important. Developments in multimodal Artificial Intelligence would enhance the proposed framework, transforming it into an advanced document management system.

References

1. J. Memon, M. Sami, R. A. Khan, and M. Uddin, “Handwritten opti- cal character recognition (ocr): A comprehensive systematic literature review (slr),” *IEEE Access*, vol. 8, pp. 142 642–142 668, 2020.
2. R. Avyodri, S. Lukas, and H. Tjahyadi, “Optical character recognition (ocr) for text recognition and its post-processing method: A literature review,” Sep. 2022.
3. K. Jung, S. Bae, N.-J. Kim, J. Lim, H. G. Ryu, and H.-J. Lee, “Enhancing asr performance through relative word frequency in ocr and normal word frequency analysis,” Apr. 2024, pp. 71–74.
4. Meruga, V. B., "LLM-Guided Multi-Agent Collaboration for Complex Task Automation." In *2025 5th International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT)*, pp. 1-7. IEEE, 2025.
5. Dhivya, R., Sagili, S. R., VamsiLala, P. N. V., & Sangeetha, A.,"Predictive Modelling of Osteoporosis using Machine Learning Algorithms." In *2024 4th International Conference on Ubiquitous Computing and Intelligent Information Systems (ICUIS)*, pp. 997-1002. IEEE, 2024.
6. M. Salıcı and Ü. E. Ö lc er, “Impact of transformer-based models in nlp: An in-depth study on bert and gpt,” in *2024 8th International Artificial Intelligence and Data Processing Symposium (IDAP)*, Sep. 2024, pp. 1–6.
7. C. C. Sai, R. Purushothaman, C. Reddy, P. R. Gangothri, V. Dhanush, and C. Bhavana, “Integrating optical characteristic recognition with conversational ai: A multimodal chatbot featuring speech and poster generation,” May 2025, pp. 1522–1529.
8. Son, L. H., Pritam, N., Khari, M., Kumar, R., Phuong, P. T. M., & Thong, P. H., "Empirical study of software defect prediction: a systematic mapping." *Symmetry* 11, no. 2 (2019): 212.
9. R. B. Sedilla, R. Joseph, J. Jeremie, Bon, J. Mariano, and M. Jane, “As- sessment of machine translation in addressing communication barriers on the perception of college students,” Aug. 2024, pp. 262–266.
10. V. K. Sharma. "AI-Based Detection for 5G Core and RAN Components-International Journal of Scientific Research in Engineering and Management (IJSREM) Volume: 06 Issue: 01| Jan-2022."
11. S. K. Assayed, M. Alkhatib, and K. Shaalan, “Transforming student advising in smart cities: A deep learning conversational ai chatbot,” in *2024 Mediterranean Smart Cities Conference (MSCC)*, May 2024, pp. 1–6.
12. D. Dixit, M. Tyagi, R. Kandoi, and R. Singh, “Enhanced multimodal conversational ai using speech and image integration,” in *Lecture Notes in Networks and Systems*, Dec. 2024, pp. 163–179.
13. M. Yang, B. Namoano, M. Farsi, and J. A. Erkoyuncu, “Named entity recognition in aviation products domain based on bert,” *IEEE Access*, vol. 12, pp. 189 710–189 721, 2024.
14. S. T. Abu-Orabi, “Artificial intelligence technologies in higher education: Reality and challenges,” Nov. 2024, pp. 479–484.
15. S. Xie, “Research on named entity recognition method based on bert model,” Jul. 2024, pp. 92–96.
16. N. Kaur, A. Saha, M. Swami, M. Singh, and R. Dalal, “Bert-ner: A transformer-based

approach for named entity recognition," in *2022 13th International Conference on Computing Communication and Network- ing Technologies (ICCCNT)*, Jun. 2024, pp. 1–7.

17. A. P. Murdan and R. Halkhoree, "Integration of artificial intelligence for educational excellence and innovation in higher education institutions," in *2024 1st International Conference on Smart Energy Systems and Artificial Intelligence (SESAI)*, Jun. 2024, pp. 1–6.
18. M. Masalkhi, J. Ong, E. Waisberg, and A. G. Lee, "Google deepmind's gemini ai versus chatgpt: a comparative analysis in ophthalmology," *Eye*, pp. 1–6, Feb. 2024.
19. K. Jung, S. Bae, N. J. Kim, H. G. Ryu, and H.-J. Lee, "Improving asr performance with ocr through using word frequency difference," in *2020 International Conference on Electronics, Information, and Communication (ICEIC)*, Jan. 2024, pp. 1–4.
20. P. K. R, B. M. G, P. Srinivasan, and V. R, "Transformer-based models for named entity recognition: A comparative study," in *2022 13th In- ternational Conference on Computing Communication and Networking Technologies (ICCCNT)*, Jul. 2023, pp. 1–5.
21. M. Liang and Y. Shi, "Named entity recognition method based on bert-whitening and dynamic fusion model," in *2022 4th International Conference on Natural Language Processing (ICNLP)*, Mar. 2023, pp. 191–197.
22. X. Wang, Y. Sun, C. Chen, and J. Cui, "A relation extraction model based on bert model in the financial regulation field," in *2022 2nd International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI)*, Sep. 2022, pp. 496–501.
23. D. R. Raman, S. Jayalakshmi, K. Arumugam, A. V. A. Raj, D. Balaji, and R. Brightsingh, "Implementation of data analysis and document summarization in social media data using r and python," in *2022 4th International Conference on Inventive Research in Computing Applica- tions (ICIRCA)*, Sep. 2022.
24. L. Hillebrand *et al.*, "Kpi-bert: A joint named entity recognition and re- lation extraction model for financial reports," in *2022 26th International Conference on Pattern Recognition (ICPR)*, Aug. 2022.
25. E. Hans, J. Laterza, M. Paulo, and M. Ladeira, "A proposal to identify stakeholders from news for the institutional relationship management activities of an institution based on named entity recognition using bert," in *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Dec. 2021, pp. 1569–1575.
26. X. Zhao, J. Greenberg, Y. An, and X. T. Hu, "Fine-tuning bert model for materials named entity recognition," in *2021 IEEE International Conference on Big Data (Big Data)*, Dec. 2021.
27. D. Christou and G. Tsoumakas, "Improving distantly-supervised relation extraction through bert-based label and instance embeddings," *IEEE Access*, vol. 9, pp. 62 574–62 582, 2021.
28. N. Begum and A. Goyal, "Analysis of legal case document automated summarizer," Oct. 2021.
29. P. Su and K. Vijay-Shanker, "Investigation of bert model on biomedical relation extraction based on revised fine-tuning mechanism," in *arXiv (Cornell University)*, Dec. 2020.
30. Y. Zhou and M. Li, "Online course quality evaluation based on bert," Jul. 2020.