Enhancing Contracts Data Accessibility through ContractBot

Cigdem Atazulal
Student ID: 22203605

School of Computing
National College of Ireland

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Cigdem Atazulal
Student ID:	22203605
Programme:	AI For Business
year:	2023
Module:	MSc Research Project
Supervisor:	Victor Del Rosal
Submission Due Date:	14/12/2023
Project Title:	Enhancing Contracts Data Accessibility through ContractBot
Word Count:	4178
Page Count:	13

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	31st January 2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Enhancing Contracts Data Accessibility through ContractBot

Cigdem Atazulal

22203605

Abstract

Amid the challenges faced by modern businesses in effectively managing contract data, this study seeks to address the inefficiencies inherent in manual data retrieval processes from contracts. Manual methods of accessing contract information are error-prone and time-consuming, necessitating the development of a technology solution. ContractBot is an AI-based and automated solution that aims to mitigate human errors and enhance the efficiency of information access, thereby empowering organisations to navigate contractual intricacies with heightened confidence and accuracy. ContractBot stands as a pivotal advancement in contract data management with its AI-based, value-adding simple architecture. Leveraged by Microsoft Power Automate OCR (Optical Character Recognition) capabilities, the system accelerates the extraction of critical data from contract documents, ensuring a level of robustness that eliminates manual errors and significantly enhances the precision and reliability of extracted information. The system seamlessly aligns with contemporary technological trends, integrating RPA, AI and virtual agents into the realm of contract management. The theoretical contributions extend beyond mere automation and time savings, preserving the notion of an interactive environment as a catalyst for advanced cognitive ergonomics and robustness within document-centric processes. While ContractBot stands as a pioneering solution for the improvement of contract management processes, ongoing exploration is worthwhile to enhance its adaptability to diverse contract formats, industry-specific idiosyncrasies, and the nuanced challenges posed by legal language interpretation. The continued refinement of the system will ensure its enduring efficacy in diverse organisational contexts.

Keywords: Contract Data Management, Robotic Process Automation, Optical Character Recognition, Artificial Intelligence Models, Virtual Agents.

1 Introduction

Motivated by the evolution of AI in the contract management field, this study identifies a critical gap in the existing solutions to extract data from signed contracts and utilise it in daily operations effectively. There is a diverse range of recent practices in contract management with AI, especially in terms of contract drafting and report generation. However, questions still exist in adapting the applications to the daily processes and securing contract data while saving time and effort.ContractBot envisages offering a solution, deriving from micro-processes of reading and extracting data from contracts and by its nature serving a data-centric strategic-level decision-making, underpinned with increased

labour quality. Extracted data can be further fed into other downstream applications from dashboards for report generation. (Tecuci et al.; 2020)

The primary research question driving this study is: How can ContractBot revolutionise contract data management in response to contemporary challenges?

To address this question, the study establishes specific research objectives:
- Investigate the state of the art in contract data management broadly.
- Design an interactive and dynamic chatbot-driven system (ContractBot).
- Implement ContractBot using AI and RPA for enhanced data extraction.
- Evaluate the performance of ContractBot in improving the efficiency of contract data retrieval.

The major contribution of this research lies in the introduction of ContractBot as a tailor-made, practical and innovative solution, through the utilisation of currently available, secure and easily integrated applications of AI and automation technologies effectively, introducing a new path for the technology solutions in the field. ContractBot not only offers an interactive environment for precise and reliable data extraction addressing the limitations of manual methods, but also shows a high contribution to cognitive ergonomics studies in work environments by decreasing the cognitive load and time consumed for contract data management. (Christy and Duraisamy; 2020)

This paper discusses related work both in the context of academic studies and the current applications in Section 2. Section 3 describes the research methodology used in this research. Section 4 presents the design specification and Section 5 summarises the implementation phase. Section 6 unveils the improvements to be achieved by ContractBot in organisations, provides outputs of the experiment held on Google Cloud Platform and discusses how ContractBot can revolutionise contract data management. Section 7 presents a conclusion and addresses future work.

2 Related Work

2.1 The Transformation of Contract Management with AI

Contracts not only fulfill the function of managing the risks between parties that relate to opportunistic behaviors of self-interested partners but also the inherent monetary and performance risks that arise even in full collaboration of parties, by providing a framework for coordination and synchronisation across companies.(Dekker et al.; 2013)

Contract management in definition, is taking all necessary actions in the life cycle of a contract from drafting to execution and termination. The philosophy behind is to harmonise contractual outputs with the organisation's existing objectives, strategies and activities, and manage the risk aspects vigilantly. Even though existing procurement and contract management tools (e.g Ariba, SAP, Oracle, etc.) support the contract life cycle, the area of contract management still remains a potential field for improvement and development of tools and services to better assist users in tasks such as defining and checking obligations, responsibilities and consequent actions. While big organisations suffer from the complexity of tools and integration problems, Small and Medium Enterprises (SMEs), usually don't own sophisticated and expensive tools and highly qualified workforce skills to face these challenges.(Bochicchio and Longo; 2011)

The current practices in the utilisation of AI in contracts include contract drafting, contract management, automated contract due diligence and AI-based contract nego-

tiations. [1] Even though more holistic approaches for automated solutions for certain elements of contract management offer an insightful evolution of companies towards AI, it is also crucial to unfold the complexities and inefficiencies around daily tasks with simple and tailor-made solutions as ContractBot.

2.2 DICR: Contract Data Management for EY

O'Leary investigated an AI application to read contracts and provide structured data to augment the strategies and labour force in EY, one of the world's leading professional services firms. Figure-1 shows the preliminary vision for the architecture of DICR ("Document Intelligence for Contract Review") system. According to the investigation of O'Leary, the "Dialogue Chat" part has not been put into use in the application phase. (O'Leary; 2022)
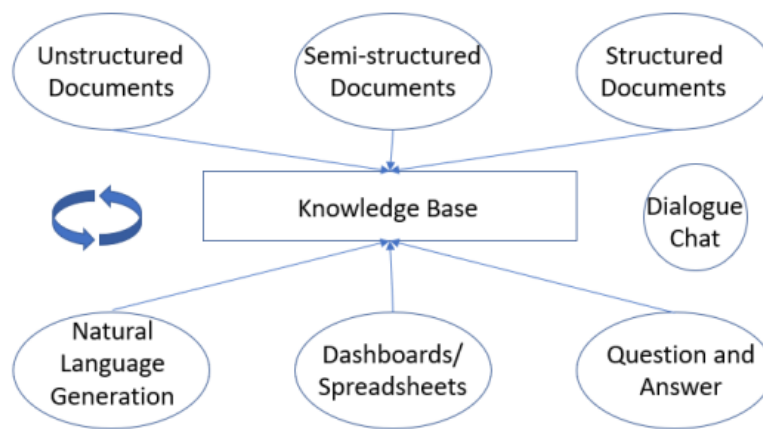

Figure 1: DICR System Elements

The DICR system guides the user through reading the contract by using an ontology. The function of ontologies is to provide a structured way of encoding domain-specific information which is helpful for context-aware NLP (Natural Language Processing) systems. Additionally, a "human-in-the-loop" is used by the system to assist in selecting which information snippets match which parts of the ontology.The system highlights the most likely value for each item in the ontology inside the text by using entity recognition and value extraction. The system is well designed to decrease the human workload and the algorithm is enriched with the human-in-the loop processes. Utilisation of Excel worksheets to store the interpreted data is the same as ContractBot. However, the user interfaces still lack the philosophy of "serve-and-enable" with complex screens and low interactivity. The fact behind this complexity is obviously, the inherent structure of real-life contracts that still requires humans to deep dive into clauses. Yet the algorithms can be further fed by information and contexts, tested and trained to understand more sophisticated concepts to serve humans with more refined results. (O'Leary; 2022)

---

[1] https://www.linkedin.com/pulse/how-ai-transforming-contracts-legamartlegal/

2.3 Current Trends in Big 3

For OCR and data extraction, some players provide exclusively SaaS (software as a service) – pay for what you use with degressive prices according to volume. (Cutting and Cutting-Decelle; 2021)Figure-2 indicates a thorough examination of the AI-based document processing tools offered by the three main competitors—Amazon, Google, and Microsoft, revealing distinctive performance features. Outstanding accuracy of 99 % for recognising key-value pairs is demonstrated by Microsoft's Azure Cloud-based solution, highlighting the efficacy of its robust and versatile cloud platform. Even though Google's solution is excellent at locating key-value pairings, its propensity to classify all data as keys and values may introduce complexity during integration. Amazon's user-friendly solution faces challenges with accurately identifying and organising information, but it uses human validation to increase precision, demonstrating its commitment to continuous improvement. (Mandvikar; 2023) Despite unique strengths across providers, Microsoft's solution stands out not only with its accuracy but also with ensuring security and privacy through on-premise deployment options which makes it a compelling choice for document management.

Figure 2: Feature Comparison of IDP Tools of Big 3

[2] To address the problems and risks in contract data management, ContractBot offers a pioneering solution to decrease the workload, and improve efficiency, data accuracy, and ergonomics in an interactive way. Currently available low-code services provide a wide range of solutions to businesses and ContractBot aims to highly benefit from these solutions to tailor a new working concept in the era of AI, magnifying the specifics of contract data management.

---

[2]https://deepcognition.ai/comparison-of-various-ai-based-ocr-for-extracting-information-from/

3 Methodology

3.1 Data Pipeline

The research methodology for ContractBot involved the generation of a synthetic data set comprising around 100 artificial contracts. This data set was meticulously curated to represent a diverse range of business scenarios, ensuring comprehensive training and testing for the AI model embedded within ContractBot.

The raw data, synthesised through artificial contracts, was subjected to analysis during the AI model training phase. The selected AI Model of MS Power Automate is a custom model to extract custom information from documents. Performance metrics of the model include confidence scores to assess the accuracy of the labelled data attributes for non-structured documents.

Figure-3 indicates the AI model performance in low training and high training cases studied. The low training included 5 artificial contracts whereas high training was performed with 100 artificial contracts. The comparison was made for the 3 of the attributes that are: Contract ID, Supplier and Penalties. Given the training is done with 20 times more amount of data, the attribute "Contract ID" gave 1 % higher confidence score where the "Supplier" attribute gave the same score and the "Penalties" attribute gave a 7 % lower confidence score. This shows that the unstructured documents are hard to measure accuracy in Power Automate and as the training includes the selection and tagging of the data in the documents, each selection area variation by the user can impact the confidence score positively or negatively regarding the location of the selection and its distance from the average selection area on PDF.

Model Performance	Confidence scores	
Attributes	Low training	High Training
ContractID	97%	98%
Supplier	98%	98%
Penalties	96%	89%

Figure 3: AI Model Confidence Scores

3.2 Tools and Techniques

Intelligent document processing (IDP) allows to capture, transform, and process data from documents (for example, email, text, Word, PDF, or scanned documents). [3]

Document processing is built by the usage of optical character recognition (OCR), machine learning, and robotic process automation (RPA) technologies. These three components are essentially able to interpret and understand information in a similar way to humans.

OCR recognises printed, written, or typed text from scanned documents or images. It identifies light and dark areas in the document and searches for letters or digits, which are then classified based on patterns or features.

---

[3] https://learn.microsoft.com/en-us/power-platform/release-plan/2023wave1/ai-builder/automate-intelligent-document-processing

Machine learning creates algorithms that learn from patterns and context in documents. By processing more and more information, it learns from all the use cases it encounters, scoring better results and being more efficient in how it makes decisions.

RPA uses bots working on rules and instructions they receive for the automation of repetitive tasks. By using advanced text recognition capabilities, RPA can swiftly process data from multiple sources.[4]

The significance of Microsoft Power Platform has gained an immense amount of recognition in the context of the evolving landscape of business. Numerous tools provide businesses with quick and efficient processes, which influences their ability to thrive in highly competitive markets. With the help of a powerful low-code programming language for logic expression, Microsoft Power Platform skillfully tackles key problems around computing power, digital transformation, and artificial intelligence. (Wang; 2023) Figure-4 indicates the AI processing in Microsoft applications that enables flexible designs and effective integration with existing applications in the organisations. The workflow for document processing steps is as follows:

An object detection model is trained in AI Builder to recognise objects that a user specifies.

The arrival of a document to a SharePoint document library, OneDrive, or Teams triggers a Power Automate event. That event:

1. Runs the AI model of Microsoft, AI Builder. AI Builder returns a JSON file with the pixel coordinates of specified objects.

2. Sends the document to Form Recogniser which uses machine-learning models to extract and analyse form fields, text, and tables from your documents, to reach a full optical character recognition (OCR) scan. Form Recogniser returns a JSON file that contains scanned-in text and pixel coordinates of the text.

3. Runs a function in Azure Functions which is an event-driven serverless compute platform. The function analyses the pixel coordinates in the AI Builder and Form Recogniser output files. If detected objects intersect with scanned text, the function returns the matched data in a JSON file.

4. Enters the metadata, or the text from detected objects, into a document library. The metadata is captured in a SharePoint search index for user search.
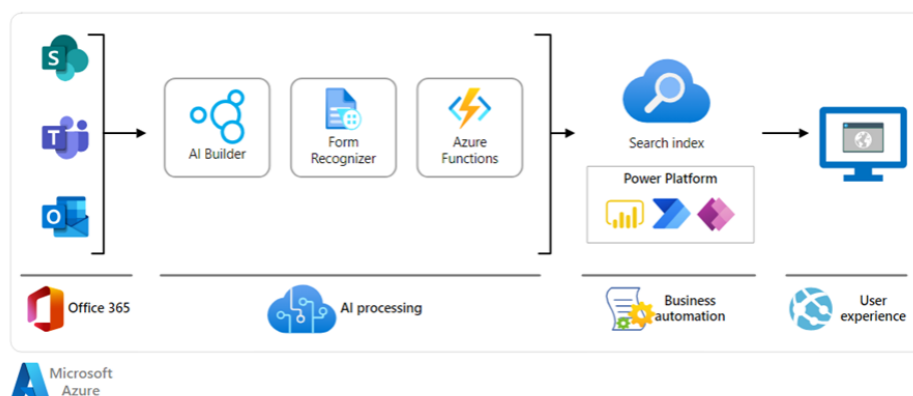

Figure 4: AI Processing in Microsoft Applications

---

[4] https://powerautomate.microsoft.com/en-us/document-processing-applications/

Azure functions are streamlined in Microsoft Power Automate in a low-code environment enabling integration with the conversational agent application of Microsoft, MS Copilot. MS Copilot Studio facilitates the creation of AI-powered chatbots that can resolve numerous issues requiring complex conversations in various user interfaces. Integrated with MS Copilot, ContractBot leverages the latest advancements in AI-driven chatbot development.

Microsoft Copilot Studio uses an underlying natural language model for understanding user's questions and determining the right topic to follow which is previously defined by developers.[6]

The integration of Microsoft Copilot with Microsoft Power Automate extends the capabilities, facilitating seamless workflow automation for AI model deployment and contract data extraction.

Microsoft Excel serves as a practical storage for the extracted data, consistent with the business applications as in the DICR product explained in Section-2.

As the user interface of ContractBot, a website is created with Google Sites which is practical to design and integrate, and is easily accessible to users. The embedding of code is simple and applicable to various interfaces for corporate use.

Microsoft OneDrive is selected as the repository for the contract documents which enables the RPA process trigger for contract data extraction.

3.3   Scenario

The primary scenario involved the signature of a contract, followed by the contract administrator's upload of the document to the MS OneDrive file. This triggers the automated data extraction process from the contracts to the Excel file.

Power Automate workflows are designed to extract predetermined contract details, introducing a dynamic element to the case studies. The predetermined data variables of contracts which are, supplier, start date, delivery date, purchased items, total amount, payment term, penalties and supplier signatory are retrieved from the PDF files with the help of a trained and tested AI model.

The ContractBot asks the users to enter the Contract ID, which they have easy access through MS OneDrive with the Contract IDs marked as the document name. Contract ID is also marked on the first page of each contract document by the contract administrator. Once the user enters the Contract ID in the conversation with ContractBot, the Power Automate Workflow retrieves relevant information from the Excel File and lists the pre-filled contract details to the user in a conversational manner.

4   Design Specification

ContractBot stands on the shoulders of Copilot, an innovative AI programming tool integrated seamlessly within Microsoft Power Virtual Agents. Figure-5 presents the system architecture of ContractBot:

Using Microsoft Power Automate enriches the architecture, providing a robust framework for workflow automation and AI model deployment. Power Automate also has other
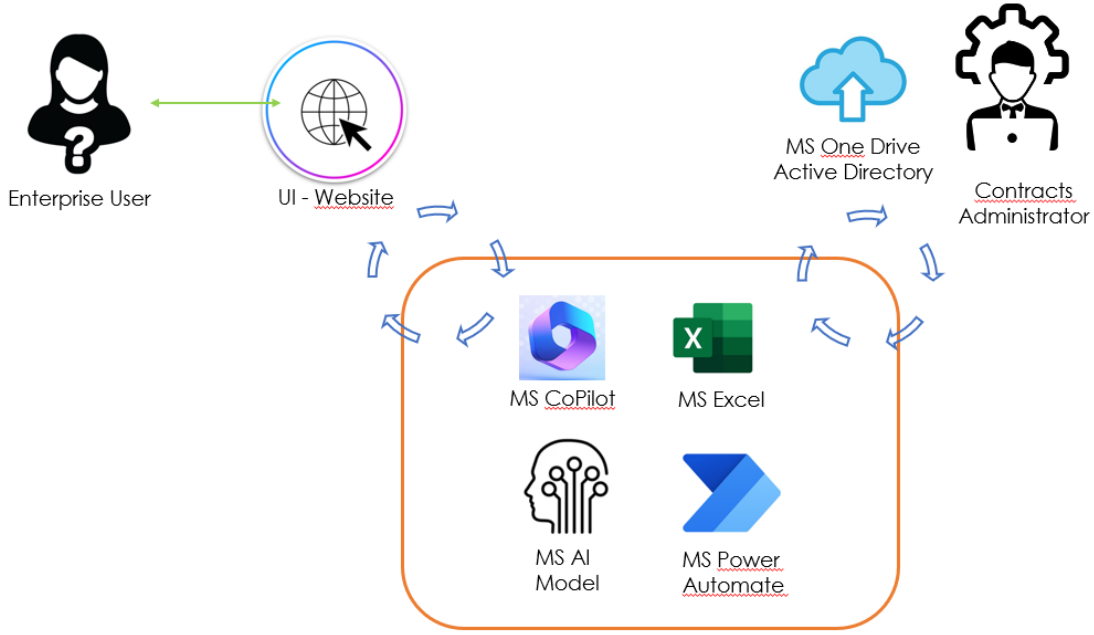
---

[5] https://tekenable.ie/azure-ai-document-intelligence-a-game-changer-for-business/
[6] https://learn.microsoft.com/en-us/microsoft-copilot-studio/nlu-gpt-overview

Figure 5: ContractBot Architecture

enriched properties such as process mining to further optimise processes. [7]

The two flows are interconnected in the power virtual agent flow by the Contract ID, which is a unique number for each contract. Contract ID is used in both flows to update a row and return values to Power Virtual Agents.

The contract document storage environment is selected to be Microsoft OneDrive which provides efficient integration with MS Power Automate tools and MS Power Virtual Agents. The choice of Microsoft Excel for the extracted data storage aligns with industry standards for structured data storage. The Excel file is managed by the contract administrator to store pre-defined Contract IDs.

At the core of ContractBot is an advanced AI model, crafted within the Copilot environment. This model exhibits a sophisticated functionality, adept at reading and extracting nuanced information from contract documents, showcasing the cutting-edge capabilities of AI in natural language processing.

5 Implementation

The culmination of ContractBot's implementation yields transformed data, from PDF documents to Excel and from Excel to the chatbot, distilling critical contract information for enhanced decision-making. Detailed outputs are presented without delving into code listings or user manual descriptions, aligning with best practices for concise reporting.

The research for the selection of the AI Model for data extraction from PDF documents was focused on the available tools of Microsoft (AI Builder) and Google (Document AI). Due to its practicality for application development and integration, as well as its

---

[7]https://carstengroth.wordpress.com/category/power-automate/

wide utilisation in organisations for the storage of contracts that skips the security aspects of confidential information, Microsoft was selected and the AI model was developed with Microsoft AI builder.

The pre-trained and tested AI model developed with AI builder is used to extract data and an MS Power Automate workflow is used to fill in the Excel file on MS OneDrive with the extracted data.

After the creation of topics for different conversations in Microsoft Copilot, the main topic for contract query with Contract ID is linked to another MS Power Automate flow created for extracting relevant data from the Excel file and transferring the data to the MS Copilot environment to be used for responding to user queries.

Crafted using MS Copilot within Microsoft Power Virtual Agents, ContractBot represents a state-of-the-art AI-driven chatbot solution. Integration of Microsoft Copilot with Microsoft Power Automate demonstrates the pragmatic use of workflow automation tools in AI implementations.

As a result, ContractBot's development, from meticulous research methodology to design specification and final implementation, encapsulates the cutting-edge landscape of AI-driven chatbot development. Utilising tools like Copilot and Power Automate, ContractBot sets a benchmark for efficient, data-driven decision support in contract management.

At the final stage, the system is integrated to a web-based user interface, a website designed with Google Sites. Figure-6 illustrates an example conversation held with the ContractBot where it answers the user queries regarding the contract with the ID 103 which is readily available in the system.
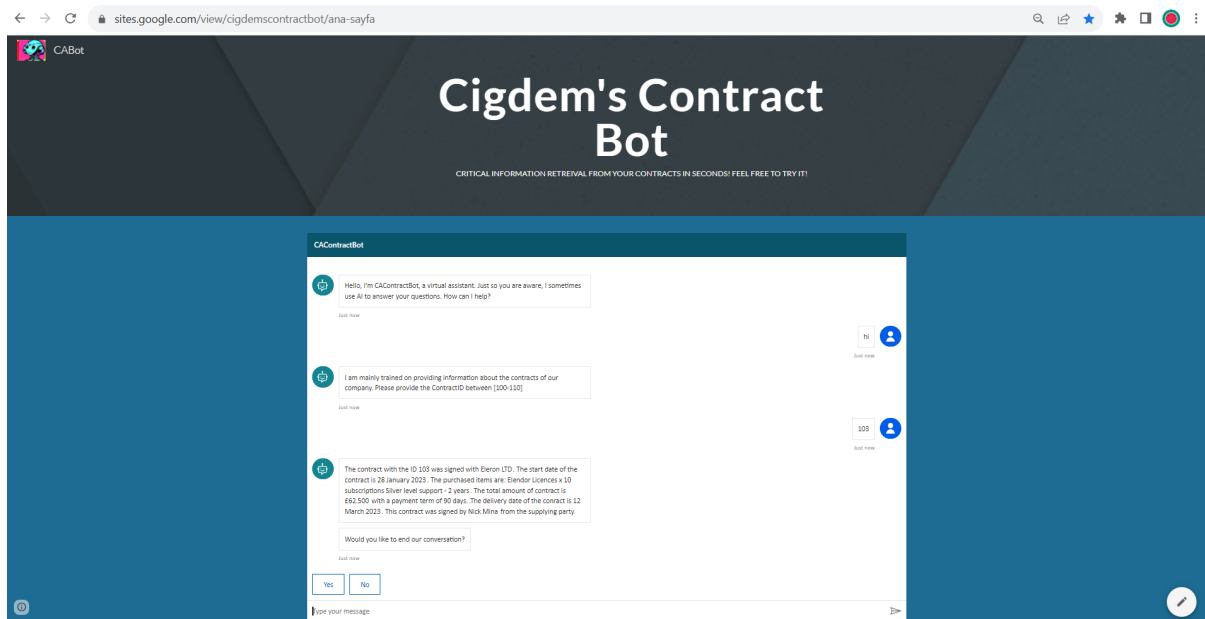

Figure 6: Outputs

6 Evaluation

6.1 Improved Process Flow

Figure-7 illustrates the process with and without ContractBot from users' point of view. The system enables users to retrieve data either in text format through ContractBot query or in Excel format through the auto-formed Excel file in MS One Drive. The most significant change in the overall process is the contract reading and putting the relevant data in writing, which is totally removed for users.
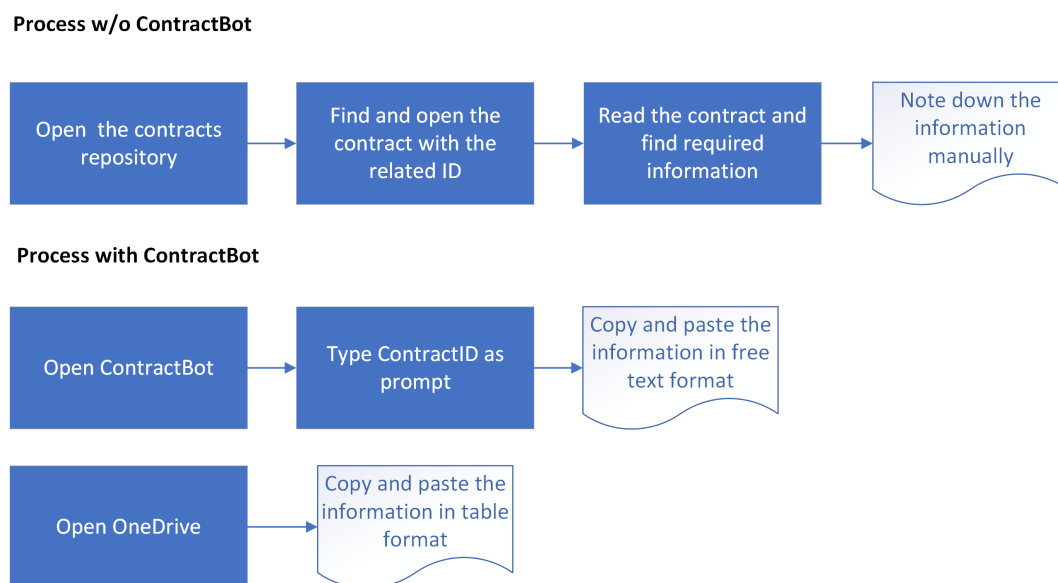
Process w/o ContractBot

Open the contracts repository → Find and open the contract with the related ID → Read the contract and find required information → Note down the information manually

Process with ContractBot

Open ContractBot → Type ContractID as prompt → Copy and paste the information in free text format

Open OneDrive → Copy and paste the information in table format

Figure 7: Current vs. ContractBot Proposed Process

6.2 Stopwatch Time Study

Working time is measured by observing processes, and recording the amount of time it takes to to complete a given task under normal conditions. It is then used to optimise the processes leading to efficient systems and increased profitability. (Budiman et al.; 2019)

The most significant added value of ContractBot is the time savings it brings to organisations. Figure-8 shows the comparison between the manual processes of extracting and noting down the related data, and the process of using Contract Bot. According to the 10 stopwatch time study measurements, ContractBot brings about 78 percent time savings in operations. Given that procurement professionals spend around 20 percent of their working hours by scrolling through contracts, the workload can be decreased by up to 15 percent with ContractBot in organisations.

6.3 AI Model Training Experiment on Google Cloud

The training and testing of the AI model was also experimented on Document AI, one of the foremost document management tools of Google Cloud.

A document processor was created on Document AI and the AI model was trained and tested with 88 artificial contracts. For the unstructured documents, Document AI
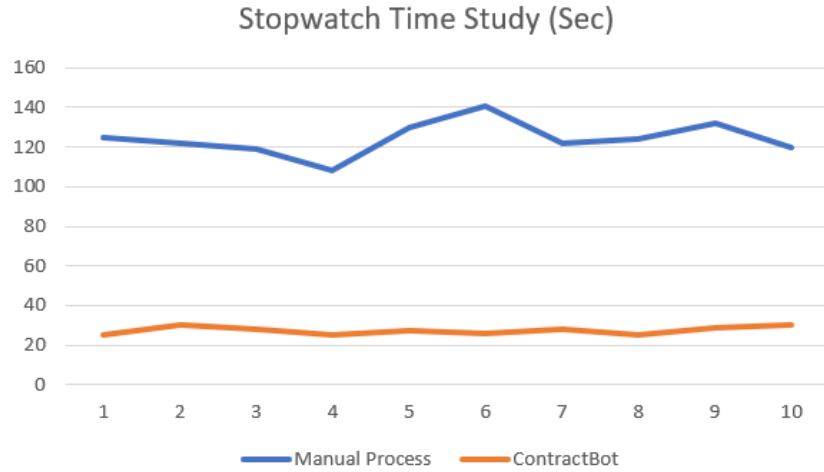
Figure 8: Time Savings with ContractBot

provided more insights into the performance outputs of the AI model than Microsoft products as seen in Figure-9. The OCR capabilities are observed to be high and more training with a large amount of data leads to better performance results. However, the difficulties around the Google Cloud Functions that require coding for the integration of Document AI with Dialogflow, the conversational agent solution of Google, and the costly structure of APIs required to run the system resulted in the choice to proceed with Microsoft Applications.

Label ↑	F1 score
All labels	0.914
delivery_date	0.964
payment_term	0.988
penalties	1.000
purchased_items	0.779
start_date	0.810
supplier	0.776
supplier_signatory	0.987
total_price	1.000

All labels

F1 score	0.914
Precision	92.4%
Recall	90.4%
Last evaluated	19 Oct 2023, 00:58:27
Test documents	43
Evaluated documents	43
Invalid documents	0
Failed documents	0

Figure 9: Document AI Processor Performance Results

11

6.4   Discussion

The successful implementation of ContractBot leads to the observation that it can revolutionise contract data management in response to contemporary challenges in several ways:

- ContractBot offers high practicality to users with its user-friendly interface
- ContractBot is easy to adapt to different organisations with different requirements of data processing for contract management, providing a readily used secure working space in terms of the confidentiality aspect of contract data storage.
- Given the limited number of currently available products in the market for contract management and their high costs, ContractBot represents the cutting-edge of utilisation of current technologies to solve business needs with lowest cost.
- Companies have to spare time for demo sessions with on-the-market products and they usually face inadequacies in meeting their specific requirements where they then have to change their way of working to equally benefit from the digital transformation world of today. ContractBot is open source, easily understandable and adoptable to various tasks, providing the companies a vision on how to cope with their daily life problems, especially in the field of document management.
- Deriving from a task-based operational monitoring approach and unveiling the critical data behind, ContractBot streamlines a robust baseline for the strategic decisions of companies in parallel with saving time and avoiding mental workload for employers. (Duan et al.; 2019)
- The prompts required for querying ContractBot is low in number, focusing on the main function it presents. It generates user responses instantly, leading to its overarching performance deriving from its robust design architecture curated with an approach of simplicity and effectiveness.

7   Conclusion and Future Work

ContractBot is a revolutionary, low-cost, low-code and easily integrated tool that utilises AI technology in the specific context of for contract data management. It is seen as a pioneering example of lean and simple utilisation of AI technologies in contract management.

The limitations of the research include training the AI model with different document variations used in real life. During the training and testing studies, it was observed that the currently available OCR technologies provide exclusive results that are worth extending the study with real data in organisations.

Future work is proposed to be broadening the study with Large Language Models (LLMs) for sentence-level semantic annotation of legal texts to resolve lawful content of contracts, (Savelka; 2023) and providing users further insights on a wide variety of actions to take in fields of negotiation, project management and risk mitigation.

Similar to the DICR application explained in Section 2, human-in-the loop is also predicted to further enhance the algorithm augmenting the precision of results.(Grønsund and Aanestad; 2020)

References

Bochicchio, M. A. and Longo, A. (2011). Modelling contract management for cloud services, *2011 IEEE 4th International Conference on Cloud Computing*, IEEE, pp. 332–339.

Budiman, I., Sembiring, A. C., Tampubolon, J., Wahyuni, D. and Dharmala, A. (2019). Improving effectiveness and efficiency of assembly line with a stopwatch time study and balancing activity elements, *Journal of Physics: Conference Series*, Vol. 1230, IOP Publishing, p. 012041.

Christy, D. V. and Duraisamy, D. S. (2020). Ergonomics and employee psychological well being, *International Journal of Management* **11**(3).

Cutting, G. A. and Cutting-Decelle, A.-F. (2021). Intelligent document processing–methods and tools in the real world, *arXiv preprint arXiv:2112.14070* .

Dekker, H. C., Sakaguchi, J. and Kawai, T. (2013). Beyond the contract: Managing risk in supply chain relations, *Management Accounting Research* **24**(2): 122–139.

Duan, Y., Edwards, J. S. and Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of big data–evolution, challenges and research agenda, *International journal of information management* **48**: 63–71.

Grønsund, T. and Aanestad, M. (2020). Augmenting the algorithm: Emerging human-in-the-loop work configurations, *The Journal of Strategic Information Systems* **29**(2): 101614.

Mandvikar, S. (2023). Augmenting intelligent document processing (idp) workflows with contemporary large language models (llms), *International Journal of Computer Trends and Technology* **71**(10): 80–91.

O'Leary, D. E. (2022). Using ai to read contracts, *Available at SSRN 4281936* .

Savelka, J. (2023). Unlocking practical applications in legal domain: Evaluation of gpt for zero-shot semantic annotation of legal texts, *arXiv preprint arXiv:2305.04417* .

Tecuci, D. G., Palla, R., Nezhad, H. R. M., Ahuja, N., Monteiro, A., Ishkhanov, T. and Duffy, N. (2020). Dicr: Ai assisted, adaptive platform for contract review, *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34, pp. 13638–13639.

Wang, X. (2023). Enhancing business processes through dynamics solutions with microsoft power platform.