Building a Platform for Intelligent Document Processing: Opportunities and Challenges

Francesco Visalli[1,*], Antonio Patrizio[1], Antonio Lanza[1], Prospero Papaleo[1], Anupam Nautiyal[1], Mariella Pupo[1], Umberto Scilinguo[1], Ermelinda Oro[2] and Massimo Ruffolo[1]

[1]altilia.ai, Piazza Vermicelli, c/o Technest - University of Calabria, Rende (CS), 87036, Italy

[2]High Performance Computing and Networking Institute of the National Research Council (ICAR-CNR), Via Pietro Bucci 8/9C, Rende (CS), 87036, Italy

Abstract

Companies of any size and industry still struggle in automatic business processes where human cognitive and contextualization capabilities are required to read and understand complex documents. Ongoing progress in the fields of Computer Vision and Natural Language Processing, where (large) language models are becoming increasingly and freely available, have made possible to create a new generation of Intelligent Document Processing technologies that allow automatically analyzing and understanding both documents layout and contents. In this paper we present an Intelligent Document Processing platform that makes use of hybrid AI techniques to allow document reading comprehension by means of a combination of Document Layout Analysis and recognition, table recognition and detection, context free grammars, and question answering techniques. Such a technology combines also no-code principles with high performance computing based on micro-services to streamline the execution of tasks such as document and text classification, document segmentation, entity extraction, sentiment analysis, question answering, and more.

Keywords

Intelligent Document Processing, Intelligent Process Automation, Hyperautomation, Artificial Intelligence, Natural Language Processing, Large Language Models, Computer Vision, Information Retrieval, Deep Learning, Knowledge Graph, Workflow,

1. Introduction

The advent of Artificial Intelligence (AI) has revolutionized the way businesses operate, with document processing being one of the many areas experiencing a paradigm shift. Document processing is the task of converting unstructured information in documents (invoices, contracts, financial and sustainability reports, orders, etc.) into a structured format, suitable for business processes automation, digital analysis and storage. Because organizations generate and manage vast quantities of documents, efficient document processing becomes vital to streamline operations, reduce manual effort, and minimize errors. However, the task is often hard, due to the diverse formats, layouts, and languages of documents, as well as the complexity of the context in which the information can be found (e.g. text, tables, charts, etc.). To tackle these challenges, there has been a surge in the adoption of intelligent automation platforms, specifically designed to effectively read and understand complex documents, extract relevant information, and perform downstream processing tasks.

In this paper, we introduce the Altilia Intelligent Automation (AIA) platform, which is a comprehensive document processing platform that empowers business domain experts to create and train AI models to automate their document processing workflows. The AIA platform comprises three modules: Teach, Automate, and Understand, each addressing different aspects of document processing. The Teach module is designed to enable business domain experts to train AI models without any coding skills. It provides tools for annotating documents, fine-tuning pre-trained machine learning models, and creating complex AI skills. The Automate module allows users to integrate multiple AIA skills into their document processing workflows and automate the end-to-end processing. The Understand module provides users with advanced analytic and knowledge graph capabilities to extract insights from their processed documents. We provide an in-depth analysis of the AIA platform, including the capabilities of each module, their underlying technologies, and their use cases.

The rest of this paper is organized as follows: section 2 presents the state of the art of technologies on which

Ital-IA 2023: 3rd National Conference on Artificial Intelligence, organized by CINI, May 29–31, 2023, Pisa, Italy

*Corresponding author.

francesco.visalli@altiliagroup.com (F. Visalli); antonio.patrizio@altiliagroup.com (A. Patrizio); antonio.lanza@altiliagroup.com (A. Lanza); prospero.papaleo@altiliagroup.com (P. Papaleo); anupam.nautiyal@altiliagroup.com (A. Nautiyal); mariella.pupo@altiliagroup.com (M. Pupo); umberto.scilinguo@altiliagroup.com (U. Scilinguo); linda.oro@icar.cnr.it (E. Oro); massimo.ruffolo@altiliagroup.com (M. Ruffolo)

0000-0002-6768-3921 (F. Visalli); 0000-0002-5529-1007 (E. Oro); 0000-0002-4094-4810 (M. Ruffolo)

Intelligent Document Processing workflows are built; section 3 describes the Altilia Intelligent Automation platform; finally, section 4 concludes the paper.

2. Related work

Recent advances in AI and machine learning, particularly in the fields of Natural Language Processing (NLP) and Computer Vision, have led to significant improvements in document processing techniques. The past few years have witnessed the development of various techniques and architectures that address different aspects of document processing, extracting and understanding information from diverse document types. Techniques include Optical Character Recognition (OCR) for text extraction, layout analysis for structural understanding, and syntactic and semantic retrieval for relevant information extraction.

Optical Character Recognition (OCR) is a fundamental technology in document processing, as it enables the conversion of scanned images or printed text into machine-readable and editable text format. OCR techniques have improved significantly over the last few years, thanks to the advancements in deep learning and Computer Vision. Recent works such as DBNet[1] and RobustScanner[2] have contributed significantly to improving the accuracy and efficiency of OCR techniques. The first work proposes a novel approach for text detection in natural scenes that uses differentiable binarization and a tailored loss function. This approach has demonstrated high accuracy and real-time performance in detecting text from complex backgrounds and challenging conditions, such as low resolution or distorted images. The second introduces a method to dynamically enhance positional clues for robust text recognition. By combining visual and textual features. Both works have achieved state-of-the-art results on several benchmark datasets and demonstrated robustness to challenging scenarios such as low resolution, significant distortion, occlusion and varying text styles and sizes. These recent advancements in OCR technology, along with the end-to-end Transformer-based OCR approach proposed by TrOCR[3], which leverages Transformer architecture for both image understanding and wordpiece-level text generation, have significantly improved the accuracy and efficiency of text recognition, enabling the extraction of textual information from various document formats and layouts.

Document Layout Analysis (DLA) is another critical aspect of document processing, focusing on identifying the structure and organization of text and visual elements within a document. Layout analysis techniques aim to understand the structure and organization of a document, which is crucial for accurate information extraction. Recently, LayoutLM[4] has emerged as a promising approach for layout analysis, using deep learning techniques to detect various components in a document, such as headings, paragraphs, tables, and images. By incorporating both textual and spatial features, LayoutLM is capable of understanding the structure and semantics of a document more effectively, thereby improving the performance of various NLP tasks. LayoutML has demonstrated its efficacy in handling complex document layouts and multi-column documents, which are often challenging for traditional layout analysis methods. Another notable work in Document Layout Analysis is DiT, a self-supervised pre-trained Document Image Transformer model for Document AI tasks[5]. DiT serves as the backbone network for various vision-based Document AI tasks, including document image classification, DLA, and table detection. The model is trained on one of the largest datasets for Document Layout Analysis by[6], the PubLayNet dataset.

Information Extraction (IE) techniques play a significant role in document processing, as they are responsible for identifying and extracting relevant information from unstructured or semi-structured documents. A popular approach for information extraction from documents is the Retriever-Reader[7] pipeline. The Retriever-Reader pipeline is a common approach for addressing document processing tasks, this pipeline consists of two main components: a retriever, responsible for selecting relevant documents or passages from a large corpus based on the given query, and a reader, which extracts specific information from the retrieved documents.

Syntactic keyword retrieval and neural retrieval are essential techniques for efficiently searching and retrieving relevant information from large document collections. Syntactic keyword retrieval methods, such as BM25[8], rely on term frequency and document frequency to identify relevant documents. These methods are often effective in situations where documents contain structured information, such as tables, where keywords play a crucial role in the retrieval process. On the other hand, neural retrieval methods, such as Dense Retriever[9], leverage dense embeddings to capture semantic relationships between queries and documents. These methods perform well in cases where the context is complex, and a simple keyword-based approach may not be enough. Neural retrieval models can effectively identify relevant documents even when the query and the document do not share exact keywords, as they are capable of understanding the underlying semantic meaning. These retrieval algorithms enable information extraction from vast collections of documents by filtering relevant ones for further analysis. Combining OCR and layout analysis with the Retriever-Reader pipeline can significantly enhance the efficiency of document processing systems. For instance, OCR technology can be used to convert scanned images into text, which can then be processed

by retrieval algorithms to identify relevant elements of the layout. Subsequently, reader models can be applied to extract information from the selected documents.

The reader component in the Retriever-Reader pipeline focuses on extracting relevant information from the retrieved documents. Some reading techniques include token classification, text classification, and question answering. These task are typically addressed leveraging Transformer architectures[10] which have enabled the development of powerful reader models. These models, such as BERT[11], RoBERTa[12], and GPT[13], have been widely adopted in these tasks due to their strong performance in capturing the semantic meaning of text, achieving state-of-the-art performance in a wide range of Natural Language Processing tasks, demonstrating their suitability for document processing. Readers can also be implemented by traditional rule-based approaches, such as regular expressions, wich rely on hand-crafted rules and patterns to extract information. However, these methods are limited in their ability to scale and adapt to new document types and variations.

Token classification techniques have been used to identify specific types of information within a document, such as dates, amounts, or named entities. These techniques leverage pre-trained language models, such as BERT, to fine-tune and adapt them to specific tasks and domains. Common token classification tasks include named entity recognition (NER), as discussed in the paper by[14]. NER automatically scans entire articles to extract and classify fundamental entities in a text into predefined categories, such as organizations, quantities, monetary values, person names, and locations.

Text classification techniques have been employed to categorize documents or sections of documents based on their content. These techniques typically involve the use of pre-trained language models and transfer learning to adapt the model to specific classification tasks[15]. Text classifiers, using NLP, automatically analyze text and assign pre-defined tags or categories based on its content. Common examples and use cases for automatic text classification include sentiment analysis, topic detection, and language detection.

Question-Answering (QA) techniques[16] have been applied to extract specific information from a document in response to a given question. Recent advancements in language models, such as BERT, have enabled the development of highly accurate question-answering models that can understand and extract relevant information from complex documents.

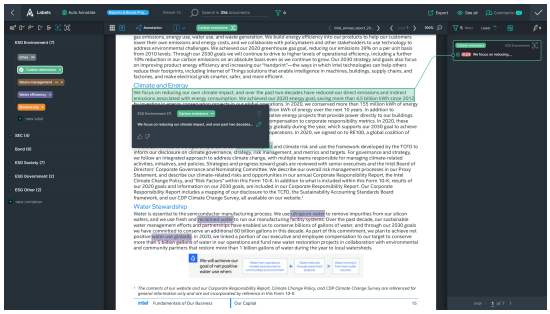

Figure 1: A snap of the platform: the Altilia Labels tool.

3. Altilia Intelligent Automation Platform

The AIA platform is composed of three sets of tools that address all document processing needs of users: Teach, Automate, Understand. Figure 1 shows a snap of the platform, in particular the Altilia Labels tool (deeply presented in section 3.1.1). On the left of the screen the user can find labels containers, a way to logically group the data that needs to be annotated and then extracted. On the right there are annotations filtered by page, the user is leveraging the auto annotation functionality (in order to speedup the data labeling process), a feature that suggests passages to annotate. Finally, in the center of the screen documents are shown along performed annotations.

3.1. Teach

The Teach module contains no-code tools that allow business domain experts to transfer their knowledge into AI models. Such tools enable users to create sophisticated AI skills for reading and understanding complex documents, starting from a set of pre-trained ML models than can be fine-tuned for the specific use case application. More in detail, tools in the Teach module allow annotating documents (Altilia Labels), training and fine-tuning ML models (Altilia Models), and creating complex AI skills (Altilia Skills).

3.1.1. Altilia Labels – Document Annotation and Dataset Creation

Altilia Labels is a user-friendly tool designed for business domain experts to conduct point-and-click document annotations as shown in figure1. It enables the creation of annotated datasets that can be used as examples to train AI models for numerous Natural Language Processing (NLP) and Computer Vision tasks. This tool

is powered by active learning and auto-labeling algorithms to reduce annotation workloads and speed up AI model deployment. A key feature of this tool is the universal document ingestion capability, i.e. users can seamlessly ingest documents of any type and format. It provides multi-language support and has the capability to recognize complex document layouts and structures (e.g., tables, lists, multi-column documents, charts, and images). Another important feature is that it enables semantic annotation which is used to enhance concepts and attributes recognition by leveraging sophisticated knowledge representation mechanisms and techniques, it also uses full-text and neural information retrieval to facilitate the recognition of document portions to annotate and provides optimized annotation techniques for different machine learning tasks.

3.1.2. Altilia Models – AI Models Training and Fine-Tuning

Altilia Models gives access to a library of out-of-the-box pre-trained models and algorithms that can be fine-tuned with the customer's datasets (based on real-world documents) according to the specific business use case and workflow requirements. Additionally, customers can import custom AI models and train them with their datasets. Altilia Models provides important features like robust transfer learning, connection with Hugging Face[1] repository, that is the largest library of pre-trained state-of-the-art models, to easily find models to fine-tuning on custom datasets to solve specific tasks. Another key feature is constant tracking of behavior and performance of models by validating on a rich set of metrics, it provides the way to visually track the training, validation, and testing process of a model. Continual learning and few-shot learning are other features used to visually review labels predicted by models and use active learning to improve model. Altilia Models provides support for over 50 languages.

3.1.3. Altilia Skills – Complex AI Skills Creation

Processing information from complex documents, regardless of format and layout, requires the use of highly sophisticated AI capabilities. This module allows combining multiple AI models, data functions, and knowledge-graph capabilities to build sophisticated AI skills, capable of recognizing and understanding all the key contents and components contained within documents. It provides a way to create advanced document processing skills, which can combine multiple AI models, predefined data manipulation functions, search queries, rules-based hybrid AI functions, and custom algorithms. Altilia Skills can be used for data and concepts identification within both standard form-like documents and complex documents with mixed layouts (including framed and unframed tables, text, charts, columns, and backgrounds).

[1] https://huggingface.co/

3.2. Automate

The Automate module of AIA allows the integration of multiple AI skills and models with RPA capabilities to deploy Intelligent Document Processing workflows. This module also allows setting up connectors to integrate the workflows with external systems and applications across the enterprise, and provides users with a review and validation tool, to give feedback and enhance the ML models' accuracy and performance. The Automate module provides three main tools to design and build process automation workflows, to connect workflows with external systems and tools, and most importantly to enable users to review and validate data extracted by AI skills. This last tool concretely enables the adaptive AI paradigm implemented in AIA because it allows the platform to leverage user feedback to re-train AI algorithms improving workflows efficiency and accuracy over time.

3.2.1. Altilia Workflows – Process Automation

By the Altilia Workflows tool users can easily create and design automated workflows that streamline complex, document-intensive processes that normally require significant human understanding and intervention. AI skills enable to process, analyze, and understand complex documents with human-level accuracy, no matter what format or layout. This tool provides workflow templates that can be customized by a few clicks. Users can build workflows by combining different tasks available in the platform. Users can configure input/output connectors, execution behavior, scheduling, accuracy thresholds, and workflow SLA, to enable execution strategies that exactly match specific business process needs. Workflow can be executed in attended and unattended mode.

3.2.2. Altilia Reviews – Human-In-The-Loop Validation

The Altilia Reviews tool provides users with a Human-In-The-Loop (HITL) AI mechanism, that allows reviewing and validating workflow results. The Human-In-The-Loop AI cycle is one of the main enabler of the adaptive AI method implemented in the platform. AI algorithm predictions can be validated manually or by specific validation rules, such feedback is used to re-train and enhance AI models. In particular, human validation and feedback are essential to enable the continuous monitoring and training of AI models to enhance models capabilities, improve data quality and accuracy. More in detail,

this tool provides a point-and-click interface to review and validate data, metadata, and concepts obtained from the execution of workflows. It uses semantic faceted search, queries in natural language, and knowledge representation functions to validate the most sophisticated data, metadata, objects, and concepts recognized in documents. This way this module provides also explainable AI features supporting trustworthy AI applications.

3.2.3. Altilia Connectors – Platform Interoperability

To execute an Intelligent Document Processing automation workflow, it is necessary to gather and exchange data between internal and external sources. AIA platform provides pre-built input and output connectors that make it possible to connect and integrate workflows with all external applications and systems, including RPA, CMS, ERP, CRM systems, and even custom-built tools. It provides support for documents and contents ingestion from any internal or external source by using predefined template connectors to FTP folders, local or remote folders, REST API, RPA, ERP, CRM systems, and any other document source. Some special connectors allow applying web wrapping and scraping techniques to gather documents and page contents of entire (deep) web sites. These type of connectors are particularly useful in ESG data collection applications where documents and contents must be gathered from company websites, online newspapers, social media, online document sources, aggregators, etc.

3.3. Understand

Ultimately, the value of modern innovative Intelligent Document Processing technology depends on the ability to analyze and understand the data contained within processed documents. The Understand module of the AIA platform allows for the extraction of meaningful information and insights from all the given document data, empowering intelligent decision-making. This module is composed of three main tools meant, to create an internal representation of the data and metadata extracted from documents, to enable keyword and semantic queries, and to synthesise information into reports and dashboards to facilitate informed decision-making.

3.3.1. Altilia Knowledge Graphs – Internal Data Representation

All the documents and data that the platform intakes and generates are indexed, stored, and handled by a knowledge base grounded on a multi-structured attribute graph that enables multi-faceted semantic searches, querying, and reporting. The Altilia Knowledge Graph simplifies semantic document indexing and data conceptualization to support intelligent decision-making. It allows an object/concept-first modeling approach that combine symbolic AI with machine learning to better address training of AI models, automation of document-driven processes, and empowerment of decision-making.

3.3.2. Altilia Searches – Search and Sort Information

All processed documents and data are indexed in the Altilia Knowledge Base to simplify the access, search, and filtering of all the available information. The Altilia Searches tool constitutes a user-friendly neural search engine equipped by powerful document filtering mechanisms to easily locate the exact information users are looking for. The tool makes use of an advanced semantic search engine that leverages both full text and neural search to help users find and retrieve documents that contain specific terms and/or concepts. It supports multi-faceted searches, to all types of documents, that combine content queries with filters applied to data, entities, and objects produced by the machine learning algorithms. For example, users can formulate queries to search a specific concept within tables. The tool also provides a way to search by similarity, and to answer to natural language questions. In this case, users write a question or insert a piece of text in the search bar and get accurate answers or a list of relevant documents.

3.3.3. Altilia Insights – Analysis and Reporting

The Altilia Insights module is a conversational decision intelligence tool that makes it possible to understand all the data and metadata available within the Altilia Knowledge Graph, thus giving a complete overview of all contents of the processed documents, to drive smarter, more efficient, and better-informed decision making. We can create multi-dimensional reports by using natural language queries. You write a query in natural language and the system creates a dashboard with multiple charts describing all aspects of the questions known in the knowledge base. It allows exporting charts and reports on your desktop, and sharing dashboards with coworkers.

4. Conclusions

In this paper, we presented Altilia Intelligent Automation (AIA), a platform that offers a promising solution for businesses and organizations seeking to streamline their document processing workflows. With its comprehensive suite of tools, AIA caters to all document processing needs of users, providing a one-stop-shop for businesses looking to automate their processes. The Teach module's ability to create sophisticated AI models for complex

document understanding, without requiring extensive coding knowledge, makes AIA accessible to a wide range of users. The Automate module's intuitive visual workflow editor and wide range of connectors simplify the automation process, enabling users to automate their workflows with ease. Lastly, the Understand module's semantic search and knowledge graph capabilities make retrieving documents and insights from large datasets fast and straightforward. Overall, the AIA platform offers a powerful combination of ease of use, automation, and intelligence, making it an excellent choice for businesses looking to boost efficiency and productivity.

The AIA platform behavior will be further improved, enhancing the Document Layout Analysis and recognition algorithms. In the future, we will extend the Platform capabilities towards a complete machine reading comprehension system providing full QA features, where we can extract the information in a question-answer manner from documents.

References

[1] M. Liao, Z. Wan, C. Yao, K. Chen, X. Bai, Real-time scene text detection with differentiable binarization, 2019. arXiv:1911.08947.

[2] X. Yue, Z. Kuang, C. Lin, H. Sun, W. Zhang, Robustscanner: Dynamically enhancing positional clues for robust text recognition, 2020. arXiv:2007.07542.

[3] M. Li, T. Lv, J. Chen, L. Cui, Y. Lu, D. Florencio, C. Zhang, Z. Li, F. Wei, Trocr: Transformer-based optical character recognition with pre-trained models, 2022. arXiv:2109.10282.

[4] Y. Xu, M. Li, L. Cui, S. Huang, F. Wei, M. Zhou, LayoutLM: Pre-training of text and layout for document image understanding, in: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery &amp Data Mining, ACM, 2020. URL: https://doi.org/10.11452F3394486.3403172. doi:10.1145/3394486.3403172.

[5] J. Li, Y. Xu, T. Lv, L. Cui, C. Zhang, F. Wei, Dit: Self-supervised pre-training for document image transformer, 2022. URL: https://arxiv.org/abs/2203.02378. doi:10.48550/ARXIV.2203.02378.

[6] X. Zhong, J. Tang, A. J. Yepes, Publaynet: largest dataset ever for document layout analysis, 2019. URL: https://arxiv.org/abs/1908.07836. doi:10.48550/ARXIV.1908.07836.

[7] D. Chen, A. Fisch, J. Weston, A. Bordes, Reading wikipedia to answer open-domain questions, in: R. Barzilay, M. Kan (Eds.), Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers, Association for Computational Linguistics, 2017, pp. 1870–1879. URL: https://doi.org/10.18653/v1/P17-1171. doi:10.18653/v1/P17-1171.

[8] S. Robertson, H. Zaragoza, The probabilistic relevance framework: Bm25 and beyond (2009).

[9] V. Karpukhin, B. Oğuz, S. Min, P. Lewis, L. Wu, S. Edunov, D. Chen, W. tau Yih, Dense passage retrieval for open-domain question answering, 2020. arXiv:2004.04906.

[10] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, CoRR abs/1706.03762 (2017). URL: http://arxiv.org/abs/1706.03762. arXiv:1706.03762.

[11] J. Devlin, M. Chang, K. Lee, K. Toutanova, BERT: pre-training of deep bidirectional transformers for language understanding, CoRR abs/1810.04805 (2018). URL: http://arxiv.org/abs/1810.04805. arXiv:1810.04805.

[12] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, Roberta: A robustly optimized bert pretraining approach, 2019. arXiv:1907.11692.

[13] A. Radford, K. Narasimhan, T. Salimans, I. Sutskever, Improving language understanding by generative pre-training (2018).

[14] R. Hanslo, Deep learning transformer architecture for named entity recognition on low resourced languages: State of the art results, 2021. URL: https://arxiv.org/abs/2111.00830. doi:10.48550/ARXIV.2111.00830.

[15] Kowsari, J. Meimandi, Heidarysafa, Mendu, Barnes, Brown, Text classification algorithms: A survey (2019). URL: https://doi.org/10.33902Finfo10040150. doi:10.3390/info10040150.

[16] P. Rajpurkar, J. Zhang, K. Lopyrev, P. Liang, SQuAD: 100,000+ questions for machine comprehension of text, in: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Austin, Texas, 2016, pp. 2383–2392. URL: https://aclanthology.org/D16-1264. doi:10.18653/v1/D16-1264.