Performance Insights: Navigating the Landscape of OCR Engines – Update

A Comparative Analysis

all updates are highlighted in orange

March 2025

Table of Contents

Executive Summary

PLANET AI's software suite, IDA (Intelligent Document Analysis), claims to offer outstanding Optical Character Recognition (OCR) accuracy by leveraging patented core technology, thereby minimizing the risk of "garbage in, garbage out". This white paper is intended to serve as a **benchmark for comparing commercial and open-source OCR solutions**. The dataset used can be made available upon request for examination and result reproduction.

Although all commercial engines surpass open-source solutions in performance, **IDA stands out by delivering market-leading OCR results**. It achieves exceptional accuracy across a variety of challenging scenarios, including distorted scans, poor-quality images, and difficult-to-read handwriting.

In this update, we evaluate the **LLM-based engines Gemini 2.0 Flash, GPT-4o, and Mistral OCR**. While promising for simpler documents, they **currently fall short of traditional OCR engines**, often struggling with complex layout and yielding higher error rates.

PLANET AI is a research-driven company dedicated to developing software products with human-inspired cognitive capabilities for information processing and understanding. IDA serves as a versatile suite for enhancing tasks within intelligent document processing value chains.

Introduction

In the rapidly advancing field of digital information processing, Optical Character Recognition (OCR) technology stands out as a cornerstone for transforming visual data into manageable, editable, and searchable digital text. The importance of OCR technology spans across various sectors where data digitization and automation are pivotal. The accuracy of OCR in the initial stages of document processing is crucial: Any errors in character recognition can significantly impact downstream tasks such as document classification and data extraction, making precise information inaccessible for these processes.

As a proclaimed leading provider for OCR engines, this white paper aims to provide a comprehensive comparison between PLANET AI's solution and both commercial and open-source OCR engines in the market. The paper will briefly introduce the diverse dataset utilized for this comparison, outline the evaluation criteria and methodologies employed, and present an initial overview of the comparative results. This overview quickly sheds light on the OCR engines' performance, paving the way for a deeper discussion.

The appendix will complement this discussion by exploring the dataset's intricacies, detailing the evaluation process, and providing in-depth descriptions of each OCR engine, including how the results were generated.

Dataset

We created a dataset originating from the Document UnderstanDing of Everything (DUDE) challenge, encompassing a **wide variety of document types, including handwritten notes and historical documents** (see Figure 1).

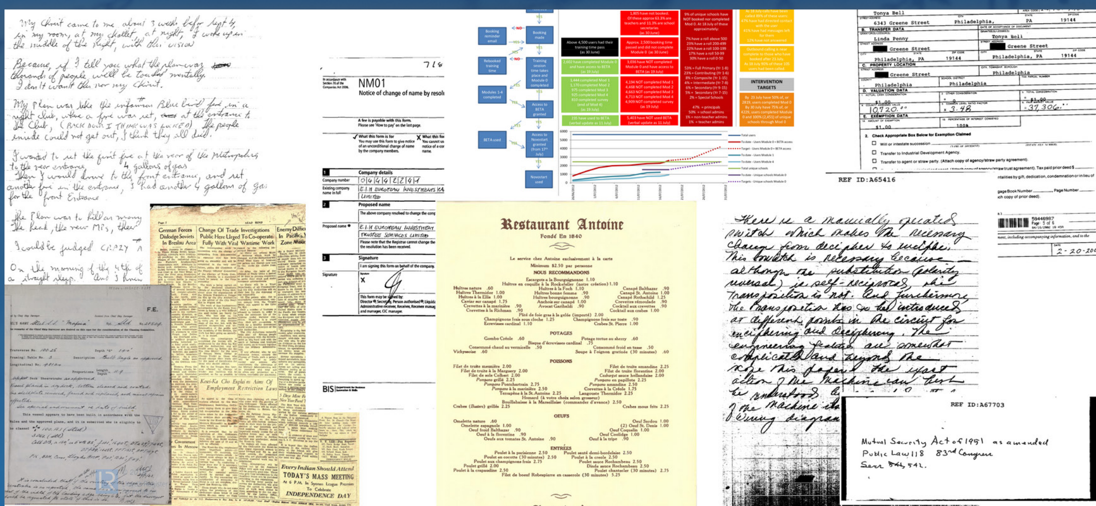


Figure 1: Examples of the dataset. The documents vary in text type, resolution, color depth, and subject.

Curated by 'jordyvl' and hosted on a public repository on Hugging Face, it is licensed under CC-BY-4.0 and was compiled on June 30, 2023. Our selection aims to present real-world OCR challenges and consists of approximately 200 images. It can be made available upon request for examination and result reproduction.

Evaluation Methodology

In assessing the quality of OCR engines within our analysis framework, we employ a meticulous and tailored approach to ensure accuracy and reliability. We opt for measuring the **Character Error Rate (CER)** over the Word Error Rate (WER), considering its superior precision in capturing the fine details of OCR performance. This choice highlights our commitment to granularity in evaluating text recognition capabilities. Moreover, we implement basic normalization techniques to address the diversity in textual elements. This includes various forms of double quotes and brackets, ensuring uniformity and comparability across different outputs.

Our methodology accommodates the non-linear nature of text organization within documents; thus, the reading order is not considered a critical factor in our analysis (see Figure 2). Additionally, we adopt a nuanced stance towards segmentation errors, recognizing that both under-segmentation and over-segmentation do not inherently detract from the overall quality of the OCR engine. This comprehensive and adaptable strategy enables us to provide a robust comparison of OCR engines, grounded in a realistic and practical understanding of document processing challenges. A more detailed description of the comparison methodology can be found in the appendix.

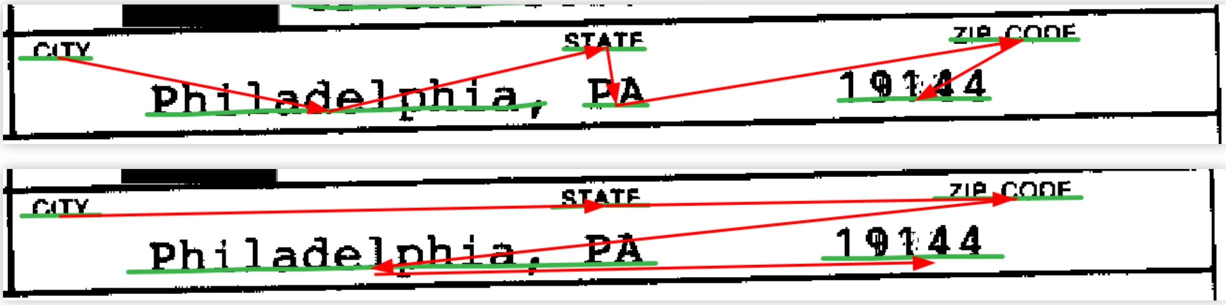
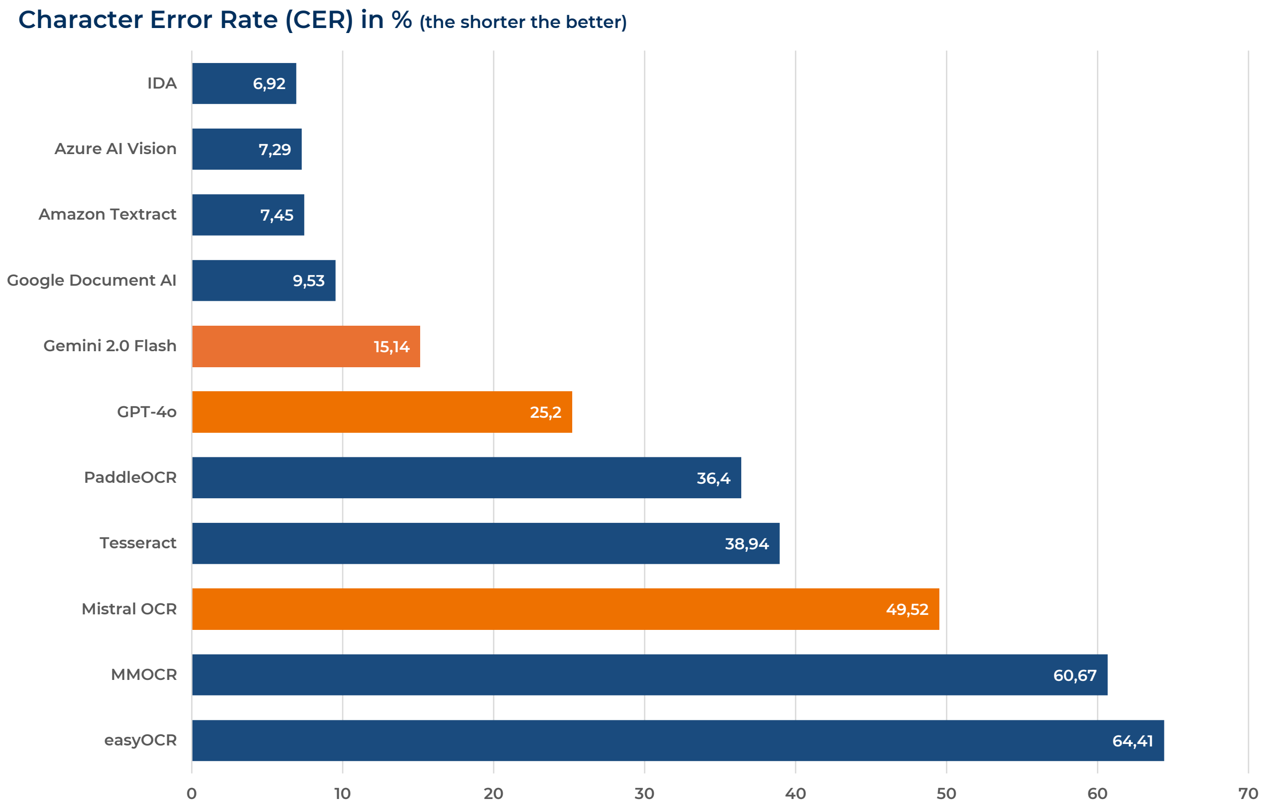


Figure 2: In our comparison of OCR results, we refrain from penalizing variations in reading order and segmentation as they often involve subjective interpretations.

Results

This section provides an overview of OCR engine performances, focusing on both the **Character Error Rate (CER)** and the **computational resources required by on-premises engines**. For deployments on contemporary 4-core notebooks, we classify the computational load as either 'low' (averaging less than 10 seconds per page) or 'high' (exceeding 40 seconds per page). This distinction guides us in understanding not just efficiency but also practical deployment implications.



Engine	CER	Used Deployment	Computational Effort	Date of Results Production
IDA	6.92%	on-premises	low	Feb 2024
Azure AI Vision	7.29%	cloud	n/a[2)]	Q1 2023
Amazon Textract	7.45%	cloud	n/a[2)]	Q1 2023
Google Document AI	9.53%	cloud	n/a[3)]	Oct 2023
Gemini 2.0 Flash	15.14%	cloud	n/a[3)]	March 2025
GPT-4o	25.20%	cloud	n/a[3)]	March 2025
Paddle OCR	36.40%	on-premises	high	Dec 2023
Tesseract	38.94%	on-premises	n/a[2)]	Q1 2023
Mistral OCR	49.52%	cloud	n/a[3)]	March 2025
MMOCR	60.67%	on-premises	high	Dec 2023
easy OCR	64.41%	on-premises	high	Dec 2023

1) Images and results are updated in many cases to obtain better outcomes. Refer to the appendix for an in-depth exploration.
2) Not applicable since OCR results were provided as part of the DUDE Competition dataset.
3) Not applicable due to cloud deployment.

Our analysis divides OCR engines into two main categories:

- **Open-source engines**: easyOCR, MMOCR, PaddleOCR, Tesseract
- **Commercial engines**: Amazon, Azure, Google, IDA (PLANET AI)
- **Multimodal LLM-based engines**: GPT-4o (OpenAI), Mistral OCR, Gemini 2.0 Flash (Google)

Commercial engines significantly outperform their open-source counterparts: In 174 of 199 document pages (~87%), any commercial engine performed better than any open-source engine. Notably, open-source engines often struggle with complex text recognition tasks, including deciphering handwritten notes and accurately reading text in unconventional orientations (e.g., rotated at 90°, 180°, or 270°) or complex layouts.

Commercial engines demonstrate robust consistency in performance. However, the quality varies across different documents, rendering it challenging to make generalized statements when categorizing these documents into groups such as letters, forms, or text styles. The following tendencies can be observed:

- Google slightly falls short in performance compared to its commercial counterparts.
- IDA and Amazon excel at handling dense text or low-resolution scenarios more efficiently than Azure and Google.
- In the specific context of forms, IDA follows a distinct approach to checkboxes that is different from any other commercial engine (see appendix for more information).
- A strength of commercial OCR technology includes adept handling of raster-based form fields; an area where only Google shows room for enhancement.

Engines based on multimodal large language models are not very consistent. They show very good results on some parts of the documents, which are in some cases even better than those of the classic engines. However, they completely refuse to output OCR results on other documents. For example, Mistral OCR claims that over 30 documents are images and thus provides no OCR results at all. A more detailed analysis can be found in the appendix.

Conclusion

In the overarching narrative of OCR engine evaluation, commercial engines undeniably set a high standard, eclipsing open-source alternatives. Notably, the multimodal LLM-based engines signal an exciting frontier with their competitive edge against open-source engines.

PLANET AI's IDA distinguishes itself further by delivering the highest overall accuracy and thus underscores its status as one of the leading OCR engine providers. By delivering outstanding accuracy, IDA effectively mitigates the risk of "garbage in, garbage out", setting a solid foundation for subsequent tasks. **All while offering both on-premises and cloud deployment – a critical consideration for privacy-conscious organizations.** Combined with sophisticated machine learning capabilities for document classification and data extraction, IDA serves as a versatile software suite for enhancing tasks within document processing value chains.

About PLANET AI

PLANET AI is a research-driven company dedicated to developing software products with human-inspired cognitive capabilities for information processing and understanding. By utilizing proprietary deep learning technology, PLANET AI empowers organizations to unlock information trapped in documents, seize digitization opportunities, and eliminate manual effort for data capture. The Intelligent Document Analysis software suite offers comprehensive capabilities for customers with the common desire for short time-to-value automation and high-quality data capture, extraction, and understanding.

PLANET AI serves a variety of customers that include Fortune 500 companies, scanning service providers, as well as software vendors in business process automation and content management. Since its beginnings in 1992, PLANET AI has has established itself as a global technology leader in cognitive computing. In 2023, German IT provider Bechtle acquired a majority share of PLANET AI.

Appendix

Dataset in Depth

To benchmark the performance of various OCR engines, this white paper is grounded on a meticulously curated, diverse, and complex dataset, forming the cornerstone of our comparative analysis. This section aims to demystify the complexities of the dataset preparation process, ensuring it is comprehensible and straightforward for all readers.

The foundation for our analysis is a dataset drawn from the Document UnderstanDing of Everything (DUDE) challenge, an ambitious attempt aimed at pushing the boundaries of document understanding technologies. This dataset provides a broad spectrum of documents in PDF format. Featuring a mix of handwritten notes, machine-generated texts, modern, and historical documents, it represents a wide array of challenges encountered in real-world OCR tasks. Notably, OCR results for Azure, Amazon Textract, and Tesseract are provided as part of the dataset, and were used here.

The raw materials for this study were sourced from a publicly accessible repository hosted on Hugging Face, under the custodianship of 'jordyvl'. The data are shared under the CC-BY-4.0 license, which ensures credit is given to contributors and permits others to use this dataset in their projects. It was collected on June 30, 2023.

Refinement Process

From the initial collection, the test partition underwent a rigorous filtering process. The goal was to distill the dataset down to an even more challenging and heterogeneous subset, consisting of approximately 200 images. To level the playing field for OCR engines incapable of processing PDFs directly, these documents were converted into PNGs and JPEGs, common image formats widely supported across OCR solutions.

Quality Assurance

Quality assurance played a pivotal role in preparing this dataset:

- **Baseline Annotation**: This foundational step was meticulously reviewed by a specialist on our team to ensure the accuracy and reliability of the baseline data.
- **OCR Reconnaissance Hypotheses by IDA**: The initial OCR results generated by IDA 5.1 underwent a thorough annotation process. Following this, at least two team members conducted a stringent review of these results, ensuring the analysis was supported only by high-quality data and eliminating any bias from the ground truth.

Through this detailed and methodical preparation process, the dataset not only embodies a wide range of difficulties present in OCR tasks but also stands as a testament to high-quality, thoroughly vetted benchmarking material. This approach ensures that the comparative analysis of OCR engine performance is both fair and grounded in real-world document diversity and complexity.

Evaluation Methodology in Depth

Traditional evaluation metrics, while useful, often fall short in providing a comprehensive assessment of an OCR system's performance, especially when it comes to the intricate interplay between text detection and text recognition. The motivation behind choosing the **Character Error Rate (CER)** as the foundation for our evaluation scheme stems from its direct relevance to the end goal of OCR – accurately converting images of text into machine-encoded text. CER offers a granular view of an OCR system's accuracy by quantifying the errors made at the character level, including insertions, deletions, and substitutions. These errors are critical in understanding the efficacy of an OCR solution, as they can significantly impact the usability of the digitized text, especially in contexts requiring high levels of accuracy such as legal documents and scholarly archives.

However, raw CER, in its traditional form, does not fully address the complexities inherent in end-to-end OCR systems. Such systems not only recognize characters but must first accurately detect text within images, a task subject to its own set of challenges like varying fonts, sizes, and backgrounds. Therefore, we utilize the **evaluation scheme proposed in [1]**. The scheme builds upon the well-established CER to include considerations for text line detection, reading order, and the geometric positioning of text lines. This comprehensive approach ensures that the evaluation not only measures the accuracy of character recognition but also how well the system can identify and process text in its varying presentations and arrangements.

Moreover, by allowing for the configuration of penalties related to reading order and the geometric alignment of text lines, the scheme offers **adaptability to different use cases**. For example, in some applications, the precise ordering of text might be critical, requiring stricter penalties for errors in reading order, while in others, the focus might be more on the accuracy of extracted content than its spatial arrangement. This flexibility ensures that the evaluation metric can be tailored to best reflect the priorities of the specific context in which an OCR system is deployed. The introduction of tolerances for over-and under-segmentation of text lines further adds to the scheme's robustness. These common issues can significantly affect the quality of OCR output, and by incorporating allowances for them, the evaluation scheme provides a more accurate reflection of a system's practical performance.

In essence, the chosen evaluation scheme is designed to offer a nuanced, comprehensive, and adaptable framework for assessing OCR systems. It acknowledges the multifaceted challenges faced by end-to-end OCR technologies and strives to provide a metric that can guide the development of more accurate, robust, and user-centric solutions.

To get the final CERs for each engine we performed the following steps:

Normalization of OCR Results: For each OCR solution tested, results were standardized into normalized text (*.txt) files. This step was crucial to ensure consistency across different OCR outputs, allowing for a direct comparison. The outputs were organized either line by line or word by word, depending on the structure of the OCR engine outputs. Furthermore, characters with similar interpretations were normalized (e.g. all characters “[{ （ (《 〈 ［ 〔 〔“ were mapped to “(“).

Character Error Rate (CER) Evaluation: The CER was calculated utilizing the above introduced end-to-end evaluation scheme. For each OCR engine, the CER was calculated across the different sample texts. The following conditions were applied for the evaluation:

- **Reading Order Irrelevance**: The sequence in which text appeared (reading order) was not considered during the evaluation, acknowledging that different OCR solutions might interpret document layouts differently. In many situations this is highly subjective.
- **Segmentation Error Correction**: Segmentation errors (incorrectly split or merged text lines) were corrected in cases where there were fewer than 256 ground truth (gt) lines. This ensured minor segmentation errors did not disproportionately affect the overall assessment. And again, these kind of “errors” are also often subjective, see Figure 2.

After implementing the above criteria, the CER for each sample was averaged on a file-wise basis. This approach provided a clear, objective, and uniform metric for comparison, enabling an accurate evaluation of each OCR solution’s text recognition capabilities across a diverse set of documents. Through this methodology, the white paper aims to offer valuable insights into the performance and reliability of different OCR technologies, guiding users and developers in their selection of the most suitable OCR solution for their specific needs.

OCR Engine Overviews

In our analysis, we not only explored the widely recognized cloud solutions provided by Amazon, Google, and Microsoft, but also incorporated a variety of open-source options. This decision was motivated by frequent inquiries we receive, regarding how commercial solutions compare to their open-source counterparts. Additionally, we have included the results of three multimodal LLM-based engines to offer a comprehensive overview.

Amazon Textract

Amazon Textract is a fully-managed machine learning service that automatically extracts text and data from scanned documents. It has been trained on a broad set of data to recognize and process various document formats and types, such as forms and tables. The engine generates a JSON file, containing bounding boxes around each line and word along, with a confidence score.

OCR Result Production

OCR results were provided as part of the DUDE Competition dataset[2] and extracted on February 20, 2024.

Azure AI Vision

Azure AI Vision is a cloud-based service provided by Microsoft that offers algorithms for processing and analyzing visual data. It enables developers to integrate capabilities in Optical Character Recognition (OCR), a well as image classification, object detection, and face recognition into their applications. The engine generates a JSON file, containing bounding boxes around each line and word, along with a confidence score.

OCR Result Production

OCR results were provided as part of the DUDE Competition dataset[2] and extracted on February 20, 2024.

easyOCR

easyOCR presents itself as a versatile and user-friendly solution designed to accommodate a wide range of users' needs in the field of Optical Character Recognition. Designed for ease of use, it combines CNN (Convolutional Neural Networks) and sequence modeling (Transformer) to recognize text. It supports multiple languages and scripts and is aimed at developers needing a straightforward, robust OCR solution. It benefits from a large and active community that contributes to continuous improvement and support. However, handwritten text recognition is not fully supported.

OCR Result Production

OCR results were produced using https://github.com/JaidedAI/EasyOCR as of December 2023.

Gemini 2.0 Flash

Gemini 2.0 Flash is an advanced multimodal Large Language Model (LLM) developed by Google. It supports various input modalities, such as text and images. For the OCR task, it was instructed with a specific user prompt to accurately extract text content from images, line by line. Interestingly, more detailed prompts describing the problem in greater depth led to poorer performance. In particular, the model tended to decline answers when guided in more detail. Therefore, a rather simple prompt was used: *Please perform an OCR. Return the recognized text line by line. Output only the recognized text.*

OCR Result Production

OCR results were produced using Google's API as of March 2025.

Google Document AI

The Google Document AI engine is a machine learning model designed to understand and process text within documents, extracting valuable information automatically. It has been trained on a diverse dataset of document types and structures to accurately interpret various forms of text, such as forms, invoices, receipts, and more. The engine generates a JSON file, containing bounding boxes around each line and word, along with a confidence score.

OCR Result Production

OCR results were produced using the Google Document AI demo as of October 2023.

GPT-4o

GPT-4o is an advanced multimodal Large Language Model (LLM) developed by OpenAI. It supports various input modalities, such as text and images. For the OCR task, the model was guided with a specific system and user prompt to accurately extract text content from images, line by line. It was necessary to encourage the system to provide results even in challenging situations, as the model tended to decline output with comments like:

I'm sorry, but the text in the image is too distorted and unclear for accurate optical character recognition (OCR).

We used the system prompt:

You are a transcription assistant performing an OCR. You return OCR results line by line. You limit yourself to the OCR results in your answer. Do not provide additional information, e.g., interpretations, backtick notation, comments like: Sure, here is the OCR result. Please try hard and even return a best guess, if the input is very difficult! Text in the input can be oriented.

And the user prompt:

Please perform an OCR on the input image provided.

OCR Result Production

OCR results were produced using OpenAI's API as of March 2025.

IDA

IDA Recognition is the core feature for data capture in PLANET AI's IDA suite. It leverages patented core technology to deliver exceptional accuracy in the most challenging scenarios, such as distorted, poor-quality scans with machine-print or difficult-to-read handwriting. This approach ensures that all possible transcriptions of a given text are preserved without any loss of information. IDA generates a JSON file, containing bounding boxes around each line and word, along with a confidence score. Please refer to our datasheets for more information.

OCR Result Production

OCR results were produced using IDA 5.2 (as of February 2024) with the following configuration:

- Textfinder: *PLANET_TEXTFINDER_MIX_BEST_QUALITY*
- Reading net: *PLANET_READING_MODERN_BEST_QUALITY*
- Decoding: *LANG_ENGLISH*

Mistral OCR

Mistral OCR is a model specifically tailored for optical character recognition (OCR) scenarios. It is based on a multimodal Large Language Model (LLM), leveraging its capabilities to handle diverse types of data inputs. Unlike other systems, Mistral OCR offers a dedicated OCR API endpoint that does not require prompting, streamlining the process for users. The output is provided in Markdown format, ensuring clear and structured results. Additionally, it converts input data into structured outputs, which include both text and images, enhancing the versatility and usability of the OCR results.

OCR Result Production

OCR results were produced using Mistral's API as of March 2025.

MMOCR

MMOCR stands out as an innovative, open-source toolbox designed for a wide range of complex OCR tasks, utilizing deep learning techniques. Developed on the frameworks of PyTorch and mmdetection, MMOCR is a vital component of the acclaimed OpenMMLab project. It is tailored to efficiently address the challenges of text detection, text recognition, and key information extraction. The engine is compatible with PyTorch 1.6 and higher versions, ensuring it leverages the latest advancements in deep learning technology.

OCR Result Production

OCR results were produced using https://github.com/open-mmlab/mmocr as of December 2023.

PaddleOCR

PaddleOCR is an open-source, deep learning-based OCR toolkit developed by PaddlePaddle, an AI platform from Baidu. It employs deep learning models optimized for efficient performance across a wide range of computing devices. It also emphasizes support for languages and scripts beyond Western languages, including Asian languages. PaddleOCR claims to provide state-of-the-art models for various text recognition scenarios, e.g., text detection, recognition, and end-to-end text spotting, including support for handwritten scripts. Its models are trained on large-scale datasets, enabling accurate and efficient text extraction in a wide range of documents.

OCR Result Production

OCR results were produced using https://github.com/PaddlePaddle/PaddleOCR as of December 2023.

Tesseract

Tesseract OCR Engine is a leading open-source OCR tool and stands out for incorporating a neural net (LSTM) based OCR engine in its 4th version, enhancing line recognition capabilities while still supporting the legacy character pattern recognition system from version 3. Tesseract is renowned for its ability to recognize over 100 languages straight off the bat, with robust UTF-8 support for comprehensive language and character set coverage. The engine processes images in formats such as PNG, JPEG, and TIFF, and exports data in multiple formats including plain text, hOCR, PDF, and more.

OCR Result Production

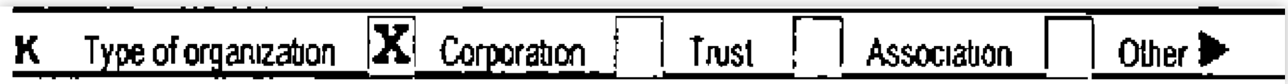
OCR results were provided as part of the DUDE Competition dataset[2] and extracted on February 20, 2024.

Focus Scenario: Forms

Forms play a crucial role in the IDP domain, which is why we pay special attention to this area. In some forms, the majority of the text is pre-typed and easy to read. The primary challenge with these documents lies in the fields filled out by individuals. Reading these fields is essential for extracting the relevant information from the document. We demonstrate how commercial engines manage two specific cases: checkboxes and raster-based fields.

Checkboxes

Forms frequently use checkboxes to solicit choices, such as the individual's sex or agreement to terms and conditions. For example, we evaluate instances where multiple checkboxes are provided. We then present the sorted raw reading results from commercial engines.



K Type of organization X Corporation Trust Association Other

Amazon:	"K Type of organization" "Corporation" "Trust" "Association" "Other"
Azure:	"K Type of organization X Corporation" "Trust" "Association" "Other"
Google:	"* Type of organization X Corporation Trust" "Association" "Other>"
IDA:	"K" "Type of organization" "× Corporation" "Ø Trust" "Association" "Ø Other"

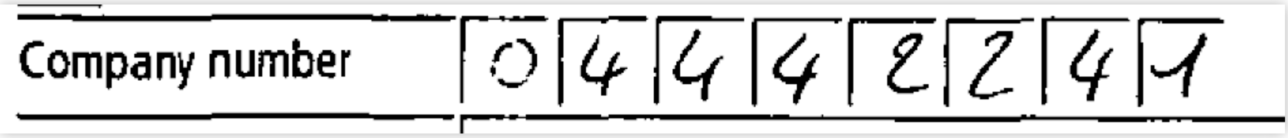

REQUEST FOR TRAVEL BY
COMMERCIAL AIR RAIL TPA
INDIVIDUAL OBJECTS TO TRAVEL BY AIR

Amazon: "COMMERCIAL AIR" "RAIL" "TPA"
Azure: "[ ] COMMERCIAL AIR" "RAIL [ ] TPA"
Google: "COMMERCIAL AIR" "RAIL [ ] TPA"
IDA: "COMMERCIAL AIR" "× RAIL" "Ø" "TPA"

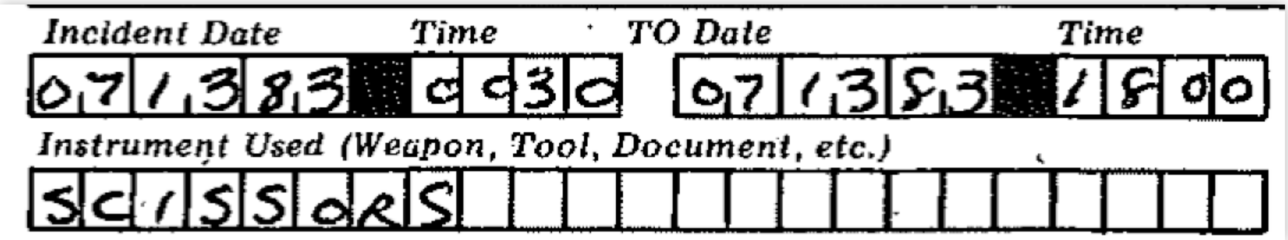
IDA employs special characters "×" and "Ø" to indicate whether a checkbox is marked; for other engines, "X" and "[ ]" symbols are utilized to denote the status of a checkbox. Generally, both methods effectively facilitate the extraction of checkbox data.

Raster-based Form Fields

Users can freely write text in many form fields. These fields are readily identifiable and accessible using commercial engines. In contrast, inputs like dates or IBANs must adhere to a predefined format. To assist users, raster-based fields are implemented, which conversely complicates the reading process for these entries.


Company number 0 4 4 4 2 2 4 1

Amazon: "04442241"
Azure: "04442241"
Google: "014/4|4|2|21414"
IDA: "04442241"


Incident Date Time TO Date Time
071383 0030 071383 1800
Instrument Used (Weapon, Tool, Document, etc.)
SCISSORS

Amazon: "071383 0030 071383 1800" "SCISSORS"
Azure: "071383 0030 071383 1800" "SCISSORS"
Google: "071383" "¢ ¢30 077383" "sc,ssodklsl"
IDA: "071383" "0030" "071383" "1800" "SCISSORS"

All commercial engines, except Google, handle raster-based text adeptly.

Multimodal LLM based Engines

In this update of the benchmark paper, Gemini 2.0 Flash, GPT-4o, and Mistral OCR were evaluated. Focusing solely on the pure Character Error Rate (CER) can be misleading. The LLM-based engines are not as bad as the results suggest throughout all images. In about 10% of the documents, none of the LLM-based engines produced feasible results. Common issues included difficulty in interpreting the text or the text being too challenging, with transcriptions sometimes stopping midway through a page. However, this is not true for all documents.

To provide better insights and explain the positive impressions one might have while testing these LLM-based engines, we limited our comparison to the 100 easier documents (roughly half of the total). By doing so, we gain a different perspective on the quality of LLM-based engines. Gemini's performance is approximately 25% less effective than that of traditional engines which is quite competitive. However, the error rates of Mistral OCR and GPT-4o are significant enough to make them unsuitable for straight-through processing, even with these easier documents. The following table shows the CER for the 100 easiest documents of the LLM-based approaches compared to IDA.

	Mistral OCR	GPT-4o	Gemini 2.0 Flash	IDA
CER	13.27%	5.58%	1.94%	1.53%

We anticipate that LLM-based engines will match or even surpass classical engines in the coming years. However, their ecological footprint and processing time are still significantly different. We will continue to monitor these trends and aim to focus on the most effective technologies when the time is right.

Sources

[1] G. Leifert, R. Labahn, T. Grüning and S. Leifert, "End-to-End Measure for Text Recognition," 2019 International Conference on Document Analysis and Recognition (ICDAR), Sydney, NSW, Australia, 2019, pp. 1424-1431, doi: 10.1109/ICDAR.2019.00-16.

[2] Jordy Van Landeghem, Rubèn Tito, Łukasz Borchmann, Michał Pietruszka, Dawid Jurkiewicz, Rafał Powalski, Paweł Józiak, Sanket Biswas, Mickaël Coustaty & Tomasz Stanisławek, "Competition on Document UnderstanDing of Everything (DUDE) " 2023 International Conference on Document Analysis and Recognition (ICDAR), doi: 10.1007/978-3-031-41679-8_24