

# DocDigitizer Launches ARENA, an LLM Benchmarking Platform That Measures Extraction Speed, Accuracy, and Cost

*DocDigitizer Launches ARENA, an LLM Benchmarking Platform That Measures Extraction Speed, Accuracy, and Cost Across Seven Major Providers*



LISBON, PORTUGAL, April 16, 2026

/EINPresswire.com/ -- — DocDigitizer, the document automation company, today announced the commercial availability of ARENA, a benchmarking platform for LLM-based document extraction accessible at [arena.docdigitizer.com](https://arena.docdigitizer.com). The platform enables engineering teams to systematically compare seven major LLM providers on real-world documents, measuring extraction speed, field-level accuracy, and cost per extraction in a single reproducible run.

## Background

Engineering teams building document extraction pipelines routinely face three interrelated questions: which LLM provider offers the lowest latency at production volume, which delivers the highest field-level accuracy for a given document schema, and which is cost-effective enough to run at scale. Current evaluation methods — ad hoc Python scripts, manual testing against individual providers, or reliance on vendor-published benchmarks run on curated datasets — produce results that are difficult to reproduce and quick to become outdated as new models are released.

The challenge is compounded by the pace of change in the LLM market. New model versions, revised pricing, and shifting capability profiles mean that the optimal provider for a given workload can change within weeks. Teams that evaluated a provider six months ago may be operating on stale assumptions.

## Platform capabilities

ARENA addresses this gap by providing a standardized benchmarking environment purpose-built for document extraction. Teams upload a dataset, define the extraction schema (invoices, contracts, identity documents, or any structured output), and run the same extraction job across every supported provider and model. The platform returns three metrics per provider, per

model, per document:

- Speed — extraction latency measured to the millisecond
- Accuracy — field-level correctness scored against ground truth on a 0.0–1.0 scale
- Cost — actual API cost per extraction in USD

Results are presented as dashboards, scatter plots, and per-dataset heatmaps, surfacing provider performance differences across document types. Benchmarks are defined declaratively using DBL (Document Benchmark Language), making them version-controllable and re-runnable as new models become available.

## Provider coverage

At launch, ARENA supports OpenAI (GPT-4o, GPT-4.1, o-series), Anthropic (Claude 3.5 Sonnet, Claude 4), Google (Gemini 1.5, Gemini 2.0), Mistral, Cohere, AWS Bedrock, and Azure OpenAI. Both text and vision extraction modes are supported, with three extraction strategies available per provider.

## Target users

The platform is designed for developers, ML engineers, and data engineers building document extraction pipelines, as well as technical leaders who need objective data to commit to a provider at scale. Typical use cases involve processing PDFs, invoices, contracts, or identity documents at volumes where provider selection has material impact on cost and quality.

"Opinions about LLMs are cheap. Field-level accuracy scores on real production documents are not. We built ARENA because every team we talked to — including our own — was wasting weeks on evaluation scripts that got thrown away the next time a new model dropped. ARENA turns that work into something a team runs once and re-runs forever."

— João Fernandes, Co-Founder and CEO, DocDigitizer

## Availability

ARENA is available now at [arena.docdigitizer.com](https://arena.docdigitizer.com). A free tier provides 100 credits with no credit card required, sufficient to benchmark a small dataset across all supported providers. Paid plans scale with credit consumption; full pricing is available on the product website.

## About DocDigitizer

Founded in 2018 and headquartered in Lisbon, Portugal, DocDigitizer is a document automation company specializing in production-grade extraction pipelines for regulated industries including finance, insurance, and legal. The company processes millions of documents annually for enterprise customers. ARENA is DocDigitizer's first developer-facing product, extending internal

benchmarking infrastructure to the wider engineering community.

Media contact

João Fernandes Co-Founder and CEO, DocDigitizer | [arena.docdigitizer.com](https://arena.docdigitizer.com) | [docdigitizer.com](https://docdigitizer.com)

João Fernandes

Co-Founder and CEO, DocDigitizer

[info@docdigitizer.com](mailto:info@docdigitizer.com)

---

This press release can be viewed online at: <https://www.einpresswire.com/article/906072549>

EIN Presswire's priority is source transparency. We do not allow opaque clients, and our editors try to be careful about weeding out false and misleading content. As a user, if you see something we have missed, please do bring it to our attention. Your help is welcome. EIN Presswire, Everyone's Internet News Presswire™, tries to define some of the boundaries that are reasonable in today's world. Please see our Editorial Guidelines for more information.

© 1995-2026 Newsmatics Inc. All Right Reserved.