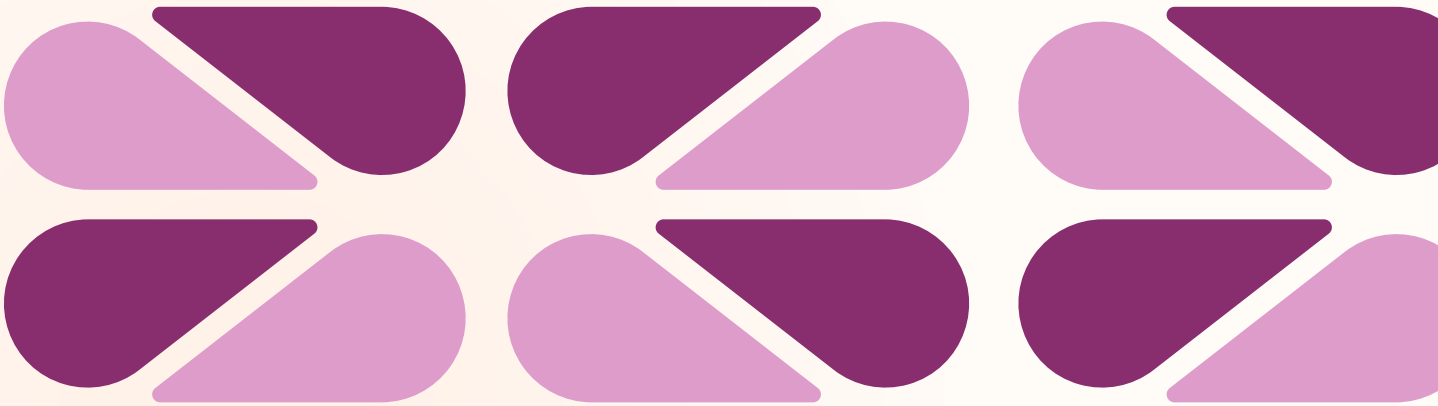
FINANCIAL SERVICES GUIDE TO

Understanding AI: the technology **behind data acceleration**



ALKYMI

Table of Contents

Introduction

Financial firms — even small ones — handle thousands or even millions of documents each day. A constant flow of PDFs, text files, images, presentations, spreadsheets, and emails moves through your firm, carrying data from one place to another. Key data within those documents is what drives your business, but accessing this information requires opening, reading, and processing every document.

With machine learning, you can extract and classify this data automatically, in the future possibly without a human ever opening the document where it's contained. You can also identify specific objects within documents, such as signatures. And, with the aid of additional products and integrations that expand the impact of your ML capabilities, you can make this information available to fuel connected applications, power data analysis, and automate business processes.

This all happens with the aid of the following technologies:

- Computer vision (CV)
- Optical character recognition (OCR)
- Natural language processing (NLP)
- New advances including GPT and LLMs

This guide will explain each of these concepts. It will also cover how you can use these technologies and techniques to automatically surface data for specific use cases in finance and beyond.

Data acceleration

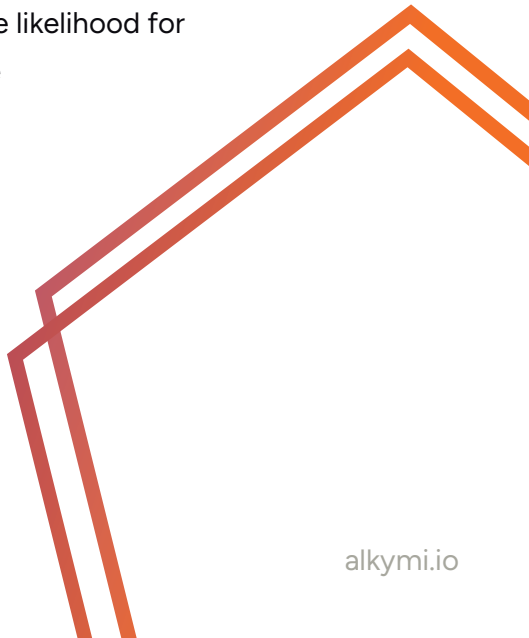
Data acceleration and document processing presents a multilayered challenge. First, documents must be automatically ingested via an API or other integration, and their format and textual contents must be recognized by your model. Next, your model must read and understand any actual text in the document. It must also identify structural elements within each document — tables, graphics, call-out boxes, and so on — and extract the relevant numbers or text, and interpret those.

The output of your model may look like this:

> Fund name: Alpha Fund
>
> Customer name: John Smith

While accurate, this data still needs to be made useful for your purposes by a software product, i.e., by putting this data into a spreadsheet or exporting it via API. For data acceleration and processing to function at scale, this crucial final step needs to work consistently and with minimal human intervention.

It is possible to build traditional software products to solve each of these steps. However, the variety and irregularity of the documents handled by financial institutions makes this impractical, for cost and resource reasons. How could software be pre-programmed to accurately process documents which vary wildly in language, structure, length, typeface, layout, color, and even writing style? As complexity increases, so does the likelihood for error and inability to handle all potential cases — which is where machine learning (ML) comes in.



In a nutshell, ML refers to an algorithm's ability to learn from a dataset and adapt itself accordingly. This is useful for data processing because it allows your model to process new or irregular inputs with accuracy by leveraging its understanding of the examples present in the training dataset. With ML, even if your model hasn't encountered a specific structure, document configuration, or type of text-based information before, it has processed enough other documents to successfully identify patterns on the fly and process the new inputs. This adaptability is what makes your data processing model "intelligent."

The following three frameworks are part of any data acceleration platform. No matter the document type or use case, processing complex document workflows requires you to use all three:

Computer vision	**CV**	How your model sees what's in a document
Optical character recognition	**OCR**	How your model identifies text and numbers
Natural language processing	**NLP**	How your model understands text and numbers

For example, when your automation model handles a bank account statement in PDF format, it:

- Uses CV to identify structures like tables and paragraphs within the document
- Uses OCR to identify text in the document
- Uses NLP to extract usable data from the text by identifying and isolating specific concepts or data elements

These technologies are what allow you to extract accurate, structured data based on documents or images that contain data in unstructured format.

Next, let's examine each of these methods in detail and learn how they can be used to solve specific business challenges.

AI Technologies

Computer vision

WHAT IS IT?

Computer vision (CV) is a branch of machine learning that allows computers to "see" digital documents, images, and videos. For the purposes of this section, we are focusing on a type of CV known as "object detection." The same way that the human eye captures visual information — for instance, when a driver observes bright headlights in the distance — CV-enabled computers capture visual information. One application of CV object detection could be to allow a self-driving car to observe other cars on the road the same way a driver does. In the context of intelligent data processing, however, CV object detection is how your model interprets the structure of the PDF, slide deck, image, or other document. Whatever is there — CV is what allows your model to interpret and organize it.

WHAT MAKES IT USEFUL?

CV is one of the key reasons that data acceleration and intelligent data processing is possible in the first place. The objective of using an intelligent model is to automate the process of extracting data from documents at least as well as if a human had reviewed the document manually and taken note of the relevant information. CV is what gives your model visual understanding, allowing it to comprehend and organize information about layout and structure the same way a real human does.

USE CASES IN FINANCE

Financial institutions handle documents that can vary widely, even within a given category, which presents a standardization challenge. Capital call notices, for example, aren't uniform across different funds or even across different investments with the same fund. But every capital call notice contains the same relevant information — e.g., amount due, due date, total committed capital, beneficiary bank account number, etc. — which a human could simply find by reading. In order for your model to interpret or extract this data, it must find it first. This usually happens with the help of CV providing valuable information about how data is organized on the page — i.e., whether it is contained in a table or a paragraph, and where it is on the page relative to other data.

In the case of more complex documents featuring tables, graphs, or other irregular features, CV is essential. It gives your model the ability to decipher the document's layout in order to find what it's looking for.

SHOULD I USE COMPUTER VISION?

It's difficult to imagine a data acceleration platform without computer vision because CV unlocks a number of useful capabilities, including:

- Optical character recognition (OCR)
- Document layout analysis
- Signature, logo, stamp and other object detection
- Table structure analysis

In short: if you're handling documents, your model needs a way to understand what's on the page, and where it is. CV solves this and more. It's also what allows your model to identify and handle document structures like tables, charts, and lists.

Optical character recognition (OCR)

WHAT IS IT?

Optical character recognition (OCR) is a category of software that allows computers to identify text and images. Similarly to computer vision, OCR enables a data processing model to "see" what's contained in a document. The difference, however, is that OCR focuses on determining the actual content of the data, while computer vision (primarily a type of computer vision known as object detection) deals with interpreting the structure and layout of the data. OCR is sometimes considered a subfield of computer vision.

WHAT MAKES IT USEFUL?

OCR is how your model "reads" text from images. The dark areas are then analyzed for patterns by a model that attempts to make a match between each shape it encounters against a known universe of characters — if your model knows what the letter "Q" looks like, for instance, then it knows how to classify each "Q" that it encounters in the document. If it can also classify other letters (along with spaces and punctuation), then it can identify full words, sentences, or even paragraphs. And finally, it's how handwritten information is deciphered.

OCR is especially relevant in the case of data extraction from documents. CV only allows your model to "see" documents; OCR is required for your model to read what it's seeing, if it is text, if your goal is to understand text-based information. In use cases where the objective is to identify visual elements — for instance, a model that looks at documents and simply calls out the presence of a signature — computer vision without OCR may be sufficient. But to extract specific data elements from documents such as names, numbers and dates, your model must be able to use OCR to recognize and identify a wide range of characters.

USE CASES IN FINANCE

As with CV, it's difficult to consider a data acceleration model focused on document workflows without OCR capabilities — this would place a whole class of document workflows out of scope, i.e., those where scanned or handwritten data is in scope. For a model to be able to process or export data, that data must be formatted as a text, not an image. To produce a string from an image of a series of letters and numbers is how OCR adds value, because CV can't do this on its own. OCR works in complement with CV which helps organize these strings into paragraphs and other page structures.

OCR's main benefit is that it converts image-based data into a machine-readable format. This unlocks other capabilities like natural language processing (NLP). However OCR by itself can be used for simple data extraction tasks if little additional processing is required. Some common documents in the financial services industry like checks or bank statements are highly standardized, and can be run through an OCR engine to extract simple elements like customer name, amount, and date — as long as their location on the document never changes. This is an approach known as "templating," which is not required if OCR is paired effectively with CV and presents some challenges in workflows with even limited variability. If the job at hand requires reading and data entry with no irregularities and low complexity, OCR may provide some help by enabling humans to copy data from documents or powering simple templates.

SHOULD I USE OPTICAL CHARACTER RECOGNITION?

OCR opens the door for other tools that can ingest its output and run additional processes. These include:

- Natural language processing (NLP)
- Digitization of paper records
- Identity document verification

With OCR, virtually all text based data in your documents is placed within reach. Your model can obtain the data that it needs, in your desired format, ready to be compiled or exported as needed.

Natural language processing (NLP)

WHAT IS IT?

Natural language processing (NLP) is how a data acceleration platform makes sense of language. As humans, we enjoy our combined abilities to see, to recognize text and characters, and to derive specific meaning. Computers, on the other hand, use CV to see, OCR to recognize characters, and NLP to derive specific meaning.

NLP works by helping your model understand language the same way a human would. We understand the relationship of "when was this person born?" and a noted "birth date" in a document. NLP helps your model understand which data points it's looking for using context that humans would usually infer to find that data.

WHAT MAKES IT USEFUL?

Data acceleration relies on your model's ability to ingest files and produce useful outputs. With NLP, those outputs can evolve beyond raw data. If you want to record every instance that a specific company or individual is mentioned in incoming documents, or create a workflow that populates a spreadsheet with company and contact information every time a document is encountered that mentions Bitcoin — NLP makes such applications possible.

USE CASES IN FINANCE

The use cases for NLP within data acceleration are extensive. A few examples to consider include:

Entity detection	Identifying specific data elements like names and dates.
Document classification	Determining what category a document falls into based on its contents.
Data classification & mapping	Classifying complex text elements against defined classes or hierarchies of classes.
Question answering models	An NLP technique powered by a large language model, capable of answering natural language queries about data.
Text summarization	Automatically produce summaries of complex documents with key elements highlighted.
Real-time news analysis	Scrape news from the web in real-time and run keyword or sentiment analysis to generate insights.

One advantage of NLP is that it can deliver results on large numbers of documents. The same model could identify specific data elements from, for example, decades' worth of customer statements or a single statement submitted live by a customer. Because NLP models can process text much more quickly than humans, it can allow you to obtain high-level insights based on thousands of documents with little time or effort required.

SHOULD I USE NATURAL LANGUAGE PROCESSING?

If you're building an application that will handle complex documents and want to extract specific information from documents, chances are a NLP model or models will come in handy. Example activities in finance that tend to deal with high-complexity documents and large amounts of data include:

- Entering capital call & distribution notices into a portfolio management system
- Creating new client proposals using brokerage account statements
- Targeting information in complex fund subscription documents

With NLP, your data acceleration platform can process a virtually unlimited number of documents and generate more precise insights and data than an actual human would. Moreover, your model can be trained to recognize new patterns across a large collection of documents which the human eye would almost certainly miss.

New advances in AI

AI technology is developing at a rapid pace, with a continued surge of investment and public attention. Next generation large language models (LLMs) like ChatGPT are generating particular interest due to their ability to summarize and synthesize text-based input as well as to answer user questions. Two notable categories of LLMs are "Generative Pre-trained Transformer" (GPT) and "Question Answering" (QA) models. GPT models are able to synthesize and compose text-based responses (with more input and output methods, such as images, close on the horizon and even available now in some cases). QA models are able to answer questions posed to them in natural language by capturing results from a source.

Both GPT and QA models can answer questions, but there are important differences in how they do so. A GPT model will abstract an answer, meaning it will come up with a response based on the context provided. This is called "abstractive question answering." A QA model will find an answer by referring to a source (for example, a document). This is called "extractive question answering." In some workflows, an extractive approach may be ideal because it only provides an answer based on a specified source document and also allows a user to see exactly where that answer came from.

Both are trained on extensive datasets to understand and generate human language, including answering questions, translating, and summarizing text—and both present new opportunities for data acceleration and intelligent document processing to innovate.

As they increase in popularity, businesses are using LLMs to offer new services and functionality to their clients. Many productivity tools, including Microsoft Teams and Notion, are being augmented with capabilities offered by these next generation models and frameworks. CRM companies are launching tools that let clients ask questions about their accounts and automate CRM tasks, including generating reports. Other companies are using LLMs to create chatbots for customer service, draft emails and marketing copy, summarize conversations, or help conduct research.

Despite their amazing capabilities, though, LLMs by themselves aren't useful for data acceleration platforms unless they're deployed into a broader workflow that allows users to engage with their output, whether that is by applying validation rules, assigning it to a specific output field, or simply by providing users with an easy way to review data output.

How could data acceleration platforms utilize LLMs in their workflows to derive greater insights out of data? One example could be by applying their capabilities directly to document workflows, i.e., by allowing users to ask questions about key points in a document on the fly or by summarizing documents. Many LLMs are already available via API, creating the opportunity for data acceleration platforms to rapidly incorporate them into their applications.

Conclusion

At a certain scale, there's no viable alternative to a data acceleration platform. The sheer quantity of documents handled by financial services firms — and the expertise required to extract and organize data contained therein — puts manual processing out of reach.

As more firms adopt data acceleration, their ability to effectively implement technologies like computer vision, optical character recognition, and natural language processing will become significant drivers of alpha. It won't be enough to merely have access to these tools; to be competitive, firms must learn to leverage data acceleration to its fullest extent and with a consistent bias toward innovation.

About Alkymi

Alkymi, the leading data acceleration platform, is the action layer between all of the data that flows into your business and the systems that support it. We help teams unlock data in any form, process and route it with tools built specifically for business users.

Valuable data flows into your business in all forms — email, PDFs, images, etc. These sources of information are not always easy to access and can slow down operations, leading to delays in your critical business functions. Liberating this data in real time, transforming it and sending it directly into your business systems accelerates your ability to make decisions.

Alkymi has built an intuitive solution that enables any business user to create end-to-end automated workflows without needing any technical expertise. We are delivering the first self-service, no-code automation tool that helps you accelerate decision-making and maintain agility.

alkymi.io

GLOSSARY

API

Short for Application Programming Interface. APIs are essentially “gateways” that allow data and commands to be shared across applications. With APIs, your data processing model can import and export data automatically with little human intervention required.

COMPUTER VISION

A subfield of artificial intelligence that deals with processing and making predictions on visual information.

DATA ACCELERATION PLATFORM

Data acceleration platforms employ both intelligent data processing and workflow management to help businesses ingest data in any form and transform it. Data is automatically exported and aggregated into connected systems, enabling immediate actions such as enriching, triggering, reporting and analysis.

DATA SOURCE

Databases, applications, or models which provide data — often via API — for other processes. Data sources can inform your data acceleration platform, and your model can also act as a data source for other processes.

GENERATIVE PRE-TRAINED TRANSFORMER MODEL

A Generative Pre-trained Transformer (GPT) model is a category of large language models that uses deep learning to generate natural language text. GPT models are able to synthesize and compose text-based responses.

INTELLIGENT DOCUMENT PROCESSING

The practice of using a machine learning-enabled model to automatically collect and process files using computer vision, optical character recognition, and natural language processing for the purpose of data extraction and export.

LARGE LANGUAGE MODEL

A large language model is a type of machine learning model trained on extensive text datasets to understand and generate human language, including answering questions, translating, and summarizing text. Examples include OpenAI’s GPT-3 and GPT-4 and Google’s BERT.

MACHINE LEARNING

A subfield of artificial intelligence that deals with algorithms that can continuously learn from data.

NATURAL LANGUAGE PROCESSING

A subfield of computer science that aims to give computers the ability to read, write, and understand language.

OPTICAL CHARACTER RECOGNITION

A subfield of computer vision concerned with computers' ability to detect text contained in images.

QUESTION ANSWERING MODEL

Question answering models are a category of large language models that are able to answer questions in natural language by referring to a source (i.e., a specific document).

STRUCTURED DATA

Data which is machine-readable, i.e., it is in a consistent, well-typed format for a database. From a business proposal email, the relevant dates, costs, and other figures can be extracted and transformed to become structured data.

UNSTRUCTURED DATA

Data which is not organized according to a structure or database schema, and is not immediately actionable. An email written in plain English which outlines a business proposal, for example, may contain relevant data — dates, costs, etc. — but cannot be actioned by a downstream system without further processing.