



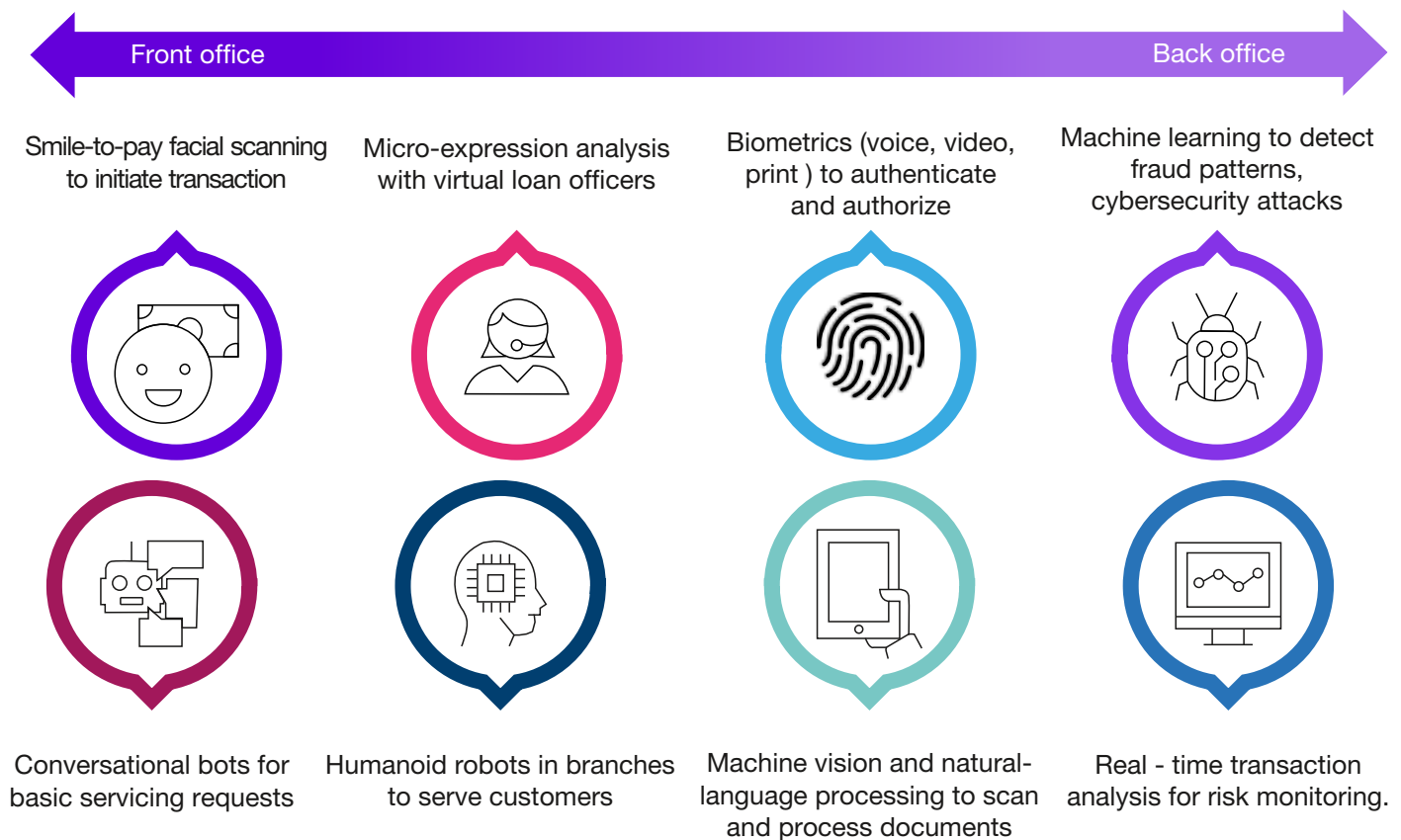
Ensuring the Insurance Sector against data Complexities

WHITEPAPER

Financial services sector is sitting on a mountain of data

It is now cliché to claim that Artificial intelligence (AI) has revolutionary potential to transform the banking, financial services, and insurance (BFSI) sector. BFSI companies are expanding their use of AI capabilities on multiple fronts. For example, banks use AI and machine learning (ML) to enhance customer experience (CX), personalize services, and automate back-office processes.

Exhibit 1: End-to-end coverage of AI/ML in banks



Source: McKinsey & Company

Within the BFSI sector, there is a technological divide between legacy or established players and recent FinTechs. FinTechs don't have the burden of legacy systems or outdated technologies and are rapid adopters of AI/ML.

FinTechs show a marked interest in using AI/ML for digital banking, customer credit scoring, and predictive analytics. Traditional players are forced to follow suit to compete and thrive. Across BFSI value chains, the use cases for AI/ML-led automation are many.

Exhibit 2: AI in BFSI—Use cases samples

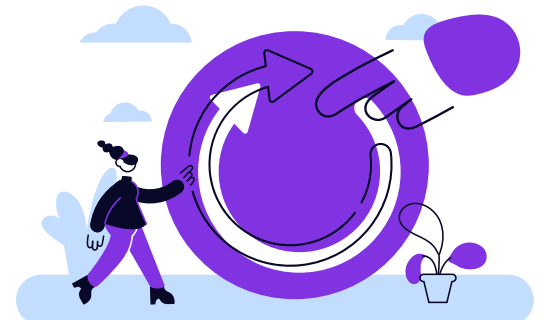
FinTech	Retail Banking	Investment Banking
Portfolio optimization	Fraud detection regarding transactions and payments	Fraud detection regarding transactions and payments
Fraud detection regarding transactions and payments	Fraud detection: AML and KYC	Fraud detection: AML and KYC
Compliance	Conversational AI	Recommender systems / next-best action
Algorithmic trading	Underwriting and acquisition	Claims processing
Fraud detection: AML and KYC	Marketing optimization	Default prediction

Source: NVIDIA

Every facet of financial services industry depends on data-driven operational models, now. Take for example the insurance space. Insurance companies require data to reach optimal levels of success. While the industry has always had a wealth of data, it's only in recent years that there has been a real push to monetize this vital asset. However, from underwriting and risk assessment to claims settlement and customer interactions, insurance enterprises today produce and deal with vast volumes of unstructured data—one of the industry's significant challenges.

Change is quick and dynamic

The financial services landscape is changing quickly. Compliance requirements are escalating. For example, in the US, banks must follow Know Your Customer (KYC) procedures that not just seek to manage risks but ensure compliance with Anti-Money Laundering Laws (AML) and the US Patriot Act's Customer Identification Program (CIP) requirement. Thus, identity verification is a critical component of KYC.



Validating every new customer in person is an expensive and time-consuming process. Digital onboarding is being adopted at a rapid rate. AI/ML technologies have made reliable facial recognition possible. Prospective customers take a selfie image that's matched against the image on a government-issued ID. Scale and speed are benefits.

AI/ML enables compelling unstructured data solutions that promote process efficiencies. Take the insurance space again as an example. Forms and claim settlements — including images, emails, and other documents — contain precious data that can empower insurers to create holistic views of their customers and optimize processes. Thus, data solutions for insurers focus on automating the capture of unstructured data residing in varied sources and diverse formats and unlocking qualitative actionable insights.

360-degree data environment

Financial services companies are part of a complex ecosystem of alliance partners, suppliers, regulators, customers, etc. The timely exchange of accurate data and information is critical. Most of the data is unstructured.



Financial services companies are part of a complex ecosystem of alliance partners, suppliers, regulators, customers, etc. The timely exchange of accurate data and information is critical. Most of the data is unstructured.

Traditional rules-based and optical character recognition (OCR) elements alone don't suffice regarding unstructured data extraction at scale with accuracy. The rising demand for processing voluminous semi-structured and unstructured documents with greater accuracy and speed has catalyzed the need for better and automated document processing. A fine blend of subject matter expertise with the latest AI/ML technologies is essential.

There are solutions to handle unstructured data complexities and implement automation at scale. Intelligent document processing (IDP) leverages AI/ML, Natural Language Processing (NLP), and/or Intelligent Character Recognition (ICR) into document processing. There are several financial services use cases possible

Exhibit 3: IDP use cases

Creating customer risk profiling by carefully running credit risk analysis of customer data from their questionnaires during KYC and Loan processing

Updating the mortgage documents by extracting data digitally with intelligent document processing from mortgage forms to initiate further processes

Expedite extraction and processing of communication documentation received via emails, texts, fax, posts, etc., and swiftly generate detailed bank reports

Capturing pertinent information by digital image processing of bank cheques in no time

Speeding up vehicle loan processing by identifying vehicle class/category from the vehicle photos given in the loan applications

Also, by bringing the paper form data into the automated digital fraud detection engine to receive a 360-degree view of potentially fraudulent activities

Performing variance analysis and form reports on planned budget vs. actual costs and many such big and small processes

Source: Straive

Some incumbent financial institutions, such as banks, spend more than 10% of their annual revenue on technology investments. Banks analyze structured data such as customer, credit, campaign, and product data. Unstructured data—which includes audio (customer interactions), video images (branch interactions), PDF documents (onboarding forms and regulatory documents), email files (communications), etc.—remain underutilized and unanalyzed.

Thus, an overwhelming quantity of financial services sector is unstructured. Compounding this are the hidden and alternative data varieties that emanate from various data sources like sensors, satellite images, traffic cameras, etc. Robotic Process Automation (RPA) does a great job with structured and even semi-structured datasets.

Exhibit 4: Understanding data types

Differences	Structured	Semi - structured	Unstructured
Storage	Structured Data is searchable and strictly defined	Semi-structured Data contains Data, however, it is not stored in a tabular or predefined format.	Unstructured Data is stored in its primitive form.
Location	Structured Data is very often stored in warehouses	Semi-structured Data is stored in Data Lake houses.	Unstructured Data is stored in Data Lake houses and Data Lakes .
Workability factor	Structured Data, hence becomes easy to search , analyze and comprehend	Semi-structured Data requires some extra work for the same processes .	Unstructured Data requires a lot of work for the same processes
Format	Structured Data is stored in predefined formats	Semi-structured Data contains tags and elements, which can be used to group Data .	Unstructured Data is stored in various formats.
Organized by	Structured Data is organized by means of Relational Database	Semi-structured Data partially organized by means of XML / RDF .	Unstructured Data is based on simple character and binary data.

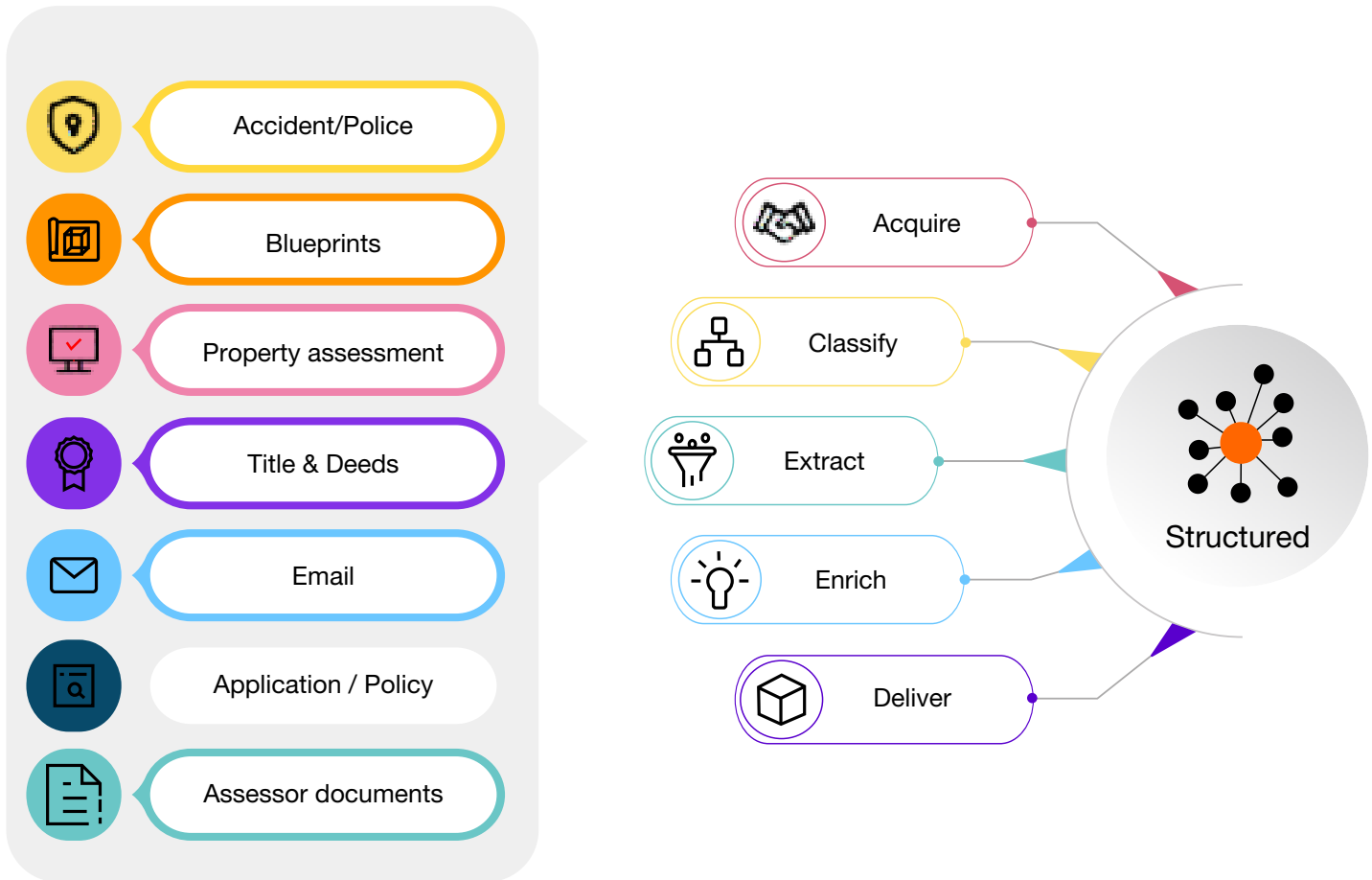
Source: Tutorialspoint, Straive

But dealing with the complexity and immensity of unstructured data requires advanced AI/ML. Going back to the IDP example – it automatically extracts data from digital information assets such as images, PDFs, and photographs.

Automation at scale for unleashing value

At a basal level, IDP can be conceived of as a sophisticated form of OCR. The critical difference is that while OCR simply focuses on extracting unstructured textual data from digital images, IDP hones the capacity to interpret and make sense of the data being extracted with AI/ML, thus advancing automated business workflows.

Exhibit 5: The journey from unstructured insurance data into structured formats



Source: Straive

Unstructured data does not have a definite format and cannot reside logically in a tabular row and column format. Consider unstructured data as "subjective" data, in the sense that it has data that you need now and which you may need later. It can be generated by both machines and humans and typically does not have a pre-defined data model.

It is not only difficult to analyze unstructured data but is also time-consuming and laborious, with manual processes limiting scalability. There is only a thin line of difference between structured and unstructured data. That is because data that seem unstructured can be processed in a structured way. That is where intelligent automation solutions that leverage AI/ML help.

Take the case of insurance claims processing. Straive's services around claims processing focus on reducing the time spent by agents on processing every claim and making data flow easy for any downstream audits. We can effortlessly extract data from forms, reports, emails, and other free-form text with our Text Intelligence solutions and Straive's proprietary data management platform, the Straive Data Platform (SDP).

The SDP can be customized either to extract sentiment or a more customized concept extraction exercise. Using SDP, we also help reduce any agent bias that may creep into claims processing. Additionally, as paperwork continues to be filed at all odd hours, SDP can also act as a classification engine, identifying the received document types, flagging important and missing documents, and identifying any potential issues at this point rather than later. Using SDP, insurance companies can ensure their agents are more focused on higher-value tasks than data extraction.

The need for a comprehensive data solutions strategy

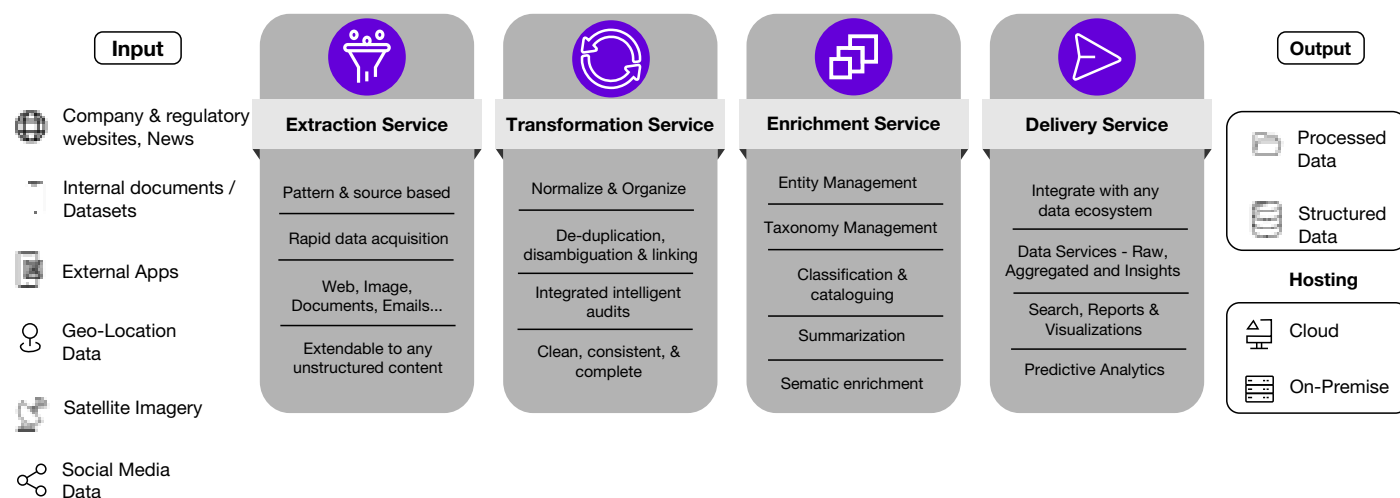
Historically, the financial services industry has relied heavily on paper processes. It's riddled with time-consuming activities, inefficiencies, and inaccuracies. AI-driven platforms effectively extract critical insights from hitherto inaccessible data, boosting efficiencies and customer satisfaction.

Text intelligence is only one part of an AI/ML intelligence ecosystem. There is also visual intelligence. We leverage computer vision technology combined with our unstructured data solutions experience to transform the way visual content is processed and consumed.



Straive's vision-led image and video processing solutions include the capability to extract information, recognize objects and entities, classify images, develop patterns, identify people, etc., on an industrial scale through the SDP.

Exhibit 6: From data to insights with SDP



Source: Straive

Unstructured data provide a layer of insights that fill the gaps in the big picture. Combining unstructured data with structured data improves business decisions. AI/ML-led platforms interface with enterprise applications to process massive amounts of unstructured data at scale, leading to smart automation.

Exhibit 7: SDP provides the data solutions edge



Source: Straive

SDP automates the data acquisition, enrichment, and delivery operations in a way that scales with minimal manual intervention. SDP's auto-extraction feature uses both a rules-based and an ML engine to deal with the data variability, sources, and volume while maintaining quality. The platform-led AI/ML approach underpins Straive's specialized data solutions.

Thus, for example, Straive enables banks to manage unstructured data from loan documents, application forms, pay slips, property documents, regulatory documents, etc., and public sources such as social media, government registries, and watchlists.

Another significant use case is anti-money laundering (AML). It's a challenging regulation for banks – even more so for large, diverse financial institutions with a global footprint. What makes complying with AML regulations a challenge is that it involves complex data, documentation in multiple formats, and significant human intervention in the form of AML analysts.

An accelerated future with new data forms

Regulations promulgated in the aftermath of the global financial crisis, technological innovations, diverse identity regimes, and the need to correlate data from multiple sources, have created a complex environment of regulatory change for the banking sector. To comply with and manage AML requirements effectively, financial need automation support to enhance their existing processes in AML and KYC. Use cases include adverse media monitoring, watchlist consolidation, and enhanced digital KYC.

Data-led analytics and automation is critical throughout the financial services value chain. Let's take the case of investment management. Investment management firms are desperate to boost their returns or alphas. In the search for alpha, data analytics is the weapon of choice for capturing competitive advantage. Investment managers can extract useful insights from various internal and external data sources, including what is dubbed as alternative data sets.



Alternative data is sourced externally regarding a particular company, and typically refers to information that's used to gain additional business insights. For example, consumer sentiment and credit card data mixed with traditional data can provide better predictions on the business performance of certain companies. Typical data types in alternative data are customer credit and debit card transactions, emails, geo-location (or foot traffic) metrics, mobile application usage, satellite or weather data, social and sentiment data, web scraped data, generic web traffic, etc.

However, alternative data analytics is only in its early adoption stage now. Some specialized data vendors sell alternative data to various enterprise stakeholders, though many companies currently only explore alternative data for potential value. The more prevalent use now is using alternative data as supplementary information to test market hypotheses derived from more conventional sources.

Exhibit 8: Major sources of alternative data



Source: Straive

Alternative data offers a quasi-real-time view of market sentiments through various routes (like social media chatter). Alternative data can also be acquired from individual or aggregate data by advanced algorithms or machine learning on traditional data sources, so that results can be used as inputs for further analyses. These sources can come from news articles, government publications, company reports, or licensing or purchasing data from third-party specialist aggregators.

Any data apart from static or conventional market data is typically classed as alternative data. By 2025, over 75% of venture capital (VC) and early-stage investor executive reviews will be informed using AI and data analytics, according to Gartner, Inc. In this context, alternative data's real-time updates is bolstered by the objectivity and accuracy of machine learning algorithms.

Technologies like Internet of Things (IoT) like smart TVs, vehicle sensors, etc., constantly provide data on user behavior. AI/ML-led analytics enable firms to analyze such unstructured data, producing a competitive edge over peers, through unconventional and nuanced data sources. The latest in AI/ML platform models allow firms to elevate subtle signals and latent trends that support insights to streamline business processes. This space will prove critical for financial services firms, especially retail banks.

The responsibility for sustainability

As the importance of the environmental, social, and governance (ESG) agenda gains momentum, the financial services industry faces unique challenges and exciting opportunities, along with dedicated responsibilities. As financial institutions embrace new investment decision-making and interacting methods with corporations, buy-side investors adjust their decision models accordingly. The result is an increase in ESG investments, research, and benchmarking to deal with the new corporate responsibility culture.



However, ESG benchmarking has been primarily qualitative with little uniformity in tracking and rating of corporations. Additionally, the data sits in dense reports and news, making it hard to track and easily consume information from an analyst's perspective. With all the growth in ESG investing and its clear financial implications, firms are pressured to track ESG at scale.

As a result, the market needs tools and partners that can bring down the average analyst reviews and structure data for a transparent metrics-driven ESG benchmarks. Advancements in AI/ML modules, NLP, and Robotic Process Automation (RPA) accelerate creation of data and metrics from sustainability reports for analyst consumptions.

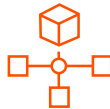
Exhibit 9: ESG Data Challenges

Challenges



Most of the data is unstructured

Data sits in sources highly textual in nature that makes it difficult to source and extract



Data is qualitative and bias in nature

Much of ESG data is qualitative in nature and has historically been analyst driven which can lead to unstandardized data due to analyst bias



The data needs to match the organizational ESG data maturity

Organizations with complex ratings framework have very different data needs than firms just starting their ESG journey. Data providers have a standard set of data that cannot accustom to the needs of all

Source: Straive

In summary, the financial services industry irrevocably and monumentally changed in the last ten years due to digital disruption. The past five years brought a higher level of innovative digital technologies than ever before. Digital transformation combined with the industry's highly regulated nature has led to an exponential increase in the content that firms deal with, such as regulatory publications, KYC docs (e.g., passports and payslips), claim forms, application forms, leases, and other documents. These are in addition to the mountains of internal documentation stuck in silos, which hinders legacy financial institutions from using it to drive business decisions.

Personalizing customer experiences

Internal data needs to be combined with public domain data to enable banks, asset management companies, and hedge funds to drive personalized customer experiences and identify ways for an alpha generation while ensuring efficient operations to reduce costs. The rise of FinTechs, which operate outside the barriers of legacy systems in a digital-only environment, has accelerated the pressure on financial institutions. While financial institutions have come a long way in implementing machine learning and AI, these are limited mainly to structured datasets.



Exhibit 10: Challenges in dealing with unstructured content

Challenge	Description
Formats diversity	Handling many unstructured documents such as identity and address proofs, pay slips, property documents, valuations, contracts, etc., in various formats with varying levels of quality.
Scalability	Manually processing the documentation limits scalability; about 30–40% of the time is spent on noncore automatable tasks.
Public data acquisition	Acquiring unstructured public data at scale within the compliance framework.
References triangulation	Information requiring manual cross-referencing with original documentation across databases, third-party providers, government agencies, and customer-supplied data.

Source: Straive

Unstructured data offer financial institutions hitherto undiscovered actionable insights—of customer preferences, unmet customer needs, and market and process gaps.

Conclusions

Acquiring, enriching, and managing unstructured data is challenging. With the emergence of solutions such as Straive's text, public, and visual data intelligence solutions and platforms like SDP, deriving insights from unstructured data becomes simple. We expect strong momentum using unstructured data and platforms such as SDP for more diverse financial sector cases.

About Straive (formerly SPi Global)

Straive is a market leading content technology enterprise that provides data services, subject matter expertise (SME) and technology solutions to multiple domains such as research content, e-Learning / EdTech and data/information providers. With a client-base scoping 30 countries worldwide, Straive's multi-geographical resource pool is strategically located in eight countries - Philippines, India, USA, China, Nicaragua, Vietnam, United Kingdom and the company headquarters in Singapore.



www.straive.com



straiveteam@straive.com



<https://www.nvidia.com/content/dam/en-zz/Solutions/industries/finance/ai-financial-services-report-2022/fsi-survey-report-2022-web-1.pdf>
<https://www.idrnd.ai/digital-onboarding/>