



# The Impact of Intelligent Document Processing

## on the Data Value Chain

October 2020

**Author:** Emily Liao, AI Consultant, Nexus FrontierTech

**Special contributors:** Terence Tse, PhD, Executive Director, Nexus FrontierTech

*Data is an enormously valuable resource for organisations seeking to stay competitive. Whereas data is becoming abundantly available, many companies are not yet maximising the value of this crucial asset. Data can gain value at many stages ranging from initial data preparation to generating useful insights. This can be represented as a series of steps, with each increasing the potential of the next, called the Data Value Chain (DVC). In response to the high volumes of documents existing in all fields of business, this paper will present Intelligent Document Processing (IDP) as a foundational way to add value to data and to enable it to reach later stages of the DVC. It will discuss some benefits of this solution, as well as some Nexus FrontierTech use cases demonstrating how IDP can augment data to boost operational efficiency.*



## **PART I**

# **The value of data**

### **Introduction**

Across all industries, data is a key strategic asset and source of value. Despite the fact that data is becoming available in exponentially larger volumes, many companies are still 'information poor', struggling to utilise data to enhance their business strategies and operations (Ladley and Redman, 2020). While there is growing recognition of the need to become more data-orientated, the NewVantage Partners' 2019 Big Data and AI Executive Survey reported that 69% of respondents have not created a data-driven organisation and 53% are not yet treating data as a business asset (Bean and Davenport, 2019).

However, especially during times of crisis - as presented by the Covid-19 pandemic - how a company uses data will differentiate between those that fail and those primed to succeed. Hence, now more than ever, it is necessary to adapt to the new business landscape by adopting innovative, data-centric solutions. Maximising the value of data is no longer merely about getting ahead; it is now a necessity to keep up with the curve and stay competitive in an unpredictable environment.

When integrated and efficiently used, data can lead to improved processes, products, and competitiveness while empowering employees and managing risk. Businesses can use data to improve operations and inform their decisions significantly through a greater understanding of customers and trends. Additionally, through digital transformation, businesses can use data to create a whole new business model as well as generate value in ways that may have been previously unforeseen. To fully understand the benefits that data can provide across all functions and departments, the Data Value Chain (DVC) should be explored. This discussion will demonstrate the importance of prioritising the early stages of the DVC, necessitating innovative solutions such as Intelligent Document Processing (IDP) to make the data available and accessible in the first place.

## The Data Value Chain

Every single organisation has data, even if they have not yet considered it as an asset. Data can be collected not only from external sources, such as client data, but is also produced in day-to-day business operations. However, data is useful only if leveraged to improve decision-making and produce actionable insights. For this to be possible, data must undergo a series of activities, each of which generates additional value wherein 'bringing together individual pieces increases the value by delivering new insights and correlations' (GSMA, 2018). This can be represented in the form of a DVC, an example of which is shown in Figure 1 and elaborated upon below.

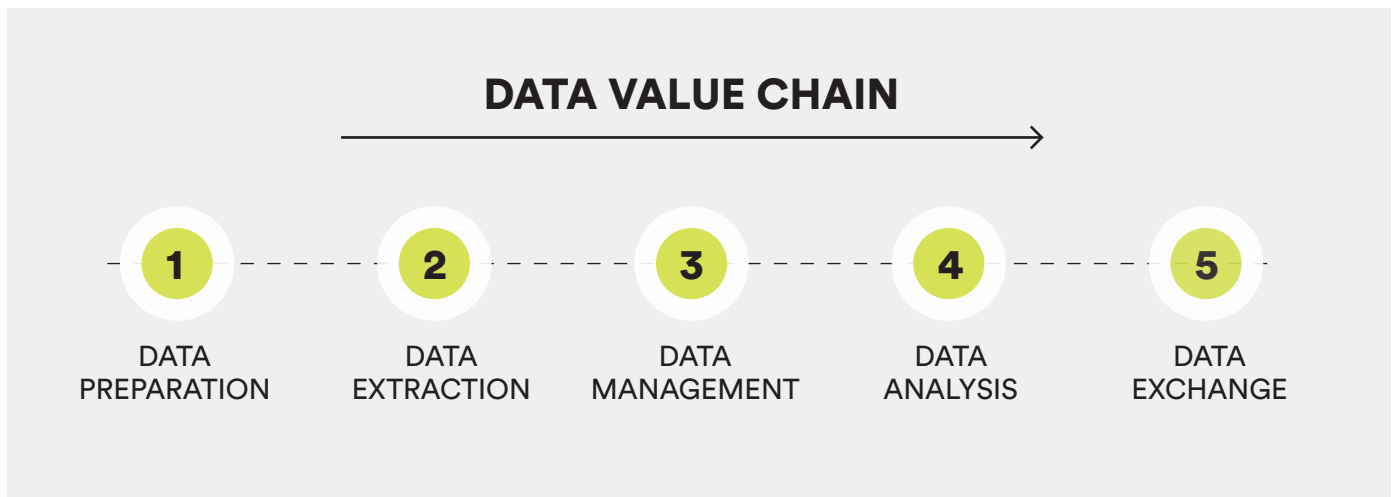


Figure 1: Example Data Value Chain

- 1. Data preparation:** Since data, which is often acquired from multiple sources, can come in varying formats and qualities, it may need to be cleaned, filtered, and standardised, to prepare it for further processing.

*Example: A bank looking to audit the quality of mortgage sales may need to acquire data from various sources, such as online portals, emails, and other avenues of communication. Organisations without client data can also collect data from operational and back-end processes. The data might then need to be preprocessed to allow for effective extraction.*

- 2. Data extraction:** This step involves the extraction and validation of the relevant data from their respective sources. Much data still exists in hard copy, which will need to be digitised and structured to move effectively along the DVC. Data can also be enriched by external sources and domain knowledge.

*Example: The bank may only be interested in checking certain data points in lengthy, paper-based documents. Digitising these documents and extracting salient information will boost the efficiency of the checking process.*

- 3. Data management:** After data is stored, which could require expensive infrastructure, it can be made far more useful and accessible when organised with metadata.

*Example: The bank may want to store their data in a database and tag documents with relevant metadata, describing the contents of documents, for ease of retrieval.*

- 4. Data analysis:** Once data is available in a structured and accessible format, analytical methods can be used to produce actionable insights and conclusions, often achieved by combining previously unconnected data sets. Usage of analytics is now widespread, with 59% of enterprises using analytics in some way (Panoho, 2019).

*Example: The bank might use data analytics tools to gain insights into customer risk appetites.*

- 5. Data exchange:** In this step, data outputs are exposed for use, for example by using data visualisation tools to reveal patterns and assist analysis of complex data.

*Example: The bank may benefit from huge gains in operational efficiency by having greater oversight of their data in the form of a customisable User Interface (UI) which can reveal previously unforeseen patterns and deliver intelligent prompts.*

Later stages of the chain are only possible when the earlier stages are performed correctly. This means that the importance of the foundational stages of the chain, data preparation and extraction, cannot be underestimated as they are pivotal for effective management of the DVC. Additionally, unlike most value chains, the resource (data) in the DVC is not consumed by the end-user; instead, it can be reused for multiple purposes, leading to a self-sustaining cycle (GSMA, 2018). As such it is a unique asset and, when leveraged, will provide organisations with a key competitive advantage.





## Three considerations to kickstart the DVC

In actual business application, kickstarting the DVC can be challenging. The three main considerations to maximise its benefits are highlighted below.

- ◆ **Unstructured data**  
Whereas structured data comes in a clearly-defined format with searchable fields (for example postscript PDFs), unstructured data (for example scans, images, and audio) lack this same structure and are more difficult to search and organise. Unprocessed, this type of data is traditionally unusable by machines, which makes it difficult to push through the DVC.
- ◆ **Multiple data sources and formats**  
As the availability of data increases, so does its diversity. Valuable information comes from increasingly varied channels, from online portals to hard copy statements. It can thus be a heavy undertaking to maintain a comprehensive overview of the data available and to ensure valuable data remains accessible and centrally collated. Additionally, differing data sources and formats are prone to cause miscommunication and duplicative efforts within an organisation.
- ◆ **Data quality**  
In a 2017 study by Nagle, Redman, and Sammon (2017), it was found that an average of 47% of newly-created data records have one or more work-impacting errors. Data quality issues can disrupt many areas of business operations and lead to worse decision-making in general. This is influenced in part due to the large volumes of unstructured data formats.

These issues can lead to data management which is error-prone, time-consuming, and operationally inefficient overall. Having an effective and automated data processing solution in place will be key in tackling these obstacles. An example of such a solution is Intelligent Document Processing, which will be discussed below.



## **PART II**

# **Intelligent Document Processing**

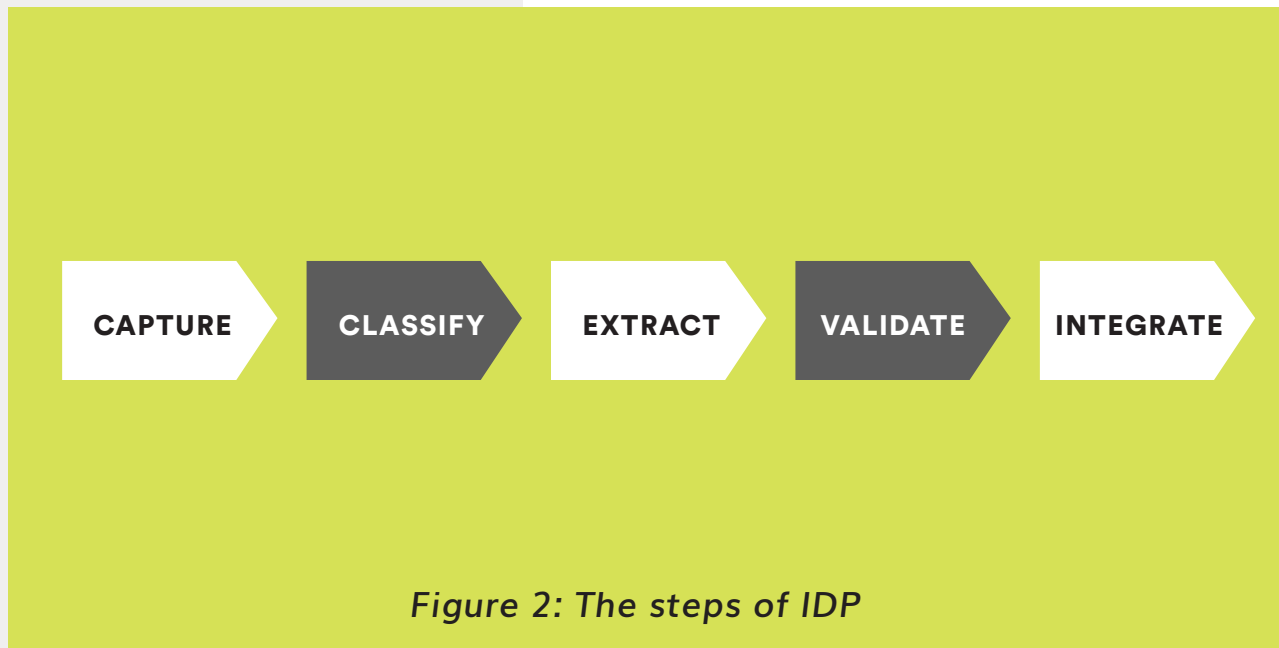
### **What is Intelligent Document Processing?**

Intelligent Document Processing (IDP) is a form of process automation which uses AI models to convert data across documents into a more useful format and to streamline document processing. This is essential for organisations to stay competitive since data is, now more than ever, a key strategic asset to inform and empower decision-making. However, a streamlined data process is difficult to achieve when documents still exist in the majority of business operations, and mostly in unstructured formats. Whilst IDP does not remove the fundamental need for companies to identify their data sources, it is a mechanism by which a company can make their data usable and, in fact, reusable. It is inefficient, time-consuming, and costly for a human workforce to manage this manually. Furthermore, manually processing documents can form an obstacle for further use of data along the DVC, especially when using AI approaches which hinge on the quantity and quality of data.

For a majority of companies, data quantity and quality problems are delaying or altogether halting companies' AI initiatives (Schmelzer, 2020). Companies are spending a significant amount of resources in getting data engineered to be AI-ready, spending up to 80% of project costs on obtaining, cleaning, and organising this data (El-Hanfy and Webster, 2020). Hence, having an automated solution to process data will go a long way in preparing an organisation for further automation and value-generating activities, leading to improved efficiency, cost savings, and data-driven decision-making overall.

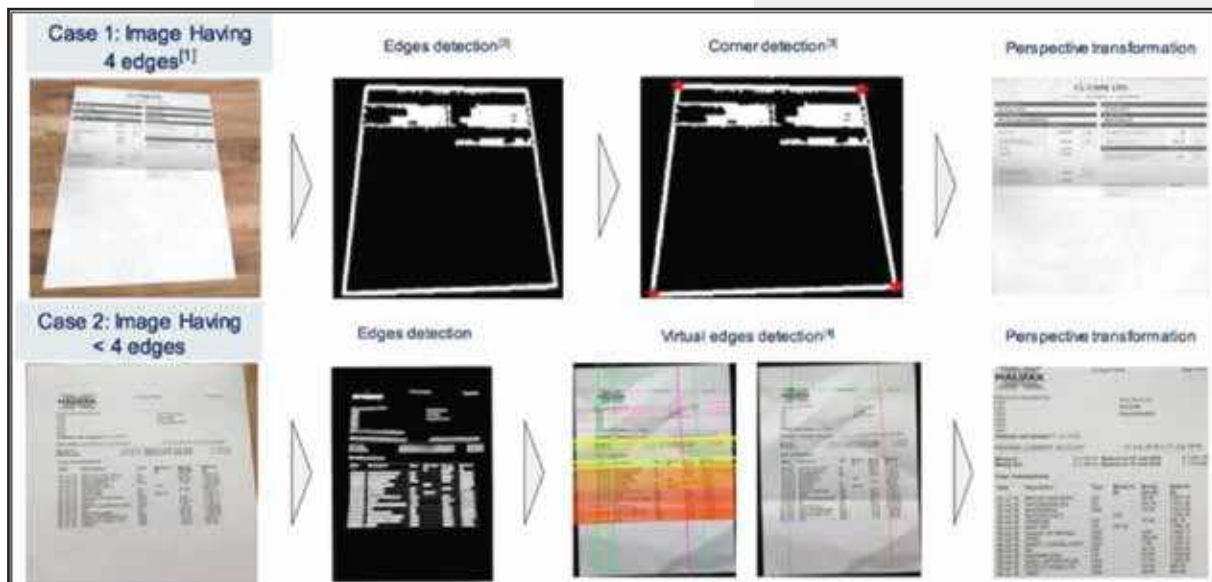
## The steps of Intelligent Document Processing

Figure 2 presents the basic steps of an Intelligent Document Processing solution. The individual steps will be explained below.



- 1. Capture:** Firstly, it is necessary to capture document inputs into the system. As aforementioned, this can be challenging as organisations typically have many channels of data. An IDP platform may also be working with multiple document types and formats, from spreadsheets to handwritten documents. Importantly, due to data quality variations pre-processing techniques may be necessary, such as the de-warping technique as shown in Figure 3.
- 2. Classify:** Documents can be classified to determine their type and where each begins and finishes. This classification process is typically done using a Machine Learning (ML) model. The document can also be tagged or indexed by reference to a pre-defined template.
- 3. Extract:** With unstructured data, it is necessary to convert the input data into a machine-usable format. Common technologies for performing this conversion include Optical Character Recognition (OCR- to convert images into text), Intelligent Character Recognition (ICR- for handwriting), Optical Mark Recognition (OMR), and Barcode Recognition. Once converted the IDP models will be used to recognise and extract individual relevant and valuable data points, which could use both ML and rule-based AI. Additionally, technology such as Natural Language Processing (NLP) can be applied to make sense of different components of the extracted data using context indicators.

4. **Validate:** Depending on the business criticality and accepted margin of error, the extracted data should be validated or confirmed by a human or against rules to confirm that it is correct. For example, this is useful for extracted hand-written text to see if the interpretations of human and machine are the same, or whether outputs need to be changed. Future improvements to and retraining of the models will be informed through these validation exercises.
  
5. **Integrate:** The extracted data can now be exported into the required format and fed into the relevant workflows or business processes automatically. The result is that this data will be available for immediate consumption, providing insights, for example, through a dashboard, and can be used at multiple points along the data value chain.



*Figure 3: The de-warping technique*

## Benefits of Intelligent Document Processing

Companies gearing for success are those automating routine activities and learning from the large data sets that enterprises increasingly depend on' (HBR, 2019). IDP is an exemplar of this and has the potential to provide many benefits to the user, including:



- ◆ **Cost savings**  
Establishing a platform to process data saves on labour and operational costs. This initial investment also enables data to be used in the future at much lower costs. One Nexus FrontierTech client estimates a 10x Return on Investment (ROI) once their IDP solution is fully implemented.
- ◆ **Risk reduction**  
Manual data extraction is normally a repetitive, mundane, and time-consuming process, making it error-prone. On the other hand, AI is terrific at preventing errors, especially in labour-intensive business activities (Tse, Esposito, and Goh, 2019). This will improve the reliability of the output.
- ◆ **Empowered staff**  
Reducing the amount of manual intervention in a workflow will boost productivity and efficiency, freeing-up human capacity for more nuanced tasks. This will allow resource effort to pivot from low-value tasks to decision-making.
- ◆ **Better customer insight**  
Having a comprehensive overview of the customer journey and relevant information will make it easier for organisations to make timely decisions and hence better engage with them.
- ◆ **Speed and efficiency**  
IDP will allow faster retrieval of relevant information. For example, one Nexus FrontierTech client managed to reduce their compliance process duration by 80%<sup>1</sup>. This will also increase the speed of business operations, as companies will have access to more organised information, allowing them to focus their efforts on more important activities.
- ◆ **Transparency**  
Data will be more visible immediately, providing a more holistic view and allowing for data visualisation tools to be used. This will provide important insight to businesses looking to make better decisions.
- ◆ **Improved ability to comply**  
Many document-heavy institutions are also heavily regulated. By reducing errors from manual data processing and providing greater coverage of compliance gaps since they will be empowered to check 100% of their cases, organisations can decrease the risk of contravening internal and external regulations.
- ◆ **Priming for more sophisticated capabilities of the DVC**  
Importantly, having a robust IDP system in place will set the foundation for further automation and unlock further steps of the Data Value Chain. Insights and data will feed into subsequent phases, which would not be possible without this initial phase.

<sup>1</sup> See: <https://www.gov.uk/government/case-studies/how-a-uk-based-bank-used-ai-to-increase-operational-efficiency>



## PART III IDP IN APPLICATION

The potential for applying IDP solutions is widespread. Below, a Nexus FrontierTech case study and various use cases will be explored wherein IDP solutions can be applied to increase operational efficiency and maximise the value of data.



### CASE STUDY

#### Mortgage Sales Quality Assurance

Nexus FrontierTech worked with a large global bank to automate their Mortgage Sales Quality Assurance process. The client's Sales Quality (SQ) team was responsible for reviewing completed sales of mortgage products. The sheer volume of applications, involving around 40,000 documents per month, led to a lengthy, complicated, and mundane checking process.

Each application took around four hours to review and required the manual extraction of 180 data points across 10 different document types. The team could only conduct a limited number of checks amounting to 10-15% of completed mortgage sales and these checks occurred on average 2-3 weeks after the point of sale, leading to compliance gaps.

Nexus worked with the bank to automate this internal compliance process by developing AI models to extract data from unstructured documents such as bank statements and payslips. Implementing Nexus' solution, the bank is now able to close compliance gaps by checking all completed sales close to real-time. A lengthy and complicated process was transformed into a speedy and simple one, wherein compliance checkers are directed to

the salient points within documents. In turn, this frees up human capacity to focus on anomalies and more complex compliance activity. As the bank moves to large-scale deployment of these AI solutions, it will benefit from a significant increase in ROI whilst increasing the accuracy of checks and reducing operational errors.



An IDP solution can be employed here to extract the necessary information from prospectuses, increasing data usability by converting unstructured data to a machine-usable format. Automation of these repetitive tasks can also lead to greater employee productivity and empowerment, allowing them to focus more on analysis and investment recommendations. Additionally, this solution can be integrated into multiple workflows to generate deeper analysis and form broader coverage; it would also be possible to set validation rules to detect missing information and compare key clauses across multiple bond deals. Finally, data visualisation tools can be used on the available data to generate deeper insights and allow for further analysis.

## USE CASE

### **Bond Prospectus Analysis**

For asset managers, the manual analysis of bond prospectuses for investment consideration can be very time-consuming and repetitive. These prospectuses, which can be hundreds of pages long and filled with complex terms, must be assessed for factors such as the structure, terms, and conditions of an offering. Especially when multiple deals hit the market simultaneously, investors may not have the luxury to go through these deal documents carefully and risk missing hidden terms and covenants - particularly for new issuers.



## USE CASE

### **Contract Management**

Manual analysis of complicated contracts, which can comprise huge volumes of detailed text, is expensive and time-consuming. As well as being highly vulnerable to human error and bias, it can lead to bottlenecks to increasing revenues and can cause legal problems due to non-compliance issues.

Furthermore, without a robust contract management system in place, it is difficult to gain oversight over the extent of the contractual obligations on both sides or to effect mass changes in clauses necessitated by changing regulations, such as the transition from the London Inter-bank Offered Rate (LIBOR).

In many organisations, contracts are still found in an unstructured, scanned format; therefore, any IDP solution must be able to convert these into a machine-usable format with sufficient accuracy. Following this, salient data points can be extracted, including contract expiry, pricing conditions, and SLAs (non-compliance of which can be more easily detected to be leveraged in negotiations or penalties). The contract analysis workflow can easily be tracked within the system with a user-friendly dashboard, complete with filtering and tagging mechanisms. Having a comprehensive document management workflow in place will be pivotal to saving time and costs, allowing the team to focus on anomalies and other nuanced activities.



## USE CASE

### Life Insurance Claims Processing

To claim life insurance, authorised relatives must typically supply the death certificate and personal identification via email or post. It is common for insurance companies to manually transfer the necessary information into their system. Hard-copy forms are then posted to relatives to be signed and returned for the claimed amount to be released.

This is a duplicative, time-consuming, and labour-intensive process for both the insurer and the claimant, which is also error-prone, resulting in the delay of releasing funds. This can add an additional burden on claimants during an already emotionally difficult time leading to poor customer relations.

An IDP solution can streamline this process by extracting the required information and directing this into the database automatically. Increasing accuracy by reducing or altogether removing manual data entry would reduce the need for unnecessary back-and-forths; the processing team would simply have to double-check details before releasing the claim. This would lead to a significant reduction of costs and time spent on manual processing, priming the organisation for a move towards a much more efficient and digital workflow, along with an improved customer experience.

## Summary

In today's business environment, merely having access to large quantities of data is insufficient to differentiate and defend an organisation. Instead, the key is to use data effectively to generate business value. Therefore, especially in document-centric industries, having an IDP system in place is essential to prime a business for maximising the value of its data. Not only is manual extraction of data, especially from unstructured sources, a cumbersome task, but it also forms a bottleneck for proceeding further along the Data Value Chain. An automated approach to processing data will be vital for organisations to stay competitive and survive.

## REFERENCES

Bean, R. and Davenport, T.H. (2019) 'Companies Are Failing in Their Efforts to Become Data-Driven', *Harvard Business Review*, 5 February. Available at: <https://hbr.org/2019/02/companies-are-failing-in-their-efforts-to-become-data-driven> (Accessed 12 July 2020).

El-Hanfy, S. and Webster, Z. (2020) 'New business horizons in the AI landscape', *Innovate UK Blog*, 3 June. Available at: <https://innovateuk.blog.gov.uk/2020/06/03/new-business-horizons-in-the-ai-landscape/> (Accessed 12 July 2020).

GSMA (2018) 'The Data Value Chain', GSMA. Available at: [https://www.gsma.com/publicpolicy/wp-content/uploads/2018/07/GSMA\\_Data\\_Value\\_Chain\\_June\\_2018.pdf](https://www.gsma.com/publicpolicy/wp-content/uploads/2018/07/GSMA_Data_Value_Chain_June_2018.pdf) (Accessed 12 July 2020).

Harvard Business Review (2019) 'The Rise of Intelligent Automation: Turning Complexity into Profit', *Harvard Business Review Analytic Services*. Available at: <https://www.oracle.com/a/ocom/docs/artificial-intelligence/hbr-pulse-survey.pdf> (Accessed 12 July 2020).

## REFERENCES

Nagle, T., Redman, C., and Sammon, D. (2017) 'Only 3% of Companies' Data Meets Basic Quality Standards', *Harvard Business Review*, 11 September. Available at: <https://www.forbes.com/sites/forbesagencycouncil/2019/10/01/the-age-of-analytics-and-the-importance-of-data-quality/#35bdd6945c3c> (Accessed 12 July 2020).

Ladley, J. and Redman, T.C. (2020) 'Use Data to Accelerate Your Business Strategy', *Harvard Business Review*, 3 March. Available at: <https://hbr.org/2020/03/use-data-to-accelerate-your-business-strategy> (Accessed 12 July 2020).

Panoho, K. (2019) 'The Age Of Analytics And The Importance Of Data Quality', *Forbes*, 1 August. Available at: <https://www.forbes.com/sites/forbesagencycouncil/2019/10/01/the-age-of-analytics-and-the-importance-of-data-quality/#35bdd6945c3c> (Accessed 12 July 2020).

Schmelzer, R. (2020) 'AI Adoption Survey Shows Surprising Results', 23 January. Available at: <https://www.forbes.com/sites/cognitive-world/2020/01/23/ai-adoption-survey-shows-surprising-results/#2fcc94c855e9> (Accessed 12 July 2020).

Tse, T.C.M., Esposito, M., and Goh, D. (2019) *The AI Republic: Building the Nexus Between Humans and Intelligent Automation*: Lioncrest Publishing.

UK Government (2019) *How a UK-based bank used AI to increase operational efficiency*. Available at: <https://www.gov.uk/government/case-studies/how-a-uk-based-bank-used-ai-to-increase-operational-efficiency> (Accessed 12 July 2020).

## About us

Nexus FrontierTech is a London-based tech company specialising in the development and integration of AI solutions that help organisations save time, money and resources by tackling process inefficiencies and data waste. Whilst industry agnostic, many clients are in the financial services space and are seeking to streamline their operations.

Sign up for a free trial today to find out how Nexus can help your organisation at <https://nexusfrontier.tech>

Visit us at:



<https://nexusfrontier.tech/>

<https://www.linkedin.com/company/nexusfrontiertech>