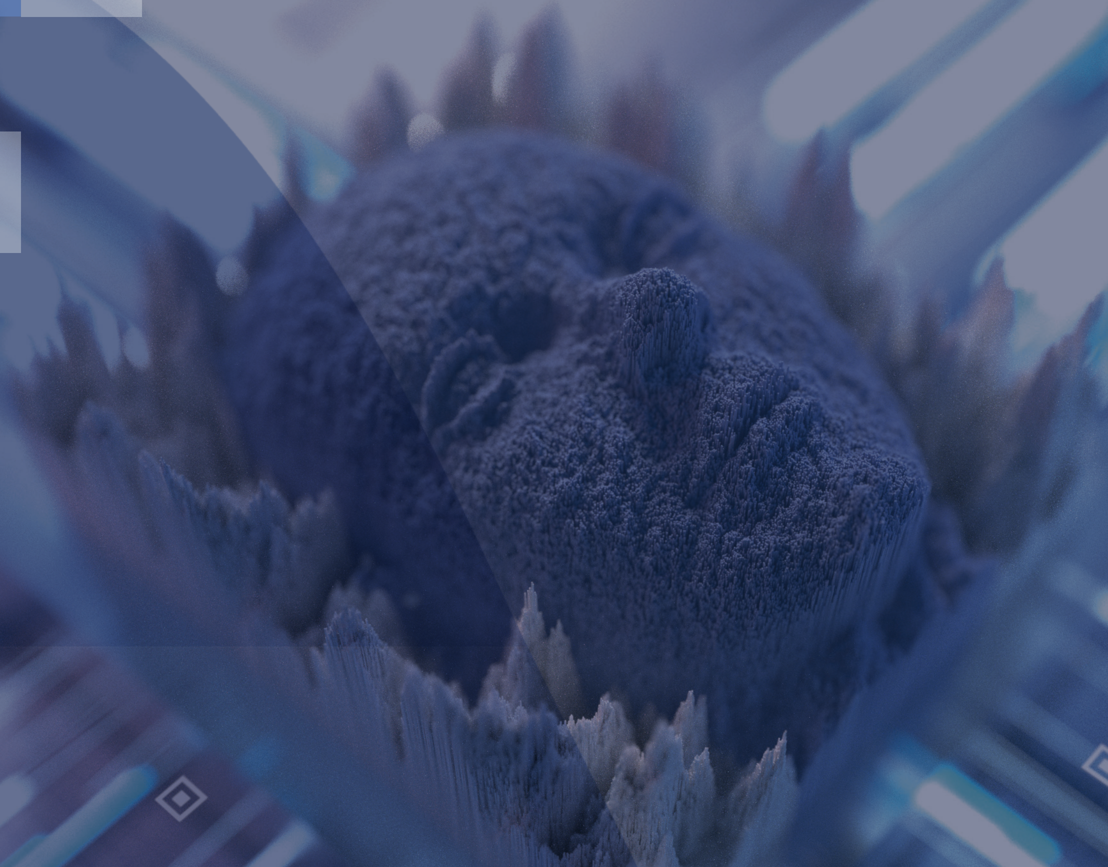
NTT DATA



GUIDE TO AI LABS

FUTURE
AT HEART

01 INTRODUCTION

The AI Cloud market is characterized by a set of big actors mobilizing massive investment to continuously extend their service catalogue for AI development. For this reason, the main cloud providers (AWS, Google Cloud, Microsoft Azure…) are currently being referred to as hyperscalers.

By leveraging the ever-increasing AI/ML cloud capabilities and its market-ready solutions that can be integrated into organization's workflows seamlessly and at a smaller cost, financial companies could reduce the cost associated to in-house research & development. This would allow banks to concentrate its talent and funding to searching for the most adequate ways to solve the business challenges they face and thrive in this new technology era.

In order to take advantage of the AI Cloud market, financial companies need an in-depth understanding of the available services, capabilities, tools, and techniques across that are offered by the different Cloud providers. It is also crucial that banks identify specific ways to integrate all relevant elements into their experimentation cycle.

In this paper we set the grounds on how companies in the financial sector can approach this change of paradigm and provide guidelines and considerations on how they can start leveraging hyperscaler's AI services.

Towards that end, we introduce the concept of AI Labs, which should be conceived as an on-going journey, where different capabilities are continuously explored for specific business purposes. This exploration would shape the overall understanding of the AI Cloud market, allowing any company in the financial sector to leverage state-of-the-art tools and techniques that Cloud providers are continuously incorporating into their portfolios.

02 APPROACH

We envision the cloud-enabled AI Lab as the environment that will bring the continuous experimentation approach into the business room, aiming at dealing successfully with the haphazard nature of experimentation, implementing well-designed, agile iterative processes that can drive experimentation and measure its progress.

In that sense, AI Labs would start from generic business needs and end prototypes, that once validated, can easily be productionized, and scaled.

The AI Lab acts as an enabler for rapid experimenting in two ways:

- Rapid experimenting of different **cloud-available techniques and tools.**
- Rapid experimenting **across different business areas,** for specific use-cases.

At the same time, the AI Lab is conceived to orchestrate the **collaboration between the variety of professionals that are involved in the successful development of AI initiatives,** from subject matter experts and business sponsors to AI professionals, including data scientists, data & ML engineers.

Building a multi-cloud AI Lab is a challenge itself, for business research and technical expertise are both needed in order to design an effective strategy that:

- Identifies which are the components that will be selected from each of the Cloud providers, analysing the technological advantage they would bring, its functional value and the comparative cost.
- Guarantees the most effective integration of the relevant Cloud components, coming from different Cloud providers, all across the AI lifecycle and across the variety of use-cases the AI Lab will execute.

In this paper, we set the grounds on how companies in the financial sector should start implementing the idea of multi-cloud AI Labs by covering those two topics, and focusing on how the industrialization of an operational model should be done to guarantee the scalability of an AI Lab (and how cloud providers enable this scalability), as well as how we can solve useful generic use cases for the financial sector using MLOps and cloud providers.

In order to perform the mentioned analysis, we have established a methodological approach that includes the definition of an operational model, identification of key use cases for the financial sector, the evaluation of the behavior of internal developments versus hyperscalers analytical services and finally some conclusions.

In the Operating Model section, we identify the different phases, roles, and procedures that the use cases should have will be defined so that they can be governed, which will allow to scale the development of said projects.

In the **Use-Cases** section, we define several use cases that will allow for the evaluation of the platforms and placing focus on the most relevant topics. The defined use cases will not need banks datasets, for its functional utility will be a secondary matter. These use cases will be implemented in both platforms, gaining the necessary experience and knowledge to create the benchmark.

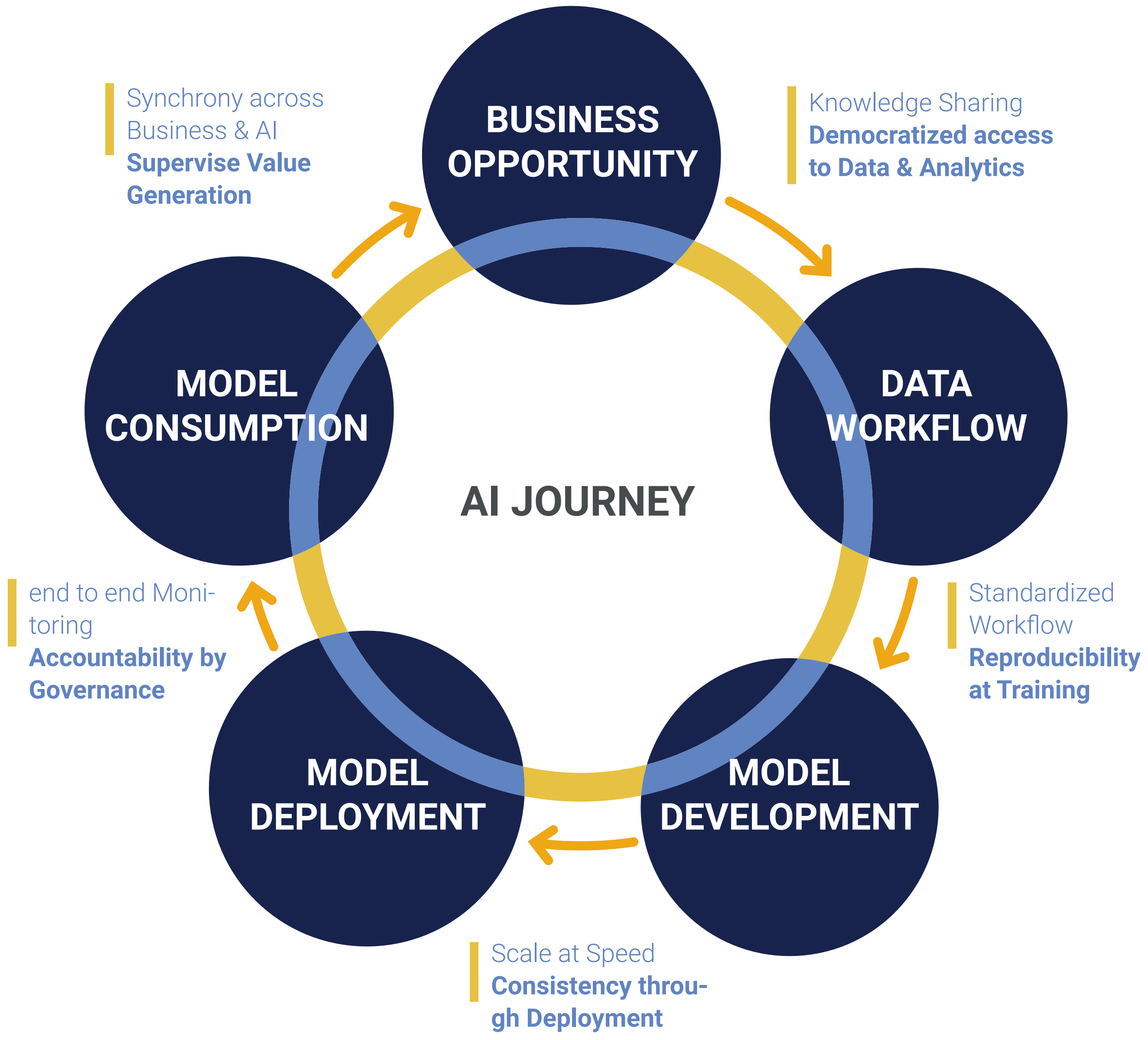
In the **Evaluation** section, the main capacities for a Machine Learning project in both platforms will be identified and placed in an evaluation map. Important questions to be asked and evaluation criteria for those capacities will be defined. Finally, the capacities will be evaluated based on the knowledge obtained during the implementation and the evaluation criteria.

Finally, in the **Conclusions** section, the different pros and cons from in-house AI development and leveraging hyperscalers AI capabilities will be discussed. Additionally, some relevant insights from this study regarding the present and future of these technologies are presented.

03 OPERATING MODEL

As a pillar for enabling the operating model, we define the phases of the AI production cycle that allow us to orchestrate, accelerate and scale the value generated by AI, unifying the development and industrialization of AI.

Throughout these phases, we identify key objectives to address the potential risks and difficulties involved in putting them into production and ensure the scalability of AI solutions.



3.1. Business Opportunity

- Collect business information to define a new Project and its scope.
- Identification and selection of data sources and dependencies.

3.2. Data Workflow

- Collect, acquire, and transform data to unlock its value.
- Create pipelines that take in raw data and output ready-for-analytics datasets.

3.3. Model Development

- Train a model whose outputs create business value for the organization.
- Create repeatable pipelines that allow for reproducible models.

3.4. Model Deployment

- Facilitate continuous deployment to production in a fast, effective and secure manner through consistent and reusable packaged pipelines.

3.5. Model Consumption

- Monitor functional and technical performance of your models in production.
- Set automated alerts to flag errors and the need of updates.

04 USE CASES

Uses cases will be presented as follows:

Introduction and Motivation:

- **Introduction** to the specific ML problems that will be tackled in the project.
- **Motivation** for dealing with such problems according to the client needs.

NLP Use Case, Documents Use Case and Classification & Forecasting Use Case:

- **Complexity** of the different scenarios that the client may encounter when facing an NLP, Documents or Classification&Forecasting problem.
- **Basic Questions and Technical Considerations** that the client should ask before selecting a specific solution.

4.1. NLP Use Case

The objective of this use case is to compare an analytical service between a pretrained model provided by the partner and a custom model trained by the bank. The analytical service to be examined will involve language models and will study the use of a provider-managed solution and a more custom, AutoML solution.

The bank's interest in this use case comes from their own experience in dealing with customer enquiries through a chat in their app. The customers chat with employees (not a chatbot), and the bank would be ultimately interested in analyzing those conversations.

The question that the bank has posed is: is it worth it to fully implement a complete analytical pipeline with a custom NLP model or is it more efficient to use one of the managed services offered by the cloud providers?

Two sections will be presented for the NLP Use Case:

- **Complexity** of the different scenarios that the client may encounter when facing an NLP problem.
- **Basic Questions and Technical Considerations** that the client should ask before selecting a specific solution.

Natural Language Processing projects can present different levels of complexity. We will consider a classification project to illustrate this idea.

	BASIC	INTERMEDIATE	ADVANCED
Subject matter	General	General but not extensively covered by hyperscalers	Specific
Language	Widely spoken (top 5)	Widely spoken (top 5)	Not widely spoken (or other conditions below)
Classification depth	Simple	Deeper, but the topics are well differentiated	Deeper, but the topics are well differentiated
Media format	Written	Written	Written or speech
Data Quality	Well annotated	Well annotated	Not well annotated, dubious quality

The objective is to **compare an analytical service** between a **pretrained model provided by the partner** and a **custom model trained by the bank**. In order to outline an NLP use case, there are some questions that need to be answered and some **special considerations** that need to be considered.

These questions and special considerations will let us choose correctly between using a cloud provider or working in local.

4.2. Use Case Documents

The objective of this use case is to process checks, invoices, receipts, and pay slips to extract data from them.

This use case is similar to the NLP one, since the goal in both is to compare an analytical service between a pretrained model provided by the partner and a custom model trained by the bank.

Since this use case can involve sensitive documents, other capabilities need to be considered , such as DLP (Data Loss Prevention).

The bank is interested in document extraction as a use case especially from the point of view of extracting key-value pairs from documents that reach their offices. Their end-goal would be to be able to extract those key-value pairs from new documents which their models have not been trained with.

The question that they have posed is very similar to the NLP one: is it preferable to develop and implement a complete pipeline with a custom document recognition model or to use one of the managed services offered by the cloud providers?

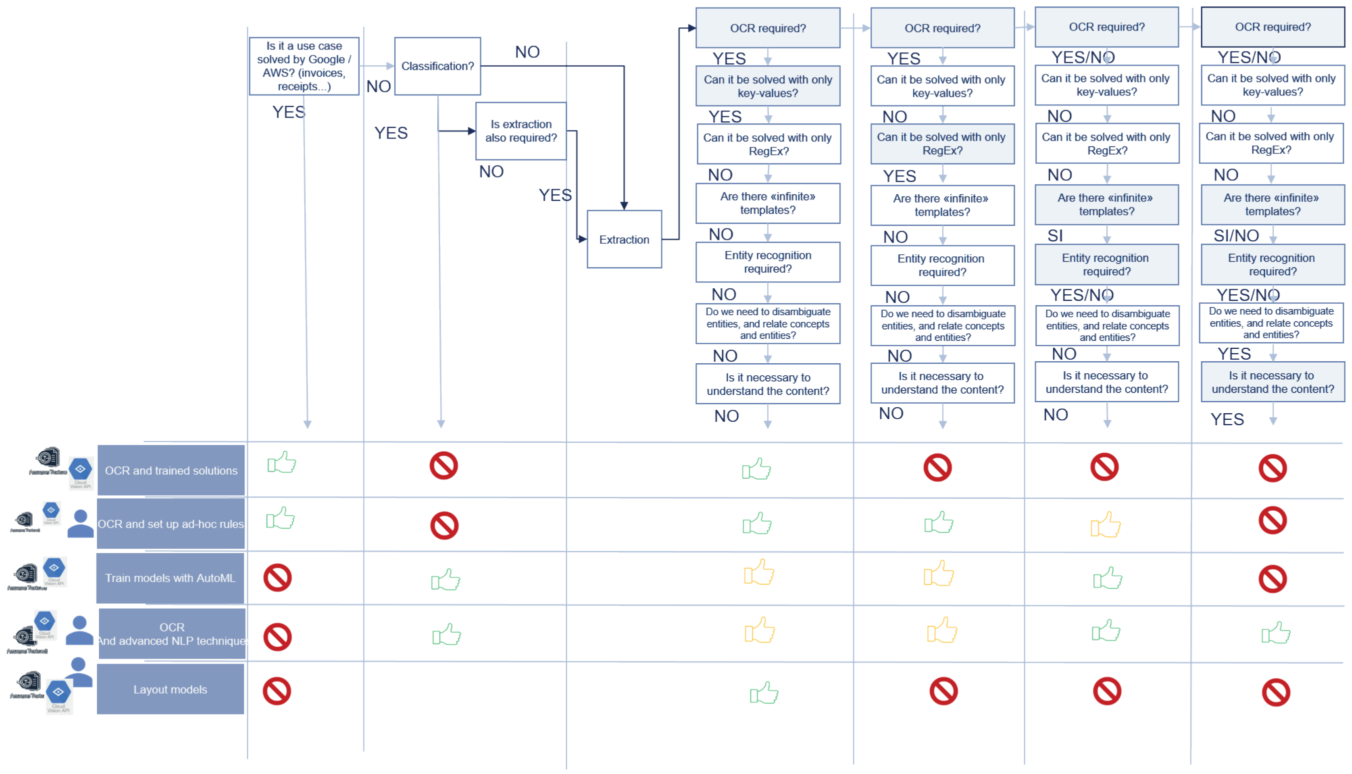
Two sections will be presented for the Documents Use Case:

- **Complexity** of the different scenarios that the client may encounter when facing a Documents problem.
- **Basic Questions** and **Technical Considerations** that the client should ask before selecting a specific solution.

Document extraction can present different levels of complexity in terms of the particular requirements of the use case. In this table we show the different scenarios we may find when dealing with this type of problems:

OCR & trained solutions (specialized Doc Ai)	OCR & configure reglas ad-hoc NLP API	AutoML for model training	OCR & NLP/ NLU advanced technique	Layout models
Pretrained models: - Receipts - Invoices - Doc. Template / forms (key value) -Doc. AI (USA lending, invoices, forms, tables)	Unstructured text: Regex, dictionaries, NER *Basic cases without disambiguation	1- Works for classification 2- Entity recognition The user can customize their forms for extraction Large number of labelled data required	Entity disambiguation Grammatical (rules, language for assigning roles, assigning facuties, etc.) Language models anchoring through ontologies	Entity disambiguation Grammatical (rules, language for assigning roles, assigning facuties, etc.) Language models anchoring through ontologies

Similar to the previous table, here we show a diagram to help us select the best option to solve our particular challenge in terms of the complexity of the use case



The objective is to compare an analytical service between a pretrained model provided by the partner and a custom model trained by the bank. In order to outline an Documents use case, there are some questions that need to be answered and some special considerations that need to be considered.

4.3. Classification & Forecasting Use Case

The objective of this use case of Classification is to explore the **options** that the **cloud providers** offer when dealing with large datasets with **hundreds of features.**

Typically, the use of a **Boosting algorithm** is appropriate for this kind of problems for which **high computational resources** are required. Possible solutions to provide these resources come from the use of **parallel computing** (horizontalization) as well as increasing a **single machine** capacity (verticalization). The bank's interest in this use case comes from the need to **find an alternative to Spark MLlib** for their data scientists to use, since it is not easy for them.

The question that the bank has posed is this: **are there any tools available in the cloud** that could serve as an **alternative to using Spark** for **working with large datasets?**

The objective of this use case of Forecasting is to explore the **options** that the **cloud providers** offer when dealing with a dataset containing over one million time-series.

A Recurrent Neural Network (LSTM – Long Short Term Memory–) will be trained for that kind of dataset to **test the efficiency** of selected instances.

Possible solutions to provide these resources come from the use of **parallel computing** (horizontalization) as well as increasing a **single machine** capacity (verticalization).

The bank's interest in this use case comes from the need to **test the goodness of Tensorflow** managed by the instances offered **inside the cloud providers.**

The question that the bank has posed is this: What is the **optimal manner to train a RNN** for a forecasting model when dealing with a large dataset containing over one million time-series?

Two sections will be presented for the Classification&Forecasting Use Case:

- **Complexity** of the different scenarios that the client may encounter when facing a Documents problem.
- **Basic Questions** and **Technical Considerations** that the client should ask before selecting a specific solution.

In order to implement the use case and perform the **binary Classification and time series Forecasting**, several steps were taken in a development pipeline. These steps are the following:

05 EVALUATION CRITERIA

The evaluation criteria will be presented on two levels, corresponding to the two main elements of the benchmark:

- The **AI Journey**, which has been previously defined.
- The **Implemented Use Cases** that will be developed.

5.1. AI Journey

The evaluation criteria for the platforms are presented in this section. The evaluation of the Journey requires the following:

- An **evaluation map**, composed of the stages of the capabilities needed by the Journey.
- A **description**, its **evaluation**, and a **weight** associated to every capability

To evaluate all platforms an evaluation map must be defined.

The evaluation map is composed of the implementation stages with all the capabilities needed by the AI Journey.

More information will be provided for all the capabilities grouped by the implementation stage.

Each capability will be defined by:

- **Description:** a small description of the capability to evaluate
- **Evaluation:** Criteria rules to measure this capability quantitatively.
- **Weight:** Capability importance for the evaluation. Greater than 100% for high importance capabilities and lower than 100% for low importance ones

5.2. Use Cases

The evaluation criteria for the platforms are presented in this section. The evaluation criteria for the **NLP, Documents and Classification&Forecasting** use cases will be explained.

For NLP:

The objective is to compare an analytical service between a pretrained model provided by the partner and a custom model trained by the bank. In order to outline an NLP use case, there are some questions that need to be answered and some special considerations that need to be considered.

BUSINESS QUESTIONS

- Objectives
- Main Specifications

TECHNICAL CONSIDERATIONS

- Voice Projects
- Non-Western Text
- Translating Data

Some of the questions to be asked are:

- What do we expect to find?
- Is it necessary to obtain results immediately or can they wait?
- What will be the order of magnitude of the data? Thousands, millions, tens of millions?
- What languages will the use case be developed in?
- What budget do we have?
- Does it make sense to create this use case based on NLP? Are the data available in any other way?
- Is it a one-off or will it be recurrent?
- Will the use case only tackle text?
- What size will the dataset be? What is our computing power?

Some of the considerations to be taken into account should be:

- A voice project uses recordings as inputs and many factors must be considered (Noise, amount of people taking part in the conversation, interruptions,...)
- Languages Non-Western tend to have a shorthand for writing with Western keyboard layouts. In this manner, it is not uncommon to see numbers or letters being used because they look similar to characters from the Cyrillic alphabet
- The main reason for using a corpus is collecting real-life language data.

For Documents:

The objective is to compare an analytical service between a pretrained model provided by the partner and a custom model trained by the bank. In order to outline an NLP use case, there are some questions that need to be answered and some special considerations that need to be considered.

BUSINESS QUESTIONS

- Objectives
- Main Specifications

- Regular document recognition
- Advanced document recognition

TECHNICAL CONSIDERATIONS

- Entity Recognition
- Entity Disambiguation
- Understanding Content

Some of the questions to be asked are:

- What would be gained from obtaining a higher accuracy? To what extent does it influence our business objective?
- Is it to feed a dashboard (and look for a basic statistic) or is it to create a report or modify a corporate strategy (further depth and precision is required for the analysis)?
- What budget do we have?
- Does it make sense to create this use case based on Documents? Are the data available in any other way?
- Is it a one-off or will it be recurrent?
- What is the document scan quality?
- How many languages are contained in the document?
- Are there handwritten documents?
- Are there signatures / stamps?
- Are there partially filled in forms?

Some of the considerations to be taken into account should be:

- **Entity recognition** deals with the detection of proper nouns and words (People names, country, cities…)
- Entities could be **ambiguous** (A person might be first introduced by name, and then referred to by a pronoun. Also, some ID numbers, bank account numbers etc might follow a specific format and be similar to each other)
- **Understanding Content** to summarize the clauses of a contract, to answer questions from information contained in a document, to perform an advanced similarity search
- Will the use case only tackle text?
- What size will the dataset be? What is our computing power?

For Classification & Forecasting:

The objective of this use case is to explore the options that the cloud providers offer when dealing with large datasets with hundreds of features. In order to outline such use case, there are some questions that need to be answered and some special considerations that need to be considered.

BUSINESS QUESTIONS

- Objectives
- Main Specifications

- Classification, a binary classification
- Forecasting, a time series forecasting

TECHNICAL CONSIDERATIONS

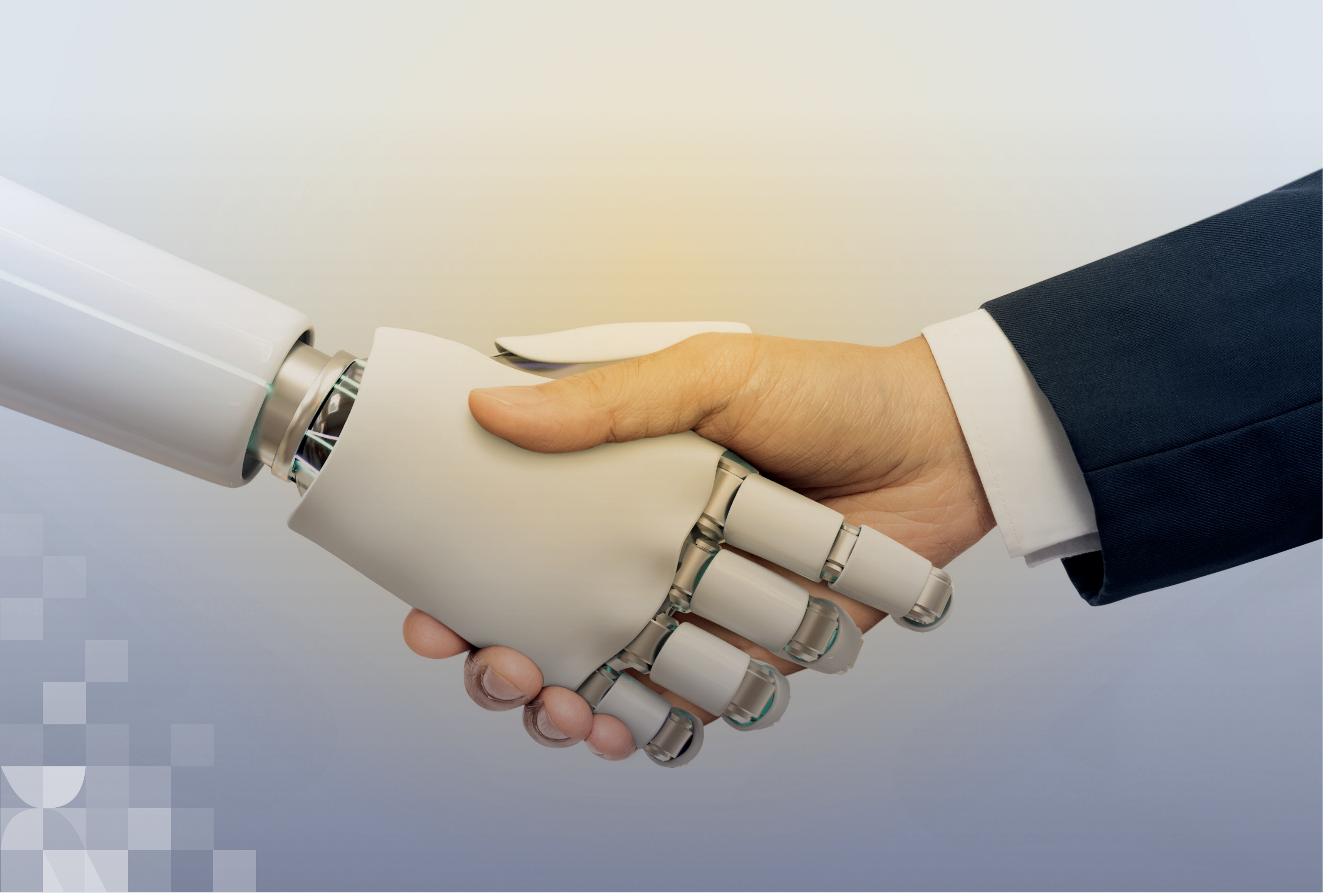
- Parallelization
- Technical Considerations
- GPUs

Some of the questions to be asked are:

- What results are we going to focus on (price, training time, metrics)?
- What is the main objective of the project?
- Is it to find out if such a project is feasible in the cloud or are we looking to optimize the process?
- What is the order of magnitude of the dataset?
- How large will the dataset be? How many rows and features are required?
- Is it a one-off process or will it be recurrent?

Some of the considerations to be taken into account should be:

- Is parallelization required? Is the dataset partitioned?
- What is the most efficient pipeline to achieve efficient parallelization using the tools of the cloud providers?
- Is the use of pre-loaded libraries, such as Tensorflow, a good fit when dealing with parallelization in a cloud-based framework?
- What is the most efficient choice for each case scenario?
- What is the best choice in each case based on a time/cost ratio?
- What are the benefits of using GPU instances in the cloud providers?



CONCLUSIONS

After carrying out different implementations of the use cases and comparing the results, we can draw the following conclusions:

PROS

- Model ownership.
- Total customization.
- There is no API cost.
- The corporation can choose to have a dedicated team.
- Better data privacy.

CONS

- **Updating** the models could be **costly** in terms of data annotation and labeling.
- The **cost of hardware** for development and training could be significant.
- **Deep expertise** is required. The area contains several sub areas and it is **hard** to **cover all** of them with equal proficiency.
- Maintenance **costs.**
- The **team** should have sufficient **bandwidth** to be able to **attend request.**
- Requires **strict cybersecurity controls.**

The pros and cons based on hyperscalers:

PROS

- Cloud providers have **huge language models.**
- There is a certain degree of subject matter abstraction **(less complexity).**
- The machines are managed by the hyperscaler, **no need** to **spend** on **hardware.**
- **Language models** are ***updated*** on regular cycles.
- **Existing NLP architecture**, given by the hyperscaler.

CONS

- The **model cannot be customized** (black box)
- Some APIs require a significant amount of **preprocessing and development**, especially when dealing with deep text analysis or voice.
- The **specialist** team must **stay updated.**
- In some scales it **can be costly.**
- When **providers change** their language model, the end **user is affected.**
- The **models** are **not owned by customer**, it´s hard to change their structure a posteriori.

Clouds solutions on the AI`s field are in constant evolution and already have a sufficient maturity level to be considered a key piece in the AI strategy of any bank.

They currently provide technological solutions that enable performing the analytical end-to-end from storage, processing/preparation and elaboration of analytical models, deployment and industrialisation. This solutions are becoming more robust, usable and efficient and allow to accelerate the adoption of this practices by taking advantage of the inner characteristics of a cloud environment (scalability, flexibility, capex minimization, etc.).

Additionally, cloud computing providers, in addition to models development tools, also offer packaged solutions and/or services that can respond to specific necessities. In this field and given the variety of casuistry's, languages, etc.., it cannot yet be considered a generalized use with good results, however, its constant evolution and continuous improvement, make it advisable to consider its use for some more common use cases and as a "commodity" part of more complex workflows.

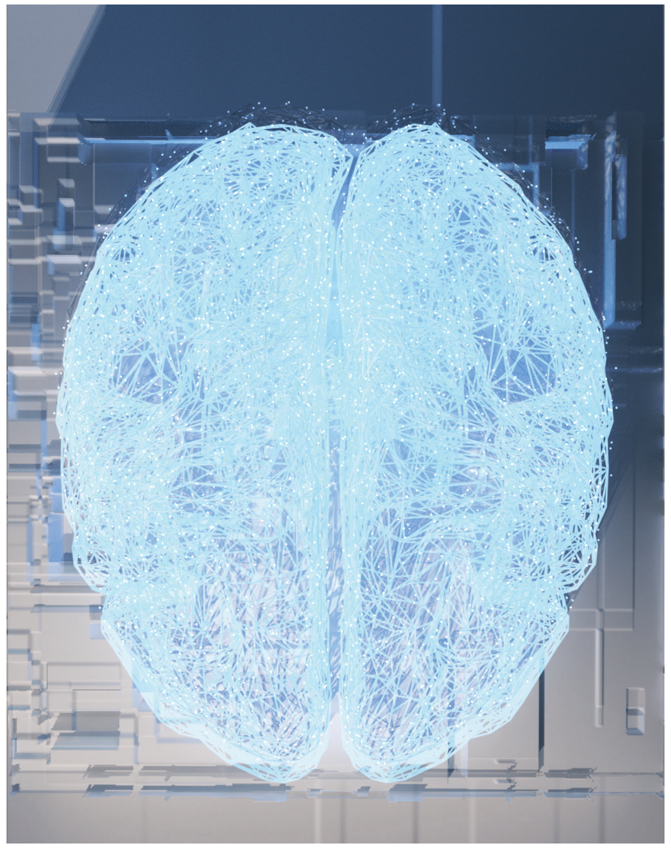
Given the constant evolution of platforms in these fields, it is expected that there will be a continuous trend towards the automation of these processes, both in terms of the development of specific models and the use and integration of packaged services/solutions. Therefore, their inclusion in the definition of corporate solutions is recommended in order not to lose these abilities.

WHY NTT DATA?

NTT DATA on AI-Driven Banking

NTT DATA, was recently named a Challenger by Gartner in its 2020 Magic Quadrant for Data and Analytics Service Providers Worldwide.

The company shares the Innovation DNA as part of NTT Group, accelerates open ecosystems and contributes to fostering the creation of disruptive AI products and solutions across the organizations.



As a trusted global innovator, our values comes from "consistent belief" to shape the future society with clients and "courage to change" the world with innovative digital technologies.

+50
Countries in which NTT DATA operates

6th
Largest company in the IT sector*

8.721
Professionals in Data & Intelligence around the globe