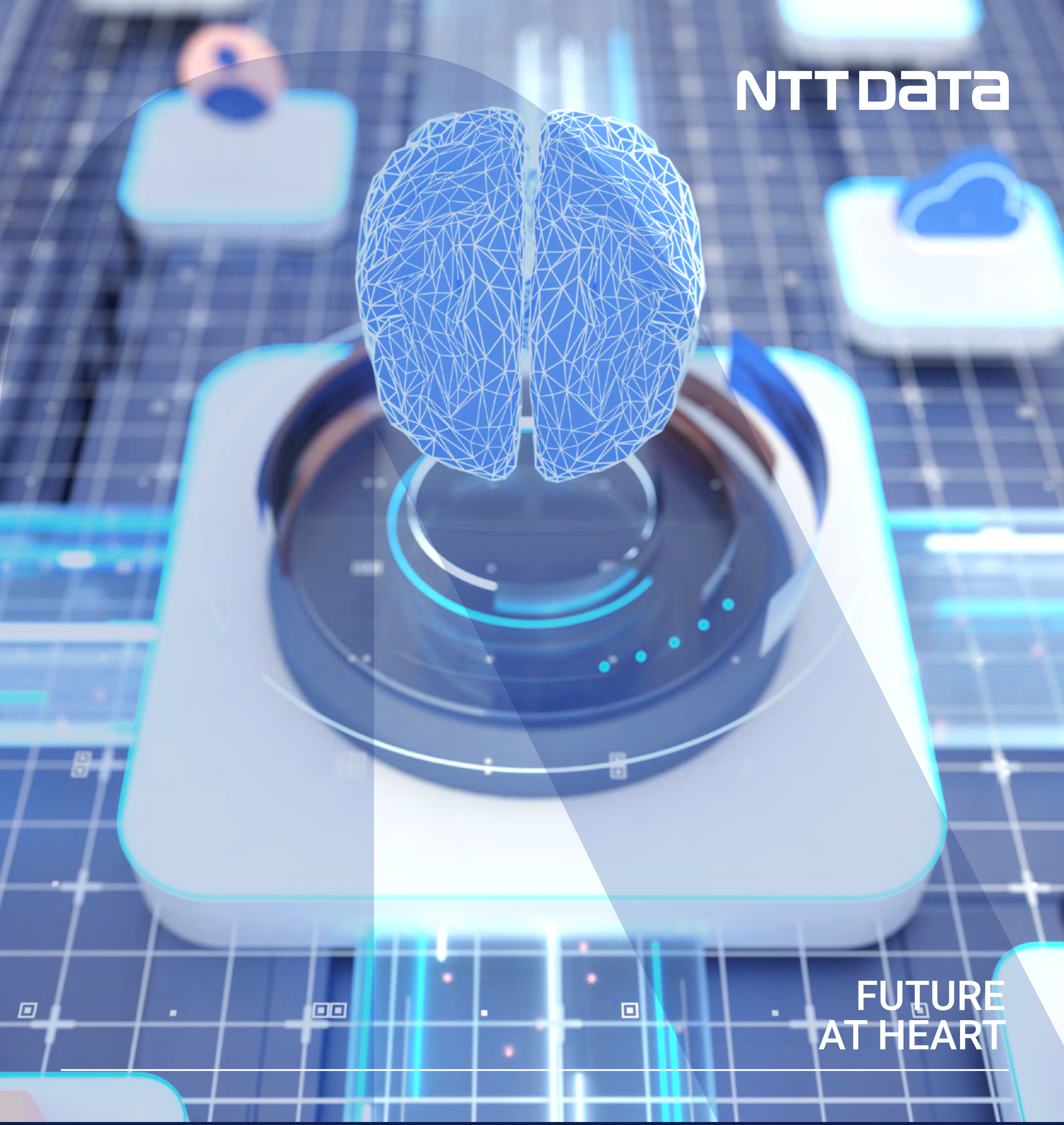
NTT DATA



FUTURE
AT HEART

Technology Stack for AI Prototyping

MARCH 2022

Technology Stack
for AI Prototyping

March 2022

CONTENTS

INTRODUCTION

In the last decade, the popularity of Artificial Intelligence (AI) has massively increased and has infiltrated into our every-day lives up to the point that we are not surprised anymore by the recommendation of related products before paying in an e-Commerce site, nor by the daily interaction with voice assistants. In this new technological era, all companies must take advantage of AI and understand how to embed it into new business models, services, and interactions with their customers, as these expect brands to deliver relevant value through differential experiences.

At NTT Data, we detected that since AI services require a lot of experimentations to ensure the expected results and the fact that this type of technology is not yet very mature, many clients prefer to have definite Proof of Concepts (PoCs) to validate AI use cases before investing more resources into a larger scope project. At the same time, in the last few years, several disruptive technologies have appeared that, if wisely combined, allow Data Scientists to rapidly build end-to-end AI prototypes. At NTT Data, we have been testing and investigating these technologies and have built a Python-centric AI Prototyping Technology specifically for that purpose.

In this guide, we present the need of prototyping AI projects, expose the game changer technologies that allow the rapid and agile prototyping of this kind of projects and introduce our very own AI Prototyping Tech Stack. Finally, we exemplify the use of this technology stack with a valuable use-case for the Insurance sector.

02 The Importance of AI Prototyping

Tech players are setting new rules on the definition and delivery of new market value proposals, embedding AI to deliver outstanding customer experiences. We are currently seeing the rise of the new digital era and in the coming years, all digital experiences and services will as well be intelligent.

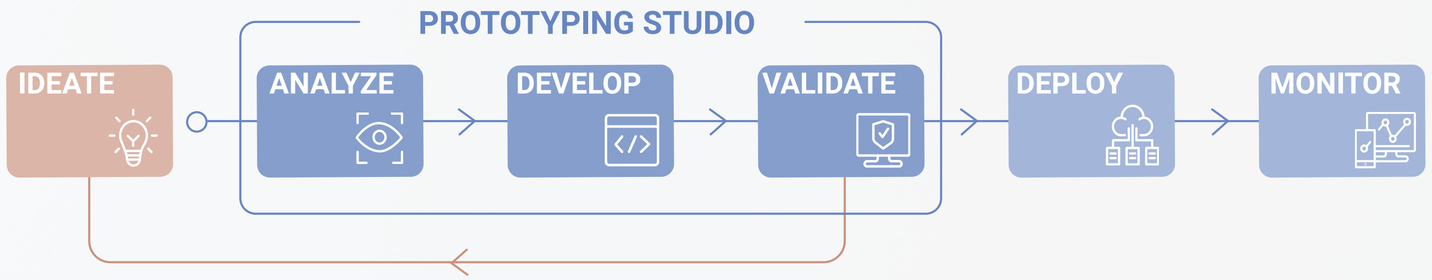
Let's say we are an insurance company that has recently become aware that we need to embed AI into our business models, services, and interactions with our customers if we want to thrive in this new technological era. But… where do we start? We know we have tons of customer data, high skilled IT professionals and brand-new IT infrastructure and so on but we do not have a clear AI-related business-valuable use-case, let alone its guaranteed success. Regardless of the ideation process, should we start devoting whole interdisciplinary teams to produce a final AI product/ service version?

Recent reports [1] show that around 87% of Machine Learning models never make it into production. Lack of data, difficulty in interdepartmental communication and lack of cross-language are some of the main contributors to the final failure in the lifecycle of these models. In addition to these reasons, the fact that AI use-cases are not always correctly validated by different levels of professionals at early stages and that the model's outputs are not always as expected result in the majority of cases in a large waste of resources.

And that is when AI Prototyping enters the scene. **With a tangible prototype, we can rapidly validate that a use-case is viable from a technological point of view and show its business value to sponsors and any other relevant stakeholders even at early stages of experimentation, thanks to an effective visual language.** Those to aspects are key to guarantee the success of an AI product or service.

But what is Prototyping anyway?

Prototyping is an experimental process where teams implement ideas into tangible products, capturing design concepts and testing on users. With prototypes, designs and products can be refined and validated so that only the right products are released. In the context of AI, we understand prototyping as the end-to-end conception and development of services and products embedding AI.



Benefits of AI Prototyping

Many are the benefits of AI Prototyping, but we go deep on the following six ones:

BUSINESS VALUE
Data Scientists will have the tools and the know-how to demonstrate to business sponsors the value that lies in the data and can be levered through AI. Prototyping will create more efficient feedback loops, as the visual representation of the power of the AI models will inspire subject matter experts in order to formulate recommendations.

CULTURE OF EXPERIMENTATION
The successful implementation of a prototyping methodology allows AI to permeate all business areas and processes, showing their potential (or not) business value rapidly and reducing the frictions that lengthy, unsuccessful projects may generate. As the value of AI can be more easily grasped, more stakeholders will understand the importance of gathering quality data to fuel the prototypes that may transform the organization.

COST
With prototyping, the use cases are validated before being deployed. This investment prevents costly errors in advance because project success is guaranteed.

ORGANIZATIONAL IMPACT
By transforming cryptic ML models written in Python code into a compelling visual storytelling, non-AI workers and business leaders will understand how AI may benefit them in their everyday tasks. A smoother, systematic dialogue will arise between AI experts and business areas, unlocking new transformation opportunities.

QUALITY
With prototyping, the Agile methodology is embraced, which leads to a high number of small iterations on the product idea. As rapid feedback is included on each iteration, the quality of the product increases over time, which leads to a high-quality product.

DELIVERY TIME
With prototyping, non-valuable use cases are rapidly stopped, so the speed to market of the valuable use cases is significantly increased, allowing fast innovation.

03 Evolution of Technologies for AI Prototyping / Where do we come from?

Some years ago (but not that many years ago), AI Prototyping was far from being a fast and agile process. It took months and different professional profiles to create a tangible prototype to validate an AI use-case. Once that prototype was developed, and in the few cases where the prototype ended up being successful, the prototype resulted in the final product for the number of resources dedicated, so it is not sure if the term prototyping applied in the first place.

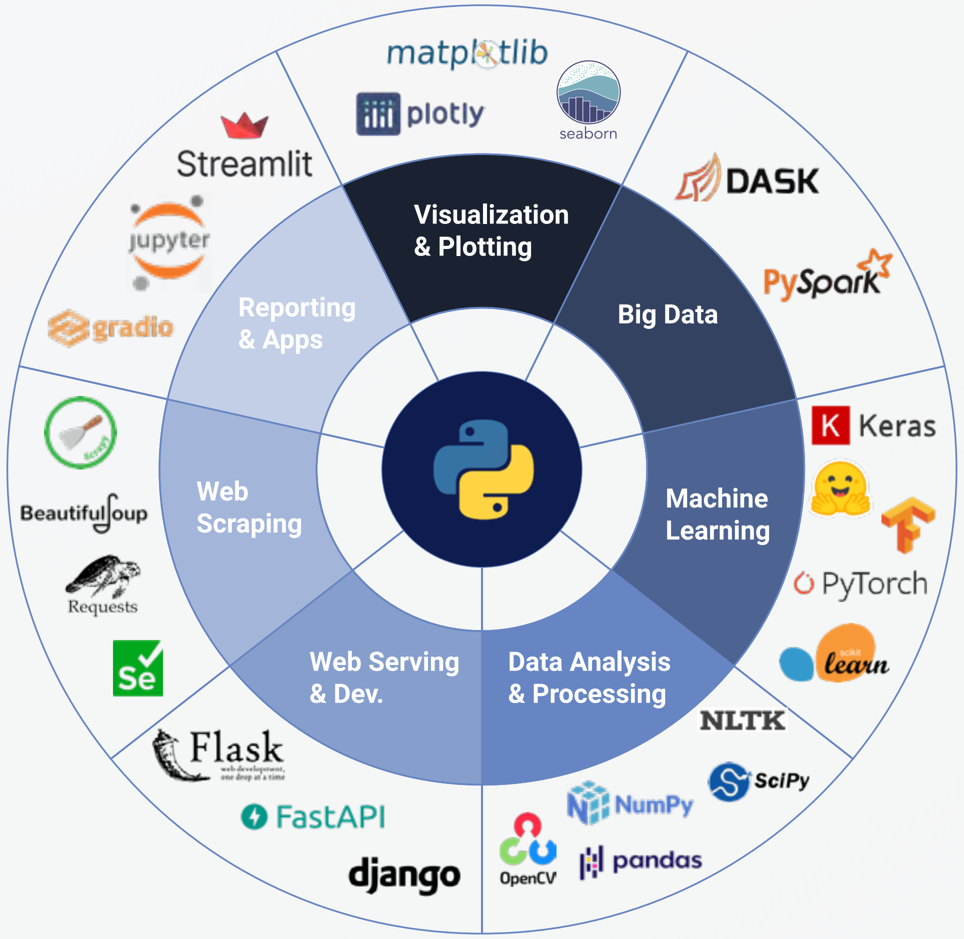
In most cases, the backend part of the prototype simply consisted of AI models built from scratch, which require a high amount of expertise, data, and time. For the front-end side, experienced front-end developers (experts on web development and UX/UI) needed to create a portal where the prototype could be showcased. Those developers should constantly communicate with the Data Scientists developing the backend to integrate both ends. Additionally, an IT architect should develop an infrastructure where the portal should be hosted.

Over the last few years, some new technologies have emerged that have ended up playing as game-changers in the world of AI prototyping. These technologies end up covering the whole full stack of the development process: the backend now can consist of combination of different ad-hoc fine-tuned models, AutoML or 3rd party APIs, while the front-end part is covered by technologies such as Streamlit. With the adoption of these technologies, a single Data Scientist can quickly create, securely deploy, and share a complex AI use-case, which results into more prototypes being implemented, higher speed-to-market and quality and so on.

We believe that there are three pillars that sustain this change of scenario, and we will following discuss them in detail:

3.1 The Almighty Python

The main thing that all these technologies have in common is that they are based on the open-source Python programming language, which has become the most popular programming language for Data Science mainly due to the following reasons:



GREAT LIBRARY ECOSYSTEM

A library is a module published by different sources like PyPi which includes a pre-written piece of code that allows users to reach some functionality or perform different actions. These libraries provide different level functionalities so developers don't have to code them from scratch every time.

Currently, Python includes a wide and ever-growing ecosystem of different libraries and frameworks that cover most programming capabilities required for performing data-related applications. This ecosystem covers from web scraping libraries such as beautifulsoup or Selenium to reporting and application libraries such as Streamlit or Gradio passing through data analysis and processing, visualization and plotting, AI and ML and so on.

EASE OF USE AND FLEXIBILITY

Python, with a simple syntax that resembles the everyday English language, is quite easy to learn and pick up, allowing Data Scientists to start using it for AI/ML development without wasting too much effort into learning the language. On addition to that, Python, is an extremely flexible language, for it can be used in a wide variety of scenarios, from simple scripting to productive Object-Oriented Programming.

COMMUNITY SUPPORT

As an open-source language, Python offers a rich number of different resources open for programmers such as a great official documentation (for the language and for most of its libraries), great blogposts, guides and so on. Additionally, as it is an extremely popular language, there exists a large supportive community that empowers the programmers in community and forums by helping each other out.

3.2 The Unbeatable Cloud Providers

Nowadays, cloud computing implies the delivery of different services through the Internet, including data storage, servers, databases, networking, and software. The term "cloud computing" was coined more than 20 years ago [2] and since its beginning it became widely used as it introduced the concept of Software as a Service, with companies such as Salesforce having a great impact on the overall industry. However, it was until 2006, with the launch of Amazon Web Services (AWS) where modern cloud computing was created and popularized. Since then, other important providers such Google Cloud or Microsoft Azure have emerged and the **technologies of cloud computing have played a very important role on the accessibility of AI, acting in most cases as an effective catalyst.**

The main contributions of cloud computing in AI can be split into the following three subsections, that resemble the different cloud delivery methods:

ML Infrastructure: Provision of a ready-to-use and pay-as-you-go powerful and specific infrastructure such as CPU, GPU, disk memory, networks and so on.

ML Services Platform: Delivery of ML platforms (such as AWS SageMaker) that provide a scalable platform to build, train and deploy ML models quickly, including a large variety of built-in algorithms and funcionalities.

High-level AI Software Solutions: Pre-trained AI models that can be easily consumed by an HTTPS API. In the recent years, with the aparition of AutoML, these endpoints can even be customized, so with little effort and practically no experience in AI, custom AI models can be easily built and consumed.

Given the conjunction and rapid evolution of these three cloud-powered AI delivery methods and their wide coverage of the AI spectrum (Computer Vision, Natural Language Processing, Forecasting, Reinforcement Learning and so on), the complexity of the use-cases that can be derived from its use is rapidly increasing. While some years ago it would take a data scientist (DS) months to create a simple use-case that included a ML classification model, nowadays the same DS can implement a complex AI use-case by combining language and vision 3rd Party APIs with fastly trained custom ML models and so on. Later in this guide we will give an example on how to prototype these kinds of complex use-cases.

Seeing the speed at which these cloud giants evolve, one may ask herself how to keep up. Or perhaps even how to get started. For those two tasks, we recommend taking a look at the following list:

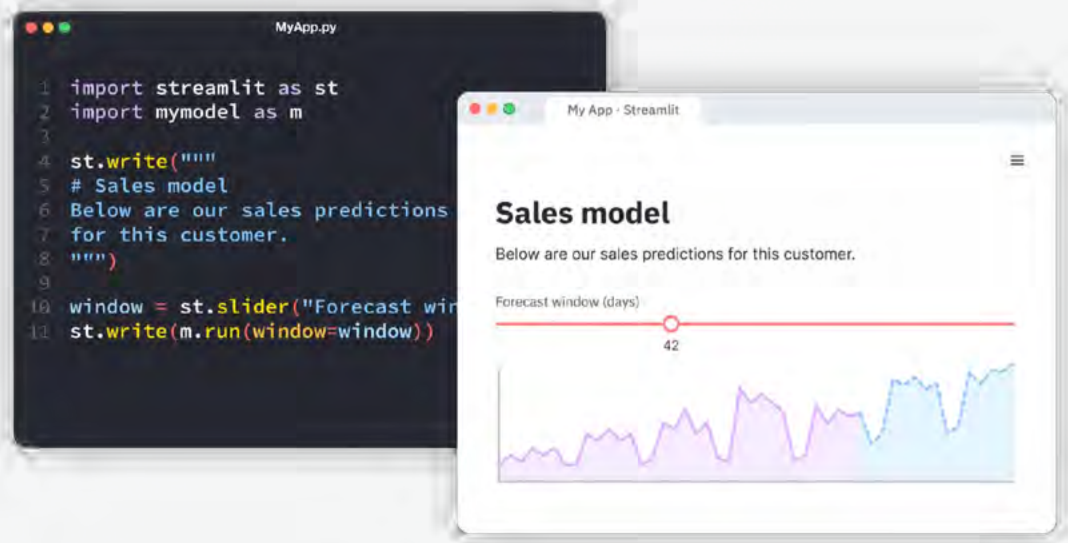
- Official and unofficial (Udemy, Udacity and so on) **training and certifications** help learners build credibility and confidence by validating their cloud expertise with an industry-recognized credential.

- Virtually (or phisically, if lucky) attend the **massive conference** that each of the main cloud vendors organize towards the end of each year, where they present keynote announcements, training opportunities and thousands of technical sessions. Currently, the main confrerences are AWS re:Invent, Google Cloud Next and Microsoft Ignite.

- Other resources such as **newsletters** (where the next events, webinars and launches are anounced), **specific blogs and forums,** and even streaming Twich chanels are currently available to broaden and deepen your knowledge of these technologies.

3.3 Streamlit: The fastest way to build data apps

The open-source Python library **Streamlit** was created by Adrien Treuille, a former employee of Carnegie Mellon University, Google, and ZOOX. It was conceived with the goal of facilitating the task of easily showing the work done by ML engineers and data scientists, without having to fall into developing web apps that would end up being unmaintainable. Since its launch in 2018, the start-up developing Streamlit has raised more than $62M in different financial rounds [3] and has reached almost 20k GitHub stars (and counting) [4], which means that Streamlit has come to stay.



```
import streamlit as st
import mymodel as m

st.write("""
# Sales model
Below are our sales predictions
for this customer.
""")

window = st.slider("Forecast wir
st.write(m.run(window=window))
```

Sales model

Below are our sales predictions for this customer.

Streamlit embraces Python scripting and treats widgets as variables, which means that creating an interactive layout becomes a really easy task. It also reuses data and computation as well as adding the ability to deploy instantly.

In brief, by just adding some lines to your already existing Python code, you can instantly obtain a high-performing data app in your browser that is easy to deploy and is responsive for all types of devices.

Moreover, Streamlit allows their users to easily create Custom Components, additional modules that extend the tools basic capabilities. Starting from an officially provided template, developers can write JavaScript and HTML code that is easily rendered in Streamlit apps to fulfill their specific use-case needs. Some examples of this kinds of components that result really useful are a component to embed a code editor in your app, a simple component to display annotated text or a component for easily cropping custom images.

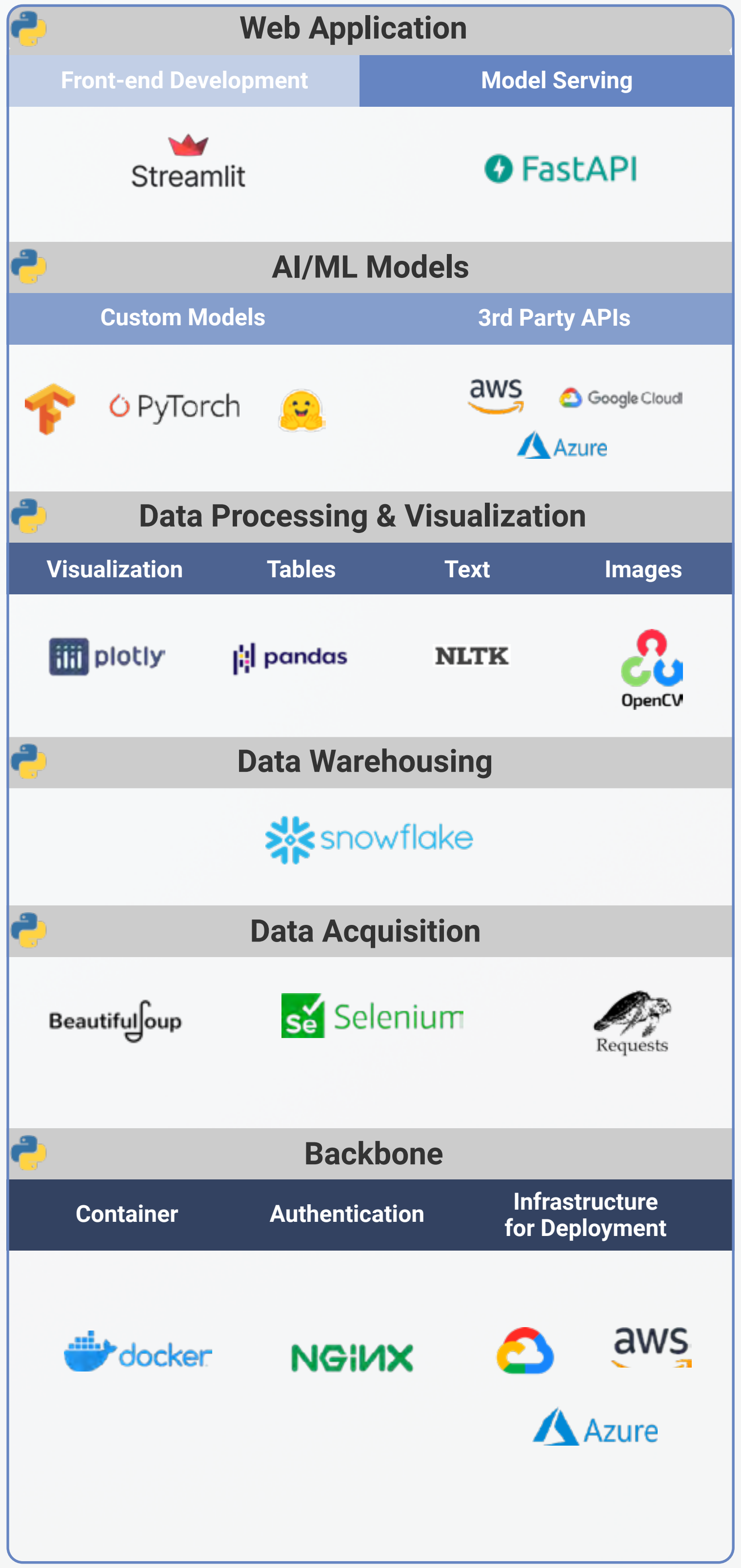
Given all of this, it is no wonder that over 50% of Fortune 50 companies such as Apple, Uber or IBM (as they claim in their website) are currently trusting Streamlit as their main tool for rapidly building and sharing data apps.

04 The AI Prototyping Tech Stack

For the last few years, at NTT DATA we have been working on multiple client projects and Proof of Concepts (PoCs) related to AI and have experimented with different technologies with the goal of reducing the time for building AI prototypes and increasing the speed of experimentation.

Based on those experimentations, we have built (and are continuously building) our own AI Prototyping Technology Stack, in which we leverage the capabilities that Python offers, using it as the main programming language, along with a set of its extensions, allowing the Data Scientist to have a wider coverage of the prototyping studio, reducing the complexity on the type of teams and delivery methods.

On the bottom of the Tech Stack, there is a **Backbone layer** that combines the different technologies that provide a fast, secure, and containerized infrastructure to deploy and be able to easily share the prototype applications. This includes the different cloud vendors for providing compute instances, NGINX for the authentication and securitization of the application and Docker for containerizing it so it can run on top of every OS.



On top of the backbone layer, there is the **Data Acquisition layer,** where we rely on beautifulsoup and Selenium for obtaining information from static and dynamic websites respectively. We also include the Requests library to easily send general-purpose HTTP requests.

Next, we find the **Data Analysis and Processing Layer,** where we include Plotly for data visualization, which integrates really beautifully with Streamlit and allows to have more interactive and responding visualizations than other similar libraries. On this layer we also find pandas, the essential library for dealing with DataFrames (tabular data) and NLTK and OpenCV for working with text and visual content in that order.

One of the key layers of the stack is the **AI/ML Models Layer,** which includes on the one hand ML frameworks for building custom ML models from scratch or customizing existing ones (TensorFlow, PyTorch and HuggingFace) and on the other the main cloud vendors that provide high-level AI solutions in the form of APIs.

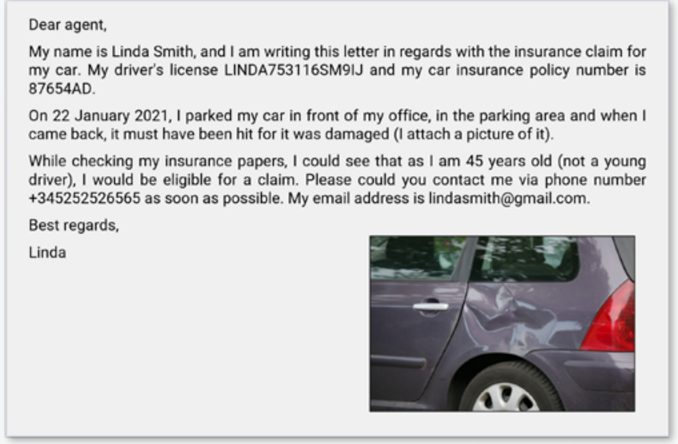
Finally, the other key layer in our Tech Stack is the Web Application Layer, which leverages the capabilities that the previously introduced Streamlit has to offer for rapidly building data apps and FastAPI, a web development framework to easily expose as independent endpoints the custom ML models.

05 A Good Example Has Twice the Value of Good Advice

Let's go back to imagining that we are an insurance company that wants to define a potential prototype for this sector. We have identified a problematic that consists of manually processing the thousands of emails that we receive from customers regarding car damage claims. These contain explanations on what happened to their vehicle together with personal information (e.g., personal addresses, driver ID, email) and some images of the damages. We would like to use AI to identify all the personal information in the text and classify the type of damage from the image.

We consider that before spending lots of efforts and resources into developing a large solution, it would be worth prototyping this use-case in order to validate its technical feasibility and its business impact by sharing it with the C-level executives.

Input: Customer Email



Dear agent,

My name is Linda Smith, and I am writing this letter in regards with the insurance claim for my car. My driver's license LINDA753116SM9IJ and my car insurance policy number is 87654AD.

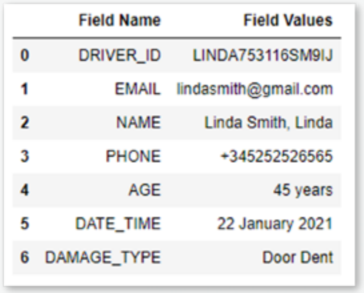
On 22 January 2021, I parked my car in front of my office, in the parking area and when I came back, it must have been hit for it was damaged (I attach a picture of it).

While checking my insurance papers, I could see that as I am 45 years old (not a young driver), I would be eligible for a claim. Please could you contact me via phone number +345252526565 as soon as possible. My email address is lindasmith@gmail.com.

Best regards,

Linda

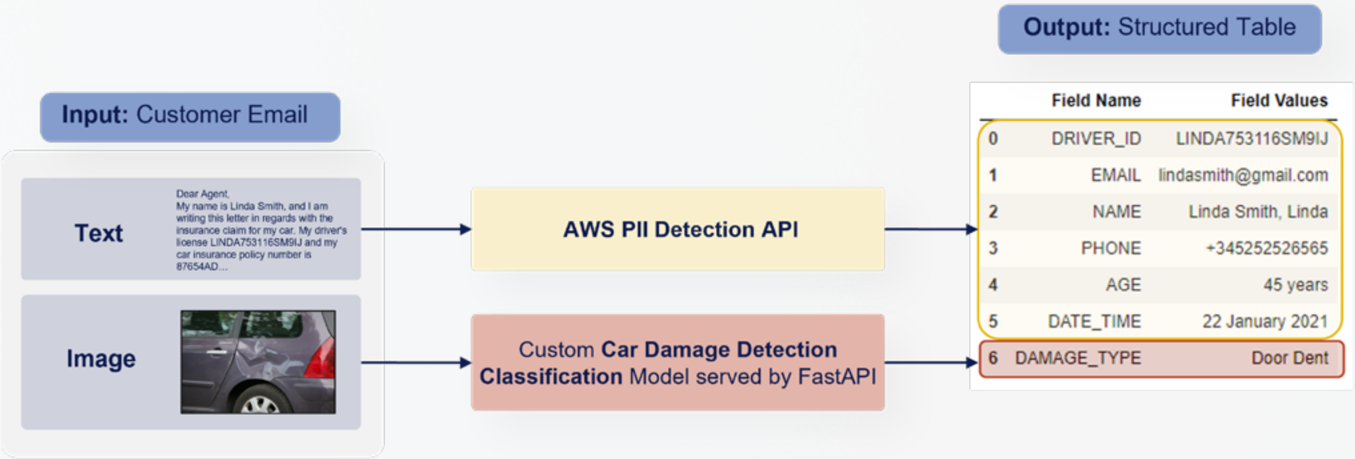
Output: Structured Table



	Field Name	Field Values
0	DRIVER_ID	LINDA753116SM9IJ
1	EMAIL	lindasmith@gmail.com
2	NAME	Linda Smith, Linda
3	PHONE	+345252526565
4	AGE	45 years
5	DATE_TIME	22 January 2021
6	DAMAGE_TYPE	Door Dent

We start thinking about how we could address this problematic from an AI perspective and determine that we could split the use-case into two subsections. On the one hand, we want to detect entities in the text that contain personal information.

There are some powerful third-party NLP APIs, such as AWS Comprehend PII Detection API, that can spot this kind of information in both real-time synchronous operations and batch asynchronous jobs. On the other hand, we want to know the type of damage that the vehicle in the image has. For this, we will need to train a custom model. We could base our computer vision model on a ResNet50, fine-tune it on the Peltarion Car Damage Dataset [5] using PyTorch, and easily wrap it inside a REST API using FastAPI.



Dear Agent,
My name is Linda Smith, and I am writing this letter in regards with the insurance claim for my car. My driver's license LINDA753116SM9IJ and my car insurance policy number is 87654AD...

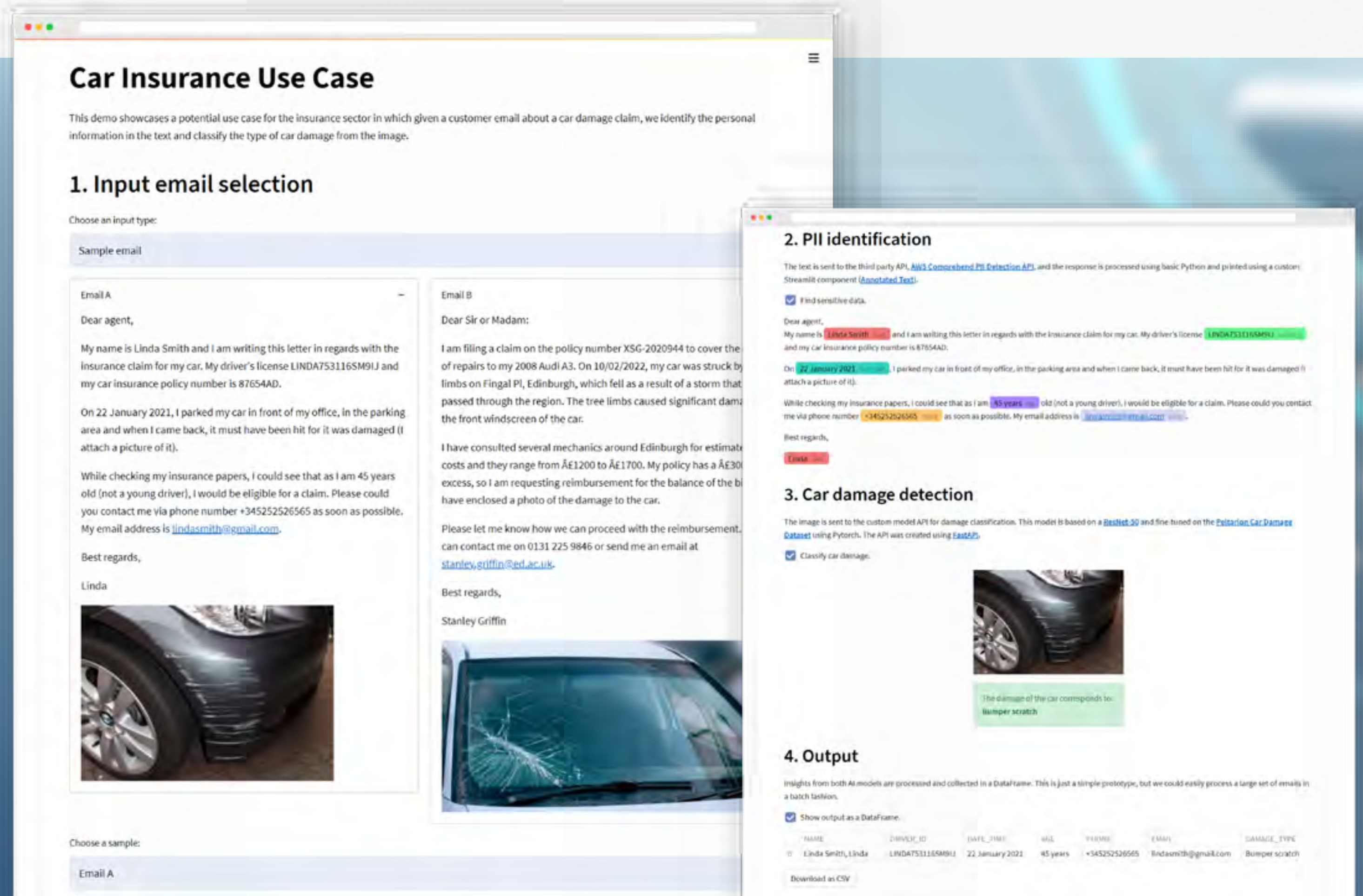
	Field Name	Field Values
0	DRIVER_ID	LINDA753116SM9IJ
1	EMAIL	lindasmith@gmail.com
2	NAME	Linda Smith, Linda
3	PHONE	+345252526565
4	AGE	45 years
5	DATE_TIME	22 January 2021
6	DAMAGE_TYPE	Door Dent

Once we have our AI models ready, we can turn our Python scripts and Jupyter Notebooks into something more interactive, shareable, and easier to present using Streamlit with little effort. We realize that we have validated the technological viability of this use-case and have something tangible that we can show to our stakeholders to refine and validate our prototype iteratively until we are satisfied with the result.

06 But How Do We Scale all This?

In the previous section, we showed how we can rapidly materialize our ideas into prototypes using the AI Technology Stack and the Prototyping Studio. However, these prototypes could be easily scaled up thanks to our **AI Labs and AI Driven Design methodology.** The former acts as an innovation center, facilitating rapid experimentation with prototypes and serving as the first step towards the creation of Intelligent Digital Services.

The latter merges business strategy with the design and development of solutions to identify, test, and scale Intelligent Digital Services. This is achieved thanks to the integration of multidisciplinary teams where specialized profiles collaborate in the different phases of the solution combining business expertise with the highest technical excellence. Allowing our customers to create value and keep accelerating their innovation process across the different stages of the service conceptualization, design, development, and deployment.



Car Insurance Use Case

This demo showcases a potential use case for the insurance sector in which given a customer email about a car damage claim, we identify the personal information in the text and classify the type of car damage from the image.

1. Input email selection

Choose an input type:

Sample email

Email A

Dear agent,

My name is Linda Smith and I am writing this letter in regards with the insurance claim for my car. My driver's license LINDA753116SM9IJ and my car insurance policy number is 87654AD.

On 22 January 2021, I parked my car in front of my office, in the parking area and when I came back, it must have been hit for it was damaged (I attach a picture of it).

While checking my insurance papers, I could see that as I am 45 years old (not a young driver), I would be eligible for a claim. Please could you contact me via phone number +345252526565 as soon as possible. My email address is lindasmith@gmail.com.

Best regards,

Linda

Email B

Dear Sir or Madam:

I am filing a claim on the policy number XSG-2020944 to cover the
of repairs to my 2008 Audi A3. On 10/02/2022, my car was struck by
limbs on Fingal Pl, Edinburgh, which fell as a result of a storm that
passed through the region. The tree limbs caused significant dama
the front windscreen of the car.

I have consulted several mechanics around Edinburgh for estimate
costs and they range from Â£1200 to Â£1700. My policy has a Â£30
excess, so I am requesting reimbursement for the balance of the bi
have enclosed a photo of the damage to the car.

Please let me know how we can proceed with the reimbursement.
can contact me on 0131 225 9846 or send me an email at
stanley.griffin@ed.ac.uk.

Best regards,

Stanley Griffin

Choose a sample:

Email A

2. PII identification

The text is sent to the third party API, AWS Comprehend PII Detection API, and the response is processed using basic Python and printed using a custom Streamlit component (Annotated Text).

☑ Find sensitive data.

Dear agent,
My name is Linda Smith and I am writing this letter in regards with the insurance claim for my car. My driver's license LINDA753116SM9IJ and my car insurance policy number is 87654AD.

On 22 January 2021, I parked my car in front of my office, in the parking area and when I came back, it must have been hit for it was damaged (I attach a picture of it).

While checking my insurance papers, I could see that as I am 45 years old (not a young driver), I would be eligible for a claim. Please could you contact me via phone number +345252526565 as soon as possible. My email address is lindasmith@gmail.com.

Best regards,

Linda

3. Car damage detection

The image is sent to the custom model API for damage classification. This model is based on a ResNet-50 and fine tuned on the Peltarion Car Damage Dataset using Pytorch. The API was created using FastAPI.

☑ Classify car damage.

The damage of the car corresponds to:
Bumper scratch

4. Output

Insights from both AI models are processed and collected in a DataFrame. This is just a simple prototype, but we could easily process a large set of emails in a batch fashion.

☑ Show output as a DataFrame.

	NAME	DRIVER_ID	DATE_TIME	AGE	PHONE	EMAIL	DAMAGE_TYPE
0	Linda Smith, Linda	LINDA753116SM9IJ	22 January 2021	45 years	+345252526565	lindasmith@gmail.com	Bumper scratch

Download as CSV

CONCLUSIONS

Speeding up innovation and unlocking the full potential of AI initiatives has become a real puzzle for companies across sectors, mainly because of the lack of environment and tools to rapidly create and easily manage the fast production of AI prototypes. At NTT DATA, we have been working on multiple AI projects and experimented with different AI capabilities and tools, letting us define an ever evolving and Python centric technology stack that increases the speed of experimentation. Prototyping let us refine and validate use cases iteratively, reducing the final costs and achieving a high-quality product in a reduced amount of time. Lastly, prototypes can be easily scaled up into Intelligent Digital Services thanks to our AI Driven Design methodology creating high-value services for our customers.

References

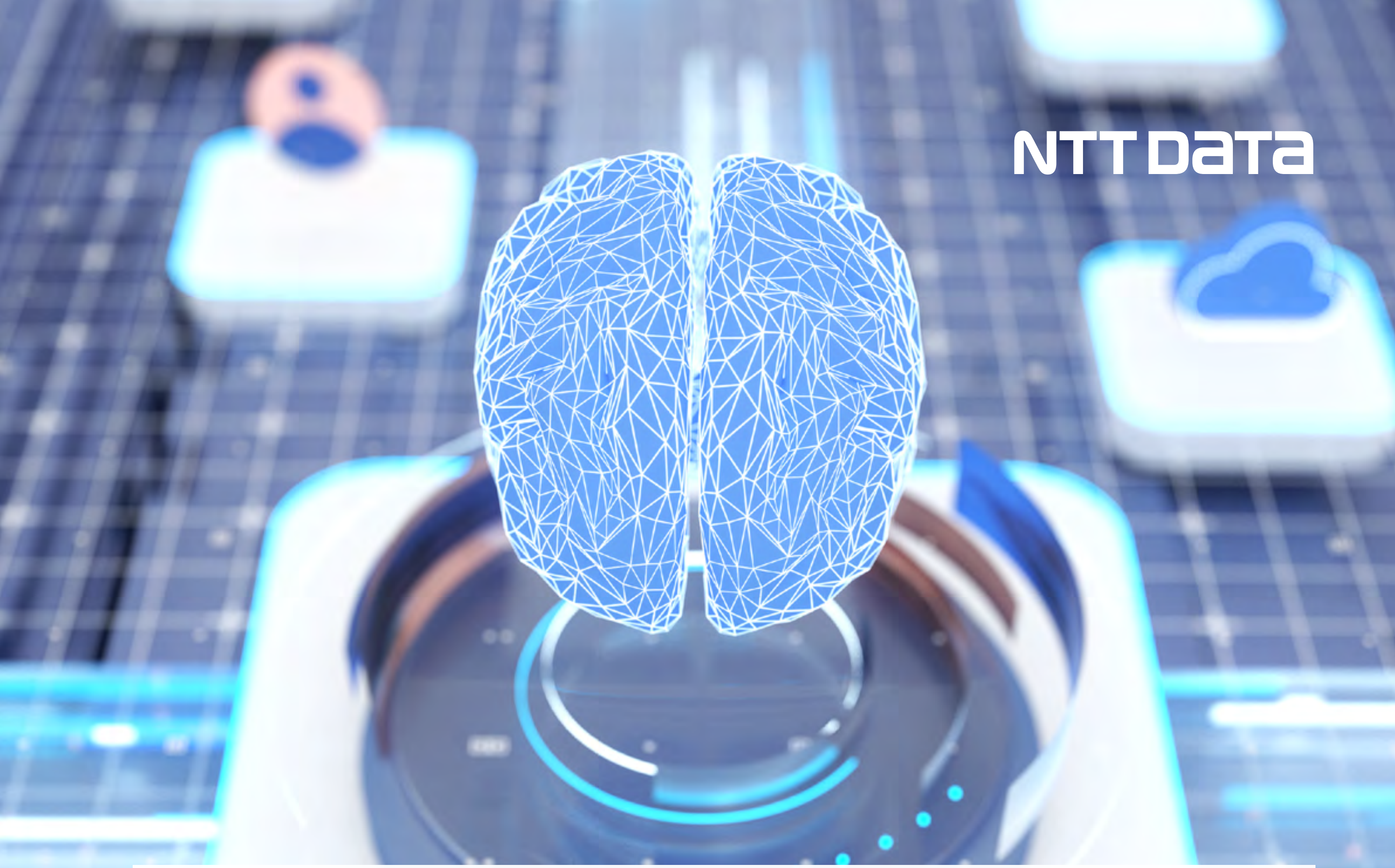
1_ “Why do 87% of data science projects never make it into production?”, Venture Beat

2_ “Who Coined Cloud Computing”, MIT Technology Review

3_ Crunchbase Streamlit

4_ Streamlit GitHub

5_ “Car Damage Assessment”, Peltarion



CONTACT US FOR FURTHER INFORMATION ABOUT THIS PRACTICE:

DAVID PEREIRA PAZ
Head of Data & Intelligence
Europe

JACINTO ESTRECHA
Head of Artificial Intelligence
Spain

ADIL MOUJAHID
Technical Manager
@ AI Center of Excellence

ANDREA CORNAVACA
Lead Analyst AI Strategy
@ AI Center of Excellence

MARCH 2022