# AI HACKATHON

November 2-27, 2023

## The challenge

Given a set of PDF documents, process them and extract the required fields.
- Initial dataset: 19 term sheets with their ground truth
- Evaluation: metrics will be extracted from a hidden test set
- Deliverable: folder with Python code

## Deliverable

In order to explain the expected deliverable, we attach an example: "deliverable_example.zip". In the example you can see two different parts: the code and the requirements file where the necessary libraries to be installed are defined. Please specify the version of each library and look out for incompatibilities among them. For the code section we give an example for those who need a more specific guide, but the design of the code and functions is more free. What is required is that the input is the location of the folder where the documents are located and that the input is in argument format. And optionally the name of the output is also an argument.

These are the exact commands we will run:

```
1   pip install -r requirements.txt
2   python main.py "test_data/"  "./predictions.csv"
```

The format of the expected output csv is similar to that of the ground truth you are asked about: each column being a field and each row being each document to be extracted.

## The fields

| Field | Format | |
|---|---|---|
| Isin | str | |
| Issuer | str | The acceptable format is both the full name (Morgan Stanley) and initials (MS). |
| Ccy | str | |
| Underlying(s) | List[str] | |
| Strike | List[float*] | |
| Launch Date | str | dd/mm/yyyy  (may also be called Trade Date) |
| Final Val. Day | str | dd/mm/yyyy |
| Maturity | str | dd/mm/yyyy |
| Cap | int | |
| Barrier | int | |

* Use dots "." for decimals: 4839.87