Sustainability Benchmarking & Predictive Forecasting

Contents

[Introduction (Business Need): 2](#_Toc155865047)

[Challenge Description: 2](#_Toc155865048)

[Tasks 3](#_Toc155865049)

[Implementation Logic 3](#_Toc155865050)

[User Interface 4](#_Toc155865051)

[Prerequisites:- 4](#_Toc155865052)

[User Stories 4](#_Toc155865053)

[Deliverables 5](#_Toc155865054)

[Expected Solution 5](#_Toc155865055)

[Solution Package: 5](#_Toc155865056)

[Design Approach 5](#_Toc155865057)

[Working code 6](#_Toc155865058)

[Playbook 6](#_Toc155865059)

[Automated Test Suite and Coverage Report 6](#_Toc155865060)

[Required Results 6](#_Toc155865061)

[Comparison Report 6](#_Toc155865062)

[Comprehensive & Brief information 6](#_Toc155865063)

[Prediction and Forecasting 7](#_Toc155865064)

[FAQs: 7](#_Toc155865065)

[How to retrieve the reporting information 7](#_Toc155865066)

[CDP 7](#_Toc155865067)

[SBTi 7](#_Toc155865068)

[TCFD 8](#_Toc155865069)

[Reports 9](#_Toc155865070)

# Introduction (Business Need):

The main goal of this problem statement is to measure the various factors concerning the ESG for given entity(Company) and currently this is manual and is chargeable to our clients by WellsFargo.

Intent is to put efforts to automate the survey benchmarking process by processing publicly available ESG reports from the respective company websites leveraging cutting edge technologies which improves operational efficiency and time-to-market.

This reports will help companies to increase their valuation & brand equity, compare and contrast the their contribution towards ESG with respect to their peers in same segment. This will also help investors to increase stake on given entity as they are adhering to regulatory and compliance and their commitment towards adopting eco-friendly, energy efficient practices.

# Challenge Description:

Design an innovative solution for **extracting and interpreting** information from **PDF, Web pages** documents using cutting-edge technology stacks. The goal is to develop a robust system capable of parsing complex financial reports and other business documents to extract relevant data related to a given company and predict the ESG scores. The extracted data should then be used to compare and analyze various indices of the target company against its competitors. Create the benchmark report in plain simple English mentioning index where company can improve for increase their valuation.

If time permits additional scope includes scrapping information available from internet for respective entities for comparison and forecasting.

# Tasks

## Implementation Logic

1. **PDF Parsing, Web Page parsing:**

- Choose a suitable library or tool for web scraping or PDF parsing based on the source of the data.

- Implement a script to extract relevant information from web pages or PDF documents.

Refer [more information](#_MSCI_Score) for details.

2. **NLP for Context Understanding:**

- Integrate NLP libraries such as spaCy or NLTK to process and understand the context of the extracted text or even suitable LLM models.

- Implement algorithms to identify key entities, relationships, and sentiment within the text. This step is crucial for extracting meaningful insights from the parsed content.

3. **Gathering Benchmarking Indices:**

- Establish connections to external data sources (e.g., APIs or databases, Webpages) to fetch benchmarking indices from designated platforms ([GRI](https://www.globalreporting.org/), [SASB](https://www.sasb.org/), [TCFD](https://www.fsb-tcfd.org/), [CDP](https://www.cdp.net/en)).

- Have a provision to upload benchmarking indices this overrides the index and metrics mentioned from framework websites (if time permits)

- Develop functions to retrieve and organize benchmarking data relevant to the entities being analyzed.

Refer [details](#_How_to_retrieve) for more information.

4. **Logic for Comparing Indices:**

- Create logical algorithms to compare the benchmarking indices obtained in step #3 with the extracted information from the target company and its competitors.

5. **Prediction and Forecasting(Good to have feature):**

The predictive aspect of the problem statement focuses on forecasting when the company is expected to achieve its sustainability goals and the amount of time required to reach those goals. This implies the need for a predictive model that takes into account historical performance data, current sustainability metrics, and potentially external factors influencing sustainability efforts.

These indices could cover a range of environmental, social, and governance (ESG) factors, forecast should not only indicate when the goals will be achieved but also quantify how much time is needed. This aspect addresses the temporal dimension of the problem, specifying the duration required for the company to meet its sustainability objectives.

6. **Report Generation with Ratings:**

- Implement a report generation mechanism that assigns ratings based on the comparison results.

- Design a rating system that reflects the performance of the target company in comparison to competitors for each benchmarking index.

7. **Verbose Report for Improvement Suggestions:**

- Develop logic to generate a verbose report that provides detailed insights into areas where the target company can improve.(Explainable AI with “citation” and “pageNumber” property).

- Utilize the comparison results and ratings to map specific recommendations for enhancement in various indices.

## User Interface

1. Use chainlit or similar framework in React for developing user interface which allows bank analyst to upload PDF as input and generate the report along with generative AI based report
2. If time permits have Q & A chatbot which retrieves information from uploaded PDF

# Prerequisites:-

1. **Sample data** – Please refer to [report](#_Reports) section for more details about the sample

2. **Access to ML tools like Azure ML studio or Google Vertex**

# User Stories

1. As a **Bank Analyst**, I want a tool to upload the PDF for all competitors and PDF of the firm which needs to be benchmarked and returns the benchmark details and generative based AI which explains how firm in the questions to be improved along with details.
2. Bank analyst needs options to change the benchmarking details as per framework so that analyst don’t need to be depend on technology team for modifying rules
3. If time permits, this tool should help banker answers basic questions available in uploaded PDF

# Deliverables

## Expected Solution

1. Develop the AI based model in Azure or GCP which identifies and retrieves the required information from PDF, Webpages based on questionnaire uploaded. You can use chatbot library like Chainlit (or) develop the web portal too which helps bank analyst to perform effectively(Additional feature)
2. Components should be modularized and exposed as API for seamless testing, This is needed even though you have user interface
3. Final Deliverables should have following artifacts
   * GIT repository without any compilations issues
   * Readme to start the application along with concise documentation which explains the core components, model, accuracy and design decisions
   * Record the working demo and store in Google Drive or GIT
   * Credentials to login to application
   * Postman collection for exposed API with authentication headers as place holders
   * Unit cases
4. Modules should be well tested with test data provided in “sample data” folder and required

Based on available time, try to implement innovative solutions like chain of thoughts, train the model from user interface based on authorization, Github actions for auto deploy and strong security layer, leverage vector DB for contextual search and embeddings, interactive chatbot functionality to query the details

## Solution Package:

Following are the tasks that need to be executed as part of this challenge:

* Provide high level user stories for the challenge.
* Provide a Design approach.
* Code that needs to be uploaded to the GIT repo (Instructions for Git).
* Provide playbook.
* Automated test suite and coverage report.

Details of each of the deliverables are as follows:

## Design Approach

The design document should detail out overall approach on solving the problem. This could include any class diagrams/sequence diagrams indicating the approach and a high level traceability matrix.

## Working code

The resultant application should have good quality and generic enough so that new rules can be added and/or removed easily and committed to respective Git repos that are assigned.

## Playbook

Provide Playbook/Runbook for deploying the solution. Provide information towards how it should be deployed & monitored.

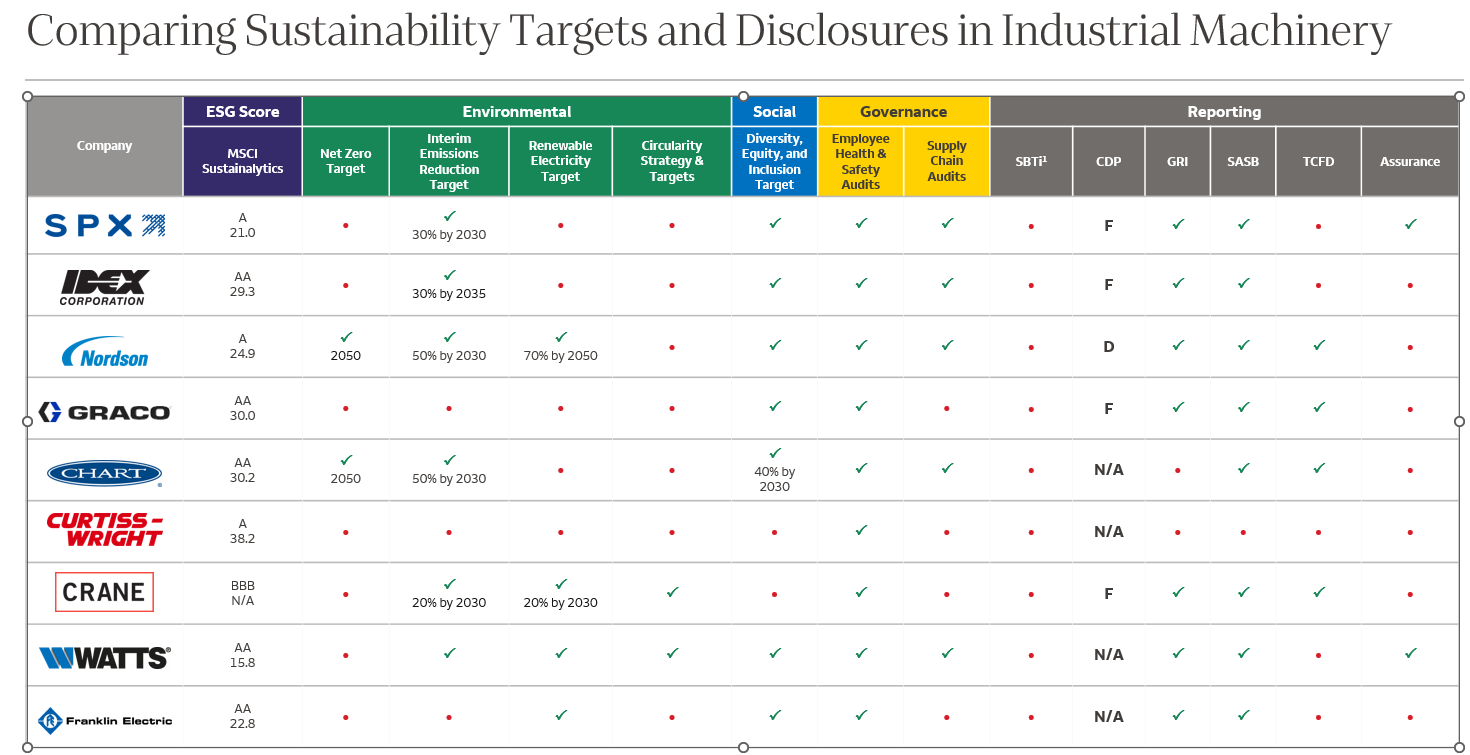
## Automated Test Suite and Coverage Report

Test cases for various edge cases should be automated. Application needs to be tested as per the test cases and the results need to be captured along with the coverage report and any excel report and graphs.

## Required Results

Please refer the respect and response schema along with same response for more details.

## Comparison Report



For example for company named “Nordson”, following details helps you to identify the response attributes.



## Comprehensive & Brief information

This information will contain following information

Above screenshot helps you to understand ESG indicators for various companies.

## Prediction and Forecasting

Report in #2 should also contain sustainability goals and the amount of time required to reach those goals. This implies the need for a predictive model that takes into account historical performance data, current sustainability metrics, and potentially external factors influencing sustainability efforts.

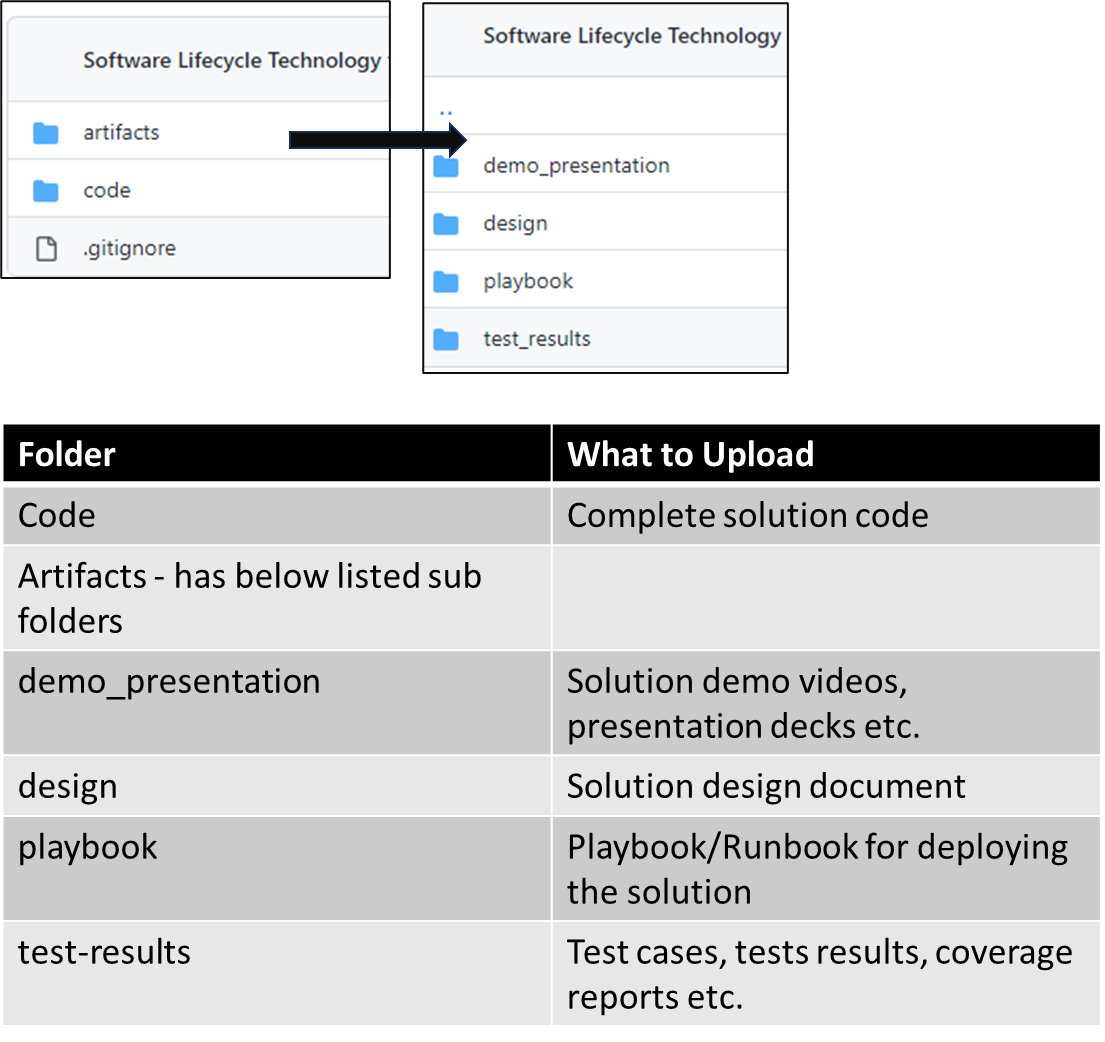
# FAQs:

Please note: We recommend you to refer to the common FAQs provided for the main challenge.

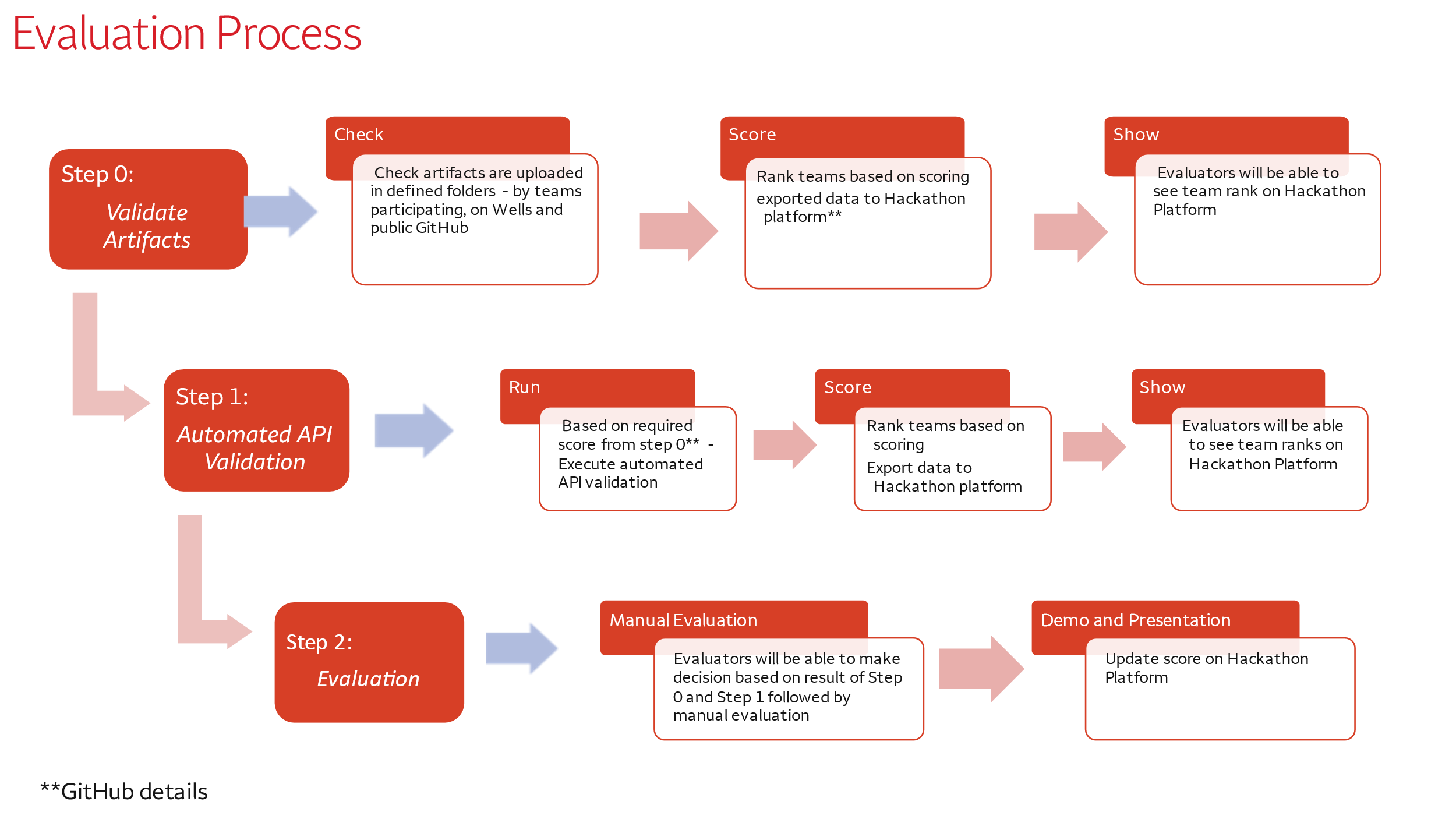
#### Artifacts to be uploaded

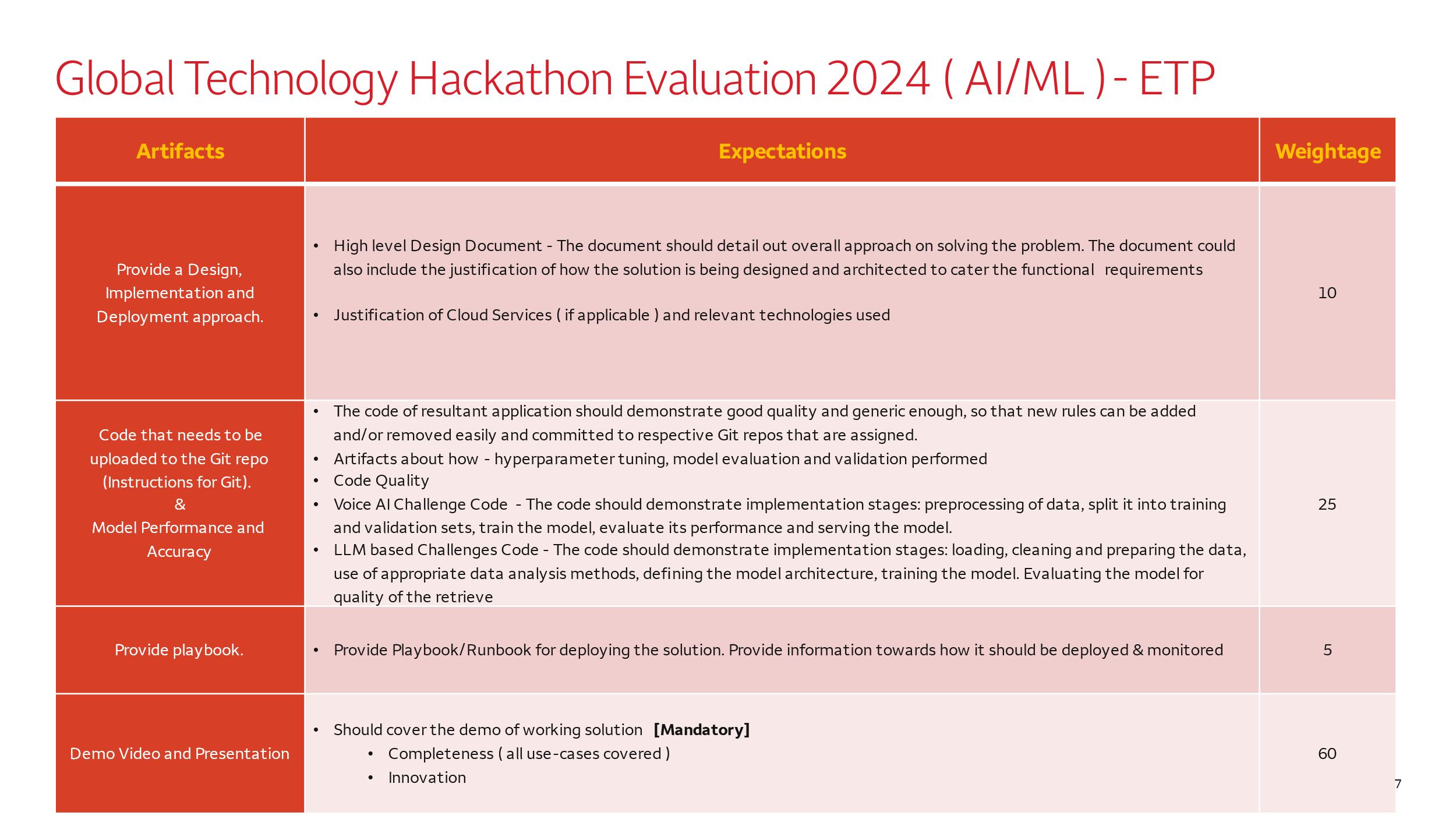
Below mentioned artifacts and code to be uploaded to Git repository following the instructions below. **It is mandatory to adhere to these instructions to be eligible for evaluation**.

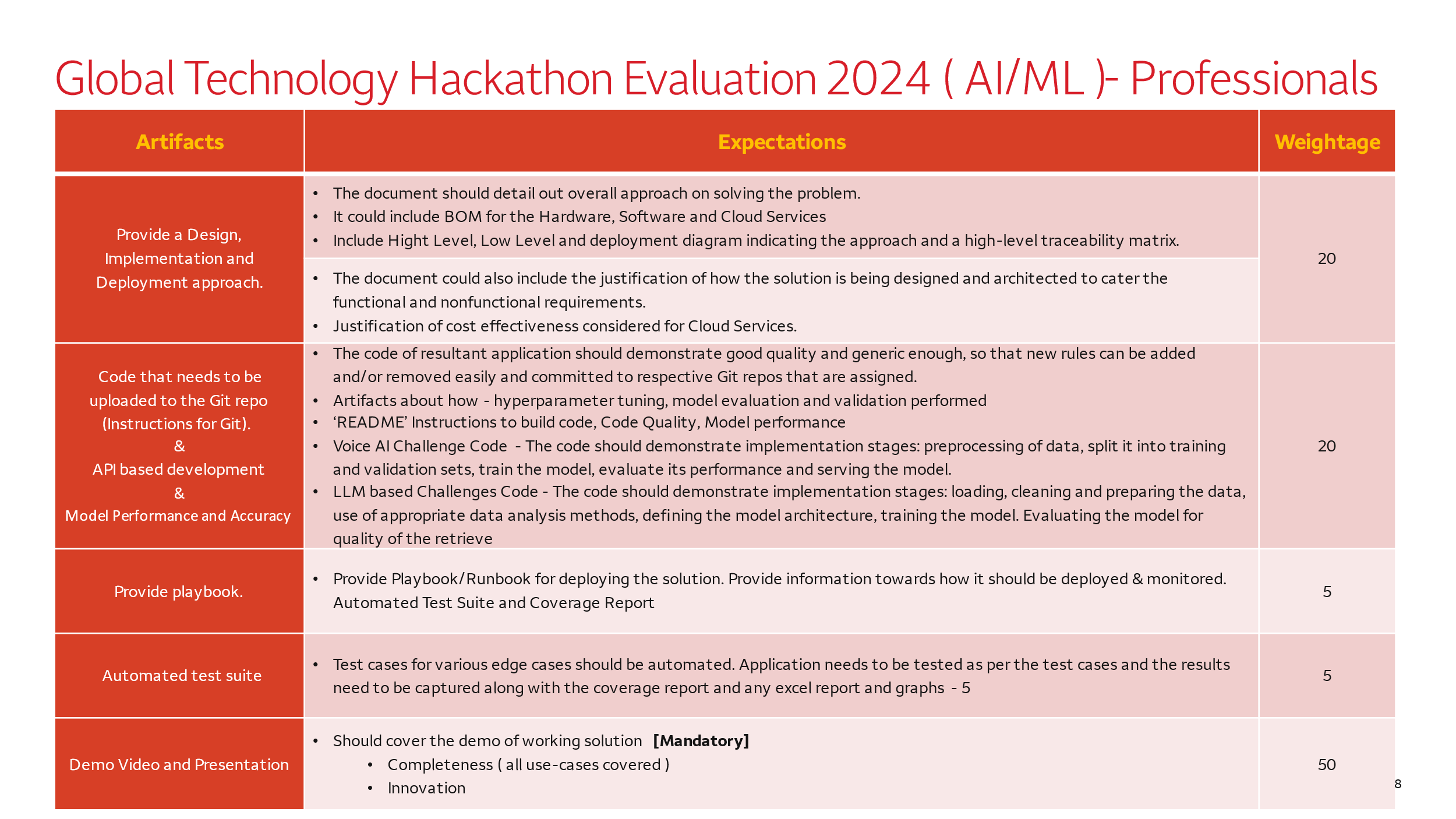
* Git repo has the following folder structure.
  + **Artifacts** 
    - **demo\_presentation**
    - **design**
    - **playbook**
    - **test\_results**
  + **Code**
* **Git folder structure**



# Evaluation Approach:



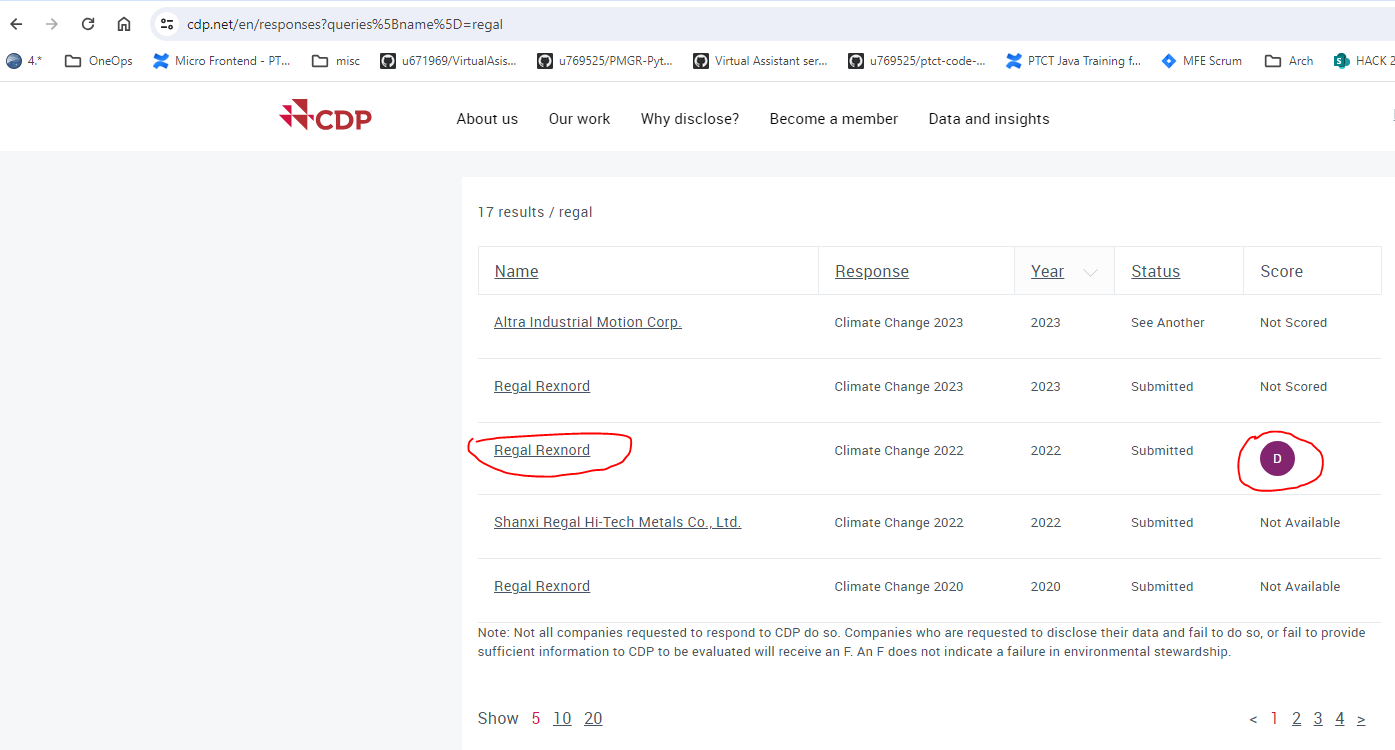




# How to retrieve the reporting information

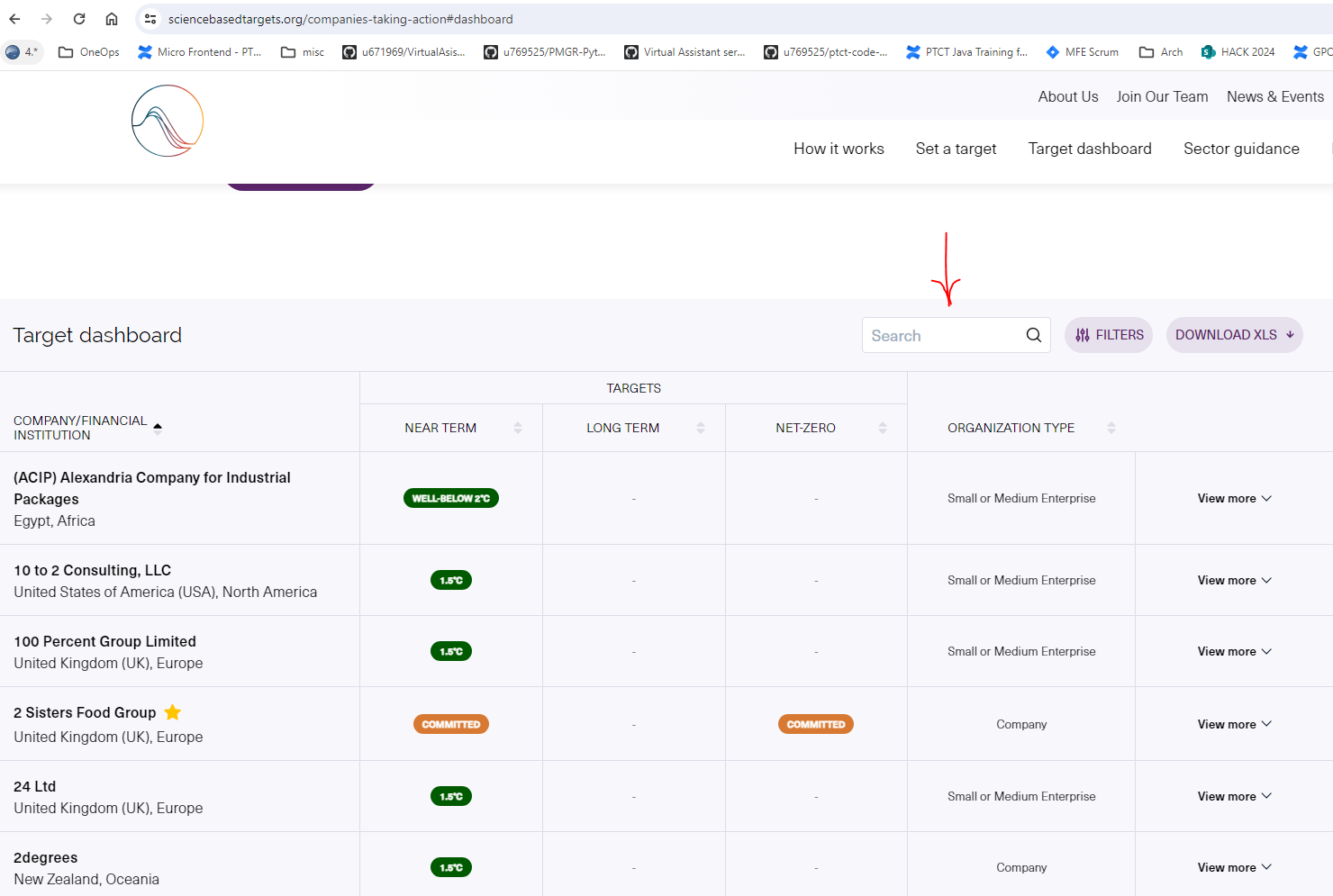
## CDP

URL : <https://www.cdp.net/en/responses?queries%5Bname%5D=graco>



## SBTi

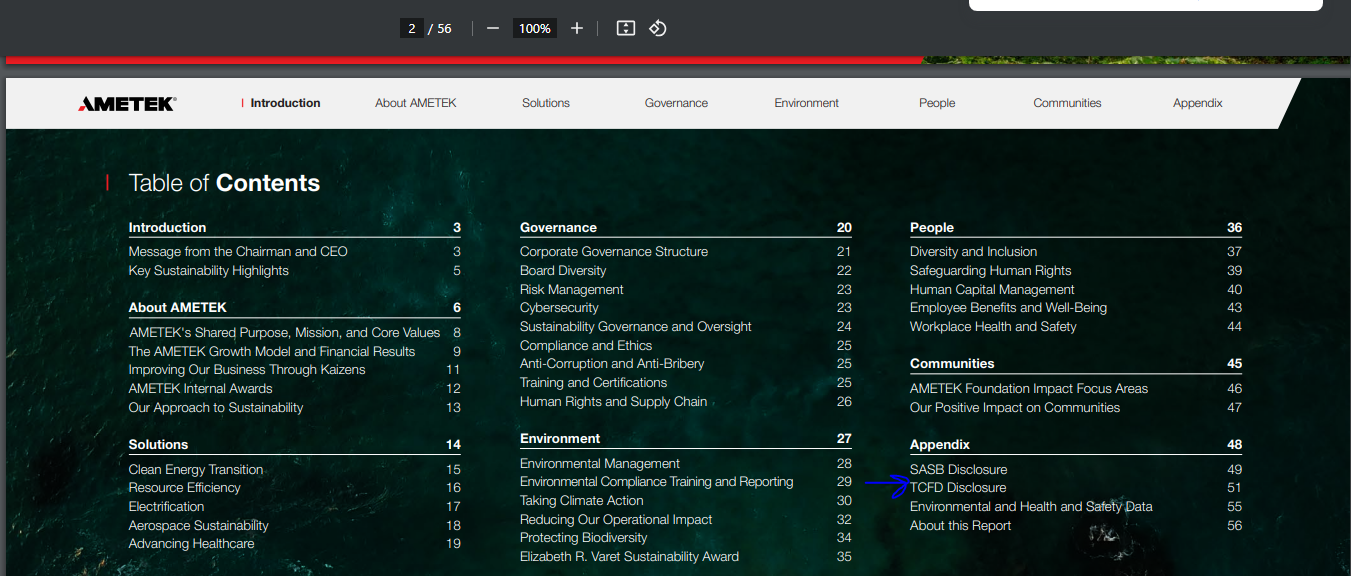
URL : <https://sciencebasedtargets.org/companies-taking-action#dashboard>

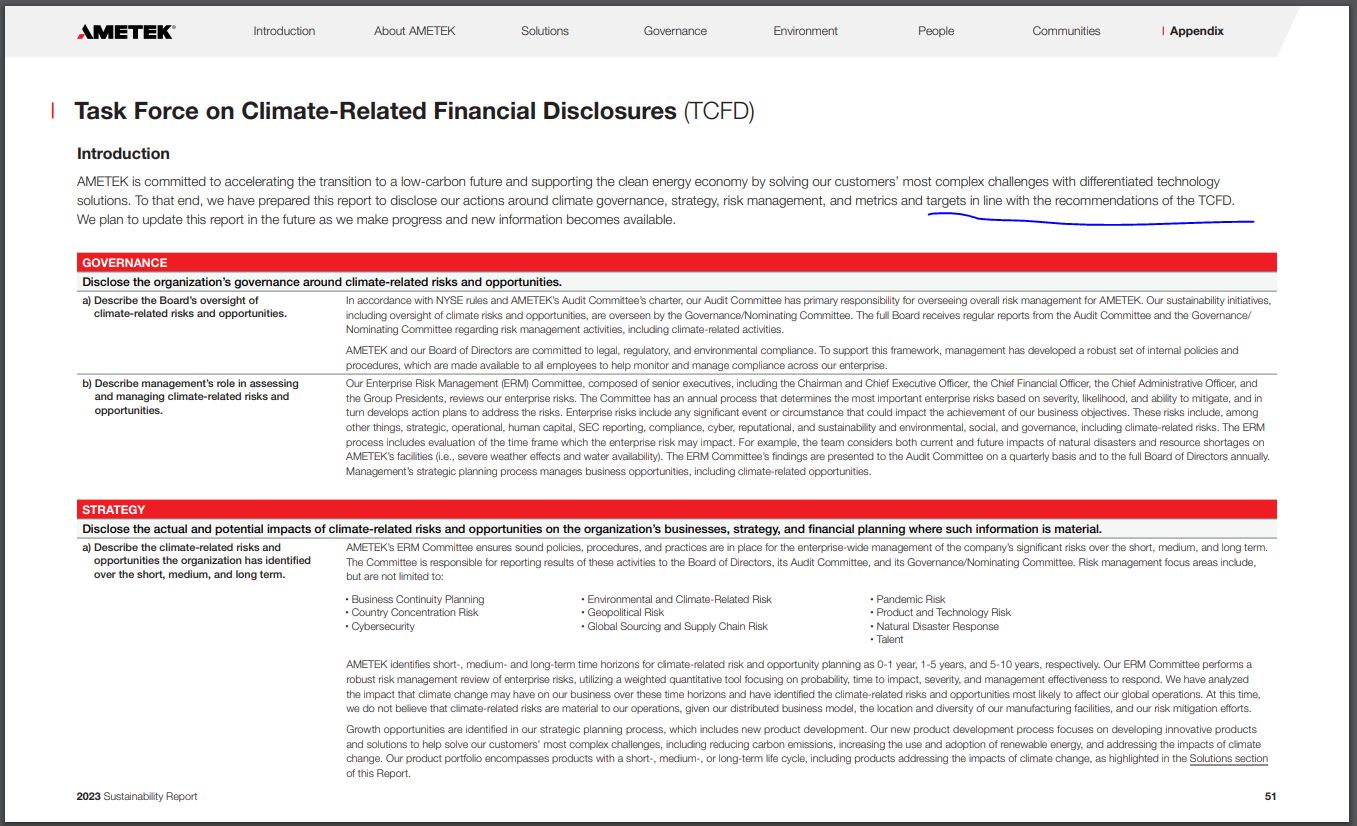


## TCFD

Navigate to company website or search <<company name>> TCFD

URL : <https://www.ametek.com/-/media/project/oneweb/oneweb/ametek/sustainability/documents/ametek_sustainability_report-2023.pdf?la=en&revision=b9ad4a8a-b02c-4828-b3d8-3c8cc68956b9&hash=5578EA881621C02B615942B893918A8E> (page 51)





# Reports

* https://www.spx.com/wp-content/uploads/2023/09/SPX\_SustainabilityReport\_2022\_FINAL.pdf
* https://www.idexcorp.com/wp-content/uploads/2023/05/IDEX\_SustainabilityReport2022\_accessible\_version.pdf
* https://nc-p-001.sitecorecontenthub.cloud/api/public/content/0b4bfc20c6934b9da852c6e38976f06f?v=a16d1c0d
* https://www.graco.com/content/dam/graco/corporate/literature/misc/Graco-ESG-Report.pdf
* https://files.chartindustries.com/ESG2022.pdf
* https://curtisswright.com/company/sustainability/default.aspx
* https://www.craneco.com/wp-content/uploads/2023/03/Crane-Co\_PSE\_2022\_2.27.23\_FINAL.pdf
* https://dam.watts.com/AssetLink/630570.pdf
* https://fele.widen.net/s/9xsgtltcfp/m1952\_sustainability\_report\_2023

**GIT HUB Link:**

[https://github.com/Hackathon2024-March/challenge-assets/tree/main/sustainability-benchmarking](https://urldefense.com/v3/__https:/github.com/Hackathon2024-March/challenge-assets/tree/main/sustainability-benchmarking__;!!F9svGWnIaVPGSwU!pDomD1i3T95RuU9UWbWFjEve2FmzsAIHnjxVODTg2O7D9NKnZwCMFF87rbc9d3kgWJMQdO3ap7KA1i5w7mZTZoSTbhx3$)