

# 永續數據分析

## (Sustainability and ESG Data Analytics)

# 永續數據分析概論

## (Introduction to Sustainability and ESG Data Analytics)

1122ESGDA01

DM4, NTPU (N4084) (Spring 2024)

Fri, 10, 11, 12 (18:30-21:15) (臺北大學民生校區 305)

## 戴敏育

Min-Yuh Day, Ph.D,  
Associate Professor

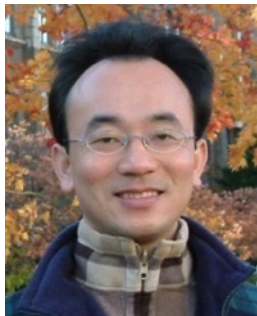
Institute of Information Management, National Taipei University

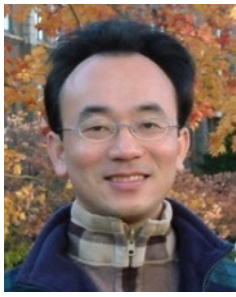
<https://web.ntpu.edu.tw/~myday>

2024-02-23



<https://meet.google.com/miy-fbif-max>





# 戴敏育

## Min-Yuh Day, Ph.D.



2020 Cohort

**Associate Professor, Information Management, NTPU**

**Visiting Scholar, IIS, Academia Sinica**

**Ph.D., Information Management, NTU**

**Director, Intelligent Financial Innovation Technology, IFIT Lab, IM, NTPU**

**Associate Director, Fintech and Green Finance Center, NTPU**

**Artificial Intelligence, Financial Technology, Big Data Analytics,  
Data Mining and Text Mining, Electronic Commerce**



2020 Cohort



Accredited Educator



Solutions Architect Associate



Cloud Practitioner



# Course Syllabus

## National Taipei University

### Academic Year 112, 2<sup>nd</sup> Semester (Spring 2024)

- **Course Title:** 永續數據分析概論  
(Introduction to Sustainability and ESG Data Analytics)
- **Instructor:** 戴敏育 (Min-Yuh Day)
- **Course Class:** DM4, NTPU (3 Credits, Elective)
- **Details**
  - **USR Course @ Digital Marketing**  
(3 Credits, Elective, One Semester) (N4084)
- **Time & Place:** Fri, 10, 11, 12 (18:30-21:15) (民生校區305)
- **Google Meet:** <https://meet.google.com/miy-fbif-max>



<https://meet.google.com/miy-fbif-max>



# 教學目標 (Course Objectives)

1. 瞭解**永續數據分析**基本概念。
2. 具備**永續數據分析**實務操作能力。
3. 整合**大數據分析**的**創新思維**，提升**永續發展**運作模式。
4. 在**永續**面向上，憑藉**數據分析**，研擬**永續議**題相關因應作為，並培養學生具有從資料中挖掘出具有管理意涵的**數據分析**基礎能力

# Course Objectives

1. Understand the **fundamental concepts of sustainability and ESG data analytics**.
2. Equip with **Hands-on practices of sustainability and ESG data analytics**.
3. Integrate **innovative thinking of big data analysis** to enhance the operational model of **sustainable development**.
4. In the context of **sustainability**, use **data analysis** to formulate responses to sustainable issues and cultivate students' ability to extract management-relevant data analysis skills from the data.

# 課程包含永續發展(SDGs)目標

1. **SDG4 | 優質教育 (Quality Education)**
2. **SDG7 | 可負擔的潔淨能源 (Affordable and Clean Energy)**
3. **SDG8 | 尊嚴就業與經濟發展 (Decent Work and Economic Growth)**
4. **SDG9 | 產業創新與基礎設施 (Industry, Innovation and Infrastructure)**
5. **SDG11 | 永續城市與社區 (Sustainable Cities and Communities)**
6. **SDG12 | 負責任的消費與生產 (Responsible Consumption and Production)**
7. **SDG13 | 氣候行動 (Climate Action)**
8. **SDG17 | 夥伴關係 (Partnerships for the Goals)**

# 課程內容綱要 (Course Outline)

- 本課程介紹**永續數據分析基本概念與實務操作**。
- 課程內容包括
  1. 永續數據分析概論
  2. 環境、社會與治理 (ESG) 淨零數位轉型
  3. 永續與ESG 資料科學
  4. Web 3.0 和大數據分析在金融科技、綠色永續金融
  5. TCFD 氣候相關財務揭露與En-ROADS 氣候變遷模擬
  6. ESG數據的收集、分析和視覺化
  7. ESG數據報告、企業永續報告書
  8. ESG數據驗證
  9. 能源之星報告與數據揭露
  10. 人工智慧物聯網在ESG永續應用
  11. 生成性AI於永續評等和報告生成
  12. 永續數據分析個案研究

# Course Outline

- This course introduces the **fundamental concepts** and **hands-on practices** of **Sustainability and ESG Data Analytics**.
- Topics include
  1. Introduction Sustainability and ESG Data Analytics
  2. Environmental, Social, and Governance (ESG) in Net-Zero Digital Transformation
  3. Data Science for Sustainability and ESG
  4. Web 3.0 and Big Data Analysis in Fintech, Green Finance, Sustainable Finance
  5. Task Force on Climate-Related Financial Disclosures (TCFD) and En-Roads Interactive
  6. ESG Data Gathering, Analysis, and Visualization
  7. ESG Data Reporting, Corporate Sustainability Reports
  8. ESG Data Verification
  9. Energy Star Reporting and Data Disclosure
  10. Artificial Intelligence of things (AIoT) in ESG and Sustainability Applications
  11. Generative AI for ESG Rating and Reporting Generation
  12. Case Study on Sustainability and ESG Data Analytics



# 課程大綱 (Syllabus)

- | 週次 (Week) | 日期 (Date)  | 內容 (Subject/Topics)  |
|-----------|------------|--|
| 1         | 2024/02/23 | <b>永續數據分析概論</b><br>(Introduction Sustainability and ESG Data Analytics)  |
| 2         | 2024/03/01 | <b>環境、社會與治理 (ESG) 淨零數位轉型</b><br>(Environmental, Social, and Governance (ESG) in Net-Zero Digital Transformation) |
| 3         | 2024/03/08 | <b>永續與ESG 資料科學</b><br>(Data Science for Sustainability and ESG)  |
| 4         | 2024/03/15 | <b>永續數據分析個案研究 I</b><br>(Case Study on Sustainability and ESG Data Analytics I)                                   |
| 5         | 2024/03/22 | <b>Web 3.0 和大數據分析在金融科技、綠色永續金融</b><br>(Web 3.0 and Big Data Analysis in Fintech, Green and Sustainable Finance)   |

# 課程大綱 (Syllabus)

週次 (Week)	日期 (Date)	內容 (Subject/Topics)
6	2024/03/29	TCFD 氣候相關財務揭露與En-ROADS 氣候變遷模擬 (Task Force on Climate-Related Financial Disclosures (TCFD) and En-Roads Interactive)
7	2024/04/05	放假 (No Classes)
8	2024/04/12	期中報告 (Midterm Project Report)
9	2024/04/19	ESG數據的收集、分析和視覺化 (ESG Data Gathering, Analysis, and Visualization)
10	2024/04/26	ESG數據報告 (ESG Data Reporting); 企業永續報告書 (Corporate Sustainability Reports)

# 課程大綱 (Syllabus)

- | 週次 (Week) | 日期 (Date)  | 內容 (Subject/Topics)  |
|-----------|------------|--|
| 11        | 2024/05/03 | ESG數據驗證 (ESG Data Verification)  |
| 12        | 2024/05/10 | 永續數據分析個案研究 II<br>(Case Study on Sustainability and ESG Data Analytics II)                            |
| 13        | 2024/05/17 | 能源之星報告與數據揭露<br>(Energy Star Reporting and Data Disclosure)   |
| 14        | 2024/05/24 | 人工智慧物聯網在ESG永續應用<br>(Artificial Intelligence of things (AIoT) in ESG and Sustainability Applications) |
| 15        | 2024/05/31 | 生成式AI於永續評等和報告生成<br>(Generative AI for ESG Rating and Reporting Generation)                           |
| 16        | 2024/06/07 | 期末報告 (Final Project Report)  |

# 課程與教學特色

1. 結合 Web3.0介紹大數據分析基本概念、研究議題與實務操作。
2. 提供數據整合溝通企劃的理論與實務概念及工具。
3. 應用於分析各領域的資料，並透過資料視覺化的方式呈現分析結果。

## 預期社會影響：

1. 由數據分析學習，培養面對永續議題與風險時，因應作為分析能力。
2. 培養具備大數據分析基本概念、研究議題、實務操作以及永續數據分析實作能力的人才。

# Course and Teaching Features

1. **Combine Web3.0 to introduce basic concepts of big data analysis, research topics, and practical operations.**
2. **Provide theories and tools for data integration and communication planning.**
3. **Apply to analyze data from various domains and present analysis results through data visualization.**

## **Expected Social Impact:**

1. **Learn from data analysis, cultivating the ability to analyze responses when facing sustainable issues and risks.**
2. **Train talents who possess basic concepts of big data analysis, research topics, practical operations, and practical abilities in sustainable data analysis.**

# 教學策略創新

1. **USR（大學社會責任） 在地連結、合作**
2. **小組學習**
3. **問題取向教學**
4. **主題教學**
5. **學思達教學**

# 校四大基本素養

## (Four Fundamental Qualities)

- 1. 專業 (Professionalism)**
- 2. 人際 (Interpersonal Relationship)**
- 3. 倫理 (Ethics)**
- 4. 國際觀 (International Vision)**

# 商學院學習目標

## (College Learning Goals)

- 1. Communication**
- 2. Analytical and Critical Thinking**
- 3. Ethics/Corporate Social Responsibility**
- 4. Global Knowledge/Awareness**



# 數位行銷進修學士學位學程 系核心能力

- 1. 具有判斷商業問題並有系統地解決這些問題 40 %**
- 2. 能展現商業教育所需的基本技能 40 %**
- 3. 具有溝通能力 10 %**
- 4. 能意識道德問題及其影響 5 %**
- 5. 透過參與相關活動了解國際議題 5 %**

# 教學方法與教學活動

## (Teaching Methods and Activities)

- **講授 (Lecture)**
- **討論 (Discussion)**
- **實習 (Practicum)**

# 評量方式 (Evaluation Methods)

- **個人報告 (Individual Presentation) 30 %**
- **團體報告 (Group Presentation) 30 %**
- **個案分析報告 (Case Report) 20 %**
- **課堂參與 (Class Participation) 10 %**
- **作業 (Assignment) 10 %**

# 指定用書 (Required Texts)

- **Cino Robin Castelli, Cyril Shmatov (2022),  
Quantitative Methods for ESG Finance, Wiley**

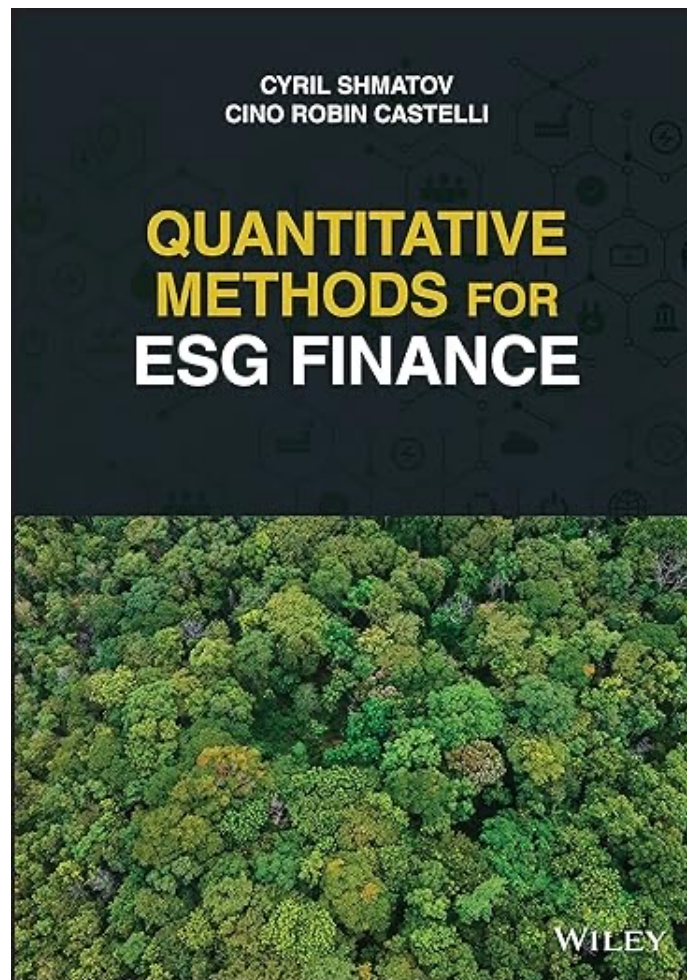
# 參考書目 (Reference Books)

1. **Simon Thompson (2023), Green and Sustainable Finance: Principles and Practice in Banking, Investment and Insurance, 2nd Edition, Kogan Page.**
2. **Chrissa Pagitsas (2023), Chief Sustainability Officers At Work: How CSOs Build Successful Sustainability and ESG Strategies, Apress.**
3. **Hariom Tatsat, Sahil Puri, Brad Lookabaugh (2020), Machine Learning and Data Science Blueprints for Finance: From Building Trading Strategies to Robo-Advisors Using Python, O'Reilly Media**
4. **Aurélien Géron (2022), Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems, 3rd Edition, O'Reilly Media.**
5. **Chris Kelliher (2022), Quantitative Finance With Python: A Practical Guide to Investment Management, Trading, and Financial Engineering, Chapman and Hall/CRC.**
6. **Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.**
7. **Abdullah Karasan (2021), Machine Learning for Financial Risk Management with Python: Algorithms for Modeling Risk, O'Reilly Media.**
8. **Yves Hilpisch (2018), Python for Finance: Mastering Data-Driven Finance, 2nd Edition, O'Reilly Media.**

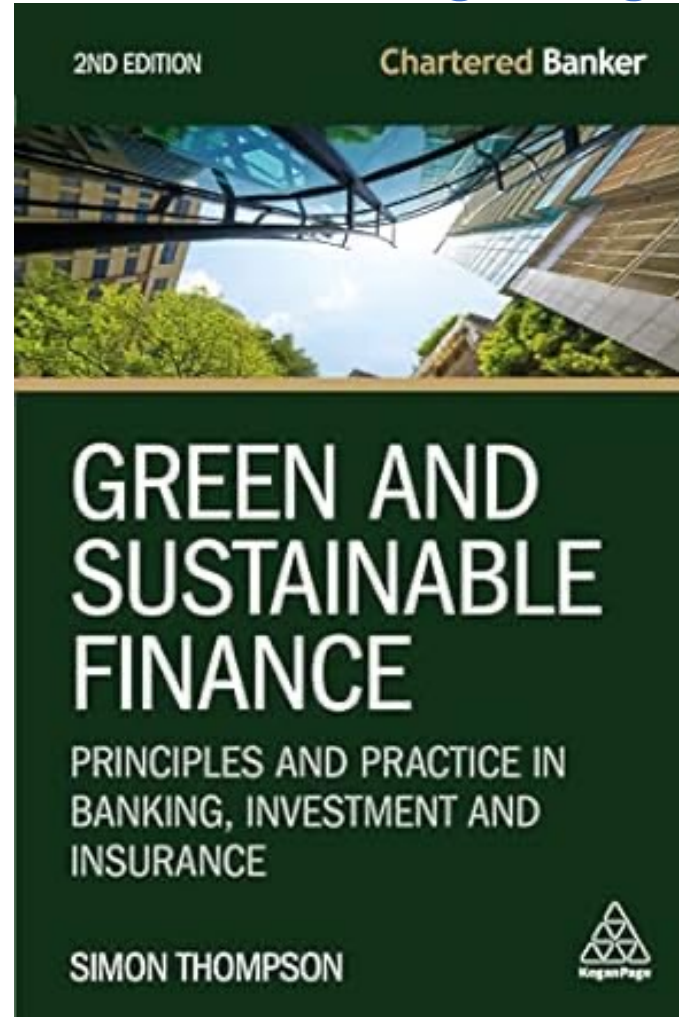
# 其他參考資料 (Other References)

1. GRI (Global Report Initiative):  
<https://www.globalreporting.org/>
2. CDP (Carbon Disclosure Project):  
<https://www.cdp.net/>
3. SASB (Sustainability Accounting Standards Board):  
<https://sasb.org/>
4. ISSB (International Sustainability Standards Board):  
<https://www.ifrs.org/groups/international-sustainability-standards-board/>
5. TCFD (Task Force on Climate-related Financial Disclosures):  
<https://www.fsb-tcfd.org/>
6. Research Papers

Cino Robin Castelli, Cyril Shmatov (2022),  
**Quantitative Methods for ESG Finance,**  
Wiley

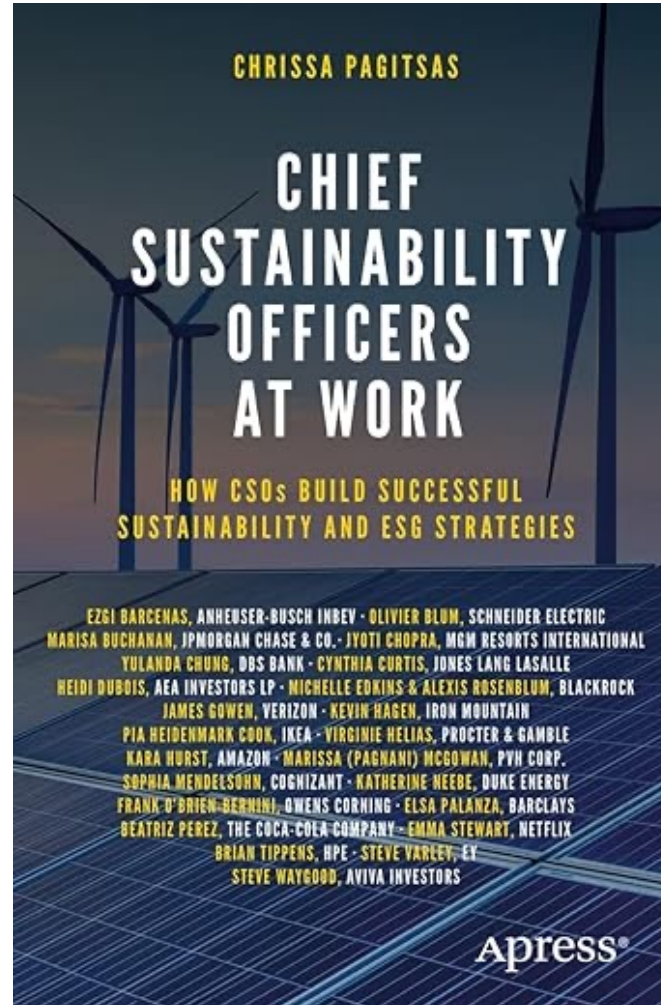


Simon Thompson (2023),  
**Green and Sustainable Finance:**  
**Principles and Practice in Banking, Investment and Insurance,**  
2nd Edition, Kogan Page.

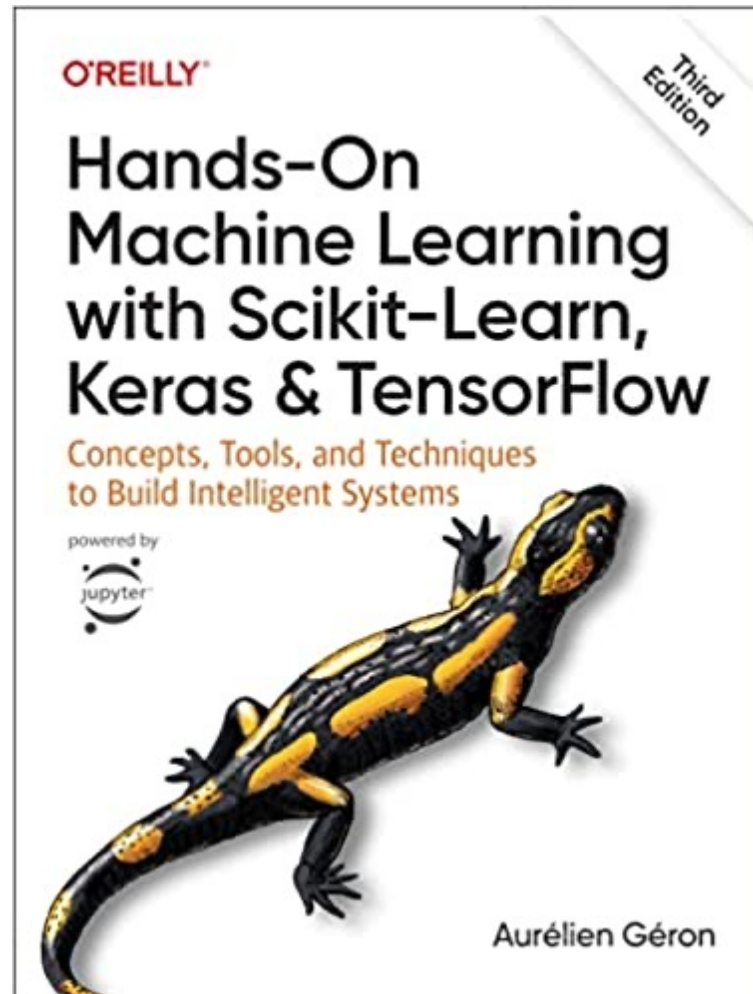




Chrissa Pagitsas (2023),  
**Chief Sustainability Officers At Work:  
How CSOs Build Successful Sustainability and ESG Strategies,**  
Apress.



**Aurélien Géron (2022),**  
**Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow:**  
**Concepts, Tools, and Techniques to Build Intelligent Systems,**  
**3rd Edition, O'Reilly Media.**



# GRI (Global Report Initiative)



Standards ▾

How to use the GRI Standards ▾

Reporting support ▾

Public policy & partnerships ▾

About GRI ▾

News ▾

Goals and targets database

Sign In

Search 🔍

Donate Now



## The global leader for impact reporting

Welcome to GRI. For over 25 years, we have developed and delivered the global best practice for how organizations communicate and demonstrate accountability for their impacts on the environment, economy and people.

We provide the world's most widely used sustainability reporting standards, which cover topics that range from biodiversity to tax, waste to emissions, diversity and equality to health and safety. As such, GRI reporting is the enabler for transparency and dialogue between companies and their stakeholders.

[Access the GRI Standards →](#)

Feedback

# CDP (Carbon Disclosure Project)



[Guidance & questionnaires](#) [Contact](#) [Regional websites](#) ▾ [Language](#) ▾

[About us](#) [Our work](#) [Why disclose?](#) [Become a member](#) [Data and insights](#)



We focus investors, companies, cities and governments on building a sustainable economy by measuring and acting on their environmental impact.

Jump to...

- Highlights
- News and events
- Where we work
- Search responses

CDP is a not-for-profit charity that runs the global disclosure system for [investors](#), [companies](#), [cities](#), [states and regions](#) to manage their environmental impacts. Over the past 20 years we have created a system that has resulted in unparalleled engagement on environmental issues worldwide. Find out more about [how we work](#).

<https://www.cdp.net/>

# SASB (Sustainability Accounting Standards Board)

IFRS Foundation

Other Resources: [The ISSB](#) [Integrated Reporting Framework](#)



[✉ Subscribe](#) [↓ Download Standards](#)

[About](#) [SASB Standards](#) [Using the SASB Standards](#) [Pathway to ISSB](#) [Education](#) [Membership](#)

An aerial photograph showing a winding river through a lush green landscape. The river flows from the top left towards the bottom center, surrounded by vibrant green fields and a dense forest of tall trees on the right side. The lighting suggests a bright, sunny day.

**SASB Standards: Your  
pathway to ISSB**

[Learn more](#)

<https://sasb.org/>

# ISSB (International Sustainability Standards Board)



ABOUT US | IFRS ACCOUNTING | IFRS SUSTAINABILITY

Home > International Sustainability Standards Board

## International Sustainability Standards Board

ABOUT

MEMBERS

MEETINGS

RESOURCES

NEWS

### About the International Sustainability Standards Board

The Trustees of the IFRS Foundation announced the formation of the International Sustainability Standards Board (ISSB) on 3 November 2021 at COP26 in Glasgow, following strong market demand for its establishment. The ISSB is developing—in the public interest—standards that will result in a high-quality, comprehensive global baseline of sustainability disclosures focused on the needs of investors and the financial markets.

Sustainability factors are becoming a mainstream part of investment decision-making. There are increasing calls for companies to provide high-quality, globally comparable information on sustainability-related risks and opportunities, as indicated by feedback from many consultations with market

#### Related information

[Sustainability FAQs](#)

[General Sustainability-related Disclosures project](#)

[Climate-related Disclosures project](#)

[Consolidated organisations](#)

<https://www.ifrs.org/groups/international-sustainability-standards-board/>

# TCFD

## (Task Force on Climate-related Financial Disclosures)



<https://www.ifrs.org/sustainability/tcf/>



ABOUT US | IFRS ACCOUNTING | IFRS SUSTAINABILITY

Home > ISSB and TCFD

### ISSB and TCFD

The Financial Stability Board has announced that the work of the TCFD has been completed, with the ISSB's Standards marking the '**culmination of the work of the TCFD**'.

Companies applying IFRS S1 *General Requirements for Disclosure of Sustainability-related Financial Information* and IFRS S2 *Climate-related Disclosures* will meet the TCFD recommendations as the recommendations are fully incorporated into the ISSB's Standards.

Companies can continue to use the **TCFD recommendations** should they choose to do so, and some companies may still be required to use the TCFD recommendations. Using the recommendations is a good entry point for companies as they move to use the ISSB's Standards.

The IFRS Foundation has **published a comparison** of the requirements in IFRS S2 and the TCFD recommendations.

#### Related Information

[IFRS Foundation welcomes culmination of TCFD work and transfer of TCFD monitoring responsibilities to ISSB from 2024](#)

[Comparison: IFRS S2 Climate-related Disclosures with the TCFD Recommendations](#)

[Resource: Making the transition from TCFD to ISSB](#)

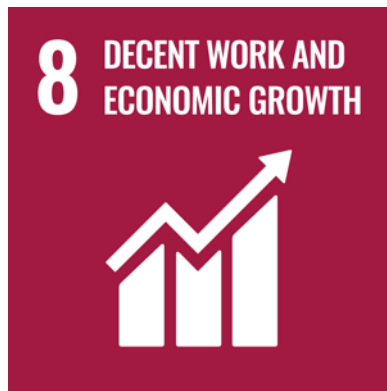
[IFRS Sustainability Standards Navigator](#)

<https://www.fsb-tcf.org/>

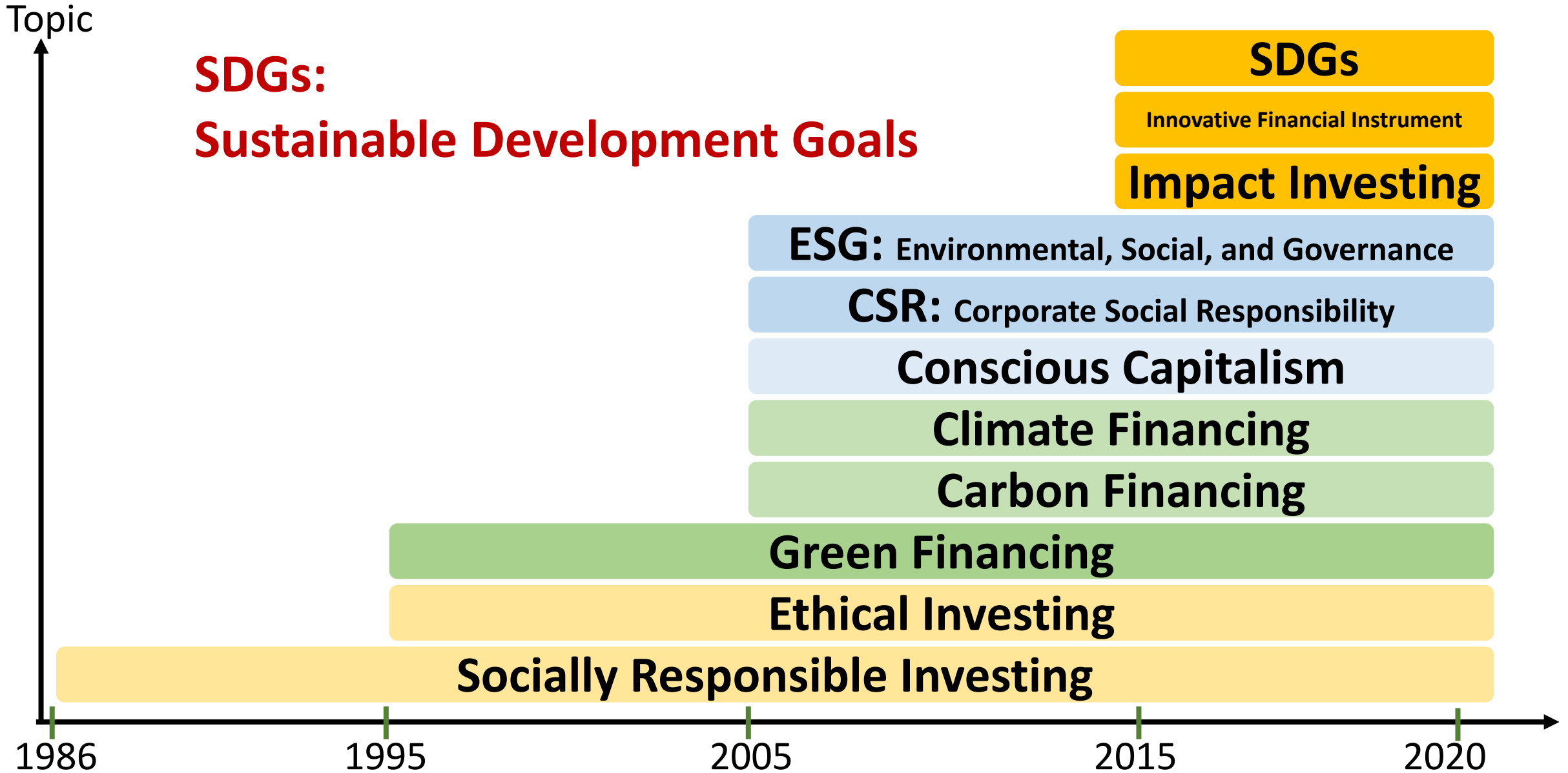
# **Sustainability and ESG Data Analytics**



# Sustainable Development Goals (SDGs)



# Evolution of Sustainable Finance Research



Source: Kumar, S., Sharma, D., Rao, S., Lim, W. M., & Mangla, S. K. (2022). Past, present, and future of sustainable finance: Insights from big data analytics through machine learning of scholarly research. *Annals of Operations Research*, 1-44.

# Sustainable Development Goals (SDGs) and 5P

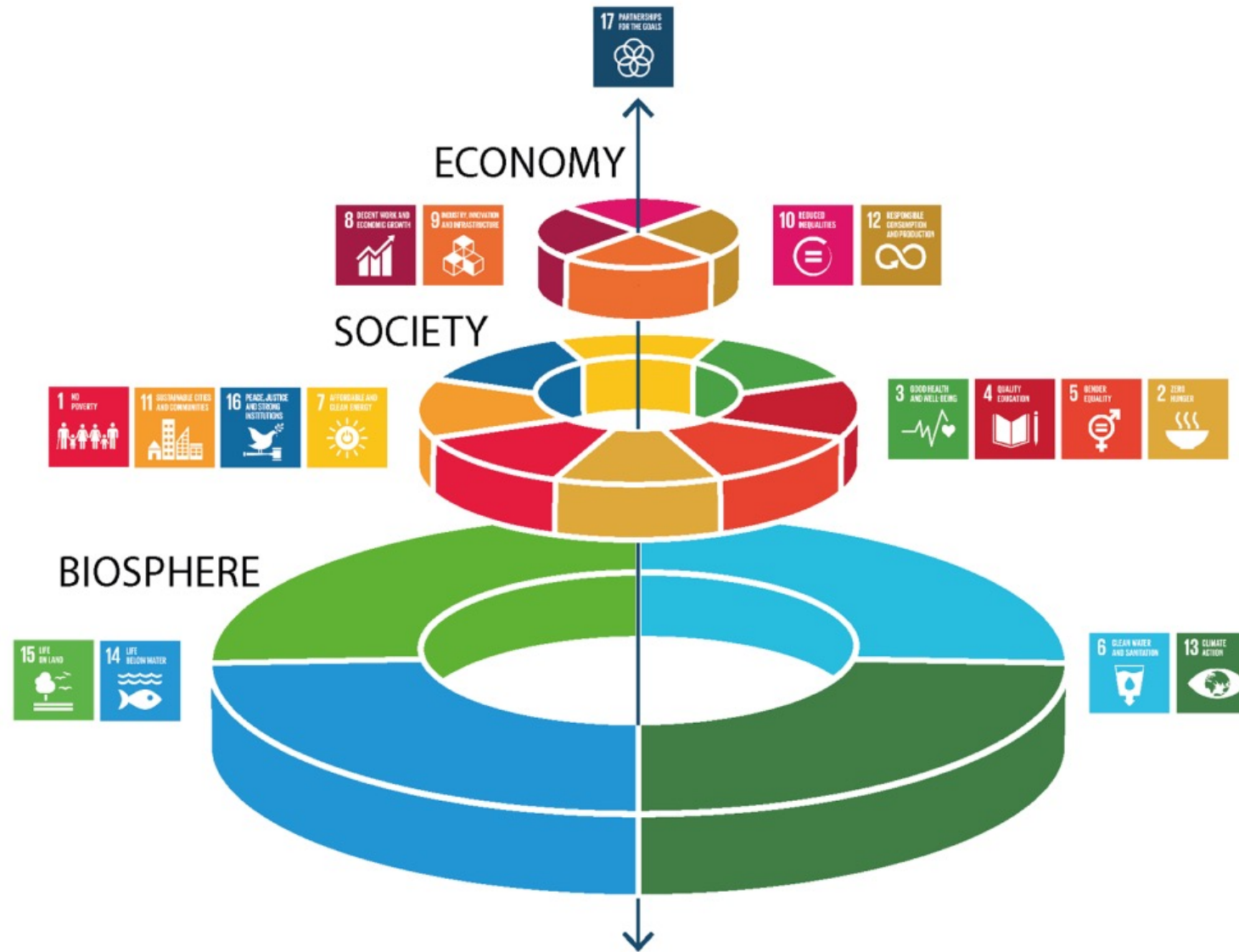
Partnership

Peace

Prosperity

People

Planet



# ESG to 17 SDGs

## ENVIRONMENT



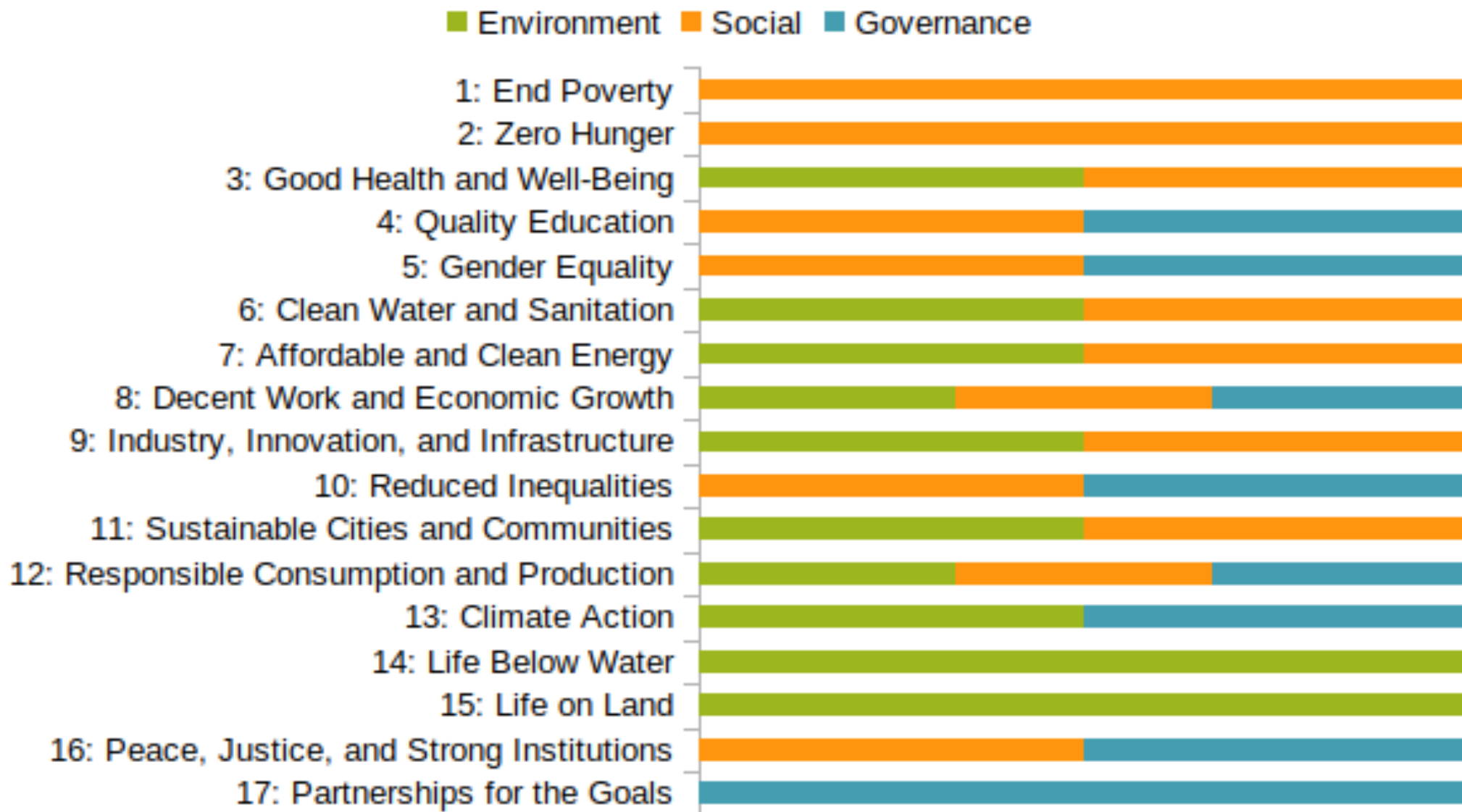
## SOCIAL



## GOVERNANCE

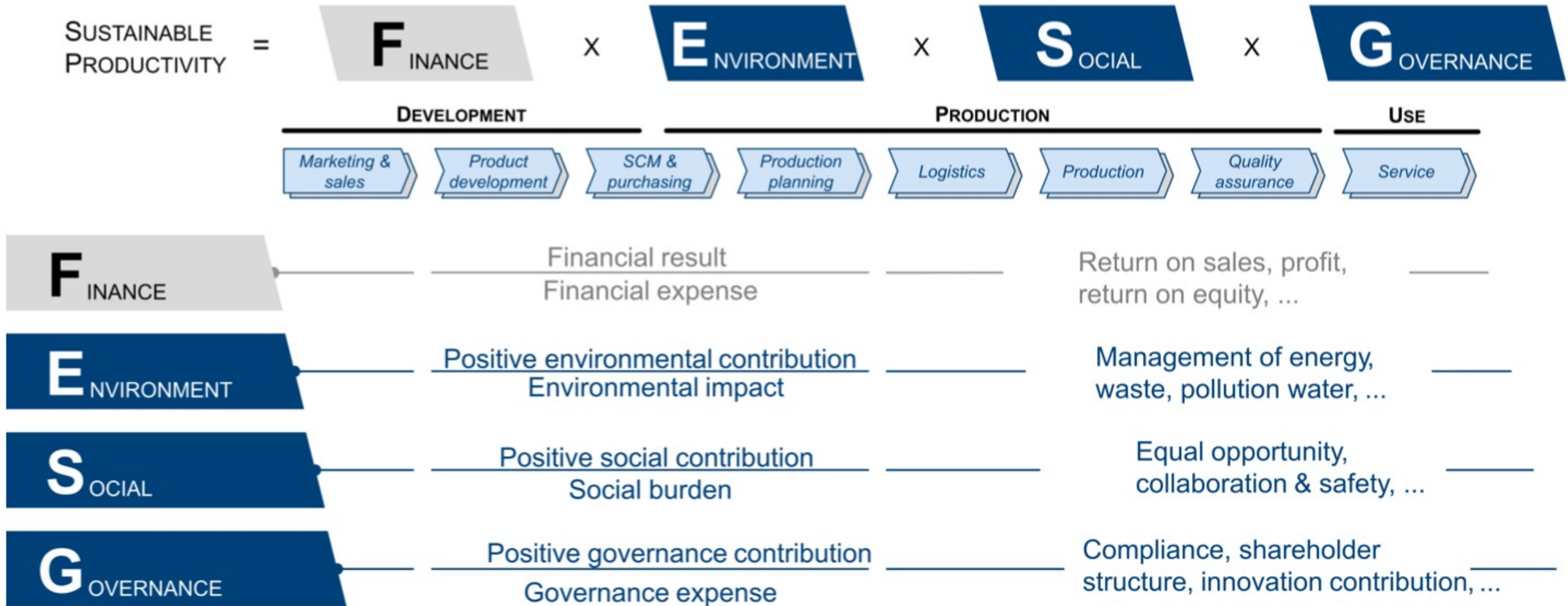


# ESG to 17 SDGs



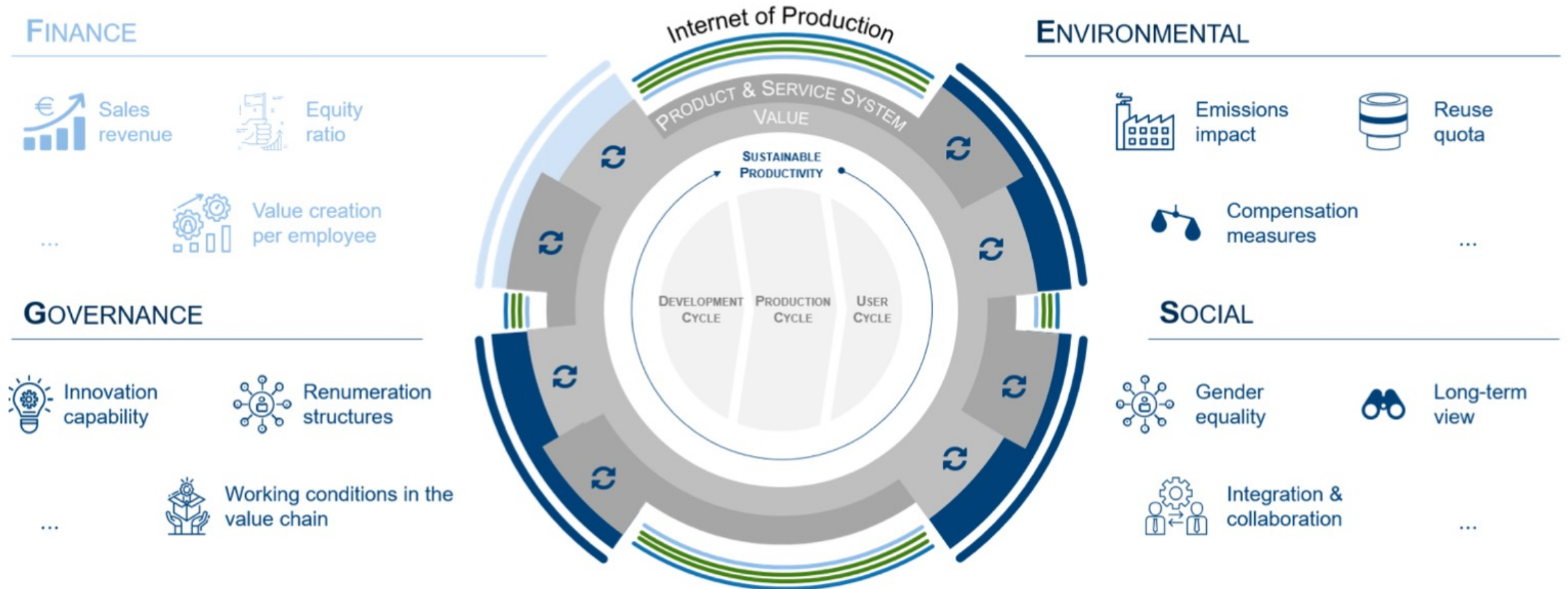
# **Generative AI for ESG Applications**

# Sustainable Productivity: Finance ESG



# Sustainable Resilient Manufacturing

## ESG

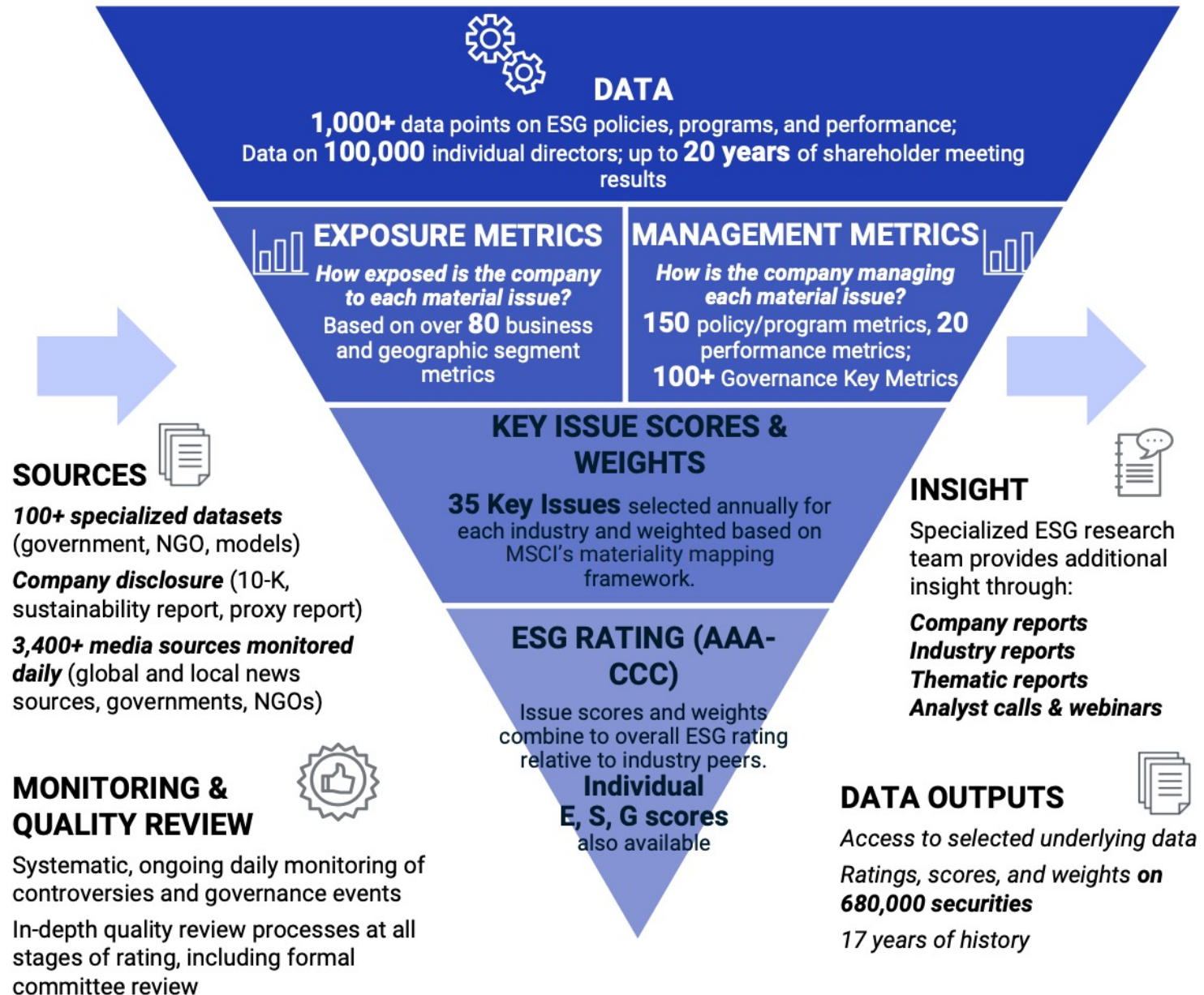




# ESG Indexes

- **MSCI ESG Index**
- **Dow Jones Sustainability Indices (DJSI)**
- **FTSE ESG Index**

# MSCI ESG Rating Framework

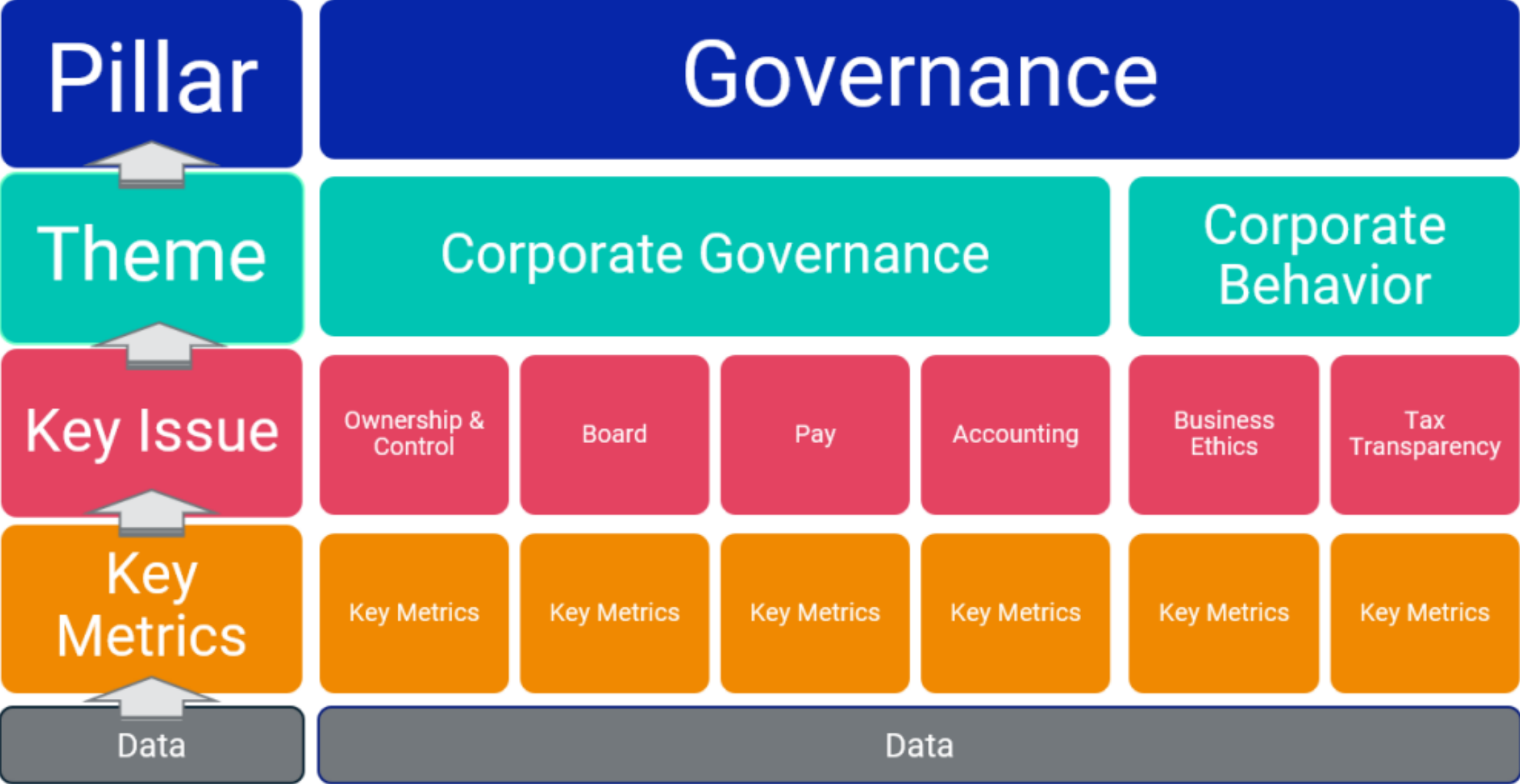


# MSCI ESG Key Issue Hierarchy

3 Pillars	10 Themes	35 ESG Key Issues	
<b>Environment</b>	<b>Climate Change</b>	Carbon Emissions Product Carbon Footprint	Financing Environmental Impact Climate Change Vulnerability
	<b>Natural Capital</b>	Water Stress Biodiversity & Land Use	Raw Material Sourcing
	<b>Pollution &amp; Waste</b>	Toxic Emissions & Waste Packaging Material & Waste	Electronic Waste
	<b>Environmental Opportunities</b>	Opportunities in Clean Tech Opportunities in Green Building	Opportunities in Renewable Energy
<b>Social</b>	<b>Human Capital</b>	Labor Management Health & Safety	Human Capital Development Supply Chain Labor Standards
	<b>Product Liability</b>	Product Safety & Quality Chemical Safety Consumer Financial Protection	Privacy & Data Security Responsible Investment Health & Demographic Risk
	<b>Stakeholder Opposition</b>	Controversial Sourcing Community Relations	
	<b>Social Opportunities</b>	Access to Communications Access to Finance	Access to Health Care Opportunities in Nutrition & Health
<b>Governance</b>	<b>Corporate Governance</b>	Ownership & Control Board	Pay Accounting
	<b>Corporate Behavior</b>	Business Ethics Tax Transparency	

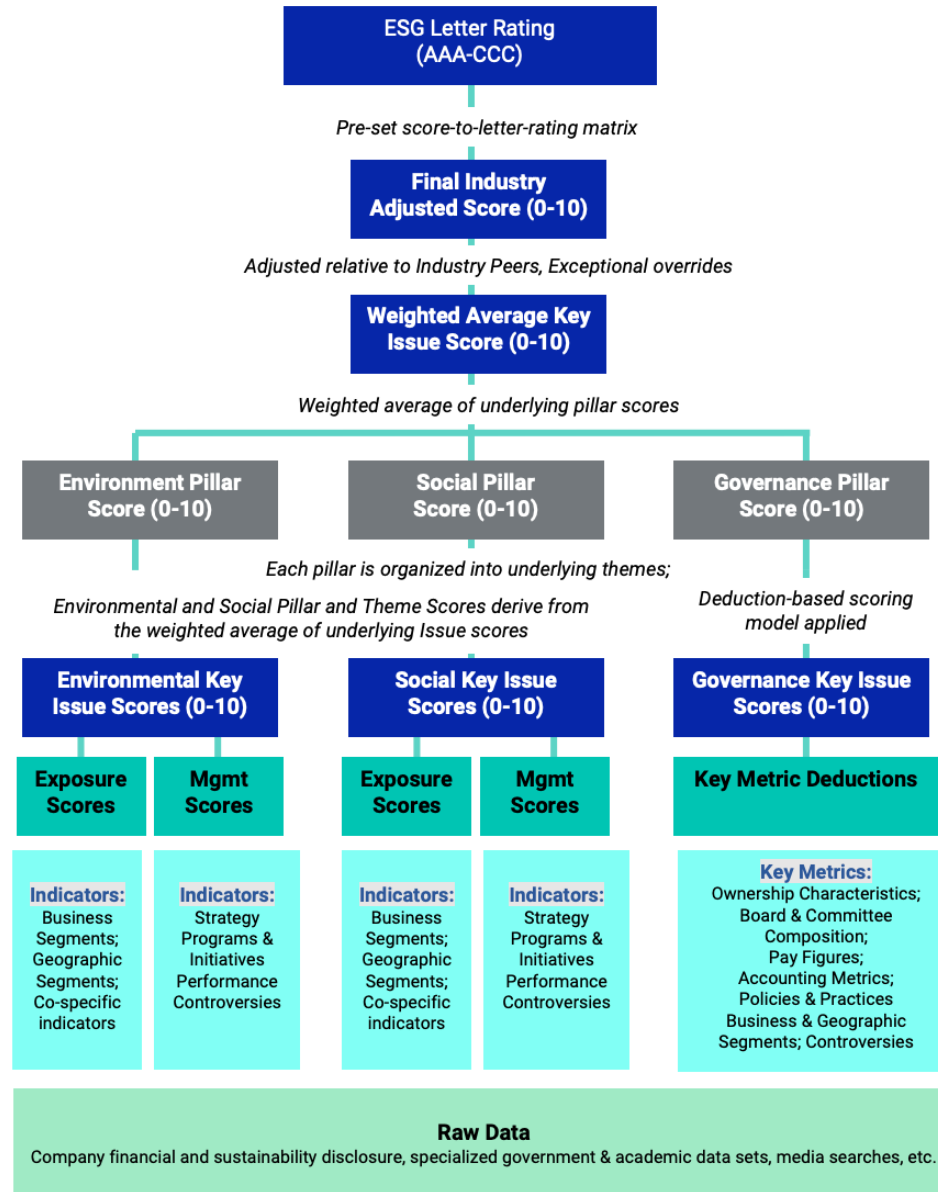
# MSCI Governance Model Structure

Deductions from Key Metrics flow up through each level to the overall Pillar score calculation



Source: <https://www.msci.com/documents/1296102/21901542/ESG-Ratings-Methodology-Exec-Summary.pdf>

# MSCI Hierarchy of ESG Scores

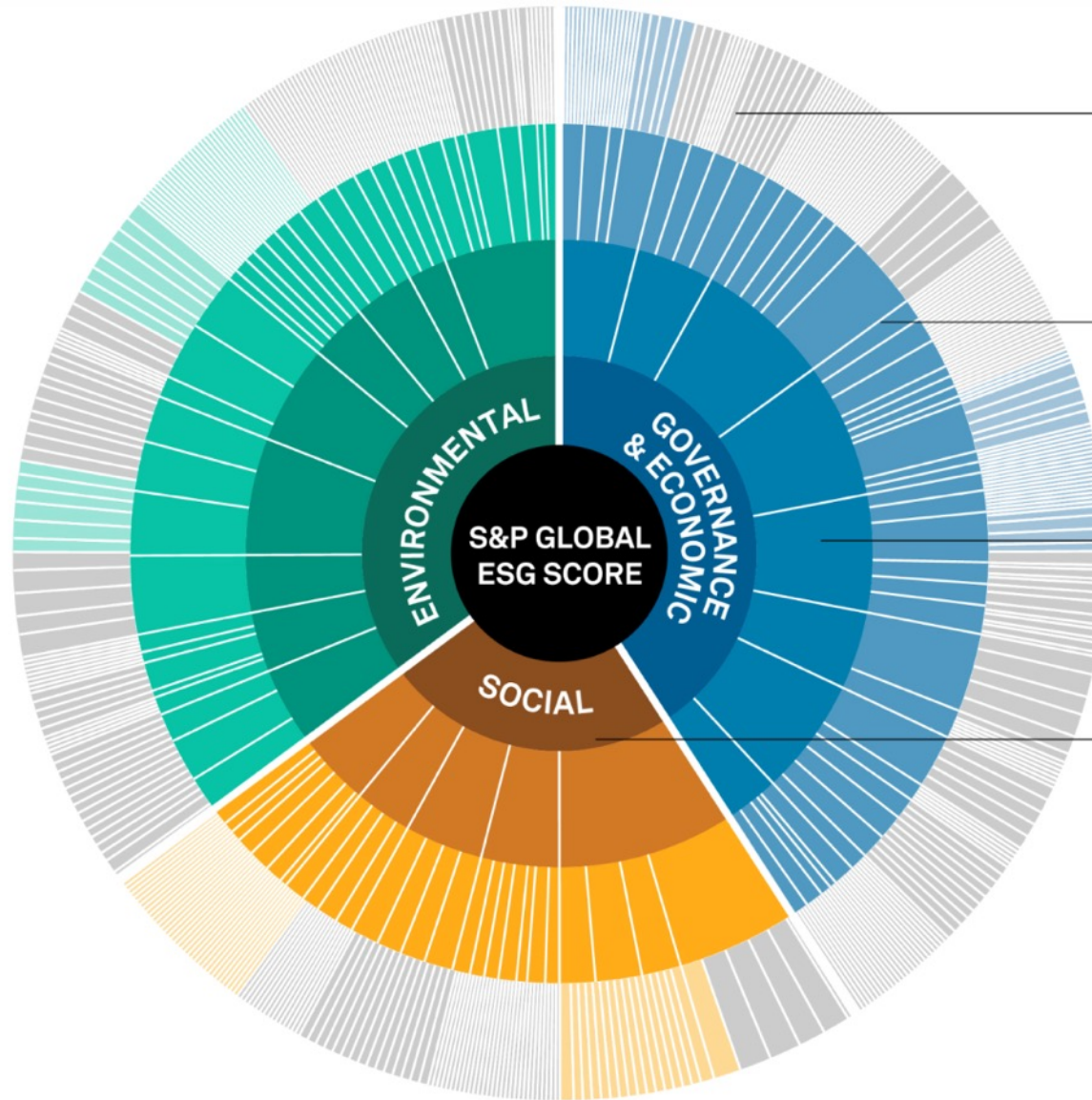


# DJSI S&P Global ESG Score

**8,000**  
Companies

**90%**  
Global market capitalization

**340,000+**  
Current Research Universe and Active Securities



Approx.  
**1,000**  
Datapoints

**Assessed values, text, checkboxes, documents**

Sources: Web-based questionnaire and company documents

**130+**  
Questions

**Weighted data point scores**

Up to 50% industry-specific

Ave.  
**30+**  
Criteria scores

**Weighted question scores**

61 industry specific approaches, with tailored questions, criteria and related weightings

**3**  
Dimension scores

**Weighted criteria scores**

Adjusted for corporate ESG controversies where applicable

**1**

**S&P Global ESG Score**

**Sum of weighted dimension scores**

# FTSE Russell ESG Ratings



# Sustainalytics

## ESG Risk Ratings

Analyst-based  
approach

Sustainalytics' ESG Risk Ratings measure a company's exposure to industry-specific material ESG risks and how well a company is managing those risks.

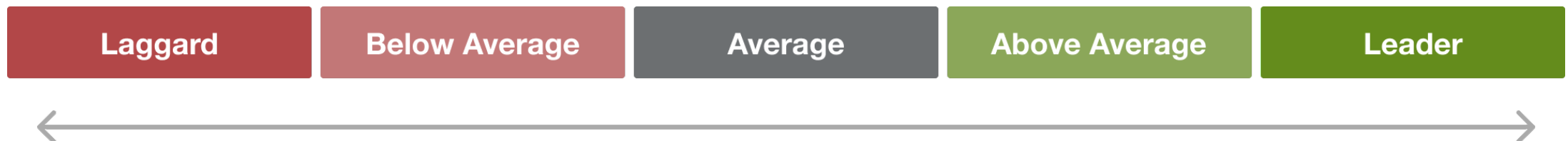
Negligible	Low	Medium	High	Severe
0 - 10	10 - 20	20 - 30	30 - 40	40+



# Truvalue ESG Ranks

Machine-based  
approach

- **Truvalue Labs** applies **AI** to analyze over **100,000 sources** and uncover **ESG risks** and opportunities hidden in **unstructured text**.
- The ESG Ranks data service produces an overall company rank based on industry percentile leveraging the **26 ESG categories** defined by the **Sustainability Accounting Standards Board (SASB)**.
- The data feed covers **20,000+** companies with more than **13 years** of history.



# Analyst-driven vs. AI-driven ESG

## Analyst-driven ESG research

Derives ratings in a structured data model

## Sustainalytics



*Analyst role at the end of the process allows subjectivity to color results*

## AI-driven ESG research

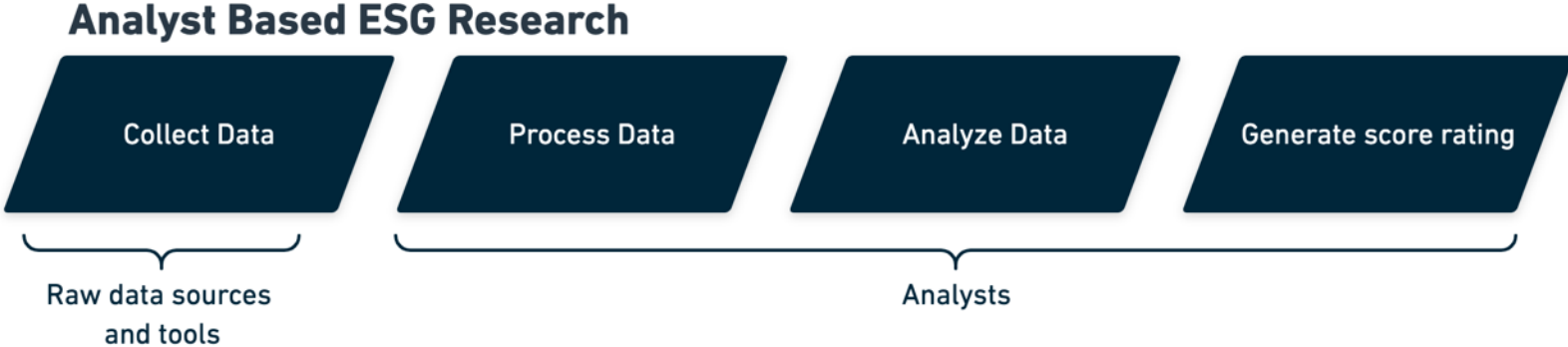
Derives signals from unstructured data

## Truvalue Labs

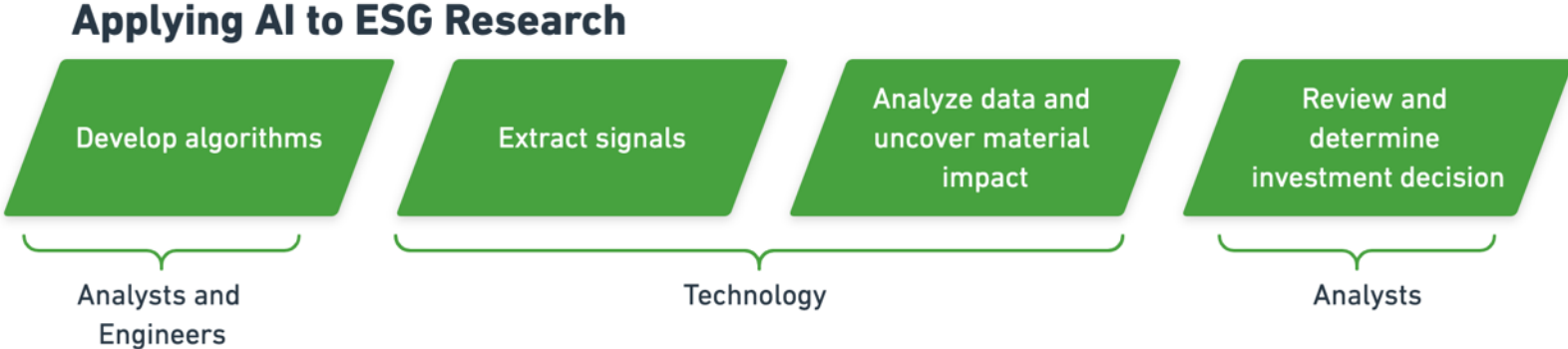


*Analyst expertise at the beginning of the process produces consistent results*

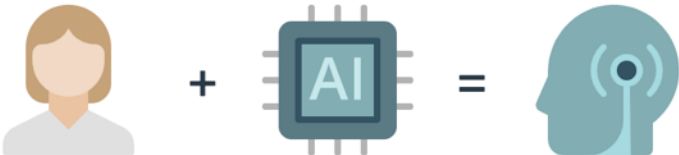
# Analyst based ESG Research



# AI based ESG Research

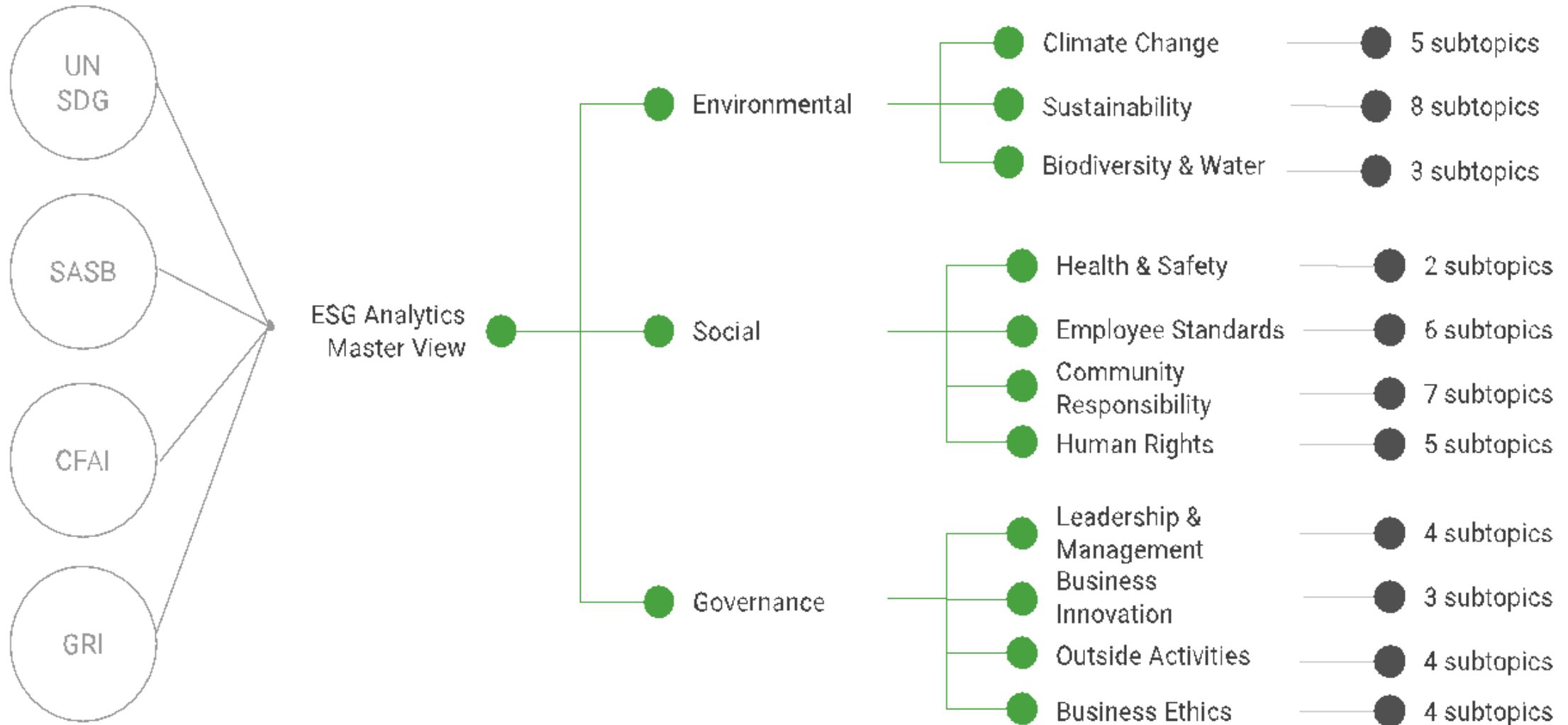


**It would take an analyst over 5 years to do what our AI can in 1 week**  
Combining analysts with AI creates gives you the full picture



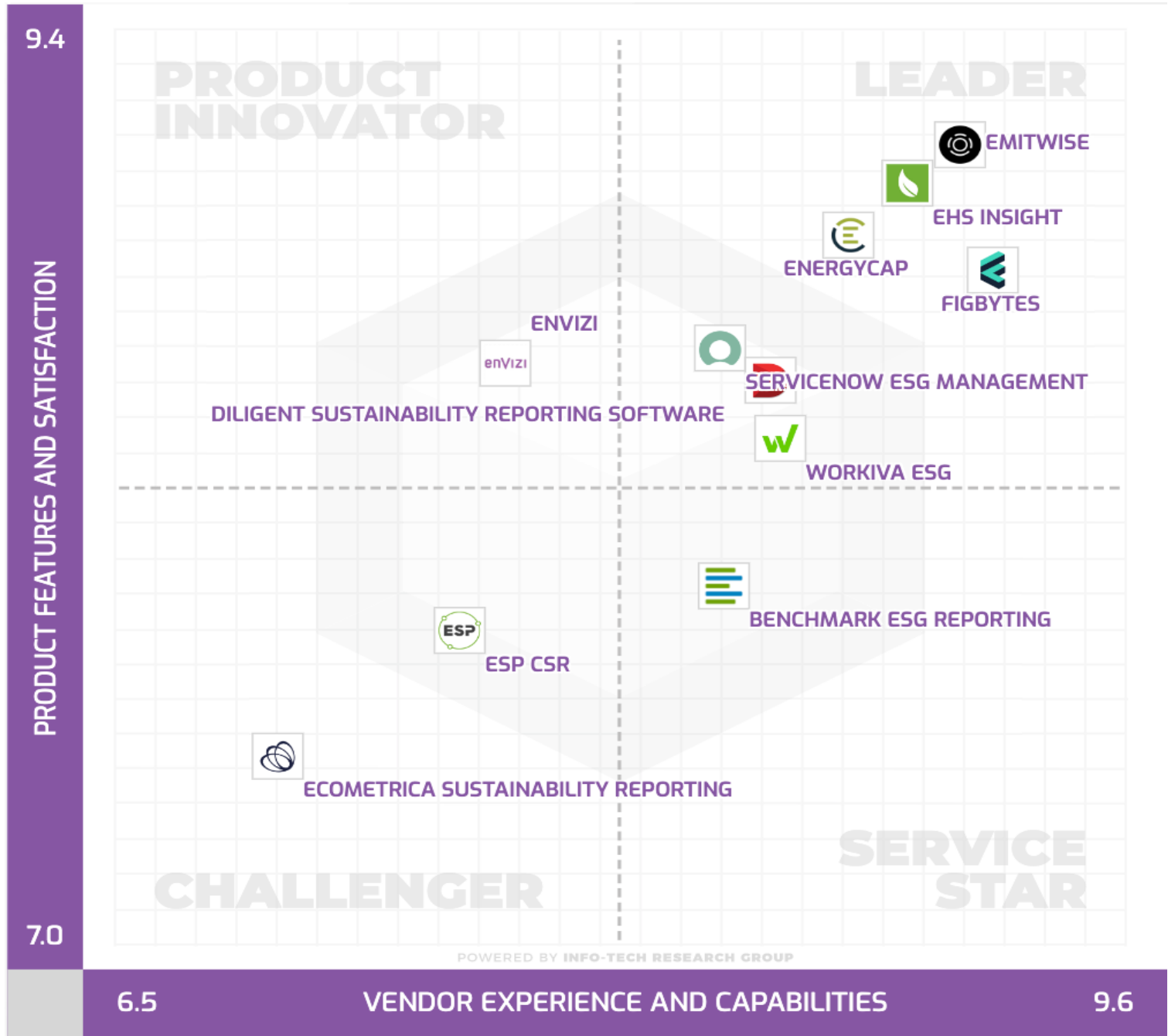
**ESG ANALYTICS**  
Invest where it matters.

# ESG Analytics: NLP Taxonomy

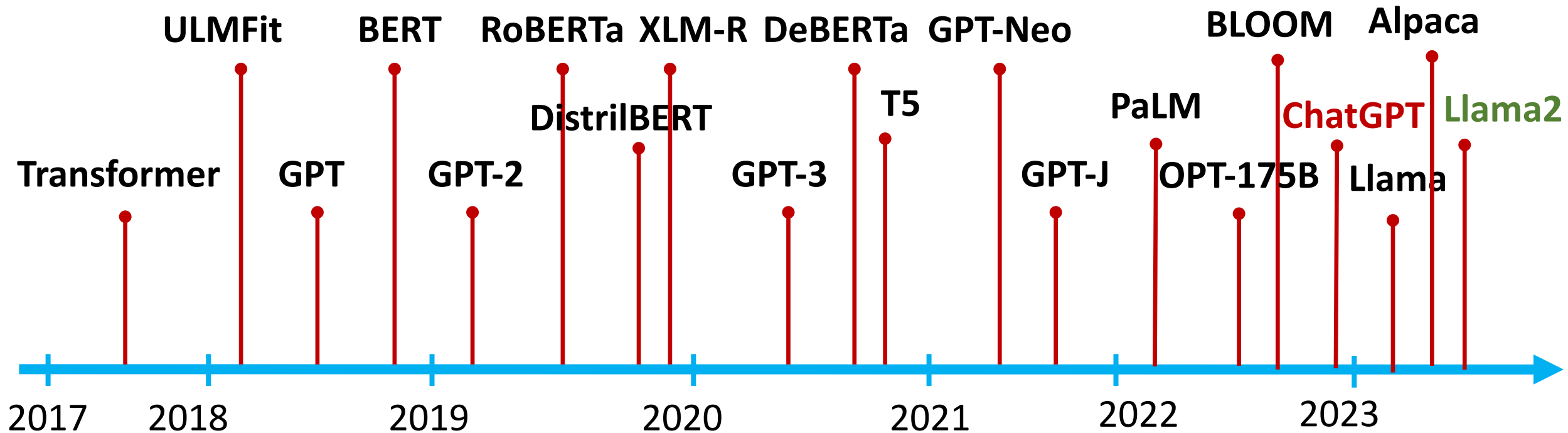


# Top ESG Reporting Software

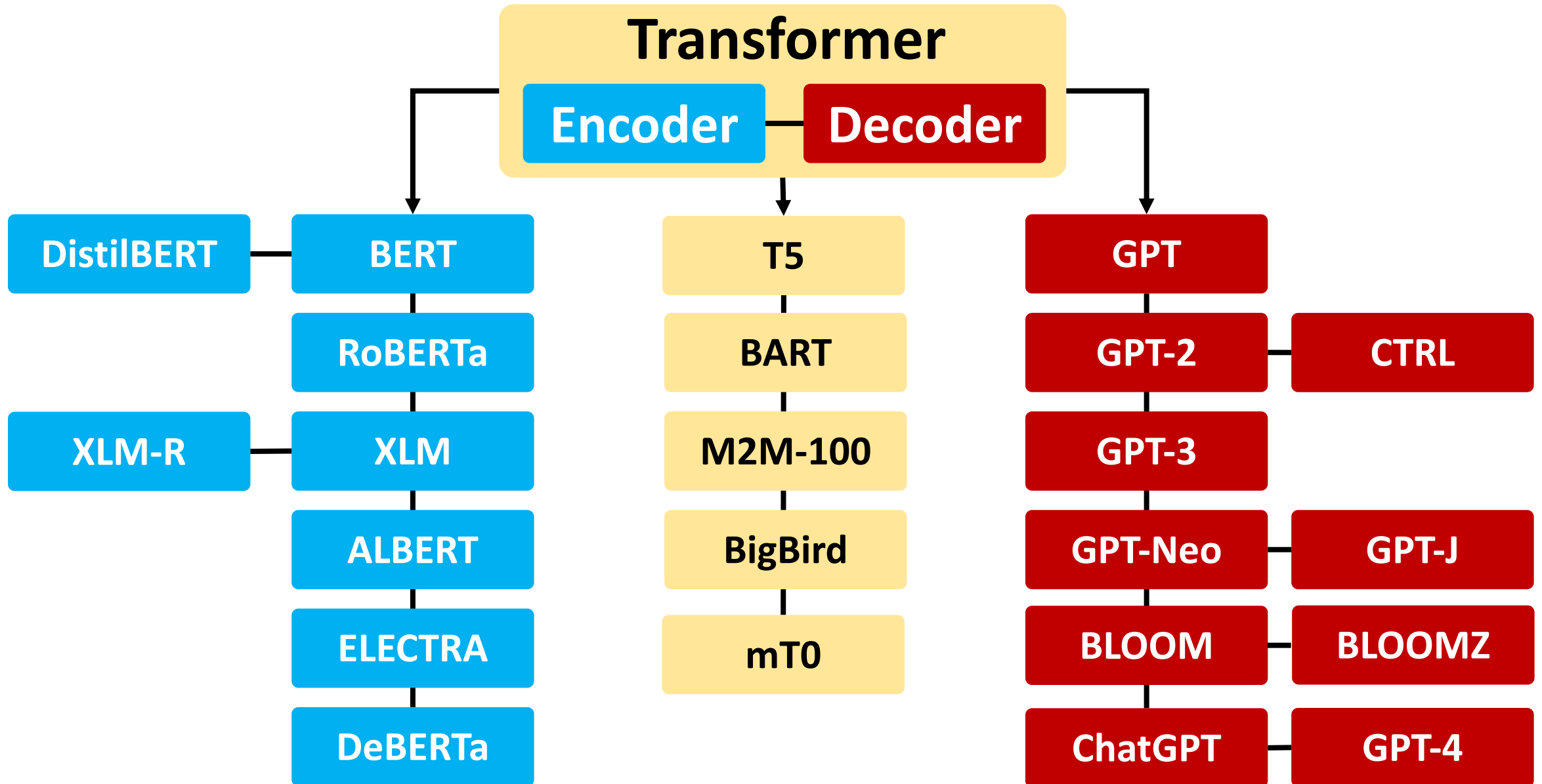
**Environmental, Social and Governance (ESG) Reporting software or Sustainability software helps organizations manage their operational data, evaluate their impact on the environment and provide reporting to perform audits.**



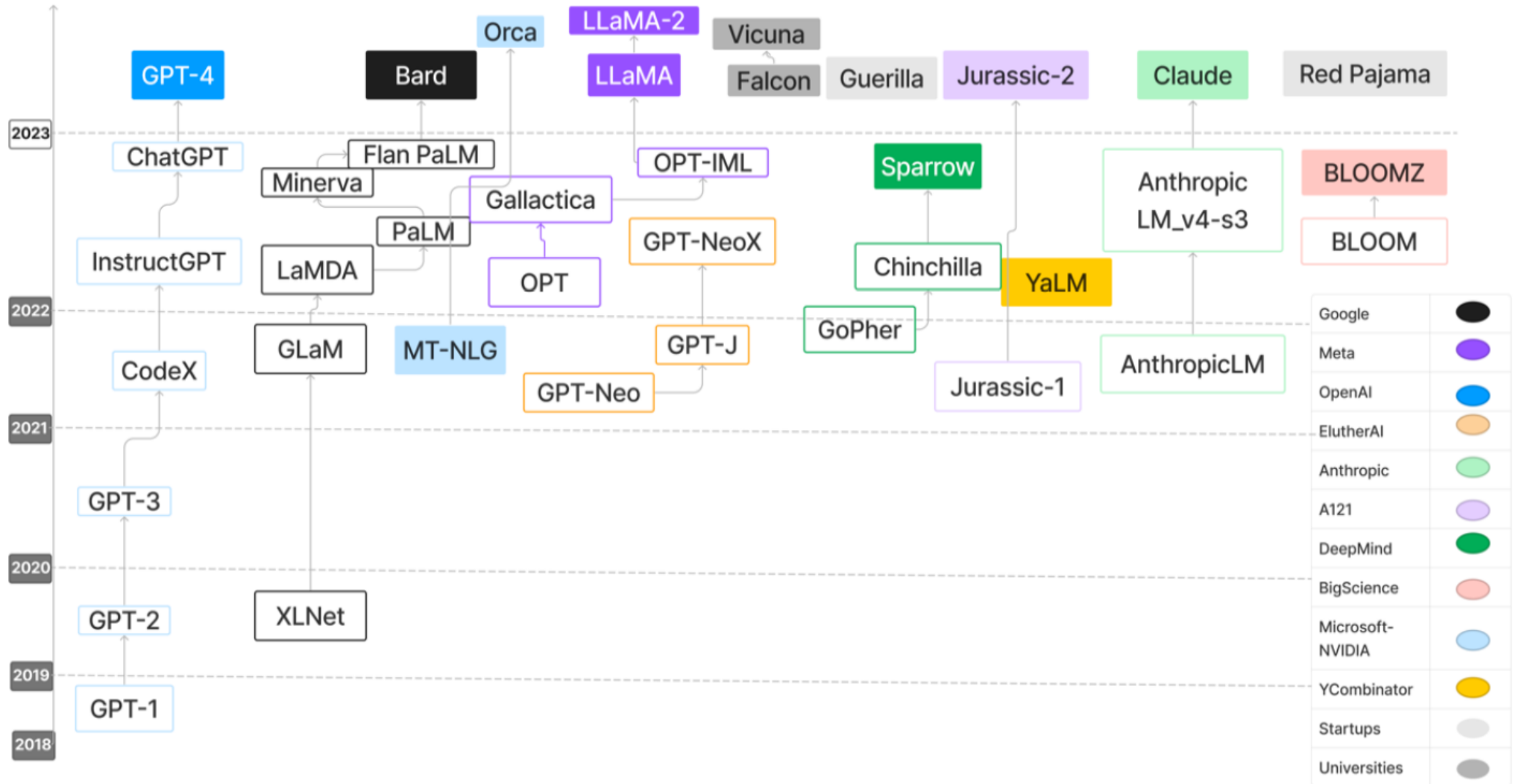
# The Transformers Timeline



# Transformer Models



# Large Language Models (LLMs)





# Four Paradigms in NLP (LM)

Paradigm	Engineering	Task Relation
a. Fully Supervised Learning (Non-Neural Network)	Feature (e.g. word identity, part-of-speech, sentence length)	
b. Fully Supervised Learning (Neural Network)	Architecture (e.g. convolutional, recurrent, self-attentional)	
<b>Transfer Learning: Pre-training, Fine-Tuning (FT)</b>		
c. Pre-train, Fine-tune	Objective (e.g. masked language modeling, next sentence prediction)	
<b>GAI: Pre-train, Prompt, and Predict (Prompting)</b>		
d. Pre-train, Prompt, Predict	Prompt (e.g. cloze, prefix)	


# Generative AI





**Text, Image, Video, Audio  
Applications**

# Popular Generative AI

- **OpenAI ChatGPT (GPT-3.5, GPT-4)**
- **OpenAI DALL·E 3**
- **Perplexity.ai**
- **Chat.LMSys.org**
- **ChatPDF**
- **Stable Diffusion**
- **Video: D-ID, Synthesia**
- **Audio: Speechify**

# OpenAI ChatGPT (GPT-4, GPT-3.5)

 ChatGPT 4 ▾

-  GPT-4   
With DALL-E, browsing and analysis  
Limit 40 messages / 3 hours
-  GPT-3.5  
Great for everyday tasks
-  Plugins



How can I help you today?

Suggest fun activities  
for a family of 4 to do indoors on a rainy day

Help me pick  
a birthday gift for my mom who likes gardening

Design a database schema  
for an online merch store

Give me ideas  
for what to do with my kids' art

 Message ChatGPT...



ChatGPT can make mistakes. Consider checking important information.

<https://chat.openai.com/>

# OpenAI ChatGPT (GPT-4) DALL·E 3

Vector graphic of a flowchart depicting the integration of generative AI in the education process, from content creation to virtual experiments, personalized learning, and innovative learning.



# Perplexity.ai



Anonymous Like



New Thread

Home

Discover

Library

Login

Sign Up

Try Pro

Upgrade to Claude-2 or GPT-4, boost Copilot uses, and upload more files.

Learn More

Download

## Large Language Models for Data Science

### Sources

What Is a Large Language Model and What Can...

snowflake · 1

What are Large Language Models? | NVIDIA

nvidia · 2

Catch Up On Large Language Models | by...

towardsdatascien... · 3



View 3 more

### Answer

Large Language Models (LLMs) are advanced AI systems designed to understand human language intricacies and generate intelligent, creative responses to queries <sup>1</sup>. They are deep learning algorithms that can recognize, summarize, translate, predict, and generate content using very large datasets <sup>2</sup>. LLMs are trained on enormous data sets typically measured in petabytes, sourced from books, articles, websites, and other text-based sources <sup>1</sup>. They are dynamic tools that are continuously updated and refined as new data becomes available <sup>1</sup>.

In the field of data science, LLMs can potentially transform how teams source, manage, and analyze data <sup>1</sup>. They are used in numerous data science applications, and their ability to process and interpret vast amounts of text data has made them an indispensable part of m

Ask follow-up...

Copilot



social media posts or customer reviews, to determine whether the overall sentiment is

<https://www.perplexity.ai/>

# Chat with Open Large Language Models

[chat.lmsys.org](https://chat.lmsys.org)

## ✂ Chatbot Arena ✂ : Benchmarking LLMs in the Wild

[Mixtral of experts](#): A Mixture-of-Experts model by Mistral AI

[Gemini](#): Gemini by Google

[GPT-4-Turbo](#): GPT-4-Turbo by OpenAI

[Claude Instant](#): Claude Instant by Anthropic

[Llama2-70B-SteerLM-Chat](#): A Llama fine-tuned with SteerLM method by NVIDIA

[Starling-LM-7B-alpha](#): An open model trained using RLAI by Berkeley

[Llama 2](#): Open foundation and fine-tuned chat models by Meta

[Code Llama](#): Open foundation models for code by Meta

[Qwen 1.5](#): A large language model by Alibaba Cloud

[DeepSeek LLM](#): An advanced language model by DeepSeek

[GPT-3.5](#): GPT-3.5-Turbo by OpenAI

[Nous-Hermes-2-Mixtral-8x7B-DPO](#): Nous Hermes finetuned from Mixtral 8x7B

[pplx-online-llms](#): Online LLM API by Perplexity AI

[Tulu 2](#): An instruction and RLHF model by UW/AllenAI

[Vicuna](#): A chat assistant fine-tuned on user-shared conversations by LMSYS

[WizardLM](#): An instruction-following LLM using evol-instruct by Microsoft

[Bard](#): Bard by Google

[StripedHyena-Nous](#): A chat model developed by Together Research and Nous Research.

[Claude](#): Claude 2 by Anthropic

[OpenChat 3.5](#): An open model fine-tuned on Mistral-7B using C-RLFT

[OpenHermes-2.5-Mistral-7B](#): A mistral-based model fine-tuned on 1M GPT-4 outputs

[Yi-Chat](#): A large language model by 01 AI

[Zephyr](#): A chatbot fine-tuned from Mistral by Hugging Face

<https://chat.lmsys.org/>

# Large Language Models for Data Science

## Llama 2-70b-chat

llama-2-70b-chat

Model A

Large Language Models for Data Science

Large language models have gained significant attention in recent years for their ability to process and analyze natural language data. These models have been successfully applied in various fields such as natural language processing, text classification, sentiment analysis, and machine translation. In data science, large language models can be used for a variety of tasks, including data cleaning, data transformation, and data analysis.

In this article, we will explore the use of large language models for data science tasks and discuss their advantages and limitations. We will also provide examples of large language models that can be used for data science applications.

Advantages of Large Language Models for Data

## Mistral-7b-instruct

mistral-7b-instruct

Model B

Large Language Models for Data Science

Large Language Models (LLMs) are becoming increasingly important in data science due to their ability to understand and generate natural language text. Here are some of the key ways that LLMs can be used in data science:

1. Text classification: LLMs can be trained on large datasets of labeled text and then used to classify new text into different categories. This can be useful in a variety of applications, such as sentiment analysis, spam filtering, and topic classification.
2. Named entity recognition: LLMs can also be used to identify and extract specific pieces of information from text, such as names of people, organizations, and locations. This can be useful in applications such as information extraction




<https://chat.lmsys.org/>

# Chat with Open Large Language Models: Chatbot Arena



# Chatbot Arena Leaderboard

## LLM Leaderboard

Rank	 Model	★ Arena Elo	 95% CI	 Votes	Organization	License
1	<a href="#">GPT-4-1106-preview</a>	1254	5/-5	38745	OpenAI	Proprietary
2	<a href="#">GPT-4-0125-preview</a>	1253	10/-8	6308	OpenAI	Proprietary
3	<a href="#">Bard (Gemini Pro)</a>	1218	8/-7	10313	Google	Proprietary
4	<a href="#">GPT-4-0314</a>	1191	6/-6	20430	OpenAI	Proprietary
5	<a href="#">GPT-4-0613</a>	1164	5/-6	32941	OpenAI	Proprietary
6	<a href="#">Mistral Medium</a>	1152	5/-7	17847	Mistral	Proprietary
7	<a href="#">Claude-1</a>	1150	7/-5	19017	Anthropic	Proprietary
8	<a href="#">Qwen1.5-72B-Chat</a>	1147	8/-8	5204	Alibaba	Qianwen LICENSE
9	<a href="#">Claude-2.0</a>	1132	6/-8	12753	Anthropic	Proprietary
10	<a href="#">Gemini Pro (Dev API)</a>	1122	7/-7	9024	Google	Proprietary
11	<a href="#">Claude-2.1</a>	1120	6/-4	27723	Anthropic	Proprietary
12	<a href="#">Mixtral-8x7b-Instruct-v0.1</a>	1120	5/-6	18410	Mistral	Apache 2.0
13	<a href="#">GPT-3.5-Turbo-0613</a>	1118	5/-5	36704	OpenAI	Proprietary
14	<a href="#">Gemini Pro</a>	1115	9/-9	6958	Google	Proprietary
15	<a href="#">Yi-34B-Chat</a>	1111	7/-8	7734	01 AI	Yi License

<https://chat.lmsys.org/>

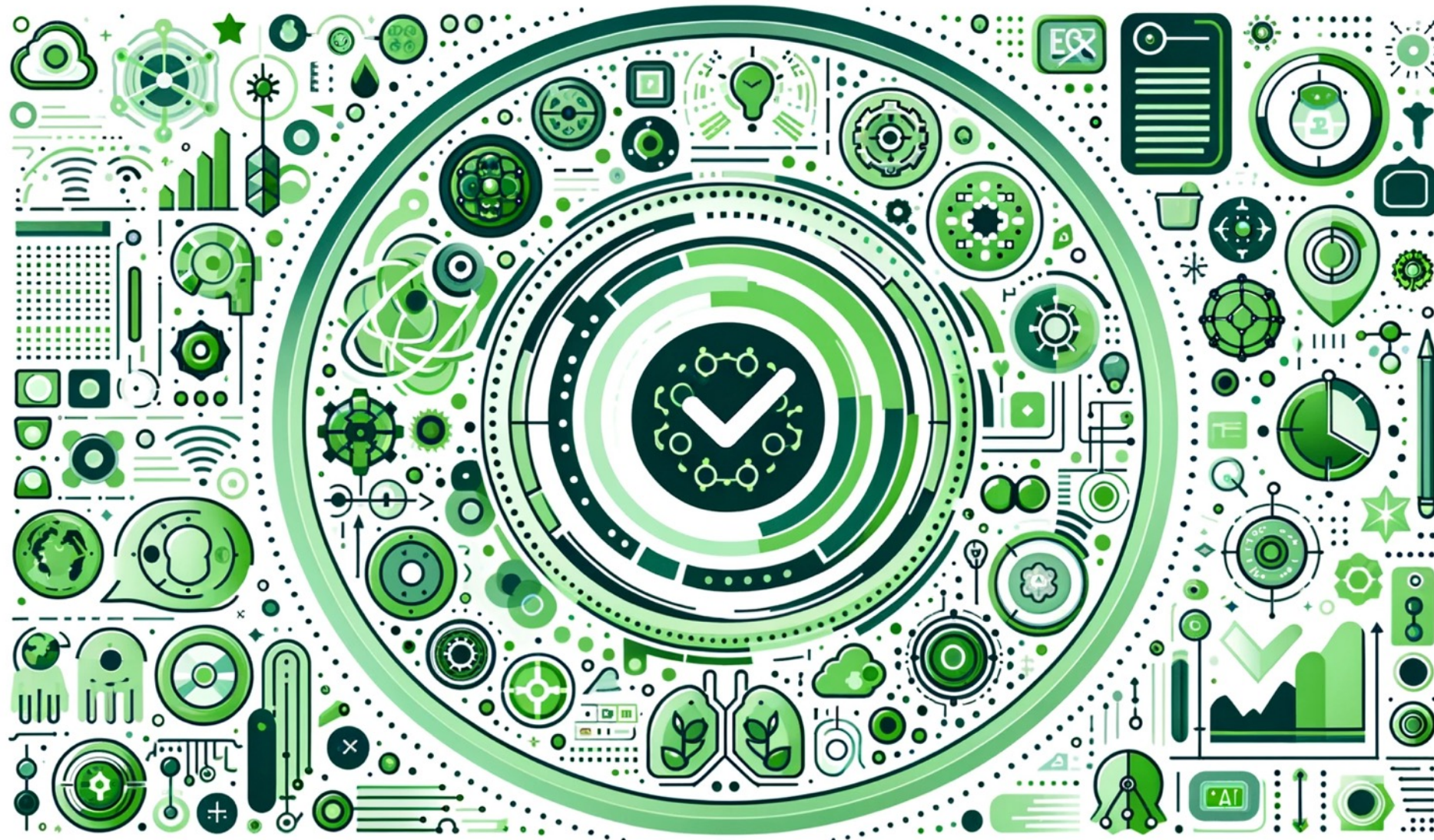
# Generative AI and LLMs for Sustainability and ESG Data Analytics



# Sustainability and ESG Data Analytics



# Generative AI for ESG Rating and Reporting Generation



# Teaching



- **Big Data Analytics**
  - Fall 2020, Spring 2023, Spring 2024
- **Software Engineering**
  - Fall 2020, Fall, 2021, Spring 2022, Spring 2023, Spring 2024
- **Artificial Intelligence in Finance and Quantitative**
  - Fall 2021, Fall 2022, Fall 2023
- **Artificial Intelligence**
  - Spring 2021, Fall 2022
- **Artificial Intelligence for Text Analytics**
  - Spring 2022, Fall 2023
- **Sustainability and ESG Data Analytics**
  - Spring 2024
- **Python for Accounting Applications**
  - Fall 2023
- **Foundation of Business Cloud Computing**
  - Spring 2021, Spring 2022, Spring 2023, Spring 2024

# Research Projects



- 1. Fintech Green Finance for Carbon Market Index, Corporate Finance, and Environmental Policies.** Carbon Emission Sentiment Index with AI Text Analytics
  - NTPU, 113-NTPU\_ORDA-F-003 , 2023/01/01~2024/12/31
- 2. Digital Support, Unimpeded Communication: The Development, Support and Promotion of AI-assisted Communication Assistive Devices for Speech Impairment (1/3).** Multimodal Cross-lingual Task-Oriented Dialogue System for Inclusive Communication Support
  - NSTC 112-2425-H-305-002-, 3 Years (2023/05/01-2026/04/30) Year 1: 2023/05/01~2024/04/30
- 3. Research on speech processing, synthesis, recognition, and sentence construction of people with language disabilities.** Multimodal Cross-lingual Task-Oriented Dialogue System
  - NTPU, 113-NTPU\_ORDA-F-004, 2023/01/01~2025/12/31
- 4. Metaverse AI Multimodal Cross-Language Task-Oriented Dialogue System**
  - ATEC Group, Fintech and Green Finance Center (FGFC, NTPU), NTPU-112A413E01, 3 Years (2023/05/01~2026/04/30)
- 5. Establishment and Implement of Smart Assistive Technology for Dementia Care and Its Socio-Economic Impacts (2/3).** Intelligent, individualized and precise care with smart AT and system integration
  - NSTC, 112-2627-M-038-001-, 2023/08/01~2024/07/31
- 6. Prospective longitudinal study on peri-implant bone loss associated with peri-implantitis**
  - USTP (NTPU, TMU), USTP-NTPU-TMU-113-03, 2024/01/01~2024/12/31

# 總結 (Summary)

- 本課程介紹**永續數據分析基本概念與實務操作**。
- 課程內容包括
  1. 永續數據分析概論
  2. 環境、社會與治理 (ESG) 淨零數位轉型
  3. 永續與ESG 資料科學
  4. Web 3.0 和大數據分析在金融科技、綠色永續金融
  5. TCFD 氣候相關財務揭露與En-ROADS 氣候變遷模擬
  6. ESG數據的收集、分析和視覺化
  7. ESG數據報告、企業永續報告書
  8. ESG數據驗證
  9. 能源之星報告與數據揭露
  10. 人工智慧物聯網在ESG永續應用
  11. 生成性AI於永續評等和報告生成
  12. 永續數據分析個案研究



# 永續數據分析

(Sustainability and ESG Data Analytics)



國立臺北大學  
National Taipei University

aws educate | Cloud Ambassador

2020 Cohort

aws academy

Accredited  
Educator

aws

certified

Solutions  
Architect

Associate

aws

certified

Cloud  
Practitioner

## Contact Information

**戴敏育 (Min-Yuh Day, Ph.D.)**

Associate Professor

[Institute of Information Management, National Taipei University](#)

Tel: 02-86741111 ext. 66873

Office: B8F12

Address: 151, University Rd., San Shia District, New Taipei City, 23741 Taiwan

Email: [myday@gm.ntpu.edu.tw](mailto:myday@gm.ntpu.edu.tw)

Web: <http://web.ntpu.edu.tw/~myday/>

