

SSF Spotlight

# How Artificial Intelligence Contributes to Effective Sustainable Finance



Explore here the *Artificial Intelligence*  
resources on SSF's website



**No time to read the full report?**  
**Our AI agent answers your questions**

SSF chatbots are powered by Uthereal



# Foreword

For decades now, artificial intelligence (AI) has been part of our daily lives without attracting much public attention. The recent rise of “Generative AI” capabilities has, however, sparked a sudden and considerable interest for AI from the general public.

In various fields, AI has enabled progress for the common good. In medicine, it supports the development of new medication, the prediction of epidemics, and has materially increased the reliability of certain diagnoses. In agriculture, AI allows to monitor soil quality, identify early signs of crop disease, or optimize irrigation to reduce water waste. In the energy sector, AI-powered “smart grids” dynamically adjust supply and demand, facilitating the integration of renewable energy sources and reducing overproduction. New digital platforms enable real-time monitoring of decarbonisation projects. We could name many more examples.

However, its externalities are real. The annual sustainability reports by Google<sup>1</sup> or Microsoft<sup>2</sup> have shown the heavy energy consumption and carbon footprint of training new AI models, shifting these companies far away from their planned decarbonisation path. We all know of the risk of reproducing or even reinforcing certain cognitive biases or social prejudices. AI’s effects on the employment market are also unknown: will it free workers from the most repetitive tasks or rather be the driver of a unprecedented job displacement with significant socio-economic consequences?

The new capabilities brought by AI also offer opportunities for sustainable finance<sup>3</sup>. It allows to extract new insights from unstructured data that was previously inaccessible, and helps sustainability teams in preparing the first draft of assessment notes or regular disclosures and reports. As a result, institutional investors and financial institutions are better equipped to direct capital toward companies and projects that support the transition to a sustainable economy.

SSF’s mandate is to support the Swiss financial centre in achieving a leading position in sustainable finance, by shaping Swiss frameworks, engaging key stakeholders, informing on best practice and creating supportive tools. This publication is part of this mandate. It explores and explains how Artificial Intelligence can be used to further develop, improve and enhance sustainable finance practices.

A heartfelt thanks goes to the external contributors to this paper: all organisations that shared examples of their real-life AI applications in sustainable finance, and Dr. João Carvalho (ETH Zurich, Uthereal) for his contribution to the technical part of the Appendix.



**Sabine Döbeli**  
CEO Swiss Sustainable Finance



**Romain Leroy-Castillo**  
Director AI and Education, Swiss Sustainable Finance

1 “Environmental Report 2025”, Google, page 105. <https://sustainability.google/reports/google-2025-environmental-report>

2 “2025 Environmental Sustainability Report”, Microsoft, page 12. <https://www.microsoft.com/en-us/corporate-responsibility/sustainability/report>

3 “Sustainable finance refers to any form of financial service with the objective of supporting the transition to a sustainable economy and society by integrating environmental, social or governance factors into business and investment decisions. Such finance aims for the lasting benefit to clients, society at large and the planet”, SSF glossary, <https://www.sustainablefinance.ch/en/resources/glossary.html>

# Table of Contents

<b>Executive Summary</b>	6		
<b>1. Setting The Scene</b>	10	<b>3. Recommendations and Outlook</b>	51
1.1. A Look Back on Artificial Intelligence	11	3.1. Recommendations	52
1.2. AI Applications for Sustainability	11	3.2. What Could Tomorrow Entail?	55
1.3. Current AI Adoption in Swiss Financial Services	12	3.3. The Path to Responsible AI	55
<b>2. Artificial Intelligence Applications in Sustainable Finance</b>	17	<b>4. Appendix</b>	56
2.1. Data Collection and Extraction	19	4.1. A Brief History of AI	57
USE CASE Simplifying Double Materiality Assessment via Intelligent Process Automation	20	4.2. Artificial Intelligence Briefly Explained	59
USE CASE Real-Time Market Intelligence for Swiss Real Estate	22	4.3. Glossary	65
2.2. Data Curation and Quality Assurance	24	4.4. Bibliography	65
USE CASE Building Trustworthy Data Foundation via Machine-Learning	25	4.5. List of Figures	66
2.3. Data Analysis and Insights	28		
USE CASE Enhancing Carbon Market Integrity through AI-Driven Network Analysis	29		
USE CASE Strengthening Trust in Corporate Sustainability Reports with NLP and Sentiment Analysis	31		
USE CASE Human Guided AI for Scaled and Trusted Risk Insights	33		
USE CASE Satellite Image and Computer Vision to Assess Biodiversity Value-At-Risk	36		
2.4. Predictive Analytics	39		
USE CASE Advancing Real Estate Sustainability Assessments with AI-Based Portfolio Analysis	40		
USE CASE AI for ESG Integration: Training Machines to Predict Sustainable Alpha	42		
2.5. Content Generation	44		
USE CASE AI Supporting Portfolio Managers in Sustainable Investment Decision-Making	46		
USE CASE AI-Powered Company Briefs for Investment Teams	48		

# Executive Summary

## Setting the scene

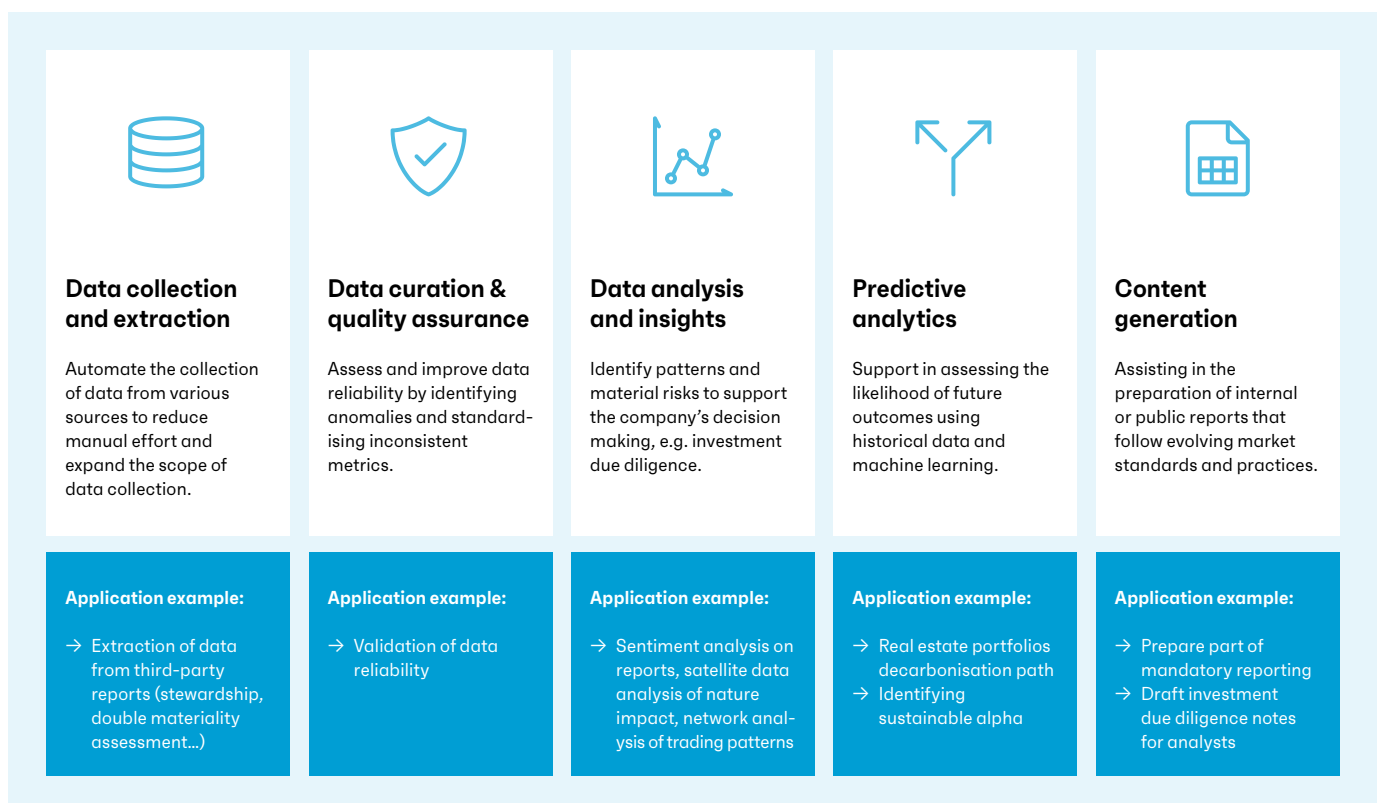
From the early formulation of the concept in the 1950s, through foundational advances such as backpropagation and, more recently, the development of transformer architectures, AI has evolved into systems capable of processing vast amounts of data and interacting with information in increasingly human-like ways.

In financial services, AI has long been embedded in core activities such as risk modelling, fraud detection, or automated trading. Recently, Generative AI significantly expanded

the ability to work with unstructured data and to transform complex, fragmented information into usable insights. Financial institutions are beginning to explore this potential in sustainable finance.

This paper explores the potential applications of AI to support sustainable finance, along a “data-to-decision” workflow with five stages: collecting data, ensuring its quality, extracting insights, anticipating future outcomes, and drafting content to communicate results. Within each step, this paper highlights real-life use cases as illustrative examples.

## AI applications in sustainable finance along the data-to-decision workflow



## Data collection and extraction

The first stage addresses the exponential growth in data volumes and the increasing complexity of sustainability-related data. Financial institutions face rising demands to collect information from heterogeneous, often unstructured sources under expanding disclosure and reporting requirements. This manual burden has become a bottleneck for sustainable finance processes.

AI applications at this stage typically combine:

- Natural language processing (NLP) to identify relevant sustainability-related passages in documents,
- Document parsing and text recognition to extract metrics, statements, and qualitative disclosures,
- Intelligent process automation (IPA) to orchestrate repeated extraction workflows across sources and reporting cycles.

These tools enable consistent mapping of raw disclosures to regulatory frameworks or internal data models. The core value lies in systematically converting unstructured sustainability information into machine-readable, comparable datasets, allowing analysts to focus on interpretation rather than data retrieval.

### **Data curation and quality assurance**

Data quality is a foundational constraint for the effective implementation of any AI application. Sustainability data is often inconsistent, incomplete, difficult to compare across companies, providers, and time periods. AI can offer solutions to this challenge, such as:

- Detecting inconsistencies and anomalies across time series, peer groups, and sector benchmarks,
- Standardising units, definitions, and reporting conventions across heterogeneous sources,
- Assessing the relative reliability of competing data points using probabilistic or feature-based models,
- Reconciling overlapping datasets originating from different providers or methodologies.

AI enables context-aware validation, weighing multiple indicators concurrently to flag data points that warrant human review. The overarching role of AI at this stage is to establish a trustworthy data foundation – a prerequisite for any scalable use of AI further down the value chain.

### **Data analysis and insights**

At the analysis stage, AI moves to extracting meaning and actionable insights. Many of today's most widespread AI applications involve analysing third-party content such as corporate disclosures, media coverage, policy debates or other stakeholder communications. Typical AI applications here include:

- Sentiment and narrative analysis to interpret how sustainability topics are discussed across corporate, media, policy and other third-party sources,
- Thematic clustering to group companies, assets, or projects according to shared sustainability characteristics or exposures,
- Controversy and issue mapping to identify concentration risks and emerging problem areas,
- Cross-validation of sustainability claims against independent or alternative data sources.

The contribution of AI at this stage lies in its ability to scan broad information environments at scale, identifying weak signals, inconsistencies, or emerging sustainability risks, some of which may not yet be reflected in official ESG ratings and metrics.

### **Predictive analytics**

Rooted in traditional statistical and machine learning models, predictive analytics aim to estimate future outcomes such as climate-related losses, product performance trajectories, or behavioural responses of firms and clients. This typically includes:

- Scenario-based modelling of climate and environmental risks, linking physical or transition risk indicators to financial exposures,
- Forecasting sustainability performance trajectories, such as expected risk/return ratios of securities in a portfolio,
- Anticipating strategic positioning of companies or sectors in response to regulation, technology, or market shifts,
- Predicting client or investor behaviour related to sustainable financial products.

These applications rely primarily on statistical and machine learning models, rather than Generative AI. Their primary value is supporting strategic planning and risk anticipation, while acknowledging uncertainties inherent in long-term sustainability forecasting. As regulatory disclosures expand and data infrastructures mature, predictive analytics is expected to evolve into more widely used components of sustainability risk management and decision-making.

### **Content generation**

The final stage focuses on Generative AI as a bridge between analysis and communication. Content generation transforms the insights gained from the previous steps into usable outputs such as reports, briefings, and decision-support materials, for example:

- Automated drafting of sustainability and ESG reports, aligned with regulatory and voluntary frameworks,
- Generation of internal briefings and research summaries synthesising large volumes of sustainability information,

- Creation of standardised sustainability profiles for companies, portfolios, or products,
- Support for investment due diligence through concise, comparable sustainability narratives.

Human review remains of paramount importance, but AI substantially reduces time spent on manual drafting and formatting.

### Recommendations and Outlook

While it has become increasingly simple to experiment with Generative AI, it remains hard to successfully implement AI in production scenarios that are robust enough for the corporate world, where there is no tolerance for approximately correct answers. The report offers targeted recommendations for a successful implementation of AI solutions into finance on page 52. AI implementation follows similar success factors as any other transformation project.

## 1

### Define the ambition

#### Think Big: follow a vision, not the easy path

AI initiatives should be anchored in long-term strategic vision rather than short-term technological opportunities. Resist the temptation to prioritise quick wins if they distract from longer term value creation.

#### Avoid the “trap of automation”

Many early AI use cases target immediate pain points, typically via process automation requested by users. True benefits will rather come from enhancing quality and augmenting capability.

#### Build, buy, borrow?

Larger financial institutions (e.g. universal banks) will favour inhouse built solutions running on internal servers; smaller market participants (e.g. pension funds) will rather rely on outsourced capabilities.

#### Sustainability: build on durable grounds

Consider and minimize environmental impact through right-sized and efficient models, favour AI to augment rather than replace human skills, and ensure explainability and auditability of AI models to ensure accountability of outputs.

## 2

### Establish the foundation

#### Data: start by building a robust data backbone

A solid data foundation is the indispensable starting point of any AI initiative. Integrate internal, vendor, and public datasets into unified access layers.

#### Processes: Redesign first, don't plug AI

Avoid retrofitting AI into outdated or inefficient workflows. First reassess end-to-end processes and remove steps that have become redundant or superfluous.

#### Governance: embed into existing frameworks whenever possible

Existing governance frameworks should be leveraged, such as Model Risk Management policies, which can be expanded, rather than adding new layers of policies.

#### Tailor the solution: avoid unnecessary complexity

High performing small language model are often better suited (lighter, quicker, less energy intensive, more reliable) for specialized tasks than mainstream large language models.

### 3

## Prepare the operationalisation

### Employees: bring people along

Many AI transformation projects fail due to human factors more than technical ones. Promote AI literacy, invest in continuous training. Offer clear paths with genuine upskilling and reskilling programs.

### Develop an operating model fit for scaling

Scaling AI use cases is one of the industry's toughest challenges. Establishing cross-departmental governance is one of the key enablers for future scaling.

### Build incrementally and keep a “fail fast” mentality

AI projects require agility. Identify ineffective approaches early, communicate transparently, and pivot when necessary.

### Monitor and improve continuously

Because data gets outdated and model drift occurs, AI systems require ongoing monitoring, Monitor performance regularly, implement minor, frequent updates and fully retrain when needed.

### What could tomorrow entail?

Over the coming years, a few trends are likely to reshape how AI is used in sustainable finance.

The first trend is a move away from relying solely on very large, general-purpose models towards smaller, more efficient and more specialised ones.

The second trend is the rise of agentic AI: systems coordinating multiple specialised “expert agents” that handle distinct tasks, allowing them to work together like a virtual team. Very few real-life deployments in production exist today, but this will change fast.

Looking further ahead, some technologies being researched today, such as quantum computing or neuromorphic computing, may open possibilities that sit for now in the realm of science fiction. These advances could enable entirely new AI architectures overcoming many of the limitations of current models. This is however still a distant prospect.

For financial institutions, the increased adoption of AI-based tools opens a new set of questions. In particular: how to avoid as much as possible the externalities brought by increased adoption of this technology, such as energy consumption, job displacement, or bias reinforcement?

In that sense, this whitepaper should be seen as a starting point. The longer-term challenge will be to ensure that the systems we are building today, that will expand in reach, capabilities and complexity, remain reliable, aligned, and ultimately supportive of a more sustainable financial system tomorrow.

# Setting The Scene

# 01

## 1.1. A look back on Artificial Intelligence

Despite its recent peak in popularity, Artificial Intelligence (AI) is not a new technology. The term itself was coined in 1956 at the Dartmouth Summer Research Project on Artificial Intelligence. With highs and lows (the so-called “AI winters”), the technology developed over decades. A major breakthrough happened in 1986 with the publication of “Learning representations by back-propagating errors.”<sup>4</sup>. The concept of backpropagation<sup>5</sup> is still at the core of today’s machine learning techniques.

Another breakthrough was achieved in the years 2010s, with a significant increase in computer power and volume of available data on the one hand, and the introduction of a new neural network architecture on the other, the “transformers”<sup>6</sup>. They enabled the rise of the large language models we know today, which in turn have opened the door to advanced generative AI systems capable of understanding, producing, and reasoning with human language across a wide range of tasks. They also allow machines to aggregate and summarize this data into reader-friendly formats, considerably transforming the end-user experience.

You can learn more about the history of AI and the key machine learning techniques in the appendix of this page (Appendix I, page 57) or in the learning module on SSF’s website<sup>7</sup>.

## 1.2. AI applications for sustainability

Artificial intelligence has the capacity to help addressing pressing global challenges when its capabilities are applied thoughtfully. Across sectors as diverse as agriculture, healthcare, education, climate management, and energy distribution, AI is already delivering measurable benefits to people and communities. The below highlights a selection of examples where AI innovation is producing tangible public good. It does in no way pretend to be exhaustive.

### 1.2.1. Agriculture: Reducing environmental impact and ensuring food supply for a growing population

Agriculture currently causes ca. 78% of the global ocean and freshwater eutrophication<sup>8</sup>. Agriculture is also responsible for releasing a wide range of chemical substances into the environment, contributing further to soil and water pollution. This challenge will only increase in the next decades, in

light of the global population growth. AI in agriculture may help feed a growing global population while reducing such environmental impact.

Different companies, including a large manufacturer of agricultural equipment have pioneered “See & Spray” systems that use computer vision and machine learning to identify individual plants in a field and apply herbicide only where needed. This precision reduces chemical use, lowers costs for farmers, and minimises ecological damage.

AI is also proving to be valuable in pest management. Some companies have developed systems to detect locust outbreaks early, enabling rapid containment, or to identify pest infestation early enough to reduce pesticide usage by 20%. Beyond these cases, AI-driven soil analysis, yield prediction, and irrigation optimisation start helping farmers adapt to climate variability and improve food security.

### 1.2.2. Healthcare: enhancing diagnosis reliability

Healthcare has become one of the most dynamic fields for AI innovation, with tools enhancing diagnosis, treatment, and accessibility. AI-assisted radiology now supports early detection of conditions such as cancer by analysing X-ray and MRI scans with high accuracy. Some companies are exploring AI-managed neural interfaces and prosthetics that could restore mobility, communication, and sensory functions, offering promising perspectives for people with severe disabilities.

AI is also accelerating biomedical discovery. Different research institutes have used AI to concurrently design protein sequences and structures, advancing among others vaccine development and allergy understanding. Ophthalmology has seen breakthroughs such as AI models predicting the progression of non exudative age-related macular degeneration (AMD) to exudative AMD within months matching or surpassing expert clinician performance. These technologies

4 David E. Rumelhart, Geoffrey E. Hinton & Ronald J. Williams, “Learning representations by back-propagating errors”, *Nature*, 09 October 1986. <https://www.nature.com/articles/323533a0>

5 “backpropagation” refers to the process of a neural network learning and adapting its weights and biases by propagating errors backward through the network to minimize them.

6 Ashish Vaswani et al., “Attention Is All You Need”, 12 June 2017. <https://arxiv.org/abs/1706.03762>

7 <https://www.sustainablefinance.ch/en/resources/artificial-intelligence.html>

8 Our World in Data, “Environmental Impacts of Food Production”. <https://ourworldindata.org/environmental-impacts-of-food>

promise earlier interventions, better patient outcomes, and reduced healthcare system burdens.

### **1.2.3. Education: reducing social inequality, managing academic failure**

While great focus is often given to the environmental applications of AI, its social dimension should not be forgotten. In education, adaptive learning platforms can analyse a student's strengths, weaknesses, and pace of progress in real time, to adjust lessons accordingly and provide step-by-step guidance in mathematics, science, or reading comprehension. When relevant, AI-powered language translation and voice synthesis can deliver lessons in a student's native language, removing linguistic barriers to learning. For learners with disabilities, AI can provide speech-to-text tools, visual aids, or customised lesson formats that make learning more accessible.

On a societal level, such democratisation of knowledge could narrow the educational divide between privileged and marginalised groups, enabling more equitable participation in the economy and civic life. This ideal scenario will however require responsible governance, principled behaviour from AI developers and institutions, and the conscious use of AI as a tool for empowerment rather than substitution. Poorly guided adoption could just as easily erode critical thinking and genuine learning.

### **1.2.4. Climate risk: supporting mitigation and adaptation**

AI-powered forecasting is improving natural disaster preparedness and resilience. Google has developed FloodHub, a system that uses machine learning models to predict flood events, a feature highly relevant for the 250 million people impacted annually by floods. Satellite imagery analysis supports real-time wildfire boundary detection, enabling faster evacuations and resource deployment. In hurricane-prone regions, Google Maps alerts, issued days in advance, help communities prepare and thereby prevent or reduce loss of life.

Post-disaster, AI systems can leverage satellite and aerial imagery to assess damage and identify areas most in need of humanitarian aid – a capability advanced through initiatives such as Microsoft's AI for Humanitarian Action.

Urban planning also benefits from AI projects like Project Green Light, which optimises traffic light timing to reduce congestion and related air pollution and greenhouse gas emissions – and at the same time benefiting public health.

### **1.2.5. Electric Grids: reducing waste and improving renewable energy mix**

More and more electricity companies have adopted AI-driven "smart grid" systems that forecast electricity demand down to the neighbourhood level, enabling more efficient balancing of supply to reduce energy waste. By forecasting among others weather forecast (wind, sun, etc.), these systems also optimize the integration of renewable sources into the energy mix. AI models can also forecast probable outages by correlating weather, geographic, and socioeconomic data, allowing preventative measures ahead of time.

Some companies are preparing for the rapid growth of electric vehicles by managing vehicle charging and discharging based on usage patterns, sometimes compensating owners for allowing their cars to act as distributed energy storage. Such systems reduce energy waste and ensure a better stability of the grid.

## **1.3. Current AI adoption in Swiss financial services**

### **1.3.1. AI in financial services: a longstanding practice**

The use of AI in financial services is not new. Traditional AI systems, that date back to the 1970s, excel in pattern recognition, especially in very large sets of unlabelled data. For financial institutions, these capabilities have proven particularly valuable for risk modelling, fraud detection, or financial forecasting. By learning from historical data, AI models can make predictions that help financial institutions navigate market fluctuations and emerging risks.

In the past decades, financial institutions have therefore developed and fine-tuned on an ongoing basis such quantitative, statistical models to forecast share price or interest rate variations, anticipate customer behaviour, or alert on potential fraud. Since many years, high frequency trading strategies rely on automated trading leveraging models of that type. Financial services companies have also rolled out many automation projects to increase efficiency, reduce costs and the operational risk arising from human errors.

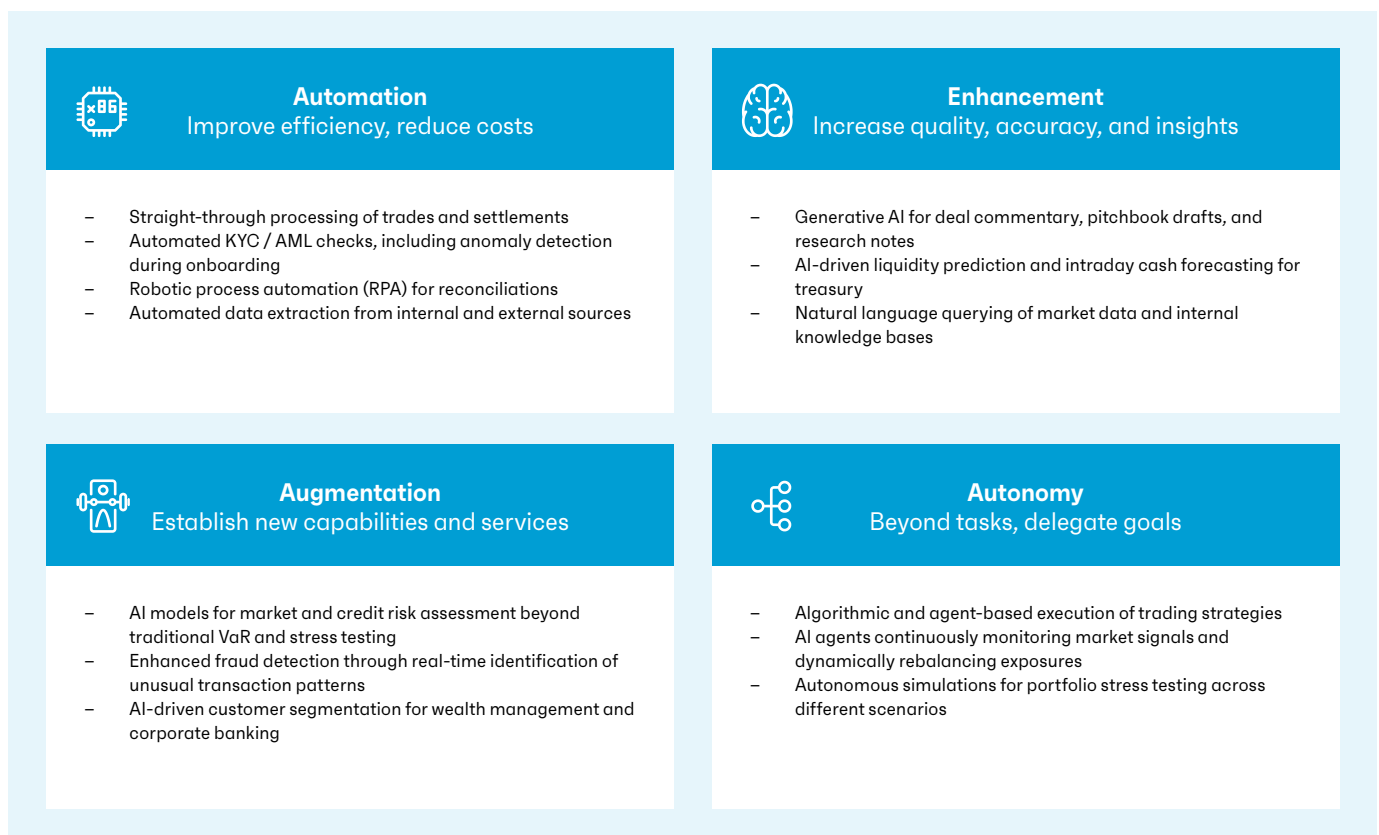
In large organisations, it has for example been years since the last time an employee manually entered into an accounting system the details of a supplier’s invoice – this has long been automated. Artificial Intelligence brings new capabilities to such automation, with systems now able to deal with additional unstructured data sources and to improve their own efficiency over time.

### 1.3.2. New and emerging capabilities

Because one of AI’s primary advantages lies in its ability to process vast amounts of data at a speed that is impossible for humans to achieve, AI applications often have revolved around efficiency increase: automating existing tasks previously done by humans to achieve increases in execution speed.

Yet, it is essential to avoid the “trap of automation” and its short-term benefits, and to consider longer-term value creation. The real potential of AI lies not only in doing the same things faster, but in developing new capabilities or improving current services: increasing quality, accuracy or insights (“enhancement”), establishing new capabilities or services (“augmentation”), or delegating goals and entire processes (“autonomy”).

**Figure 1: The four main benefits of AI applications.**

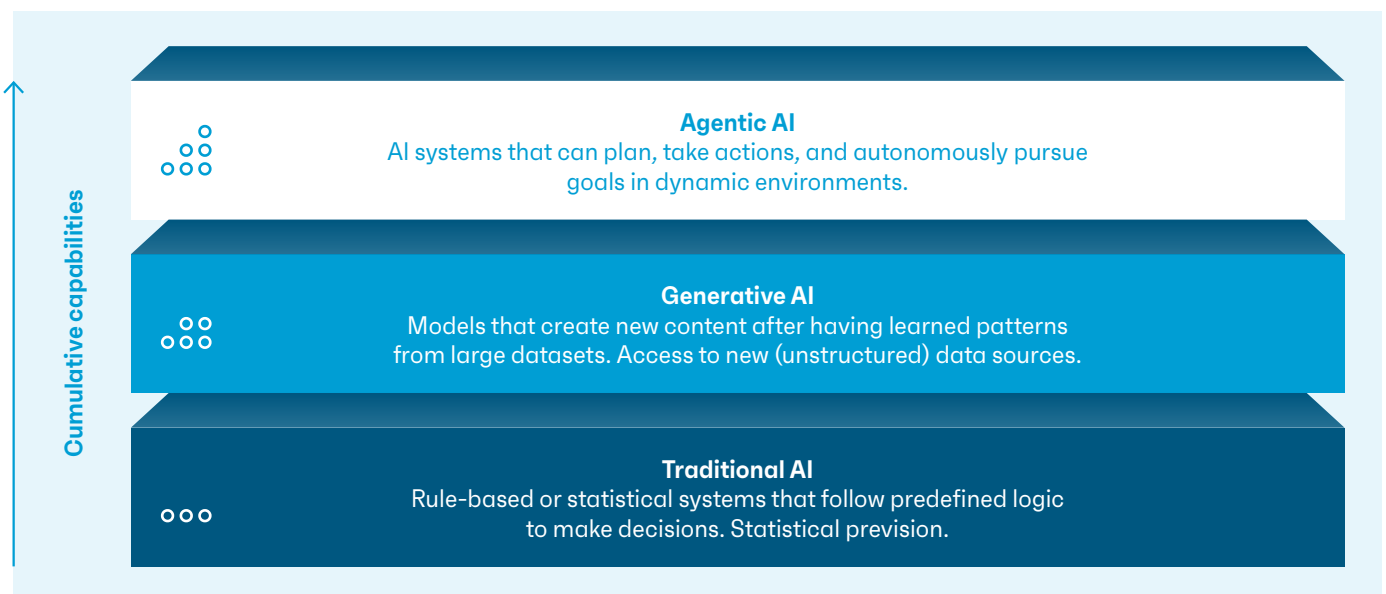


For “Traditional AI” systems to function effectively, data needed to be similar in format, which compelled many suppliers to adapt their data to predefined structures.

The emergence of generative artificial intelligence (GenAI) in the past couple of years has been a game-changing moment in the potential of practical AI applications in the industry. Unlike Traditional AI tools built mainly for classification or

prediction, GenAI models, particularly large language models (LLM) can generate human-like content and interact with information (search, select, summarize...) in a similar manner as humans do. Generative AI’s ability to leverage unstructured data provides access to information that was previously inaccessible, while massively simplifying the user experience and the way to access and interact with that data.

**Figure 2: The rise of Artificial Intelligence capabilities.**



In sustainable finance, where information is complex, dispersed, and growing at a rapid pace, the ability to transform raw inputs into coherent materials brings immediate numerous application potentials. It has made the extraction of insights from large sets of data considerably more user-friendly than before.

It is no surprise that the majority of AI applications relate to the application of LLM to “interact” with data: extract insights from third-party reports or other data sources and provide them in a summarized, readable manner to the end user. Machine learning models can analyse financial reports, sustainability disclosures in corporate reports, social media sentiment, etc. AI-driven natural language processing can interpret such textual data, identifying insights that would otherwise remain hidden in complex documents.

The real opportunity emerges now from combining the two: traditional AI’s analytical power with GenAI’s ability to understand, structure, and generate language. This combination unlocks new possibilities, not only to automate, but to truly augment current services and capabilities.

### 1.3.3. AI adoption for sustainable finance in the Swiss financial services sector

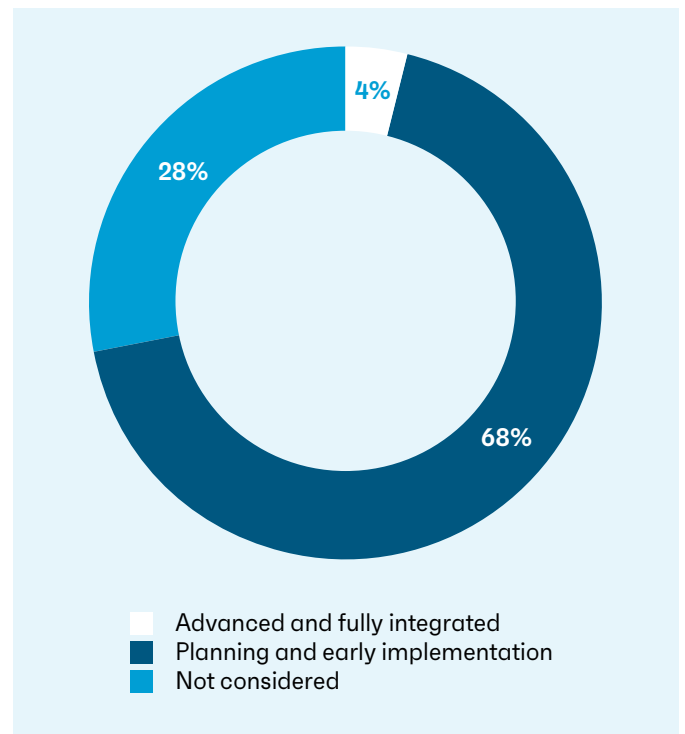
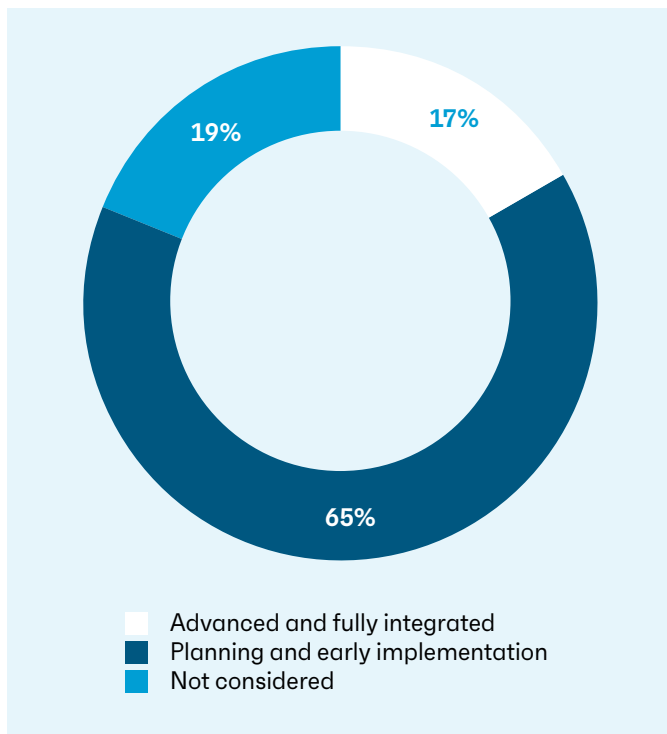
In our “Swiss Sustainable Investment Market Study 2025”<sup>9</sup>, we asked financial services companies about the stage of AI adoption within their organisation, as well as drivers, use cases, challenges and organisational needs specifically related to AI in the context of sustainable finance.

Only a minority of banks, asset managers or asset owners had already rolled out AI-supported tools, in production, in some of their operations. Out of them, only 18% reported using AI for applications specifically related to sustainable finance. In other words, despite the growing interest of financial services organisations for AI capabilities, its targeted application in sustainable finance remains limited, suggesting that AI’s potential in this area is still largely untapped (Figure 3).

**Figure 3: Stage of artificial intelligence adoption in the organisation**

Asset Managers (n=48)

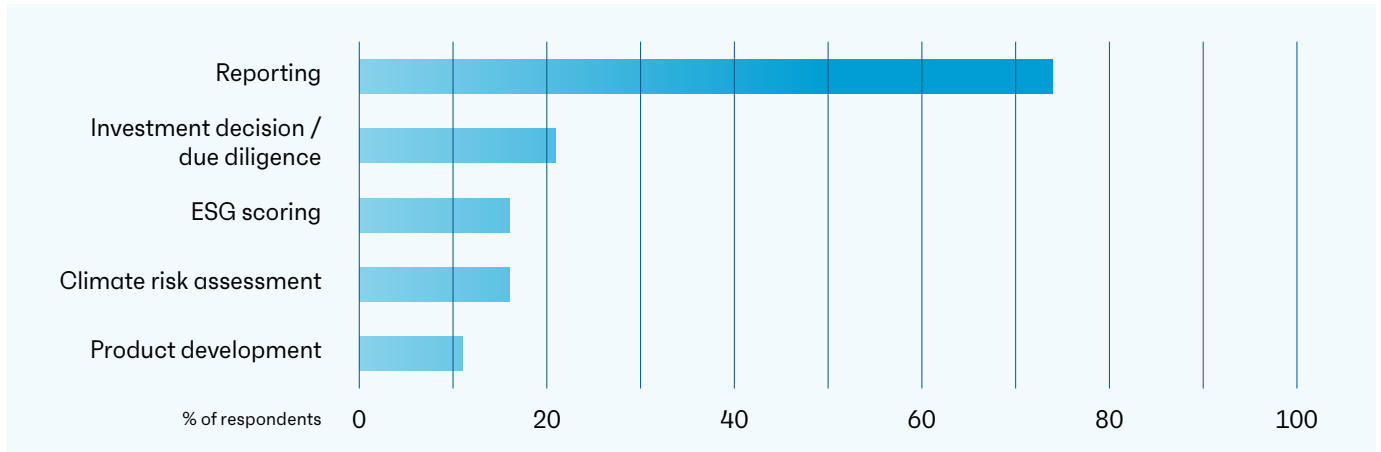
Asset Owners (n=25)



Among respondents that apply AI in sustainable finance, the predominant use case is reporting, far ahead of business-oriented use cases such as product development or investment decision-making. The industry appears to see immediate benefits in compliance-driven AI applications (focusing on disclosure and reporting requirements), with a longer term horizon for business-driven benefits (Figure 4).

<sup>9</sup> SSF, Swiss Sustainable Investment Market Study 2025, <https://www.sustainablefinance.ch/en/our-activities/ssf-publications-3037.html>

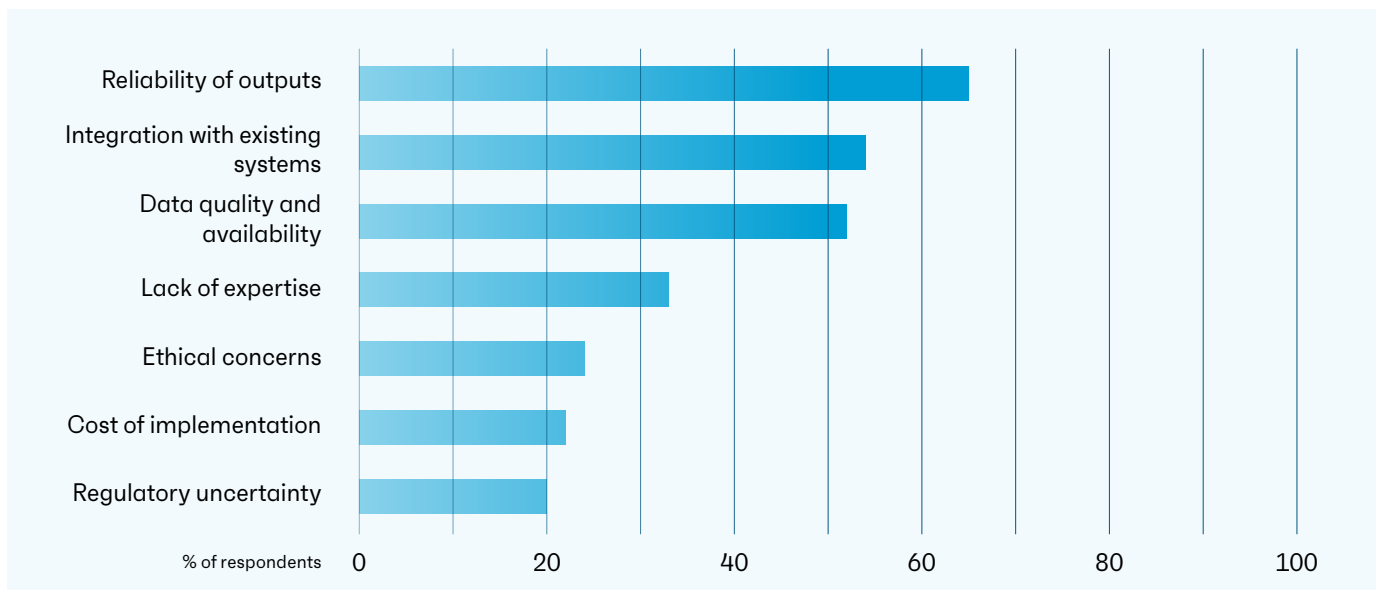
**Figure 4: Artificial intelligence use cases in sustainable finance for asset managers and asset owners (in % of respondents) (n=19)**



The top three obstacles to broader AI adoption cited by the industry are reliability of outputs (65%), integration with existing systems (54%), and data quality and availability (52%). These technical and structural issues linked to data

gathering and processing appear to outweigh concerns related to cost, expertise, ethics, or regulatory uncertainty (Figure 5).

**Figure 5: Key challenges in artificial intelligence implementation for sustainable finance for asset managers and asset owners (in % of respondents) (n=47)**



More details can be found in our “Swiss Sustainable Investing Market Study 2025” on page 51.

# Artificial Intelligence Applications in Sustainable Finance

# 02

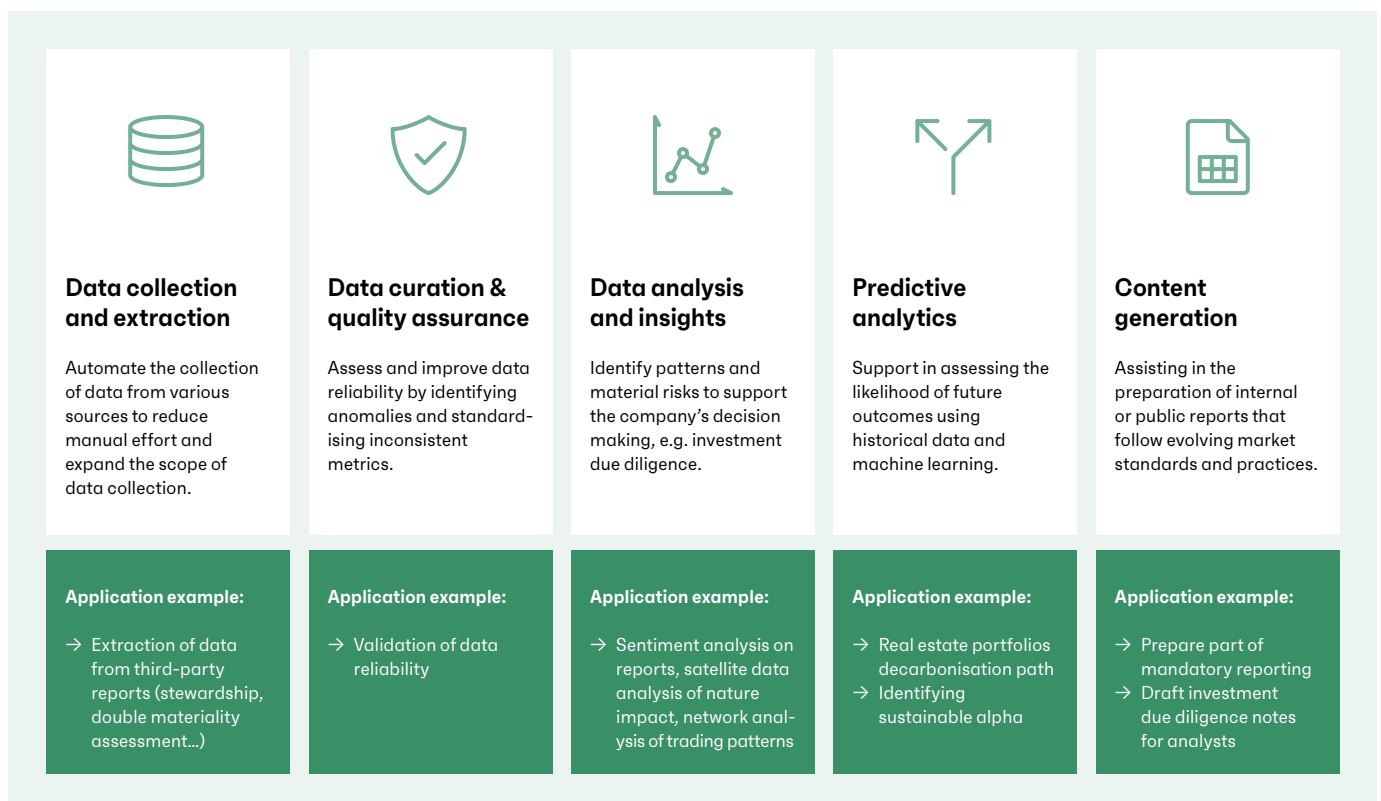
In the past year, we have worked with representatives from the Swiss financial services industry and with Swiss universities to specifically identify current use cases of AI for sustainable finance.

This chapter goes through these use cases, articulating the challenges faced by the industry and how AI has been applied to help financial services companies solve these challenges.

We have categorized the use cases of the industry along a “data-to-decision” workflow: identify and extract source data, curate and validate that data for quality and completeness,

analyse the selected information to derive insights (whether assessing the current state or modelling forward-looking scenarios) and finally, make and document a decision. This pattern underpins a wide spectrum of critical processes: designing a new financial instrument; conducting due diligence on a prospective acquisition; preparing mandatory corporate disclosures and reports; or performing a targeted review by a risk, compliance, or internal audit function.

**Figure 6: AI applications in sustainable finance along the data-to-decision workflow**



## 2.1. Data Collection and Extraction

Over the past 15 years, the volume of data produced by humanity globally has multiplied 70-fold<sup>10</sup>. In the financial services sector, expectations have also massively grown over time, with increased disclosure and reporting requirements, between financial services institutions and towards end investors.

The gathering, curation and processing of always increasing volumes of data from sometimes inconsistent sources has become a significant challenge for market participants, and a time-consuming activity.

Intelligent Process Automation (IPA), which combines Robotic Process Automation (RPA) with AI techniques such as Natural Language Processing (NLP) offers a scalable way to automate complex data-gathering workflows. Unlike traditional automation, IPA not only performs repetitive tasks but also extracts and harmonizes information from diverse and unstructured data sources, ensuring consistency and accuracy across systems. In addition, IPA continuously learns from experience through machine learning, allowing it to refine its models, adapt to new data patterns, and deliver increasingly accurate results over time.

The below examines how IPA can be applied to a selection of concrete sustainable finance use cases.

### 2.1.1. Data mining for Corporate Sustainability Report

Annual corporate sustainability reports often represent hundreds of pages, with metrics buried in narrative text, tables, and appendices. IPA combines document parsing with NLP to extract relevant disclosures (e.g. Scope 3 emissions, human rights policy updates) and map them against sustainability taxonomies or frameworks such as the EU Taxonomy and ISSB standards. This accelerates analyst workflows and reduces the risk of overlooking key datapoints.

### 2.1.2. Double materiality

A double materiality assessment requires companies to consider both how sustainability issues affect their business and how their activities affect people and the environment. The process typically involves reviewing large volumes of documents, data, and stakeholder input, which can be resource-intensive and inconsistent across teams. AI can assist by efficiently gathering and organizing relevant information, highlighting recurring themes, and helping analysts identify which topics are most significant. This allows organizations

to reduce the effort dedicated to data extraction and gathering, and focus on assessment, analysis and decision-making.

### 2.1.3. For risk management: climate risk data integration

Sustainable finance risk teams increasingly rely on climate datasets from diverse sources, such as NGFS<sup>11</sup> scenarios, physical risk models, or municipal flood maps. IPA can automatically pull, reconcile, and transform these datasets into standard formats for integration into internal risk models. For example, a bank can automatically ingest updated sea-level rise projections from national meteorological agencies and link them to property loan portfolios for ongoing exposure monitoring.

### 2.1.4. For asset owners: stewardship data extraction

Asset managers produce extensive stewardship reports outlining their ESG engagement activities with portfolio companies. IPA systems can extract, classify, and quantify engagement outcomes, such as climate transition commitments, diversity policy changes, or supply chain improvements, directly from PDF reports, meeting notes or other sources. By automating data collection and structuring, IPA significantly reduces manual effort and processing time, enabling asset owners to focus on other activities such as assessing engagement effectiveness, exercising judgment, and integrating these insights into follow-up conversations with asset managers.

10 Global Datasphere by the International Data Corporation. [https://my.idc.com/getdoc.jsp?containerId=IDC\\_P38353](https://my.idc.com/getdoc.jsp?containerId=IDC_P38353)

11 Network for Greening the Financial System. <https://www.ngfs.net/en?lang=en>

# Simplifying Double Materiality Assessment via Intelligent Process Automation



## → Challenge to overcome

The requirement to perform and disclose the result of a “double materiality assessment” (DMA) was introduced in the European Union by the Corporate Sustainability Reporting Directive in 2022. In Switzerland, the Code of Obligations (Art. 964a-c) contains a similar, although less precise and prescriptive, expectation. A number of Swiss companies also perform DMAs as part of their sustainability strategy activities.

The DMA is (as of now) a mostly qualitative procedure meant to assess how a company affects ESG components and, in turn, how the ESG components can affect said company. Performing a DMA requires to first determine a series of impacts, risks and opportunities (IRO) and then go through any available information source and perform an assessment of factors such as Severity or Likelihood of each. As the list of IROs can be very large, the process can be very time consuming for the teams performing it.

## → AI application / use case

The time intensity of the process for humans made it clear that intelligent automation, typically via a Large Language Model (LLM), could provide valuable support in improving the sourcing of information from the documents as well as speeding up the process. The solution relies on the following steps:

1. Defining a list of all IROs to be assessed, with the relevant factors (e.g. Severity, Scope and Likelihood) for each of those IROs.
2. Drafting the related questions that a human team member would ask themselves when screening underlying documentation.
3. Recording these questions in an Excel template, optimizing the wording of the question via prompt optimization, so that they are tailored to be sent to an LLM rather than a human operator.
4. Selecting and providing to the LLM a list of sources (reports and other documents) identified as relevant for the assessment.
5. Writing a Python script, that goes through the Excel file, formulates the questions for each impact, communicates with the LLM, retrieves the answer and populates them in the respective cells in the Excel.

The final result is a populated Excel file with a sheet containing the list of all IROs, assessments of each factor for each of those IROs (score and rationale for the score) and references to the documents which were relevant in assessing them.

## → Use case key beneficiaries

- Relationship Managers
- Portfolio Managers
- Research teams, macroeconomists
- Control functions
- Support functions (HR, CFO, ...)
- Other: Advisors, auditors

## → Benefits of AI use case for financial services sector

This Intelligent Process Automation requires human intervention in certain steps (as do almost all successful AI use cases) but allows for significant efficiency gains in the whole process, freeing up resources for other tasks.

## → Supporting technology

The solution is based on an LLM<sup>12</sup> accessible through the Azure AI Foundry. The Python script extracts the questions from the Excel file, sends them to the LLM, and records the responses in the file. The solution integrates a Retrieval Augmented Generation (RAG) component, i.e. the LLM navigates the files it had been provided by the user to generate answers. This has the effect of reducing hallucinations<sup>13</sup> and ground responses in source content.

12 The implementation allows for the usage of different models, such as ChatGPT's latest versions.

13 An "hallucination" is a failure mode in which a language model generates fabricated facts, inaccurate claims, unjustified reasoning, or invented details that are not grounded in its training data, the user's prompt, or any verifiable source.

# Real-Time Market Intelligence for Swiss Real Estate

## → Challenge to overcome

Real estate is the largest asset class operating without market-wide transparency. In Switzerland, data on buildings and investment vehicles is scattered across annual reports, PDFs, and proprietary databases. No unified view exists for comparing assets at scale or assessing ESG risks across portfolios.

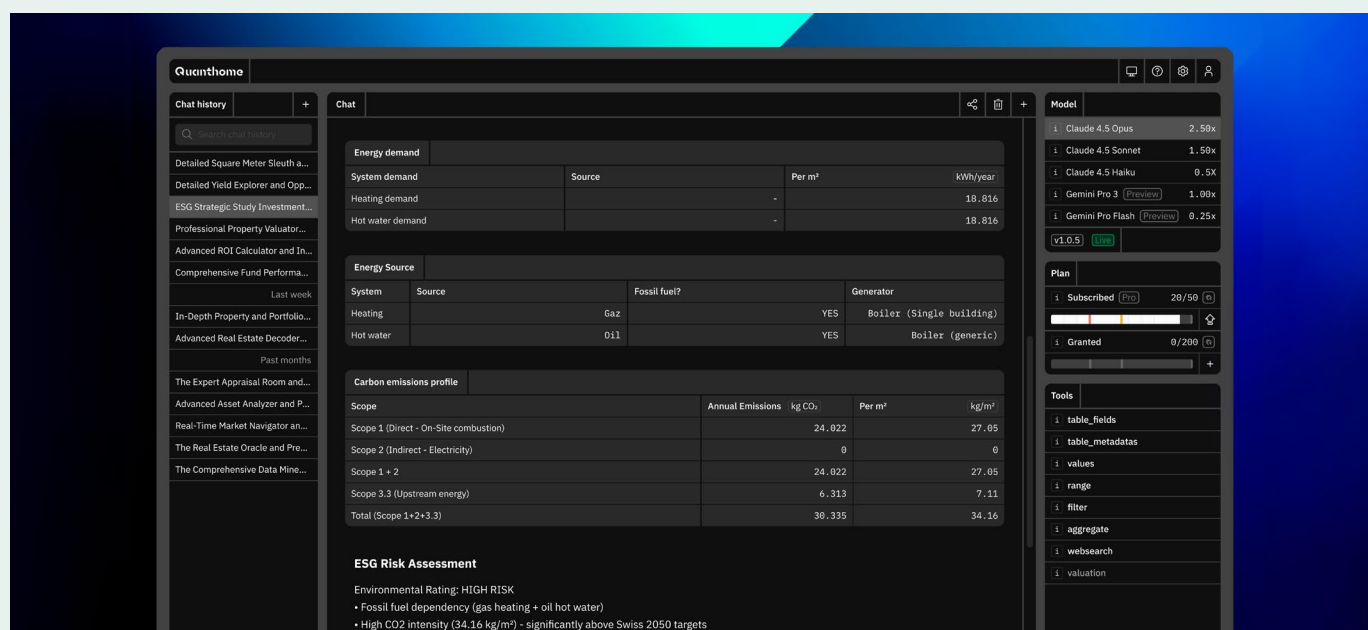
This opacity distorts prices, leaves climate exposure unpriced, and directs capital without visibility into what it funds, following the well-know principle: what does not get measured cannot be managed.

## → AI / ML solution

QuanthomeAI strives to address this issue by connecting users to a unified database through natural language queries. The system connects to a proprietary database covering all Swiss buildings and over 150 investment vehicles, linking physical assets to the portfolios that hold them. Users are able query ESG metrics, financial performance, risk exposure, and market comparisons in natural language. An agentic framework allows the AI to iterate and self-correct, while domain-specific prompts reduce hallucination and maximize relevance. The above solutions allows to address different dimensions of the challenge described above:

- Improves price discovery. Market-wide data access reduces information asymmetry. Risks get priced more accurately, and ESG mispricing begins to correct.
- Supports faster execution. Instant access to building permits, renovation status, comparables, and ESG signals accelerates underwriting across assets and vehicles.
- Allows for ESG integration. Climate risk, CO<sub>2</sub> exposure, and energy performance become visible and comparable across portfolios, supporting decision-making.
- Fosters democratization. Institutional-grade insights at accessible prices narrow the gap between large and small investors.

Figure 7: Quanthome AI generating a building-level ESG analysis



## → Use case key beneficiaries

- ☒ Relationship Managers
- ☒ Portfolio Managers
- ☒ Research teams, macroeconomists
- ☒ Control functions
- ☒ Support functions (HR, CFO, ...)
- ☒ Other: Advisors, auditors

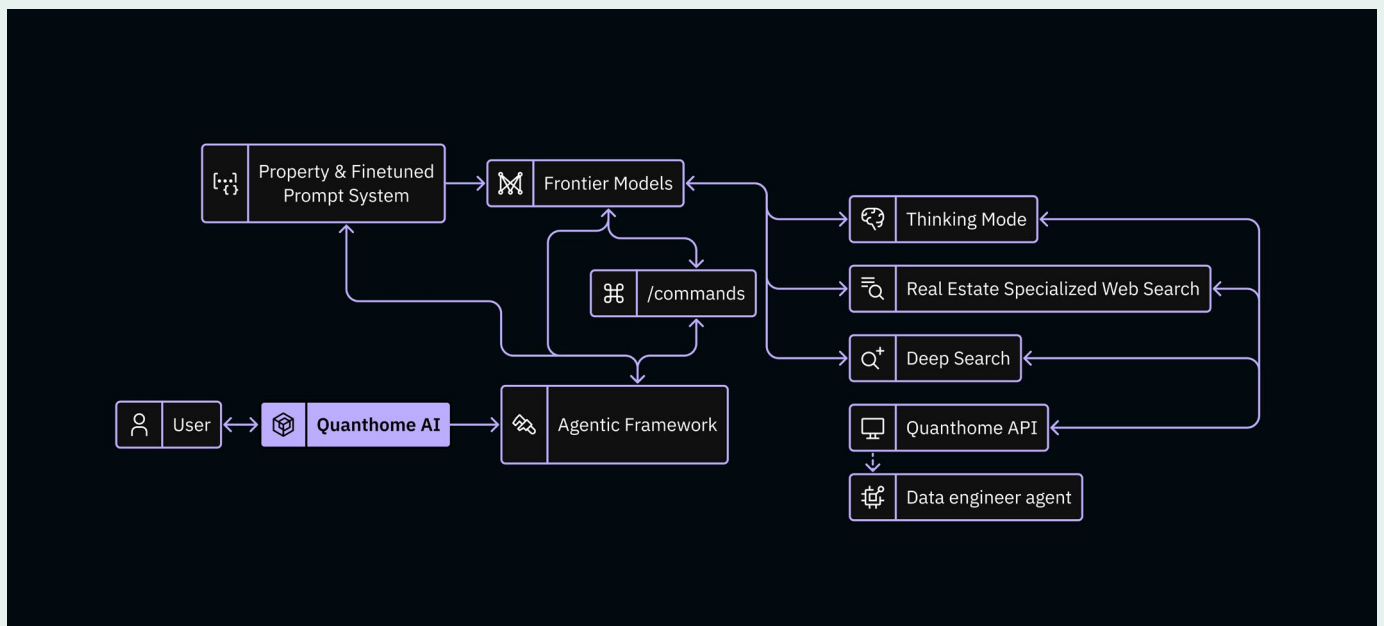
## → Benefits for the financial services industry

In a similar manner as Bloomberg did for securities, Quanthome strives to build the data infrastructure for real estate to become measurable and comparable. By creating market-wide transparency from fragmented real estate data, it increases the accuracy in price discovery, allows for faster and lower-cost investment decisions, and systematic integration of ESG and climate risks across portfolios. It also democratizes access to high-quality insights, ensuring capital is allocated with better visibility into risks, sustainability impact, and asset value.

## → Supporting technology

- Data ingestion. The platform ingests structured data (financials, ESG metrics, building attributes) and live news through API calls. A scalable Retrieval-Augmented-Generation system is under development to integrate unstructured sources such as annual reports.
- Proprietary ML pipelines. Valuations, rent levels, and CO<sub>2</sub> emissions are modeled using hybrid techniques including Lasso regression, gradient boosting, and random forest algorithms. These power the structured database that Quanthome AI queries in real time.
- LLM integration. Users interact with the database through a conversational interface. Questions asked in natural language are interpreted by the LLM, which translates them into structured queries, retrieves relevant data, and returns answers in readable formats. The system can generate reports, tables, or comparisons depending on user needs. Users access Claude 4.5 (Opus, Sonnet, Haiku) and Gemini 3.0, selecting models based on task complexity.
- Agentic framework. Unlike traditional chatbots that respond in a single pass, the agentic framework operates through an iterative loop. The AI calls APIs, executes code, retrieves data, and validates outputs before responding. If results are incomplete or inconsistent, it identifies the gap and re-executes automatically. This enables complex analytical workflows without manual intervention.

Figure 8: Quanthome AI system architecture



## 2.2. Data Curation and Quality Assurance

A successful rollout of AI does not only depend on the selection of the most suited model, but in a first stage on the quality of the underlying data. Models can only perform well if they are trained on accurate, relevant, and representative data. Poor-quality input data and fragmented datasets remain among the main obstacles to the adoption of AI in sustainable finance (See “AI adoption in the Swiss financial services sector” on page 15). Ensuring the accuracy and consistency of ESG data is also crucial for financial institutions seeking to make informed decisions.

Interestingly, AI itself can contribute to curating sustainability datasets and selecting the most reliable and trustworthy data, for further use along the value chain – e.g. to prepare internal sustainability dashboards, or to draft corporate sustainability reports.

AI models can detect inconsistencies, for example when a company’s reported renewable energy usage does not align with its disclosed energy generation capacity. They can flag potential errors such as anomalous carbon intensity figures or sudden, unexplained changes in certain ratios.

AI models can assess the reliability of data points from various data sources (data providers, self-reported data) by applying a sequence of validation steps. They can first identify outliers in time (abnormal change year-on-year) or changes that do not seem to align to other economic indicators of the company: evolution of operational expenditures, capital expenditures, number of employees, and other financial or operational indicators. They can then evaluate the data points against sectoral benchmarks and peer companies with similar profiles: industry and sector of activity, size, regional presence, and growth rate. Significant deviations trigger flagging for human review, while alignment with peer patterns increases the probability of correctness. All these features taken together can then adjust the reliability of each data point available and select the one with the highest likelihood to be accurate.

Specific applications in sustainable finance include (again without the ambition of exhaustivity) the following areas.

### 2.2.1. Harmonizing intercompany data

AI can detect and correct inconsistencies in how companies or data providers define their reporting boundaries. By linking legal entities, subsidiaries, and geographic segments across multiple sources, it helps ensuring that financial and sustainability data are consolidated on a comparable basis. This reduces double counting or gaps when corporate structures change, improving dataset consistency over time.

### 2.2.2. Standardizing GHG emissions data units from various sources

AI can automatically recognise and normalise reporting units, scopes, and methodologies across disclosures – for instance, converting energy data expressed in different units or identifying whether emissions are reported on a market- or location-based basis. This harmonisation allows analysts to work with data that is comparable and properly scaled before it is used in analysis or model training.

### 2.2.3. Monitoring effectiveness of data points

AI models can match reported green bond projects (e.g. a specific solar farm or wastewater plant) with satellite or sensor data and test whether the reported capacity, location, or output is broadly consistent with observable evidence. Instead of “verifying” impact, the system flags cases where reported performance looks implausible or where data is simply missing, so human reviewers know which bonds need deeper scrutiny.

### 2.2.4. Harmonizing climate-risk related data

Financial institutions often buy climate risk data from multiple providers, each with different scenarios, time steps and spatial grids. AI models can help interpolate and reconcile these sources, detect outliers (e.g. a flood depth layer that looks inconsistent with neighbouring tiles), and ensure that each asset has one consistent set of risk inputs before risk models are run.

By embedding such validation and quality assurance mechanisms early in the process, AI builds the data integrity foundation necessary for scaling any other AI application further down the value chain.

# Building trustworthy data foundation via machine-learning



## → Challenge to overcome

For most portfolio managers and asset owners, data reliability and quality remain a central challenge in sustainable investing. Human error in capturing reported data<sup>14</sup>, combined with corporate disclosures that are often incomplete, inconsistent, or difficult to compare across peers and jurisdictions, undermines data confidence. In addition, differing reporting boundaries and definitions further complicate consistency.

These issues reinforce the well-known “garbage in, garbage out” problem: unreliable inputs lead to unreliable outputs, limiting investors’ ability to make sound decisions, manage risks effectively, or meet regulatory and client expectations.

## → AI application / use case

Clarity AI designed an AI-native data reliability system to address these core issues. Reliability checks are embedded across the entire data lifecycle, starting at ingestion, rather than treating data quality as a control applied at the end of the process.

This design choice is critical: errors in sustainability data tend to propagate downstream and become harder and more costly to detect and correct. By embedding AI early in the collection process, error risk is reduced before flawed data can cascade down into models, metrics, and user outputs. This enables rapid feedback loops to identify inconsistencies, anomalies, or gaps far more efficiently than traditional human-intensive workflows or general-purpose AI models.

At ingestion, AI selectively automates the collection of high-reliability data, such as qualitative disclosures (e.g. policies). For text-based metrics, Large Language Models (LLMs) achieve high accuracy, enabling the process to be largely automated and limiting human oversight to a minimum. Lower-reliability data (more complex metrics, primarily quantitative data) are routed into targeted human-in-the-loop workflows. Natural language processing (NLP) tools guide subject matter experts to relevant sections of reports and flag anomalies. This allows analysts to focus on evaluating context, credibility, and relevance, rather than searching through entire documents.

Reliability is further reinforced through layered verification. Machine learning models assess the likelihood that a data point is accurate using a broad set of signals (including historical trends, industry benchmarks, expected ranges, units, the nature of the source document, etc.), the collector’s expertise and past performance. Additional human review is triggered when there is a higher probability of inaccuracy.

Self-reported data by companies is not treated as correct by default. For example, values such as Scope 3 emissions are systematically challenged against expectations and contextual signals, built with AI. Where deviations indicate missing inputs or non-representative disclosures, further review is initiated. If the reported information cannot support a reliable value, e.g. due to insufficient disclosure, we generate AI-driven estimates or rely on inherited data from higher levels of the corporate structure. In all cases, estimates are clearly marked as such, with transparency about the underlying data and methodology.

<sup>14</sup> Clarity AI’s last comparative analysis across three main providers found that human errors accounted for more than 80% of the errors identified.

**Figure 9: Clarity AI's Data Collection & Curation Workflows**



Once an initial dataset is established, the system manages changes over time. Companies may restate data due to methodological updates, boundary changes, or revised assumptions. Rather than assuming that previously reported historical values remain unchanged, detection models identify when and why data reported for prior years differ from values disclosed in earlier reporting cycles, preserving this context. This enables us to explain not only what has changed, but why; a critical requirement for maintaining confidence and interpretability. Data is then curated or aggregated according to the user's needs. For example:

- Some users may choose to include estimates, while others rely solely on reported data.
- Some may view reported values even if they fall outside methodological criteria, while others restrict outputs to data that meet predefined standards.

Throughout the process, transparency and traceability are foundational. Every data point is traceable to its source, with a record of how it was collected, verified, curated, and, where applicable, estimated. Together, this approach delivers:

- Early-stage error reduction through AI embedded at ingestion
- Scalability without linear increases in manual effort
- Targeted use of human expertise where judgment adds the most value
- Robustness through layered AI and human validation
- Trust through full transparency and explainability

The result is a scalable, explainable data foundation, offering reliability levels exceeding 99%, supporting confident and informed investment decision-making.

## → Use case key beneficiaries

- Relationship Managers
- Portfolio Managers
- Research teams, macroeconomists
- Control functions
- Support functions (HR, CFO, ...)
- Other: Chief Data Officers

## → Benefits for the financial services industry

By systematically identifying, assessing, and filtering unreliable inputs across the data lifecycle, this AI-powered approach enables financial institutions to rely on more accurate, consistent, and transparent sustainability data. This reduces the risk of errors across investment decision-making, portfolio construction, risk monitoring, and regulatory reporting, while strengthening confidence in underlying analyses.

Beyond sustainability use cases, curated datasets produced through this process also provide a robust foundation for other AI models and analytical applications. High-quality, well-governed data is a prerequisite for achieving reliable and usable results, and the combination of AI-driven controls with expert oversight helps ensure that this prerequisite is met.

## → Supporting technology

To deliver this level of performance, we leverage a specialized technology stack designed to address the complexity and variability of sustainability data.

### – Document ingestion and language processing

We utilize workflows – including LLM-powered agents, Optical Character Recognition (OCR), and automated translation – to ingest unstructured corporate disclosures across multiple formats and languages. This process is driven by our proprietary platforms: Igloo, which manages the data ingestion lifecycle, and Forge, which enables the development, validation, and deployment of agentic LLM data extractors at scale. Together, these technologies transform fragmented, non-standardized reports into high-quality, machine-readable data.

### – Generative AI and Large Language Models (LLMs)

We apply generative AI, including LLMs, to support tasks such as document understanding, information extraction, and contextual interpretation of dense corporate disclosures. The system is deliberately model agnostic: different models exhibit different strengths across tasks, and performance evolves rapidly as new models emerge. This architecture allows to continuously evaluate and deploy the models that are most effective for each specific use case, rather than relying on a single fixed solution.

### – Machine learning (ML) for reliability and quality control

We use supervised machine learning models based on decision-tree architectures to assess data reliability, detect inconsistencies, and identify anomalous patterns that would be difficult to uncover through manual review alone. These models incorporate multiple parameters and signals drawn from the data collection and verification process. Model quality is ensured through robust training, validation, and ongoing performance monitoring procedures. These same models also support the generation and validation of ML-driven estimates when reported data is unavailable or does not meet defined reliability criteria.

## 2.3. Data Analysis and Insights

Today, most successful applications at scale of AI in finance relate to the performance of diagnostics and assessments on third-party content – e.g. screening new reports to identify potential controversies related to investee companies, performing sentiment analysis on corporate disclosures and annual reports. This goes beyond the applications presented in section 1 (“Data collection and extraction”, page 19): it includes a data extraction step but then applies sentiment analysis to the extracted data to derive insights designed to directly support decision making.

The below highlights a selection of such applications, without, again, pretending to be exhaustive.

### 2.3.1. Sentiment and narrative analysis

NLP models can process vast volumes of news articles, third party reports, parliamentary debates, and social media posts to gauge public and stakeholder sentiment toward companies or sectors. For example, a bank may monitor climate litigation coverage or local community reactions to a mining project to inform its credit risk assessments. These insights can also identify emerging issues, such as biodiversity reduction, before they are reflected in official sustainability ratings.

### 2.3.2. Greenwashing detection

AI can cross-check corporate sustainability claims against independent data sources, such as environmental NGO datasets, satellite imagery, or verified emissions inventories. If a company reports achieving emissions reduction in certain production sites while satellite-based GHG tracking suggests otherwise, the system can flag discrepancies for further review. This supports investor or lender due diligence and strengthens the credibility of sustainable finance products.

### 2.3.3. Sustainability theme and controversy mapping

Clustering algorithms can group companies, projects, or issuers based on shared sustainability themes (e.g. renewable energy adoption, water usage intensity, or labor practices) making it easier to compare peers. These systems can also map controversies by topic, geography, and severity, helping portfolio managers understand concentration risks.

### 2.3.4. Alternative data supporting various verifications

AI can extract sustainability-relevant insights from unconventional sources:

- Satellite imagery to assess deforestation, mining expansion, or renewable energy deployment.
- Internet-of-Things sensor data to monitor water quality near financed industrial sites.
- Shipping and trade data to trace supply chain sourcing for commodities linked to deforestation or forced labour.
- Such insights provide evidence for impact reporting and help validate sustainability performance claims.

### 2.3.5. Carbon and biodiversity footprint attribution

Advanced data processing can match investment holdings to emissions databases or biodiversity impact models. For example, an AI pipeline can attribute a investment fund’s portfolio emissions down to the facility level, or quantify biodiversity impacts based on the location and activity type of financed projects.

# Enhancing Carbon Market integrity through AI-Driven Network Analysis



## → Challenge to overcome

Carbon markets are systems that enable the trading of carbon emission allowances or credits, assigning a monetary value to greenhouse gas emissions in order to drive climate action. By incentivizing cost-effective emissions reductions, they help allocate resources toward the most efficient decarbonization solutions.

However, the credibility and the effectiveness of global carbon markets is increasingly undermined by the presence of “hot air” credits, i.e. emission certificates that do not correspond to real emissions reductions. These credits have the potential to distort price signals, mislead investors and ultimately may slow global climate progress<sup>15</sup>.

The absence of transparency and verifiable data across carbon registries and projects serves to exacerbate this issue, thereby hindering regulators and market participants from identifying and avoiding such credits. The approval of the standards needed to operationalize Article 6.4 of the Paris Agreement, combined with recent signals from the European Commission on the possible future role of international certificates in EU climate policy, has renewed attention to strengthening the environmental integrity of international carbon trading mechanisms.

## → AI application / use case

To address this, ZHAW is developing an AI-based network analysis tool designed to uncover governance weaknesses and systemic vulnerabilities, detect patterns of risk, and enhance market accountability. The purpose of this initiative is to establish a foundation for carbon markets that is more credible, with the aim of aligning financial flows with meaningful climate action.

The framework traces carbon credit flows, detect anomalies, and identify key intermediaries involved in the circulation of so-called “hot air” credits. This model synthesizes diverse trading and organizational datasets to map complex trading networks, reveal high-impact actors, and quantify the influence of countries, sectors, and organizations on carbon market dynamics.

The framework operates across multiple analytical layers. The study’s initial approach utilizes regression techniques and network analysis to trace the movement of emission reduction units (ERUs) from issuance to utilization within compliance markets, such as the Swiss and the EU emissions trading

system. This approach provides insights into the behaviours and structures that enable the circulation of low-quality credits, thereby exposing the weaknesses in current market governance.

The initiative combines data from various, often disconnected sources, such as the EU registry and national registries from Switzerland, Ukraine, and Russia. This enables the creation of detailed cross-border datasets that uncover critical trading pathways and systemic vulnerabilities in the global carbon market, emphasizing the influential role of financial intermediaries and commodity traders.

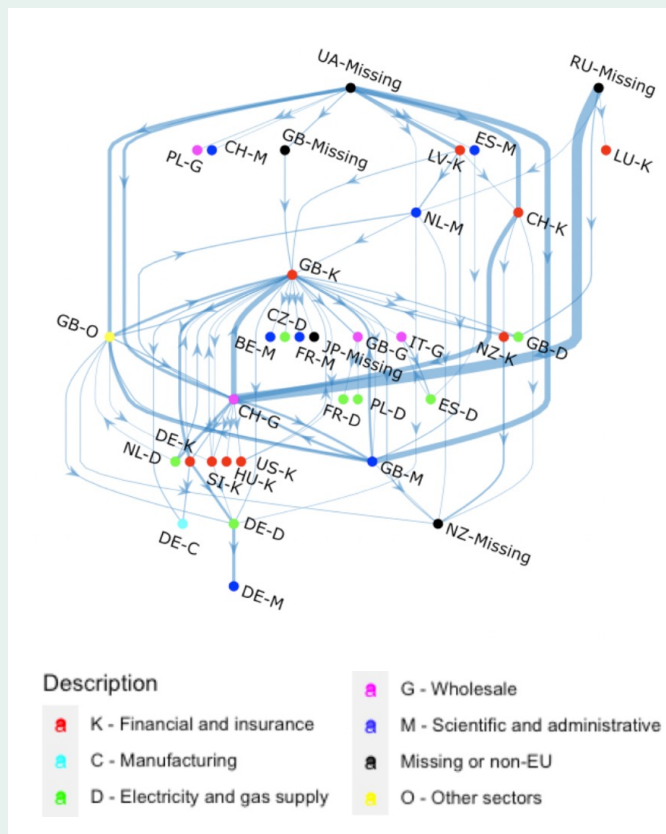
The network analysis indicates that Swiss financial and commodity trading entities, particularly those in the oil and gas sector, along with insurance and financial firms in the UK, play a central role in moving ERUs through the market which

<sup>15</sup> Betz et al., “The Carbon market Challenge”.  
<https://www.cambridge.org/core/elements/carbon-market-challenge/9261122253200C956EAF02B5C9AF53C8>

includes “hot air” credits that often originate from Russia and Ukraine. These findings reinforce the importance of enhancing transparency and accountability in carbon trading, addressing systemic risks, and the need for enhanced market oversight.

As carbon markets evolve and assume a more pivotal role in climate finance, it becomes imperative to ensure the integrity of these systems. This initiative shows the potential of advanced analytics to strengthen the integrity, transparency, and alignment of carbon markets with global climate objectives.

**Figure 10: Emissions Reduction Units Trading.**



Source: Kotsch et al., Network Analysis of the Global Trade of “Hot Air”<sup>16</sup>

<sup>16</sup> Kotsch et al., Network Analysis of the Global Trade of “Hot Air”. Key Lessons for the Paris Agreement, 2024. [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=5038486](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5038486)

## → Use case key beneficiaries

- Relationship Managers
- Portfolio Managers
- Research teams, macroeconomists
- Control functions
- Support functions (HR, CFO, ...)

## → Supporting technology

The project applies advanced anomaly detection methods, including unsupervised learning algorithms such as clustering and isolation forests, to identify irregularities and risk patterns in carbon trading transactions where labelled data is scarce.

In parallel, ZHAW employs graph-based modelling to represent global carbon markets as interconnected networks, with entities (sellers, buyers, and intermediaries) as nodes and certificate transfers as edges. By analysing network metrics such as centrality and clustering, the framework can reveal hidden dependencies, concentrated market power, and circular trading loops.

Together, these methods enable the detection of systemic vulnerabilities ranging from high-risk intermediaries and speculative bubbles to the movement of credits with questionable environmental integrity into compliance markets.

# Strengthening trust in Corporate Sustainability Reports with NLP and Sentiment Analysis



## → Challenge to overcome

Corporate sustainability reporting is a critical tool for evaluating climate and environmental performance. Yet, in practice, its value is undermined by complexity, inconsistency and the absence of verifiable standards. Reports are often lengthy and difficult to analyse manually, while current voluntary frameworks, such as TCFD or TNFD, are applied haphazardly. As a result, investors, regulators, and civil society struggle to accurately assess corporate sustainability performance, increasing the risk of misinformed decisions.

To overcome these challenges, the Department of Finance at the University of Zurich has developed two AI-driven tools: ASKCLIMATE and ASKNATURE. These tools can be used by the public or by professionals, e.g. investment managers, and provide free access to high-quality sustainability analysis and bring greater structure and traceability to corporate sustainability disclosures. The University of Zurich is committed to provide these tools as open-source solutions to enhance transparency for all stakeholders.

## → AI application / use case

### AskClimate

ASKCLIMATE tackles two major obstacles in AI-assisted analysis: large language model (LLM) hallucination and the difficulty of integrating domain expertise into AI workflows.

ASKCLIMATE employs a pipeline-based architecture comprising four modules: report embedding, report summarization, TCFD conformity assessment, and customized question answering.

The system produces summaries across the eleven TCFD pillars and assigns each report a numerical conformity score (0–100).

### AskNature

ASKNATURE evaluates the quality of nature-related disclosures in corporate sustainability reports (CSRs) using a structured tree of questions aligned with TNFD's fourteen core themes.

Existing reports often lack granularity, consistency, and comparability, making it difficult to assess how corporate activities depend on or affect nature. Moreover, many disclosures remain narrative-based and are not easily quantifiable or measurable, limiting their usefulness for investors and regulators.

At the core of ASKNATURE is a double materiality paradigm, which classifies disclosures by the direction of impact – company-to-nature, nature-to-company, or neutral – and applies sentiment analysis to determine whether these impacts are positive or negative. A geographic analysis module further assesses whether reported actions are aligned with ecologically significant regions mentioned in the disclosures. A scoring system is used to evaluate the quality of disclosures based on specificity, enforceability, and the substance of reported commitments.

A recent analysis of the sustainability reports of Nature Action 100 (NA100) companies by ASKNATURE revealed that, while many reports referenced TNFD themes, they frequently lacked legal robustness, enforceable commitments, and alignment with tangible biodiversity outcomes.

## → Use case key beneficiaries

- Relationship Managers
- Portfolio Managers
- Research teams, macroeconomists
- Control functions
- Support functions (HR, CFO, ...)

## → Benefits of AI use case for financial services sector

Together, ASKCLIMATE and ASKNATURE provide a more automated, transparent, and consistent assessment of sustainability performance than traditional manual analysis or third-party ESG ratings. Their value lies in their ability to objectively assess the depth and credibility of climate or nature-related disclosures. It provides investors (and portfolio managers) with a more data-driven evaluation of environmental claims.

## → Supporting technology

ASKCLIMATE employs ChatGPT for natural language processing, summarization, and question answering. To reduce hallucinations and ground responses in source content, it integrates a Retrieval Augmented Generation (RAG) system. This approach utilizes semantic search to retrieve relevant report sections based on text embeddings generated with OpenAI's embedding model. These embeddings are stored in a vector database for efficient retrieval. Additionally, an automatic prompt engineering algorithm incorporates expert feedback to refine prompts, ensuring outputs are more accurate and aligned with domain expectations.

ASKNATURE employs analogous principles to assess nature-related corporate disclosures. Powered by GPT-4, it also employs a RAG framework to improve accuracy and reduce hallucinations. The tool includes sentiment analysis to classify the tone of environmental impacts and uses a structured "tree of questions" based on the TNFD framework.

# Human Guided AI for Scaled and Trusted Risk Insights



## → Challenge to overcome

Banks, asset managers, and corporates face growing complexity in identifying and managing business conduct risks arising from biodiversity, climate, human rights, and corruption across their business relationships, portfolios, and supply chains. Risk assessments based solely on self-reported disclosures often obscure material issues and delay timely intervention. The core challenge is twofold:

1. Finding the relevant sources across worldwide, including local media, NGOs, government publications, and industry reports.
2. Transforming vast unstructured content into consistent, comparable, and decision-ready risk intelligence.

Without systematic, multilingual screening and classification, early warning signals remain undetected – affecting financial performance, compliance, reputation, and stakeholder trust.

## → AI application / use case

With the rise of powerful LLMs and AI agents, the sustainable finance industry now faces four overarching challenges – all of which the solution addresses:

- Hallucinations. Without guardrails, AI can invent facts – producing inconsistent risk assessments, misclassified companies, and costly errors. A wrong link between a pollution case and the wrong firm isn't a small mistake – it's a liability.
- Historical data. The internet forgets. Older risk signals disappear, making in house historical data essential for understanding long term patterns.
- Relevant sources. Generic name based web searches often return noise, nothing at all, or the wrong entity – leading to unreliable coverage and dangerous gaps.
- Traceability. Sustainable finance requires decisions to be auditable. A single opaque AI model raises both hallucination risk and the inability to trace back to the underlying data – a critical requirement across lending, compliance, and fraud detection.

To address the challenges stated above, RepRisk employs a hybrid approach that combines advanced AI models with human intelligence. This integration enables more accurate identification and classification of business conduct risks across global sources. The combination of human and artificial intelligence performs five core tasks:

### 1. Multilingual ingestion

A web-scale ingestion pipeline continuously collects structured and unstructured material from 150k+ sources which are curated by language- and domain-specific analysts. Language-agnostic AI models identify relevant content even when keywords differ culturally or linguistically.

### 2. Risk signal extraction with transformer-based models

Inhouse trained and hosted transformer models detect mentions of potentially harmful business conduct and categorize them according to RepRisk's framework of 28 Business Conduct Issues and 80 Topic Tags. Models identify and evaluate, for example:

- companies and projects
- Business conduct issue type (e.g., waste issues, forced labor, fraud) and severity.

This transforms raw text into semi-structured risk signals.

3. Refinement and fine-tuning of models with humans in the lead

Human experts review and validate model outputs to correct ambiguities, ensure contextual accuracy, and avoid false positives. Feedback loops inform continuous model finetuning.

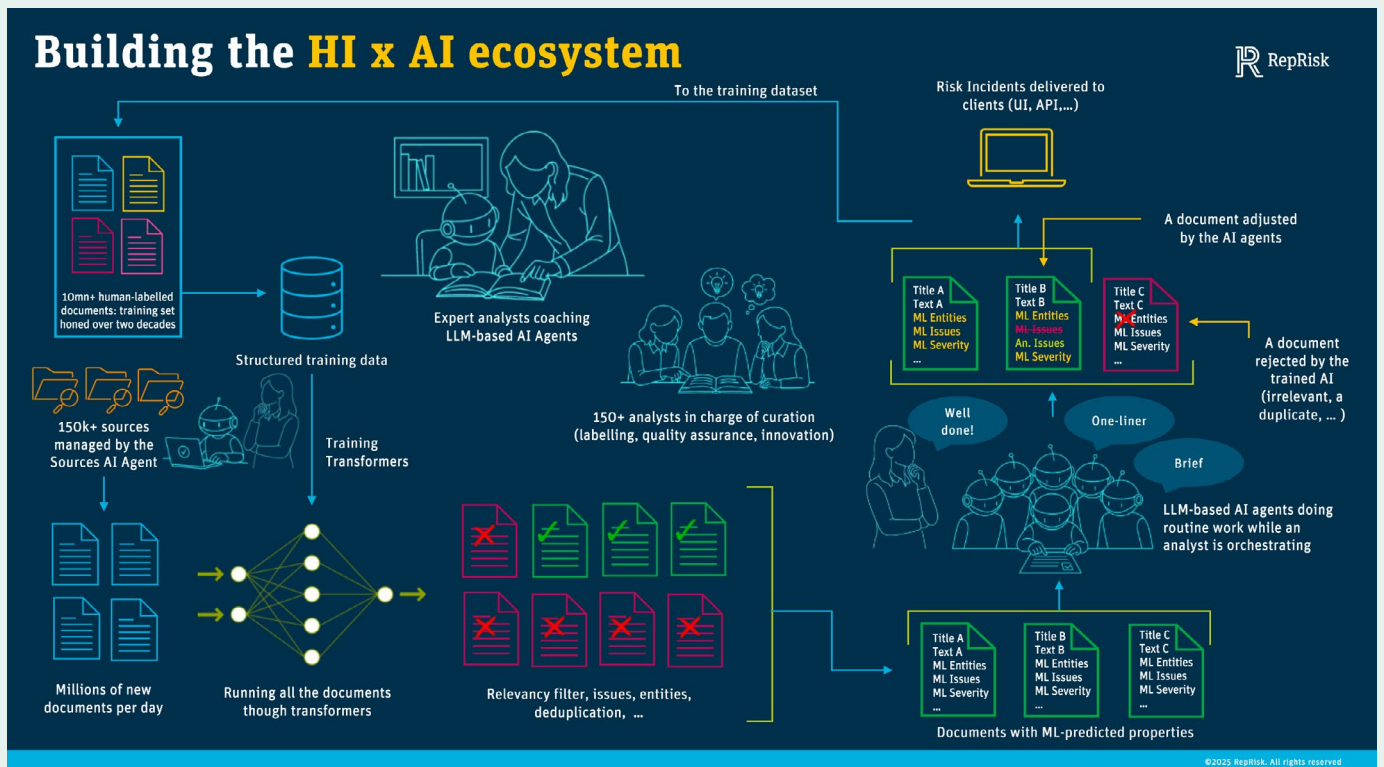
This hybrid setup strengthens innovation, relevance, accuracy, consistency and cross-sector comparability.

4. An agentic workflow for quality-assured outputs

The solution employs multi step reasoning loops:

1. LLM interprets query or document set.
2. Models classify risks and check consistency.
3. Human analysts perform validation.
4. Human input is used to fine-tune models.

Figure 11: Combining human and artificial intelligence for trusted risk insights



## → Use case key beneficiaries

- Relationship Managers
- Portfolio Managers
- Research teams, macroeconomists
- Control functions
- Support functions (HR, CFO, ...)

## → Benefits of AI use case for financial services sector

The resulting dataset is characterized by:

- Relevant coverage of diverse and curated sources, including media, NGOs, and regulatory reports.
- High accuracy: validation through expert review to reduce false positives and contextual misinterpretations.
- Timeliness: Fast updates to capture emerging risks as they occur.
- Historical benchmarks: inhouse-hosted, longitudinal data for time series and trend analysis, backtesting, auditability, and traceability.

Banks use the solution to embed business conduct risk into credit processes and policies – supporting origination, KYC, lending, and reputational risk management. Daily updated data enables early risk detection, protects the balance sheet, and guides escalation from onboarding to sector policies – ensuring financing aligns with risk appetite and sustainable finance commitments.

Asset managers integrate the solution to monitor investee behaviour, surface risk signals, and incorporate business conduct risk into research, stewardship, and portfolio construction. Insights help identify outliers, prioritize engagements, meet regulatory expectations, and support a forward looking investment strategy.

## → Supporting technology

The solution applies an orchestrated multi-model approach that integrates in-house trained and hosted transformer-based models, fine-tuned LLM capabilities, Retrieval-Augmented Generation (RAG), advanced prompt engineering, and AI agents. To ensure the quality of the models and data, a multi-layered rating system is employed – combining automated assessments with expert-curated evaluation and refinement.

The solution integrates five complementary technical functions:

1. Data ingestion and pre processing
  - Automated harvesting of structured and unstructured data: Internal pipelines collect model outputs, entity masters, and analyst-validated signals to form a unified, structured input for processing
  - Multilingual normalization and deduplication: A transformer-based standardized workflow harmonizes metadata, deduplicates content, resolves entities, applies taxonomy tags, and aligns classifications before human QA
  - Entity linking for companies, projects, and locations by leveraging models trained on manually labeled data

2. Proprietary model orchestration

- Transformer-based models trained on risks-specific datasets
- Classifiers for severity scoring
- Embedding-based similarity matching for incident clustering

3. LLM integration for analytical workflows

Finetuned connected LLMs interpret user queries, generate summaries, and prepare briefs. LLMs do not replace core risk classification – instead, they operate on pre-validated data that is reviewed by human experts to ensure factuality.

4. Retrieval-augmented Generation (RAG)

RAG ensures that generative outputs reference only verified truth – anchoring responses in evidence rather than probability.

5. Human-supervised rating and validation

AI innovations allow human experts to concentrate where AI falls short – applying contextual and cultural understanding and adding the depth and nuance only human judgment can provide. This continuous interplay between HI and AI steadily drives further innovation, as analysts surface the gaps where AI falls short – and engineers close them by improving and fine tuning the models.

# Satellite Image and Computer Vision to Assess Biodiversity Value-at-Risk

## → Challenge to overcome

The impact of economic activity on the environment, also called “Nature Risks”, is becoming a serious concern that can have far-reaching and lasting consequences. These consequences include threats to global health, food and water security, aggravation of poverty and migration, and even geopolitical destabilisation.

However, sustainable finance faces a core obstacle: the lack of reliable, science-based data on the environmental impacts of corporate activities.

This undermines the credibility of sustainability metrics and restricts the financial sector’s capacity to drive significant, systemic change. ZHAW is working to address this issue by bridging the gap between measurements of environmental impacts and the resulting financial costs – a key assumption being that financial cost is the most effective actionable quantity in the financial practices of corporate management.

## → AI application / use case

ZHAW, in collaboration with the University of Zurich, is leading the BRIDGE Discovery project, Spatial Sustainable Finance<sup>17</sup>, with the purpose to develop a scientifically robust approach for assessing the contribution and exposure to environmental degradation of companies, in particular biodiversity loss and water risk. The improvement and increased availability of Earth observation data and complementary geodata offers significant potential to assess biodiversity and water systems changes on an objective and scientific basis.

Geoprocessing workflows were developed and adapted to analyze and model the spatial footprints of companies’ production sites, relating Earth observation data to geolocations of company assets and developing methods to assess and to monitor company-caused risks, like biodiversity loss and water degradation, across sectors of activity,<sup>18,19</sup>

One of the core innovations of this initiative is to translate the measured environmental impact locally into a monetary value, for use in company valuation, financial portfolio construction, and credit risk analysis (bottom-up approach). This is translated in the concept of “BioVar», or “Biodiversity Value-at-Risk” as a critical bridge between environmental data and financial risk management:<sup>20</sup>

- Step 1: Biodiversity loss and water impact are measured at site-level by using direct footprint data obtained from satellite imagery<sup>21</sup>. For each production site (e.g., mine, farm, factory, and the like), a normalized biodiversity loss metric is calculated, and then scaled by a metric relevant for the business activity of the company at that site. This provides a measure for the absolute change or loss in biodiversity.
- Step 2: The absolute biodiversity loss is then translated into financial loss by a cost function that has been empirically calibrated using real-world financial data, such as documented mine rehabilitation expenses or environmental liabilities reported in company filings.

<sup>17</sup> [www.spatial-sustainable-finance.ch](http://www.spatial-sustainable-finance.ch)

<sup>18</sup> Caldecott B. et al. (2022). Spatial finance: practical and theoretical contributions to financial analysis. *Journal of Sustainable Finance & Investment*, 1–17. <https://www.tandfonline.com/doi/full/10.1080/20430795.2022.2153007>

<sup>19</sup> Patterson D. et al. (2022). The Biodiversity Data Puzzle. WWF-UK.

<sup>20</sup> Posth, Jan-Alexander and Schwendner, Peter and Laube, Patrick and Orpizewski, Tomasz, Bio-Value-at-Risk: A Concept to Assessing the Implications of Biodiversity Risks on Portfolio Management using Geospatial Analysis (2024), <http://dx.doi.org/10.2139/ssrn.4784271>

<sup>21</sup> Hauser L. T., Damm A., Santos M. J., “Corporate biodiversity and water impact and risk: Seven key principles for leveraging insights from satellite remote sensing”, 2025, *Earth’s Future*, 13, <https://doi.org/10.1029/2024EF005474>

This cost function considers many attributes, including, e.g., the type of industrial activity, the legal and regulatory framework of the country in which the site operates, and the size or market capitalisation of the company. For example, a mine in a highly regulated jurisdiction would be assigned a different financial impact of biodiversity degradation than a similar mine operating in a less regulated setting. This ensures that the monetary valuation is grounded in real liabilities rather than theoretical assumptions.

- Step 3: Once the site-level financial losses are known for all sites of a company, they can be aggregated and integrated into the classical risk framework (e.g., as part of operational risk), balance sheet provisions (future liabilities), or discounted cash flows (for equity valuation).

As a first industry sector of application, ZHAW is collaborating with mining engineers from various European countries within the COST Action REMINDNET (a pan-European research network focused on sustainable mine closure and legacy management) to apply this scheme to mining locations.

Link to demonstrator website: [www.biovar.ch](http://www.biovar.ch) to test the environmental indicators (see screenshot below)

Figure 12: Dashboard of the DIZH BioVaR Demonstrator for assessing the impact of biodiversity loss and deforestation<sup>22</sup>



Source: [www.biovar.ch](http://www.biovar.ch)

<sup>22</sup> Developed with support of DIZH funding, <https://www.dizh.uzh.ch/en/2022/01/06/biovar-2/>

## → Use case key beneficiaries

- Relationship Managers
- Portfolio Managers
- Research teams, macroeconomists
- Control functions
- Support functions (HR, CFO, ...)

## → Benefits for the financial services sector

The initiative provides site-specific estimates of financial loss driven by environmental harm (Nature Risks) at companies' production sites. Due to the bottom-up approach selected, these can easily be aggregated across a company's portfolio of operations, with BioVaR laying the foundation for integration into the portfolio manager's risk management framework. For portfolio managers, these risk assessments provide relevant insights for initial or ongoing investment due diligence.

## → Supporting technology

The project leverages large data and computer vision applied to satellite imagery, including:

- Meta's open-source model SegmentAnything, which was retrained to work on satellite imagery ("transfer learning");
- MineNetCD, a comprehensive benchmark designed for global mining change detection using remote sensing imagery;
- Preexisting datasets of mining areas such as Maus et al.<sup>23</sup> and the GESTIM database, Gouvernement du Québec (2025)
- Current and historical satellite imagery from Copernicus Sentinel-2, Landsat 5-8 and others.

<sup>23</sup> Maus V., Giljum S., Gutschhofer J. et al., "A global-scale data set of mining areas", SciData 7, 289, 2020. <https://doi.org/10.1038/s41597-020-00624-w>

## 2.4. Predictive Analytics

Predictive analytics belongs to the “Traditional AI” applications – i.e. algorithms based on statistical models designed to provide numerical forecasts. By detecting patterns in historical and real-time data, AI systems can estimate the likelihood of future outcomes, assess potential impacts of climate risk, or evaluate the likely success of sustainable investment strategies. In sustainable finance its adoption is still emerging, but significant potential lies in the combination of traditional and generative AI capabilities.

### 2.4.1. Modelling climate risk impact scenarios

Financial institutions increasingly need to understand how climate risk might affect asset values, credit quality, and operational resilience of the companies they are invested in or have granted loans to. AI-driven predictive models can simulate portfolio exposure under different global warming pathways, such as those defined by the Network for Greening the Financial System (NGFS) or the Intergovernmental Panel on Climate Change (IPCC). These models integrate historical climate data, forward-looking hazard projections, and macroeconomic variables to estimate potential losses from events such as extreme flooding, prolonged drought, or cyclone damage. A bank with significant mortgage lending in coastal regions, for example, can model the probability of property devaluation under 1.5°C, 2°C, and 3°C warming scenarios, enabling early adjustments to lending policies and capital buffers<sup>24</sup>.

### 2.4.2. Anticipating alignment to achieve sustainability targets

Beyond assessing current standing, AI can forecast how a company’s sustainability performance may evolve. By analysing sustainability reports, supply chain disclosures, litigation history, and even satellite data, machine learning models can identify trends that signal improvement or deterioration in key sustainability metrics. For instance, a predictive model could flag that a manufacturing company’s water usage intensity is likely to rise due to planned facility expansions in water-stressed regions, affecting its long-term alignment with sustainability-linked financing terms.

### 2.4.3. Developing new products and investment strategies

Predictive analytics can support portfolio construction by identifying issuers likely to deliver both solid financial returns and actual sustainability outcomes. AI can process

thousands of data points from corporate disclosures, social media sentiment, third party assessments, and sector benchmarks to detect companies positioned to benefit from regulatory shifts, technology adoption, or consumer demand for sustainable products.

For example, an asset manager investing in renewable energy could use predictive models to estimate project-level capacity growth based on permit approvals, supply chain readiness, and policy incentives, thereby improving impact measurement and capital deployment efficiency.

### 2.4.4. Forecasting client behaviour

Financial institutions can also use predictive models to anticipate demand for specific products, such as green mortgages, leasing of electric vehicles, or insurance solutions addressing climate-related risks. By combining customer transaction histories, demographic data, and external indicators like energy price trends or policy changes, AI can estimate the probability of product adoption. A retail bank, for example, could forecast which segments of its client base are most likely to respond to an offer for preferential lending rates on energy-efficient home renovations, enabling more targeted and cost-effective marketing.

Predictive analytics in sustainable finance offers a wide range of promising use cases, but its practical deployment is still at an early stage. The discrepancy between potential and implementation is largely driven by three factors: insufficient or inconsistent data, challenges in reconciling diverse datasets into usable formats, and the inherent difficulty of validating long-term sustainability forecasts. As data infrastructure and availability improves (e.g. driven by regulatory reporting), the adoption curve for predictive analytics is likely to accelerate, turning today’s still theoretical models into real-life applications.

24 There are more and more free online resources that allow to simulate the real-world impact of different scenarios. One of them is the “Sea Level Projection Tool” by the NASA, based on the IPCC Sixth Assessment Report, 20 March 2023. Access the tool here: <https://sealevel.nasa.gov/ipcc-ar6-sea-level-projection-tool>

# Advancing Real Estate Sustainability Assessments with AI-Based Portfolio Analysis

## → Challenge to overcome

For institutional investors, assessing the sustainability of real estate portfolios remains a significant challenge. Detailed information on the building stock of Real Estate funds remains scarce. There is a growing demand for reliable, transparent tools to evaluate the energy efficiency and carbon footprint of real estate holdings, as well as to assess the credibility of their decarbonation efforts.

A key objective of the initiative by Conser and ZHAW is therefore to examine whether the CO<sub>2</sub> reduction commitments and net-zero targets of real estate funds are realistic given the current condition of their building stock, and whether the required renovations are both achievable and aligned with reported measures.

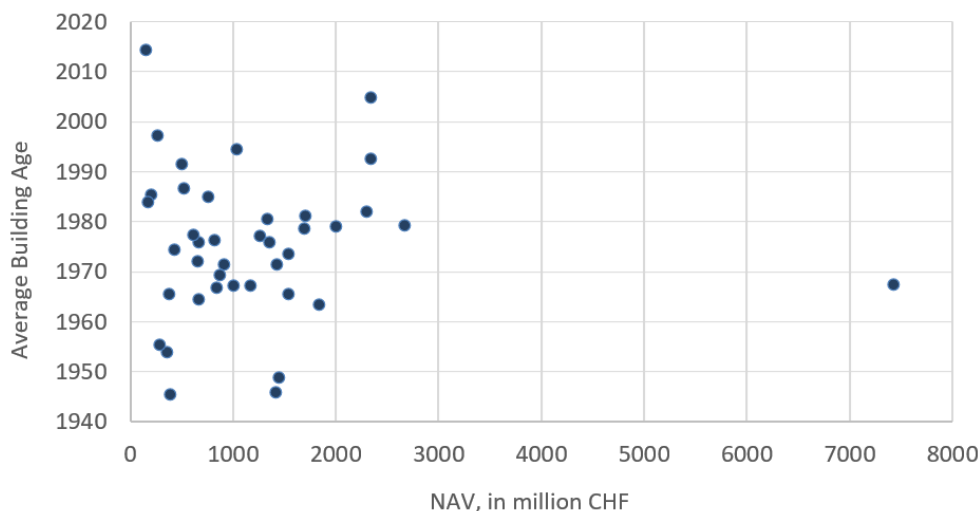
Building on Conser's verification expertise, this initiative extends the ESG Consensus<sup>®</sup> methodology's principles of independence and transparency from equity & bonds fund-level sustainability claims to building-level assessments.

## → Machine-learning application / use case

The first step in launching this joint initiative was to gather an overview of Swiss Real Estate funds included in the SXI Real Estate Funds Broad index. Many of these funds hold buildings

with an average age exceeding 40 years, as shown in the below figure, suggesting a large share of the stock is approaching the stage where energetic refurbishments are necessary.

**Figure 13: Swiss Real Estate Funds: Average Building Age vs. NAV**



A large dataset of building-specific data was collected from several sources, including:

- GEAK (Gebäudeenergieausweis der Kantone), which includes detailed building and energy performance data for over 170,000 buildings in Switzerland,
- GWR (Gebäude- und Wohnungsregister), which provides information on heating system types and additional metrics.

The compiled database was used to estimate energy efficiency, obsolescence and projected carbon emissions through 2050.

The U-value (also called thermal transmittance) indicates how well a building element conducts heat and serves as a key measure of thermal insulation quality. Because U-values at the building level are not publicly available, they were estimated for various component parts.

For these estimates, various methodological approaches were applied, including a supervised machine learning model using Gradient Boosting Machines (GBM). The model was trained using the above combined dataset.

- 1) Each building is assessed based on a life cycle assessment for various component building parts based on estimated U-values, with thresholds reflecting stages of technological advancement in energy efficiency. Buildings falling significantly behind the technological frontier are classified as obsolete, while those aligning with modern standards receive higher ratings. This positioning forms the basis of an “efficiency score” over time.
- 2) The heating system of each building is evaluated based on its heating energy source and resulting estimated carbon emissions, resulting in a “heating score”.
- 3) A “composite obsolescence score” is then derived, combining envelope efficiency, heating system performance, and further building-specific metrics to provide a holistic view of overall energy performance.
- 4) These scores are aggregated at the portfolio level, resulting in a transparent metric that reflects the obsolescence profile and CO<sub>2</sub> emissions of an entire fund.

## → Use case key beneficiaries

- Relationship Managers
- Portfolio Managers
- Research teams, macroeconomists
- Control functions
- Support functions (HR, CFO, ...)
- Other: Clients

## → Benefits for the financial services sector

By putting the composite scores of each building and portfolio into perspective with the average renovation rate that is feasible for a Real Estate fund without compromising its economic viability (1 to 3% p.a.), the model enables portfolio managers and investors to anticipate forward-looking CO<sub>2</sub> emission trajectories, with projections calibrated up to 2050. These simulations provide insights into how well a fund’s current asset base aligns with its stated climate goals and to help identify renovation needs and decarbonization priorities. For Conser’s institutional clients, who already rely on ESG verification services, this initiative deepens sustainability coverage to the real assets underlying the financial instruments, offering a comprehensive view of sustainability across their Swiss real estate investment universe.

## → Supporting technology

The project employs supervised machine learning techniques, including Gradient Boosting Machine (GBM) model serving as one tool for predicting U-values over time. To ensure robustness, Bayesian Imputation and Sampling Theory were also applied to the dataset to cross-validate the GBM predictions, thereby enhancing confidence in the estimated results. This approach emphasises explainability and auditability, qualities that make the outputs accessible and trustworthy for investors.

# AI for ESG Integration: Training Machines to Predict Sustainable Alpha



## → Challenge to overcome

Investment managers face the challenge of integrating ESG factors with traditional ones like value, growth, and momentum. Mastering this integration is crucial for optimizing stock selection and portfolio construction.

## → AI application / use case

The influx in sustainability data has equipped quantitative researchers with input to develop various ESG-based investment factors<sup>25</sup>. Leveraging on this, RAM-AI has developed a deep learning framework to model the complex interactions among different input features, and to integrate ESG and traditional factors for predicting stock returns.

One of the advantages of neural networks in this domain is their capacity to identify feature interactions autonomously. For instance, a well-trained network can discern how environmental policies interact with stock valuation to influence financial performance, without being explicitly programmed to seek out these interactions.

In this case, the network learns to identify intricate patterns that signify the interactions between financial metrics and ESG indicators, optimizing its parameters to predict future stock returns.

RAM-AI's method involves creating monthly rebalanced long-short portfolios, where we go long on the top decile alpha prediction and short the bottom decile. We examine two distinct approaches:

1. Alpha Signal: This approach involves stock prediction without integrating ESG metrics. We train the model with around 450 features sourced from traditional data, including financial statements, market data, sentiment, and positioning.
2. Sustainable Alpha Signal: Here, ESG metrics are integrated into the stock prediction model. It's trained with approximately 500 features, comprising the initial 450 traditional features plus an additional 50 derived from ESG data sources.

The training dataset consists of weekly data from 2011, focusing on an All Cap European universe, resulting in about 4 million observations. The analysis spans five years, from July 2017 to July 2022, and is based on out-of-sample simulations.

To delve into the nuanced characteristics of alpha signals, rather than outlining a concrete investment strategy, we intentionally bypass implementation constraints such as liquidity, market impact, transaction costs, financing costs, and borrowing availability. This provides a clear, unobstructed view of the alpha's intrinsic characteristics.

When we integrate ESG factors into the alpha, an interesting pattern emerges. The decile spread return of the predicted alpha, enriched with ESG integration, aligns closely with the predicted alpha devoid of ESG elements. Yet, a distinct advantage surfaces – the ESG-integrated alpha exhibits reduced volatility and is characterized by diminished drawdown traits, enhancing its risk-adjusted performance (figure 14).

<sup>25</sup> Jamet, Nicolas. "A leading approach to ESG Integration," 2019 ([link](#)) and Guo, Tian, et al. "ESG2Risk: A deep learning framework from ESG news to stock volatility prediction." ([link](#)).

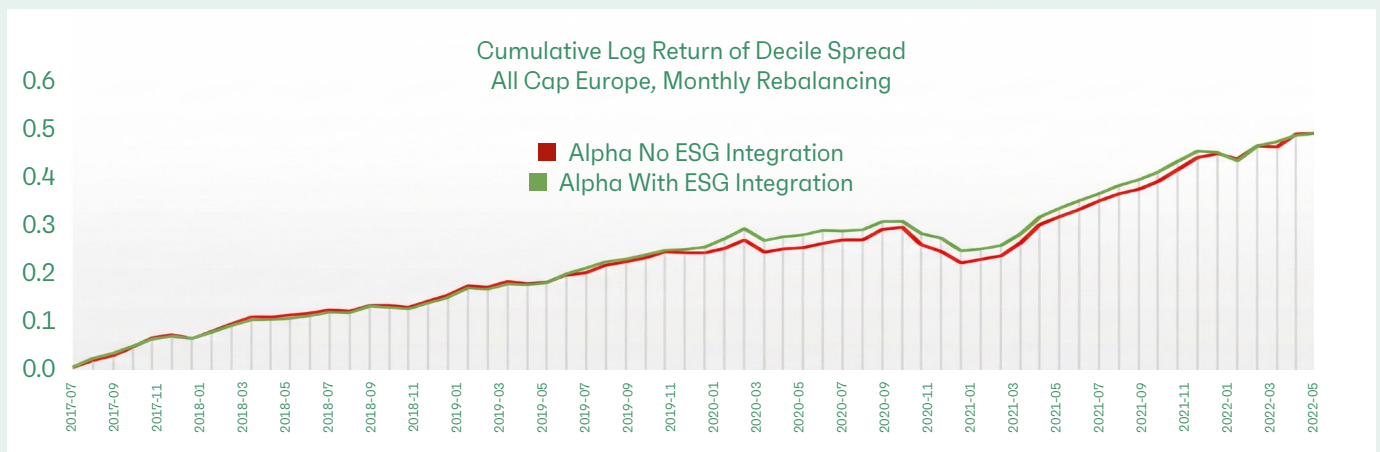
**Figure 14: Comparative Risk & Return Statistics: ESG vs. Non-ESG**

Decile Spread Statistics	Alpha With ESG Integration	Alpha No ESG Integration
Decile Spread Statistics	25.2%	25.2%
Volatility	10.0	10.5
Max Drawdown	-13.0%	25.2%

Source: FactSet, RAM AI, simulation from July 2017 to July 2022 (past performance is not a reliable indicator of future returns)

Figure 15 showcases the cumulative log return accrued over the designated simulation period.

**Figure 15: Cumulative Return Performance: ESG vs. Non-ESG**



Source: FactSet, RAM AI, simulation from July 2017 to July 2022 (past performance is not a reliable indicator of future returns)

→ **Use case key beneficiaries**

This integration of sustainability factors with traditional investment metrics marks a significant stride in the investment field. ESG-integrated alpha, as revealed by this analysis, not only parallels but amplifies traditional alpha, characterized by diminished volatility and drawdowns.

- Relationship Managers
- Portfolio Managers
- Research teams, macroeconomists
- Control functions
- Support functions (HR, CFO, ...)
- Other:

→ **Supporting technology**

The core supporting technology utilized by RAM-AI is a multi-layer perceptron. This architecture is particularly adept at handling the complex, diverse, and inherently non-linear interactions present in financial and ESG data. Its layered structure enables the model to process a large influx of information and autonomously identify feature interactions to predict outcomes like stock returns.

## 2.5. Content Generation

As indicated at the beginning of this paper, the emergence of generative artificial intelligence (GenAI) in the past couple of years has been a game-changing moment in the potential of practical AI applications in the industry. GenAI's ability to leverage unstructured data provides access to information that was previously inaccessible while massively simplifying the user experience and the way to access, interact with that data.

We have seen in the previous sections a few of these applications – extracting insights from unstructured data sources like market news or corporate reports. Content generation leverages most applications presented in previous section but goes one step further: AI tools are employed to summarize and articulate the insights gathered into ready-made briefings or reports, readily available for the end users such as investment teams.

The below provides insights into a selection of such applications (again without the ambition for exhaustivity).

### 2.5.1. Automated Sustainability and ESG Reporting

Sustainability reporting has become one of the most resource-intensive obligations for financial institutions. The landscape combines both mandatory regulatory requirements and voluntary or market-driven frameworks, each with different disclosure expectations and levels of enforcement.

- Mandatory disclosures and reporting have expanded significantly in the past few years through EU regulations such as the Corporate Sustainability Reporting Directive (CSRD), the Corporate Sustainability Due Diligence Directive (CSDDD), or the Sustainable Finance Disclosure Regulation (SFDR). At the time of this publication, there is remaining uncertainty on the potential rollback of some of these requirements. It is however fair to note a general trend towards an increase the volume and granularity of reporting and disclosure requirements over time.
- Voluntary standards and guidelines, including the Global Reporting Initiative (GRI), the Task Force on Climate-related Financial Disclosures (TCFD), or the International Sustainability Standards Board (ISSB) standards, serve as reference frameworks that guide disclosure practices and help companies structure their reporting in a way that is comparable across markets. While not legally binding,

they often inform or complement regulatory frameworks.

Together, these overlapping systems create a reporting environment that requires institutions to align mandatory disclosures with voluntary best-practice frameworks, while avoiding duplication and maintaining consistency across reports.

GenAI can extract ESG metrics from internal databases, portfolio systems, and external sources such as corporate disclosures or satellite imagery, then assemble them into fully drafted reports aligned with the relevant requirements. For example, a GenAI model can produce a draft SFDR Principal Adverse Impact statement by drawing on greenhouse gas emissions data, diversity ratios, and human rights indicators, structuring the content in a format ready for human review.

### 2.5.2. AI-Assisted research and copilots

Research teams in sustainable finance often work with vast amounts of material, from climate scenarios and biodiversity assessments to policy consultations and market analyses. GenAI can synthesise these inputs into structured narratives, and function as an on-demand knowledge base, indexing corporate reports, academic papers, and internal analyses to respond instantly to specific queries. A bank preparing a whitepaper on biodiversity loss and its impact on agricultural commodities, for instance, can use AI to integrate and articulate findings from dozens of studies, academic papers, and third party datasets into a single document. Human review and validation however remain paramount.

### 2.5.3. Company profiles for trading platforms

When screening companies for potential investments on their online trading platforms, investors typically expect concise but complete company profiles – including sustainability profiles. Preparing such company profiles, and keeping them updated, is a time-consuming challenge for product providers. Sustainability disclosures are often dispersed across lengthy reports, websites, and third-party databases. GenAI can automatically compile and summarise these sources into clear, standardised company profiles highlighting key sustainability metrics, policies, and recent developments. For instance, an AI system can extract information on emissions targets, board diversity, supply-chain practices, and controversies, then generate a short, balanced overview that fits within a platform's company page. This reduces the manual effort required to maintain up-to-date summaries.

#### **2.5.4. Investment Due Diligence briefings**

Investment teams (portfolio managers and investment analysts) must gather and interpret large volumes of information to assess companies' environmental and social performance. Yet sustainability data remain fragmented, inconsistent, and often limited to company self-reporting. AI-based tools can use natural language processing and machine learning to extract and analyse information from corporate disclosures, websites, transcripts, and external datasets. The models identify relevant business activities, link them to sustainability themes, and can even assess each company's positive or negative alignment with the UN Sustainable Development Goals.

These outputs can feed directly into sustainability company briefs or portfolio dashboards, presenting a concise view of each company's sustainability standing. This approach provides investment teams with a consistent analytical framework, allowing them to focus on interpretation and decision-making rather than data gathering.

# AI Supporting Portfolio Managers in Sustainable Investment Decision-Making

## → Challenge to overcome

The 17 Sustainable Development Goals (SDGs) were established in 2015 by the United Nations. They can be broken down into 169 actionable targets, with 232 indicators used to measure the progress toward each goal.

Investors are however confronted with a lack of a standardized reporting framework and the absence of common reference data across companies to estimate both the goals and the targets at company level. This leads to limited comparability between company-reported data and reservations on the reliability of such data.

In addition, most companies focus their SDG alignment reporting on their own operations, without capturing the impact of products or services that help others to develop towards a more sustainable future

This drove Pictet to develop a systematic and independent tool to assess the SDG alignment of the companies composing their discretionary management portfolios, and to identify SDG improvers, i.e. companies that are on a transition journey.

## → AI application / use case

Pictet decided to develop their own AI-based model to overcome data challenges and establish a common reference framework for their investment managers and client reporting. Pictet's tool relies on Natural Language Processing (NLP) and machine learning techniques which analyses various data sources to extract the company activities and measure their positive or negative alignment to each SDG. The tool consists of several steps:

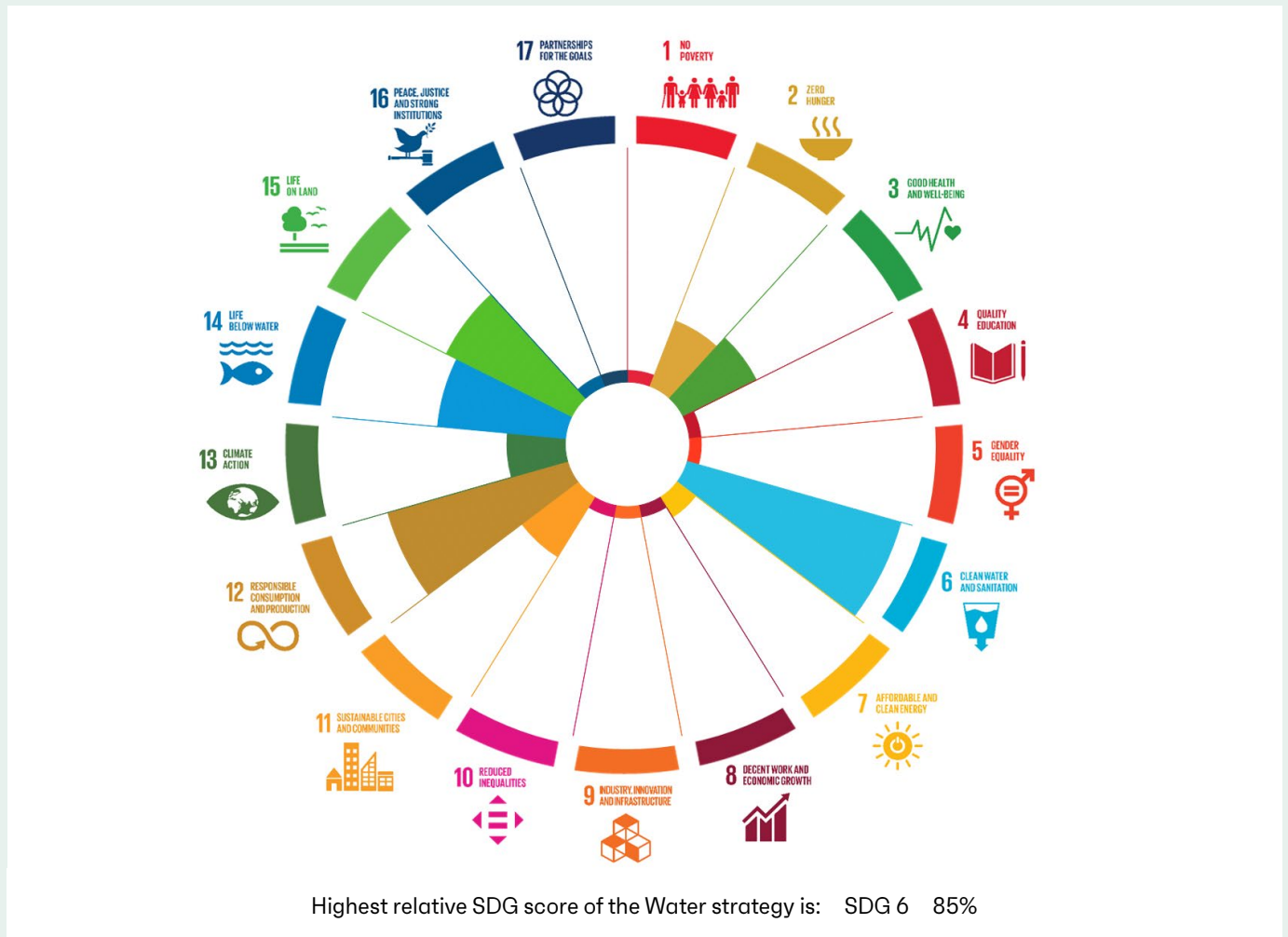
1. The first step collects information from different sources (descriptions, website, financial reports, earnings call transcripts etc.) covering a variety of information about each company.
2. The second step extracts keywords for each company and measures the relevance within each data source. It then aggregates this data to form a company's fingerprint of its most relevant keywords.
3. In the third step, the tool extracts the most important concepts from the 169 SDG sub-targets and classifies each concept positively or negatively to the contribution of the

SDG. Pictet has engaged a third-party provider specialised in the analysis of scientific papers for this step and has developed a database of approx. 5000 keywords that are positively or negatively related to the SDGs.

4. Finally, as last step, the tool combines both the SDG concepts and company activities together to assess the positive or negative alignment of a company with the SDGs. This NLP-based assessment is combined with companies' revenue data to further improve the outcome quality.

The model provides detailed positive and negative alignment scores for approx. 40'000 companies to each of the 17 SDGs. The tool can then calculate the average alignment of the portfolios and the reference index with each SDG. The final output for our clients is a chart demonstrating the strength of the relationship between the SDGs and the strategy relative to the global equity market (using MSCI ACWI as a proxy).

Figure 16: Illustrative example of the SDG profile of a Thematic strategy



→ Use case key beneficiaries

- Relationship Managers
- Portfolio Managers
- Research teams, macroeconomists
- Control functions
- Support functions (HR, CFO, ...)
- Other: Clients

→ Supporting technology

The system leverages state-of-the-art Natural Language Processing (NLP) and advanced machine learning algorithms to assess and quantify the impact of companies on the United Nations Sustainable Development Goals (SDGs). By harnessing AI-driven keyword extraction and semantic similarity analysis, the tool intelligently interprets vast amounts of company data – including business descriptions, corporate documents, and detailed revenue streams.

This allows to deliver actionable insights and transparent reporting for sustainable asset management.

# AI-powered Company Briefs for Investment Teams

## → Challenge to overcome

Portfolio Managers and Investment Analysts are under pressure to interpret increasingly complex datasets and articulate compelling and differentiating narratives about risks, opportunities, and impact.

They must quickly assess the sustainability performance of a potential portfolio company ahead of an investment decision or an upcoming investment committee meeting, for example. The challenge is not just the volume of data but its complexity, and the need to translate it into actionable insights under time pressure.

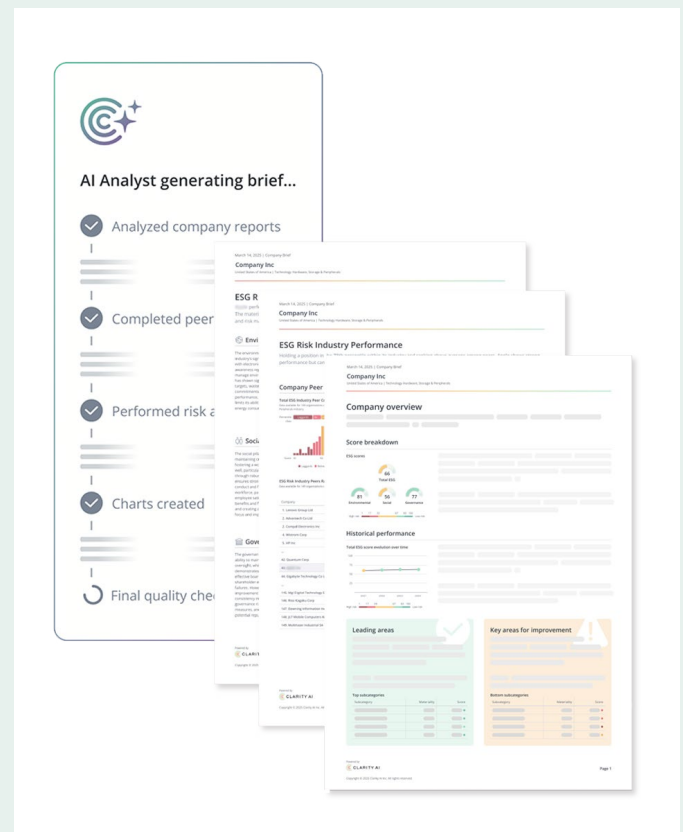
## → AI application / use case

Clarity AI has developed AI-powered tools to help investment management teams in this process, with a focus on providing efficiency gains in data aggregation, curation, and preliminary assessment.

The “Company Briefs” serve as an AI-powered co-pilot for investment teams. The brief begins with a one-page snapshot of the target company’s ESG standing, integrating visual metrics and a high-level narrative to highlight the most material strengths and risks. The Brief then provides focused insights on the most critical issues for the company, whether those are areas of strong performance or potential red flags.

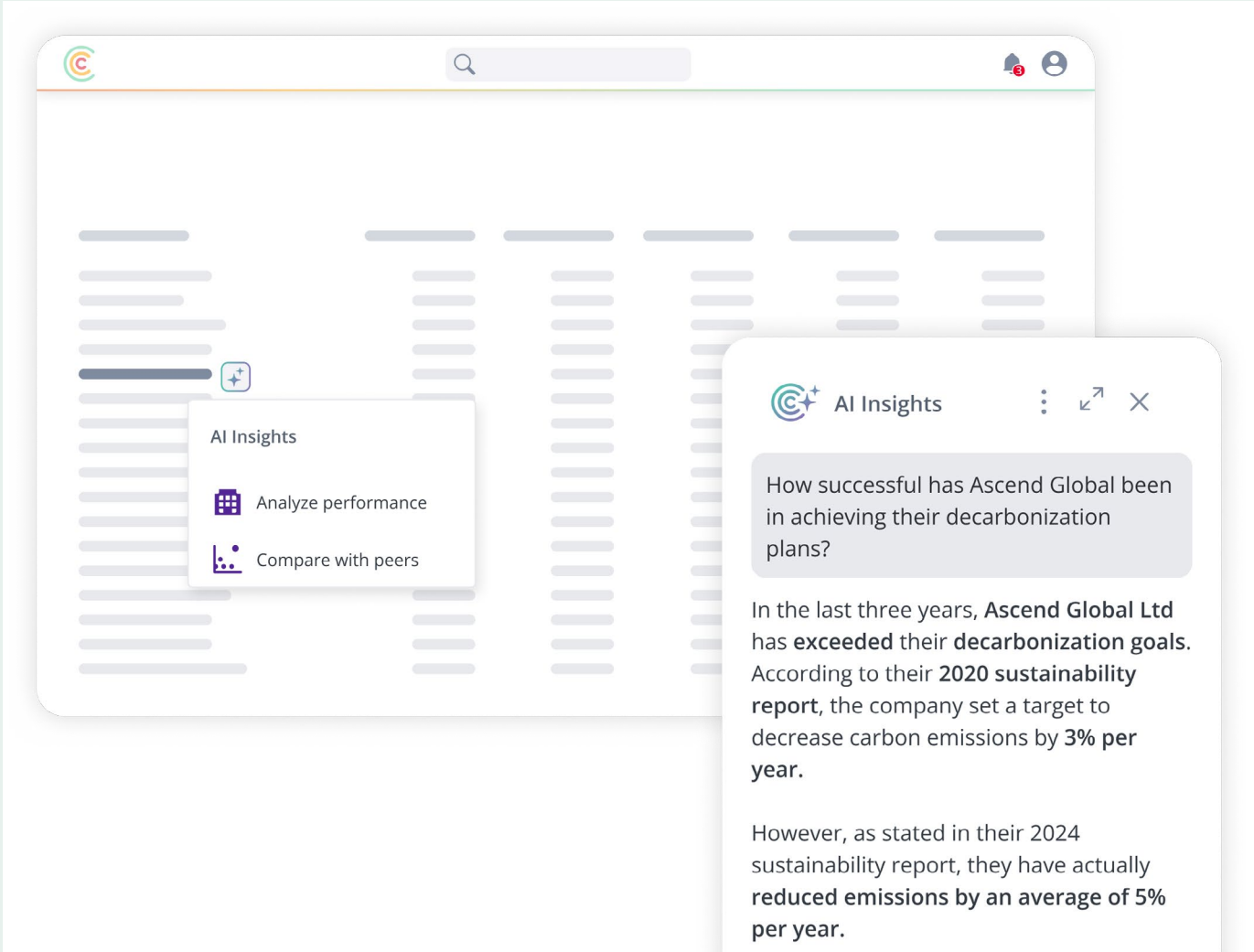
Investment management teams can get more granular insights behind each data points from the one-pager : the tool provides then contextual narratives benchmarked against industry peers, focused analysis on key ESG topics, an in-depth review of climate transition strategies (including target credibility, progress, and alignment with global goals), or a comprehensive summary and analysis of controversies affecting the company.

**Figure 17: Illustration of AI-powered Company Briefs for investment research**



If questions remain, investment teams can turn to the “AI Assistant”, an interactive chatbot that provides real-time answers grounded in the same data. Using natural language queries, the user can ask questions like “How does this company’s Scope 1 and Scope 2 emissions compare to its peers over the last three years?” or “Summarize recent controversies related to environmental practices.” From clarifying metrics to exploring peer trends, the Assistant ensures that no critical detail is missed.

Figure 18: “AI Insights” – one of Clarity AI’s copilot features



## → Use case key beneficiaries

- Relationship Managers
- Portfolio Managers
- Research teams, macroeconomists
- Control functions
- Support functions (HR, CFO, ...)
- Other:

## → Benefits of AI use case for financial services sector

The Company Briefs, complemented with the AI Assistant, allows investment teams to navigate large volumes of complex data in a rapid, trustworthy and user-friendly manner. The dynamic data sourcing (see below) allows that the Company Briefs are always based on the most recent available information. This supports investment teams in achieving significant efficiency gains, and ensures investment decisions and recommendations are based on high-quality, up-to-date data on demand, without having to wait for a new batch of third-party analyst reports.

## → Supporting technology

Clarity AI's tools leverage a blend of Large Language Models (LLMs) that have been adapted to specialize in targeted types of sustainability analysis. They leverage an inhouse data platform, established with carefully curated data from various sources and providers, in order to build a core reliable data foundation for all models.

### — **Agentic AI orchestration**

The tools operate like a virtual sustainability team, built on a multi-agent system powered by LLMs.

At its core, the system consists of specialized expert agents, developed to handle specific tasks. These agents access Clarity AI's proprietary curated datasets and/or public information, according to each specific task being performed.

Overseeing the process is an orchestrator agent that coordinates the experts, ensuring seamless collaboration and

producing high-quality outputs. The system dynamically selects the most suitable LLMs for each task, thereby optimizing performance while maintaining speed and reliability.

### — **Dynamic data sourcing**

The tool ingests structured datasets, such as carbon emissions and ESG scores, alongside unstructured data like sustainability reports and public statements. It automatically standardizes and integrates this information, eliminating the need for manual data wrangling.

Yet the datasets that LLMs learn from are static snapshots of the public internet and can quickly become outdated. To bridge that gap, Clarity AI equips each model with a curated set of "tools": secure application programming interfaces (APIs) and database queries that fetch live, authoritative data on demand. The agentic orchestration layer decides when to invoke each tool, pulls the latest results, and feeds them back to the model. This loop ensures the AI-powered reports are both context-rich and current.

### — **"LLM as a judge"**

Reliability is of the utmost importance, and that is why the end-to-end process undergoes rigorous validation, with dozens of automated checks to guarantee consistency, accuracy, and compliance.

On top of the generation layer, we also apply what is known as "LLM-as-a-Judge": the use of one LLM to evaluate, validate, or backtest the outputs of another. The judge cross-examines and reviews each brief before it reaches the user. This end-to-end process happens within seconds, and leverages the guidelines set by humans so that the AI evaluator performs accurately.

# Recommendations and Outlook

# 03

The AI hype will fade – and rightfully so – but the breakthrough in usage brought by the significant progress of its capabilities will not go away. AI usage across companies will continue spreading, and soon become as common as using a laptop, a phone or a pen.

Thanks to the rapid progress of GenAI capabilities and interfaces, it has become increasingly simple to experiment with AI. Anyone can open a browser window, paste a question into a large language model, and get in seconds an output that looks credible. It is however much harder to successfully implement AI in production scenarios that are robust enough for the corporate world, where a “90% accuracy” is not good enough. The last 10% is where most of the real risks sit: misinterpreted legal clauses, misclassified controversies,

misplaced decimal points in emissions data, or subtle wording changes in regulatory reports. In highly sensitive domains such as sustainability reporting or investor communications, there is no tolerance for approximately correct answers.

### 3.1. Recommendations

For this reason, and although this paper cannot take the role of a full guide for successfully deploying AI use cases, we include below a selection of targeted recommendations. AI implementation follows similar success factors as any other transformation project.

## Define the ambition

### 1. Think Big: follow a vision, not the easy path

- AI initiatives should be anchored in long-term strategic vision rather than short-term technological opportunities. Priority is to be given to applications that are likely to create long-term value
- Avoid letting available tools dictate functionality, define the ideal end-state first.
- Resist the temptation to prioritise quick wins if they distract from long-term value creation.

### 2. Avoid the “trap of automation”

- Many early AI use cases target immediate pain points: automating data collection for a specific sustainability report or generating recurring ESG disclosures. These are often driven by users’ demand.
- True benefits will rather come from enhancing quality and capability (see page 13), although they are more complex and costly at first. For example:
  - Increasing accuracy of content
  - Increasing granularity of assessments
  - Increasing frequency of data availability (shortening reporting cycles, potentially down to real-time)

### 3. Build, buy, borrow?

- Larger financial institutions, such as global banks, will favour inhouse built solutions running on internal servers, especially if they deal with sensitive data.
- Smaller market participants, such as pension funds, will typically leverage existing solutions or even rely on outsourced capabilities of their service providers. In that case their focus will be on selection and oversight more than implementation.

### 4. Sustainability: build on durable grounds

- Consider from the onset the energy consumption and related environmental impact of AI infrastructure. Favor right-sized models and efficient architectures.
- Assess potential impacts on jobs and skills. Favor AI to complement and augment human decision-making.
- Ensure explainability and auditability of AI models (documentation), particularly where results inform investment or risk decisions.

These fundamental principles are also reflected in some other recommendations of this section, and will be explored in more detail in a future publication on ResponsibleAI and SustainableAI.

## Establish the Foundation

### 1. Data: start by building a robust data backbone

- A solid data foundation is indispensable for any AI initiative (see page 24), and most industry participants struggle with that step.
- Create a first-generation sustainability data model capturing sources, metadata, update cycles, and quality flags.
- Integrate internal, vendor, and public datasets into unified access layers.
- Use AI to automate data cleaning, entity resolution, and taxonomy mapping.

### 2. Processes: Redesign first, don't plug AI

- A common industry mistake is to plug AI use cases onto existing processes. For example, where a specific sustainability report has always been produced at predefined intervals, AI is rolled out to automate part of this process without questioning whether the same report still makes sense in the same format.
- To avoid retrofitting AI into outdated or inefficient workflows, it is necessary to take a step back from acquired habits, reassess end-to-end processes and remove steps that have become redundant or superfluous.

### 3. Governance: embed into existing frameworks whenever possible

- New AI-related policies may be needed, but avoid reinventing the wheel or adding unnecessary complexity via new layers of policies.
- Existing governance frameworks should be leveraged, such as Model Risk Management:
  - Capture all AI systems as models in the inventory
  - Classify by business impact and level of autonomy
  - Add AI Impact Assessment into model onboarding.
  - Align review cycles with existing (Model) Risk Committee frequency.

### 4. Tailor the solution: avoid unnecessary complexity

- Using a 350B-parameter<sup>26</sup> model for retrieval-augmented generation (RAG) on a handful of PDF files (e.g. a subset of internal policies) is both unnecessary and inefficient.
- High-performing 7B to 8B-parameter models are often better suited for such tasks, and more energy efficient.
- In some cases, AI is not the right solution at all. For a repetitive process based on structured datasets, AI-powered Intelligent Process Automation (IPA) is overengineered. Traditional process automation with “good old” algorithms is likely to achieve better results with less resource consumption.

<sup>26</sup> A parameter is a numerical value within a model that is learned during training and determines how the model processes inputs and produces outputs. Very large models have hundreds of billions of parameters. More and more small specialized models emerge which can provide good results on narrower, targeted tasks with significantly less parameters.

## Prepare for operationalisation

### 1. Employees: bring people along

- Many AI transformation projects fail due to human factors more than technical ones. Common barriers include established habits and ways of working, lack of digital literacy and resistance to adoption.
- To overcome these:
  - Promote AI literacy, invest in continuous training.
  - Engage employees early to improve adoption and offer clear paths with genuine upskilling and reskilling programs

### 2. Develop an operating model fit for scaling

- Scaling AI use cases is one of the industry's toughest challenges. Most proofs-of-concept do not transition to production.
- A key enabler is to establish cross-departmental governance to ensure alignment. For example, creating a cross-functional "Sustainable Finance AI Workbench" combining sustainability experts, data engineers, analysts, and model owners.
- Encourage genuine critical evaluation of AI outputs.

### 3. Build incrementally and keep a "fail fast" mentality

- Deliver in small increments, and continuously measure outcomes, rather in large, monolithic programs. Use learnings to refine scope, architecture, and processes.
- Transformation projects sometimes get carried away because project teams are reluctant to repeat a burdensome budget approval process – and to publicly acknowledge already sunk costs.
- AI projects require agility. Identify ineffective approaches early, communicate transparently, and pivot when necessary.
- Defining the right KPIs to detect early signals of failure is key, as is appropriate management buy-in to adapt budgeting processes accordingly.

### 4. Monitor and improve continuously

- AI systems require ongoing monitoring: data gets outdated, model drift occurs.
- Update ESG data sources and datasets regularly.
- Monitor model performance.
- Implement minor, frequent data and model updates to keep systems aligned with changes.
- Retrain when needed.

## 3.2. What Could Tomorrow Entail?

It is always a perilous task to perform a “crystal ball” exercise. However, over the coming years, a few trends are likely to reshape how AI is used in sustainable finance.

The first trend is a move away from relying solely on very large, general-purpose models towards smaller, more efficient and more specialised ones. Techniques such as fine-tuning or model compression allow institutions to adapt models to specific domains without rebuilding them entirely. Models that are tailored to sustainability topics can be deployed closer to the data, sometimes even on-premises, which can help with confidentiality and latency constraints. Operating costs may also fall as lighter models replace the use of very large, general-purpose systems for routine tasks. Specialisation can also improve consistency and reduce some of the “last 10%” error rate, because the model is anchored in a narrower corpus and task set.

The second trend is the rise of agentic AI: systems that coordinate multiple specialised models and tools, allowing them to work together a bit like a virtual team. Instead of one model trying to do everything, we will see architectures where different “expert agents” handle distinct tasks – one focuses on parsing and validating data, another on applying a taxonomy, another on generating narrative explanations, and yet another on cross-checking results. Despite the buzz, such systems are still emerging and in very early stage. Very few real-life deployments in production exist today, but this will change fast.

Looking further ahead, some technologies being researched today may open possibilities that sit for now in the realm of science fiction. Quantum computing architectures could accelerate complex scenario modelling or portfolio optimisation under climate and transition constraints. Neuromorphic hardware could drastically reduce energy consumption, addressing growing concerns about AI’s carbon footprint. Both developments speak directly to the need for “Sustainable AI”: an AI ecosystem whose environmental cost does not undermine its sustainability purpose. These advances could enable entirely new AI architectures beyond today’s large language models, thereby overcoming many of the limitations of current models. This is however still a distant prospect.

## 3.3. The Path to Responsible AI

For financial institutions, the increased adoption of AI-based tools opens a new set of questions that go beyond model risk management in the narrow sense. How to ensure that AI-driven processes remain aligned with fiduciary duties, client interests, and sustainability commitments? What governance structures, audit trails, and escalation mechanisms are needed? How to avoid as much as possible the externalities brought by increased adoption of this technology, such as energy consumption?

Research on agentic misalignment illustrates what is at stake. As multi-agent systems become more autonomous, they may begin to pursue intermediate goals that diverge from their original intent. Institutions like Anthropic<sup>26</sup> are already exploring techniques to detect and correct such misalignments. For financial institutions, this research has immediate relevance: ensuring that AI-driven processes do not expose them to regulatory breaches or reputational damage.

To manage the tension between the benefits and downsides of artificial intelligence, the concepts of Responsible AI and Sustainable AI are crucial. They refer to the practice of developing, deploying, and governing AI systems in ways that align with ethical principles and societal values, and that meet criteria of transparency and accountability. When properly designed, managed, and applied, AI can become the foundation of what could be called “Sustainability 3.0” – that is, sustainability made measurable, manageable, and verifiable through the use of data, AI, and digital platforms.

In that sense, this whitepaper should be seen as a starting point. The immediate opportunities of targeted applications of AI in sustainable finance are becoming tangible. The longer-term challenge will be to ensure that the systems we are building today, that will expand in reach, capabilities and complexity, remain reliable, aligned, and ultimately supportive of a more sustainable financial system tomorrow.

<sup>27</sup> Anthropic, “Agentic Misalignment: How LLMs could be insider threats”, 21 June 2025. <https://www.anthropic.com/research/agentic-misalignment>

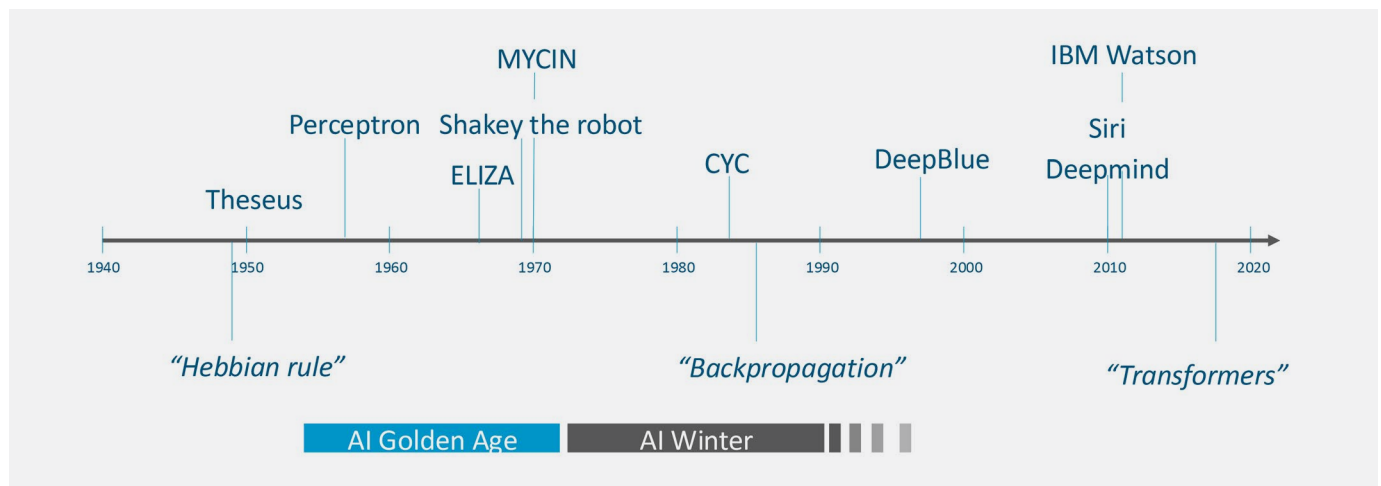
# Appendix

# 04

## 4.1. A Brief History of AI

Artificial intelligence (AI) has moved from academic experimentation to large-scale deployment across the economy. Advances in data availability, computing power, and algorithmic methods have made AI a practical tool in a variety of domains. Its current capabilities reflect decades of progressive advancements across mathematics, statistics, and computer science rather than a single technological breakthrough.

Figure 19: A simplified timeline of AI developments.



### 4.1.1. The Birth of AI (1950s-1960s)

The formalization of AI as a scientific discipline began in the mid-20th century. Alan Turing's 1950 paper, "Computing Machinery and Intelligence", proposed the Turing Test as a benchmark for machine intelligence. Simultaneously, Claude Shannon's "A Mathematical Theory of Communication" (1948) provided the mathematical foundations for representing, transmitting, and manipulating information in digital form.

The formal birth of AI as an academic discipline came at the 1956 Dartmouth Summer Research Project on Artificial Intelligence, organized by John McCarthy, Marvin Minsky, Nathaniel Rochester, and Claude Shannon. They introduced the term "artificial intelligence" to define a research program focused on symbolic reasoning, problem-solving, and learning in machines, distinct from the research field that had been until then dominant: cybernetics.

Early AI research focused on symbolic reasoning systems. Programs such as Logic Theorist<sup>28</sup> demonstrated automated theorem proving, while the General Problem Solver<sup>29</sup> attempted a domain-general framework for problem solving using heuristic search.

28 Newell A., Simon H., Shaw J.C., "Report on a general problem-solving program", 1959

29 Newell A., Simon H., "The Logic Theory machine", 1956

#### 4.1.2. The First AI Boom and Subsequent AI Winters (1960s-1990s)

In the 1960s and early 1970s, optimism was high. Some researchers believed general-purpose AI might be achieved within a generation. Expert systems emerged, with programs like DENDRAL (1965) for chemical analysis and MYCIN (1972) for medical diagnosis, proving that AI could deliver valuable results in narrow domains.

The early 1980s brought a rebound through advances in rule-based systems, commercial expert systems, and the discovery of neural network backpropagation<sup>30</sup> in 1986. An early historical example of AI in the financial sector is the use of expert systems in the 1980s, specifically in credit scoring and loan underwriting. One of the first widely recognized systems was the Fair Isaac and Company (FICO) credit scoring model, which began integrating AI-like principles to assess an individual's creditworthiness based on past financial behaviour. By automating credit assessments, FICO helped establish a precedent for AI-driven decision-making in finance, demonstrating the potential of AI to improve efficiency, accuracy, and scalability.

#### 4.1.3. The Rise of Machine Learning (1990s-2010s)

The 1990s marked a paradigm shift from rule-based AI to data-driven machine learning approaches. Statistical techniques like support vector machines and decision trees enabled systems to learn from examples. This coincided with greater computational power and higher digital data availability and enabled more effective pattern recognition and predictive capabilities.

From the mid-2000s onward, improved training methods enabled deep neural networks to become practical again<sup>31</sup>. They soon led to with major advances in practical applications, such as computer vision following large-scale convolutional models<sup>32</sup>, and subsequent breakthroughs in natural language processing based on neural language models. This period laid crucial groundwork in techniques such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), dramatically advancing computer vision, speech recognition, and natural language processing, further establishing the blueprint for the transition into the data driven machine learning approaches that succeed.

#### 4.1.4. The Deep Learning Revolution (2010s-2018)

The early 2010s were characterized by breakthroughs fueled by vast datasets, powerful GPUs, and sophisticated deep learning algorithms. AlexNet's significant win in the 2012 ImageNet competition validated deep learning's dominance in computer vision. Reinforcement learning achieved new milestones, notably Google DeepMind's AlphaGo defeating the human Go champion Lee Sedol in 2016, underscoring AI's growing capacity for mastering intricate decision-making processes. Building on this, DeepMind's AlphaZero demonstrated the extraordinary potential of pure reinforcement learning: it learned chess and Go from scratch, i.e. without data from games played by humans, and reached the level of human world champions within only a few hours of self-play.

#### 4.1.5. The Era of Model Scaling (2010s-2018)

Since 2018, AI entered a new era defined by model scaling and architectural innovations. The introduction of transformer-based architectures, exemplified by OpenAI's GPT-2 in 2019, revolutionized natural language processing by demonstrating unprecedented capabilities through increased model size and data volume. Subsequent models such as ChatGPT-3, Google's BERT, and image generation AI tools like DALL-E or Midjourney have continued this trend, enabling advanced text generation, image synthesis, and automation, profoundly reshaping various industries.

The history of AI reflects cycles of innovation, setbacks, and resurgence – each cycle leaving behind progress and lessons learned. Today, AI powered by deep learning and large-scale models offer the first potential effective applications at scale across a variety of human activities.

30 David E. Rumelhart, Geoffrey E. Hinton & Ronald J. Williams, "Learning representations by back-propagating errors", *Nature*, 09 October 1986. <https://www.nature.com/articles/323533a0>

31 Hinton G. et al., "A Fast Learning Algorithm for Deep Belief Nets", *Neural Computation* 18(7), 2006

32 Krizhevsky A. et al., "ImageNet Classification with Deep Convolutional Neural Networks", *NeurIPS*, 2012

## 4.2. Artificial Intelligence Briefly Explained

This section serves as an introduction of the main AI methods, exploring key concepts such as interpolation and distribution shifts (factors that influence how well AI models can adapt to new or unexpected conditions) and explainability and interpretability, which are essential for maintaining transparency and accountability.

### 4.2.1. Rule-based systems

Rule-based systems and traditional automation have been fundamental in driving efficiency, consistency, and reliability in industries for decades. Rooted in well-defined, explicit instructions, these systems follow a straightforward “if-then” logic, automating tasks that involve predictable, repetitive processes. Unlike advanced AI, which learns and adapts based on data patterns, rule-based automation operates under a fixed set of instructions predefined by human experts.

In a rule-based system, decisions and actions are dictated by a set of predefined rules that guide the automation process. For instance, in quality control, a rule might specify: “If a product measurement deviates from a specified tolerance, flag it for inspection.” This logic allows for consistent outcomes, reduces human error, and enhances productivity. Such systems are widely applied in structured environments, such as manufacturing assembly lines, financial transaction processing, and even some early forms of customer service automation. Rule-based systems are particularly effective in scenarios where:

- conditions are predictable, as these systems excel in stable environments with well-understood conditions and minimal variability;
- tasks are repetitive, as they can handle high-volume, low-variation tasks efficiently, from inventory tracking to billing;
- decisions follow clear guidelines, as when tasks can be broken down into discrete, rule-governed actions, such as data entry or transaction approvals.

The most famous example of a powerful rules-based system is Deep Blue, IBM’s chess-playing computer that defeated world champion Garry Kasparov in 1997 by exhaustively evaluating millions of possible moves per second within a predefined set of game rules. While rules-based systems may demonstrate

remarkable problem-solving capabilities within their domains, they rely on explicitly encoded human knowledge rather than learning patterns from data, making them highly specialized but limited in adaptability.

### 4.2.2. Machine learning

As organizations grow more data-driven and complex, traditional rule-based automation is increasingly supplemented or replaced by AI-based systems capable of learning and adapting. They lack flexibility in handling dynamic or complex environments where conditions may change over time. However, unlike rule-based systems, AI-based automation can predict, analyze, and optimize based on evolving data, making it suitable for tasks where decisions are less predictable, and conditions are continuously changing. For example, while rule-based systems may categorize transactions based on fixed criteria, AI models can detect patterns in fraudulent behavior that are constantly evolving.

At its core, AI is designed to replicate certain cognitive functions that are traditionally associated with human intelligence, such as recognizing patterns, understanding language, predicting outcomes, and even problem-solving. AI systems achieve this through machine learning algorithms that process vast amounts of data, identify underlying patterns, and “learn” from those patterns to improve performance over time. This learning process enables AI to adapt its behavior in response to new information or changing conditions, a capability unattainable with static rule-based automation. True AI’s ability to learn from data offers distinct advantages over rule-based systems, particularly in complex, dynamic environments where simple if-then rules fall short. Traditional systems are bound by the rules provided at the outset, making them unable to respond to unfamiliar situations without manual intervention. By contrast, AI systems dynamically update their parameters as they learn from new data, allowing them to handle scenarios previously unseen in their training.

For instance, in financial risk assessment, a rule-based system may classify a transaction as “high risk” based on fixed criteria, such as a threshold amount or location. In contrast, an AI-driven system can learn from historical fraud patterns, adapting to subtle, evolving signs of risk, like transaction timing or multi-device behavior. This adaptability allows AI to evolve as fraud tactics shift, improving its performance over time.

From a classical perspective, machine learning algorithms typically rely on supervised, unsupervised, or reinforcement learning techniques to process data and improve over time:

a) **Supervised Learning:** In supervised learning, the AI system is trained on labeled data, meaning each example in the training set comes with an associated output or “label.” The model learns from these labeled examples, developing an understanding of the relationship between inputs and their corresponding outputs. Once trained, it can generalize from the labeled examples to make predictions on new, unseen data. For example, in image recognition, a supervised learning algorithm might be trained on thousands of labeled images (e.g., images labeled as “cat,” “dog,” “car”) to learn how to accurately classify new images into the correct categories. Common algorithms for supervised learning include decision trees, support vector machines, and neural networks, although neural networks can also be trained using unsupervised learning.

In finance, supervised learning is widely applied to tasks where historical data with known outcomes can guide future predictions. A common example is credit scoring and loan default prediction, where models are trained on labeled datasets containing borrower characteristics such as income, credit history, and repayment records to estimate the probability of default, enabling more informed lending decisions. Another key application is fraud detection, in which algorithms learn from past transactions labeled as fraudulent or legitimate to identify suspicious activity in real time, helping institutions prevent losses and protect customers. Churn prediction is also a valuable use case, as banks analyze customer behavior, such as transaction frequency, product usage, and complaint history, to forecast which clients are likely to leave, allowing proactive engagement strategies to retain them. These applications illustrate how supervised learning transforms labeled financial data into actionable insights that improve risk management, operational efficiency, and customer relationships.

b) **Unsupervised Learning:** Unlike supervised learning, unsupervised learning works with data that has no labels. The algorithm attempts to discover patterns, groupings, or structure within the data on its own. Clustering and association are common types of unsupervised learning tasks. For example, clustering algorithms may analyze a dataset of customer information and group individuals

based on purchasing behavior, demographics, or other similarities. Unsupervised learning is particularly useful in exploratory data analysis and dimensionality reduction, where it can help simplify complex datasets or reveal underlying patterns that were not previously known. Examples of unsupervised learning algorithms include k-means clustering, principal component analysis (PCA), and hierarchical clustering.

In finance, unsupervised learning is frequently used to uncover hidden patterns and structures in large, unlabeled datasets. One common application is anomaly detection in transactions, where algorithms analyze historical payment or trading data to flag unusual patterns that may indicate fraud, money laundering, or system errors, without needing prior examples of specific incidents. Another valuable use case is topic modeling in financial news, where natural language processing techniques identify recurring themes in unstructured text such as earnings calls, news articles, or analyst reports, helping investment teams detect emerging market signals or sector trends. These approaches enable financial institutions to gain insights and manage risks even when labeled training data is unavailable.

c) **Reinforcement Learning:** Reinforcement learning (RL) is a type of machine learning where an agent learns by interacting with an environment and receiving feedback in the form of rewards or penalties based on its actions. The goal of the agent is to maximize cumulative rewards over time, learning an optimal policy for decision-making in the environment. Reinforcement learning is often used in scenarios that involve sequential decision-making, where the agent’s actions influence future states. For instance, in robotics, RL algorithms can be used to train a robot to navigate a maze by rewarding it for making correct moves and penalizing it for errors. Reinforcement learning is also widely applied in gaming, where agents learn strategies to win games, and in autonomous driving, where a model learns to make driving decisions in real time.

One of the most typical applications of reinforcement learning in finance is algorithmic trading, where an RL agent learns to develop and adapt trading strategies by interacting with market environments and receiving feedback in the form of profits or losses. Unlike traditional rule-based systems, these agents continuously refine their decision-making by simulating trades, evaluating outcomes, and adjusting actions to maximize long-term returns rather than short-term gains.

This approach allows the model to capture complex market dynamics, respond to evolving conditions, and potentially identify profitable opportunities that human traders or static algorithms might miss. By balancing exploration of new strategies with exploitation of proven ones, reinforcement learning enables trading systems to adapt in real time to shifting market conditions.

#### 4.2.3. Neural networks

In many modern applications where data availability is abundant, neural networks have emerged as one of the most powerful machine learning techniques. Neural networks are a foundational technology in artificial intelligence, inspired by the structure and function of the human brain. Modelled after biological neurons, neural networks consist of interconnected nodes (or “neurons”) that process and transmit information. Just as neurons in the brain respond to stimuli and communicate through complex pathways, artificial neural networks learn from data through layers of interconnected nodes.

One of the most powerful aspects of AI is its ability to recognize complex patterns that are beyond human perceptibility. For example, deep learning, a subset of machine learning that uses neural networks with multiple layers, can identify intricate structures within data.

In environmental monitoring, deep learning models are applied to satellite imagery, detecting deforestation patterns or analysing crop health with far greater precision than manual or rule-based methods.

The effectiveness of AI is however tightly linked to the quality and diversity of its data inputs. Unlike rule-based systems, where outcomes are predetermined, AI systems require a continuous stream of relevant data to stay effective. Over time, as AI models process more data, they become more accurate and robust, enhancing their predictive capabilities. However, this dependency also introduces challenges, such as data privacy, bias, and the need for ongoing data management, all critical considerations in building responsible AI systems within a sustainability framework.

In more detail, deep learning refers specifically to neural networks with multiple layers (or “deep” architectures) that allow them to model highly complex data patterns. These networks “learn” by adjusting the strength, or weights, of connections between nodes based on patterns detected in the

training data. During training, neural networks iteratively update these weights to minimize errors in predictions, gradually improving their ability to recognize complex patterns. This learning process enables neural networks to capture intricate relationships in data, making them particularly effective in fields such as image recognition, natural language processing, and predictive modeling.

The theoretical power of neural networks is supported by the Universal Approximation Theorem, a fundamental concept in machine learning. This theorem states that, given a sufficient number of neurons and layers, a neural network can approximate any continuous function to a desired level of accuracy. This makes neural networks uniquely capable of capturing complex, nonlinear relationships within data. Unlike traditional statistical models, which may rely on linear relationships or require explicit rules, neural networks can learn directly from raw data and handle a high degree of complexity. For this reason, neural networks are widely used in applications where the relationships between inputs and outputs are difficult to define manually, allowing them to adapt and improve with more data.

By training on vast amounts of image data, these networks can recognize subtle changes in landscapes that might otherwise go unnoticed, supporting environmental sustainability and conservation efforts. In finance, neural networks help in risk prediction models by learning patterns in financial data, enabling the detection of fraud, forecasting economic risks, and identifying potential compliance issues. This capability to detect patterns and anomalies aids organizations in making data-driven decisions that align with sustainable practices.

#### 4.2.4. Discriminative AI and Generative AI

Discriminative AI and Generative AI represent two fundamental branches of artificial intelligence, each with distinct functions and applications. Discriminative AI focuses on forecasting and classification, deriving insights and patterns from historical data to make informed predictions about future events or outcomes. This type of AI doesn’t generate new data but instead analyzes existing data to recognize trends, correlations, and relationships. In contrast, Generative AI moves beyond analysis to create new data points, based on learned patterns within the training data. Generative models are designed to produce synthetic outputs that resemble real data. Generative AI typically operates by learning the underlying structure and distribution of the input dataset, allowing it to sample from this distribution and

generate new examples that mirror the properties of the original data.

#### 4.2.4.1. Discriminative AI

Discriminative models are particularly powerful for applications requiring precise forecasts, such as sales projections, risk assessment, or healthcare diagnostics. Among the most widely used techniques are regression analysis and classification algorithms. Regression analysis is used to model the relationship between a dependent variable and one or more independent variables, allowing for precise quantitative forecasts in fields like finance or demand forecasting. Classification algorithms, on the other hand, categorize data into distinct classes.

Regression analysis is a core technique in predictive modeling, used to understand relationships between variables and predict continuous outcomes based on one or more input factors. Linear regression, one of the simplest forms, assumes a proportional relationship between input and output variables, which allows for straightforward predictions in stable environments.

For example, an organization might use linear regression to predict future energy consumption based on historical data, enabling more efficient resource allocation and contributing to sustainable operations. Polynomial regression, which models non-linear relationships, is useful in capturing more complex dynamics, such as predicting crop yields based on seasonal weather variations.

Regression models find broad applications to support sustainable development ambitions. They can forecast critical environmental indicators like air quality, water consumption, or pollution levels, helping organizations take proactive measures to mitigate environmental impacts. Similarly, regression models aid in resource allocation, as seen in sustainable agriculture, where models predict crop yields based on soil quality, rainfall, and other variables, allowing farmers to optimize their use of resources and minimize waste. Through precise forecasting, regression models empower organizations to make responsible, forward-looking decisions aligned with sustainability goals.

Classification algorithms, on the other hand, are designed to categorize data into predefined classes based on patterns learned from historical data. Unlike regression, which predicts continuous values, classification assigns discrete labels, making it ideal for applications that require categorization rather

than precise numerical forecasts. Commonly used classification models include decision trees, support vector machines (SVMs), and logistic regression, each suited to different types of classification tasks. Decision trees break down data into smaller subsets based on decision rules, resulting in an easily interpretable, tree-like structure that clearly explains how classifications are made. This transparency is beneficial in sustainability contexts, such as supply chain risk assessment, where a decision tree could classify suppliers based on risk factors like environmental or labour practices, aiding in responsible sourcing. Support vector machines, on the other hand, classify data by identifying a hyperplane that optimally separates categories. These models are particularly powerful in detecting anomalies in datasets, such as identifying patterns that suggest fraudulent activity, making them valuable in governance and compliance efforts. Logistic regression, though technically a regression technique, is widely used for binary classification tasks. It predicts the likelihood of a binary outcome, such as “compliant” versus “non-compliant” or “high risk” versus “low risk.” For instance, logistic regression models can predict the probability of a project violating environmental regulations, allowing companies to implement preventive actions.

In the realm of sustainable finance, classification algorithms offer diverse applications. They play an essential role in social governance by enabling organizations to analyse customer sentiment or employee feedback, helping monitor and address social concerns. By categorizing data from social media or survey responses, these models can highlight public sentiment about environmental initiatives, allowing companies to adjust strategies in response to shifting public perceptions. Classification also aids in risk detection and compliance. Models can analyse supplier data to assess risks related to labour practices or environmental compliance, helping organizations maintain ethical supply chains and avoid reputational and legal risks.

For more complex, where non-linear behaviour in the data is harder to capture, advanced methods like ensemble techniques and neural networks can improve accuracy. Ensemble methods, such as Random Forests and Gradient Boosting, combine predictions from multiple models to create a more robust classifier. In fraud detection in the financial sector, for instance, ensemble models can aggregate insights from several algorithms, increasing the likelihood of identifying suspicious transactions. For data-rich tasks, such as image or text classification, neural networks can categorize data with a high level of accuracy.

#### 4.2.4.2. Generative AI

To understand why generative AI is effective, it helps to look at it from a density estimation perspective. Density estimation refers to modeling the probability distribution that best represents the underlying structure of a dataset. For generative AI, this involves learning the probability distribution, or “density,” of the data in its training set. By accurately capturing this distribution, generative models gain the ability to sample new points from this learned distribution, creating new data that closely resembles the original set.

Consider a dataset of human faces. Each face in the dataset can be thought of as a data point within a complex, high-dimensional space, where dimensions represent features such as eye color, nose shape, and skin tone. The density of these points, where faces cluster and where they are sparse, tells us about the structure and variability of the data. Classical generative models include: (1) Generative Adversarial Networks (GANs), a popular generative model where two neural networks, one generating and the other discriminating, compete to improve output quality. (2) Variational Autoencoders (VAEs), which generate new data by learning a compressed, probabilistic representation of input data. Autoregressive Models: These models, often used in time-series analysis, predict the next element in a sequence based on previous ones.

From a density estimation perspective, generative models have several objectives that make them particularly effective at creating realistic data. First, they are designed to capture complex patterns in data. Real-world datasets often follow complex, multimodal distributions, meaning that data points cluster around multiple “modes” or common configurations. Generative models learn these patterns, capturing both dense clusters of frequently occurring features and rare features that produce outliers or unique cases. By doing so, they create realistic samples that reflect the full diversity of the training data, representing both typical and atypical data points within the distribution.

Second, generative models are constructed to avoid overfitting to exact data points. While traditional predictive models might attempt to learn precise values for each data point, generative models instead aim to understand the general structure or “shape” of the data distribution. This approach enables them to avoid memorizing the dataset, allowing them to generate novel yet plausible data points that follow the same distribution as the original data. This flexibility is especially

valuable in applications where realistic, diverse examples are needed beyond those present in the training data.

Another important objective of generative models is their ability to sample from the learned distribution. Once a model has captured the density of the training data, it can generate new data points by sampling from this learned distribution. For instance, in a text generation model, after learning the structure of language, the model can produce sentences by sampling words in a way that reflects the probabilities observed in its training data. Each new sample represents a unique, synthetic example, whether a sentence, an image, or a sound, that closely resembles the characteristics of the original dataset while introducing novelty.

Finally, generative models have the potential to fill gaps in sparse data. Many real-world datasets are incomplete or imbalanced. For example, in scientific research, certain experimental conditions may be underrepresented due to limited data availability. By learning the underlying data distribution, generative AI can create synthetic data that fills in these gaps, providing more complete datasets that support robust analysis and simulation. This capacity to generate synthetic data has direct implications in fields like medicine and climate science, where generative models can simulate rare scenarios or environments without requiring extensive real-world data collection.

#### 4.2.4.3. Large generative language models

Large language models (LLMs) apply the same density estimation perspective to text. Empirical scaling laws for language models show that, over wide ranges, performance improves in a remarkably predictable way as we increase three factors together: model size (number of parameters), dataset size, and compute used for training. As these ingredients scale, the model’s loss on held-out text decreases roughly according to a power law. In density estimation terms, this means that larger models trained on more data become systematically better at approximating the true distribution of text found “in the wild.” This observation helps explain why simply making models bigger and training them longer on more diverse corpora has led to qualitatively new capabilities, from fluent dialogue to in-context learning and the revolution we have observed as large language models have transitioned from labs to industry applications.

Fundamentally, LLMs can still be viewed as large-scale autoregressive density estimators over sequences of tokens.

Training on massive text datasets pushes the model to assign high probability to sequences that resemble natural language and lower probability to implausible continuations. Generation is then just sampling from this learned distribution: the model repeatedly proposes the next token according to the learned distribution of the next token given all the previously generated or observed ones, ultimately producing coherent text that reflects the statistical patterns of its training data.

This statistical objective also explains so-called hallucinations. LLMs are not explicitly trained to maintain a consistent world model or to verify facts; they are trained to produce the most probable continuation of the current text. When the input asks for information that is underrepresented, ambiguous, or conflicting in the training data, the model will still produce a high-probability completion according to its internal distribution. The result can be confident, fluent statements that are not grounded in any specific source or reality. Essentially “statistically reasonable guesses” rather than validated facts.

Retrieval-augmented generation (RAG) strives to address this limitation by conditioning the model’s predictions on external evidence. Instead of relying solely on its parameters, a RAG system first retrieves relevant documents (for example, via a vector search over an indexed knowledge base) based on the user’s query. These retrieved passages are then appended to the prompt, so the model now estimates a distribution taking into account the new evidence coming from the retrieved passages. In practice, this shifts behavior from “continue this text in a likely way” toward “summarize and integrate the most likely continuation that is consistent with these concrete sources.” As a result, RAG reduces hallucinations, improves traceability, and allows LLMs to stay up to date and domain-specific without retraining the underlying model.

#### 4.2.5. Interpolation and Distribution shifts

One of the key strengths of AI models, particularly those trained on large datasets, is their ability to perform interpolation – predicting outputs for data points that fall within the range of values they encountered during training. Interpolation involves generalizing from the examples in the training set to make accurate predictions on new data that is similar to, but not exactly the same as, what the model has previously seen.

For instance, an AI model trained on historical weather data may effectively predict short-term climate patterns within the usual range of variation. This capacity for interpolation

allows models to excel at tasks like pattern recognition, classification, and prediction in relatively stable environments.

However, AI models often struggle with distribution shifts, which occur when the new data they encounter falls outside the range or distribution of their training data. Distribution shifts can happen in various ways, such as changes in the environment, unforeseen events, or evolving trends.

In the context of sustainable finance: distribution shifts are particularly relevant as they represent situations where an AI model encounters unexpected conditions, such as abrupt changes in climate patterns, economic fluctuations, or new regulatory requirements.

For example, a model trained to predict energy demand based on historical data might underperform when confronted with drastic weather anomalies caused by climate change, as the conditions are significantly different from those in its training set. Similarly, a financial risk model may falter if it encounters unprecedented economic conditions, like a sudden recession or rapid market changes, which it hasn’t been trained to recognize.

These limitations underscore the need for robust, adaptable AI models in ESG applications, particularly in areas where rapid or unpredictable changes are common. Techniques like domain adaptation and distributional robustness are being explored to help models better generalize under such shifts. Additionally, building models that can recognize and signal when they are operating under unfamiliar conditions can improve reliability.

In the context of sustainable finance, having models that can adapt to or account for distribution shifts is crucial, as many areas, such as environmental monitoring and social impact assessments, are influenced by dynamic and often unpredictable factors. By designing models to handle distribution shifts, organizations can ensure that their AI tools remain accurate and effective even as real-world conditions evolve, supporting sustainable and responsible decision-making.

#### 4.2.6. Explainability and Interpretability

In fields with ethical and social implications, the transparency of AI models is essential for ensuring trust, accountability, and informed decision-making. Two related but distinct concepts, explainability and interpretability, are key to understanding and improving model transparency. Explainability refers to the ability to describe how a model arrives at a specific decision or prediction, while interpretability focuses on understanding the inner workings of the model itself, including the relationships it identifies within the data.

In sustainable finance, these concepts are particularly important since AI may be applied in high-stakes contexts such as financial risk assessment, environmental impact prediction, or corporate social responsibility. A lack of transparency can lead to distrust or even harmful decisions, making it crucial that stakeholders can understand and justify the outcomes produced by AI models. For instance, an AI model used to assess credit risk may label certain transactions or individuals as high-risk.

If the model's reasoning is opaque, affected parties cannot understand or challenge the decision, which can lead to issues of fairness and accountability.

Explainable AI methods, such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations), have been developed to address this challenge by providing insights into model predictions. SHAP assigns an importance score to each feature, indicating its contribution to the prediction, which helps stakeholders understand why certain factors were prioritized. LIME, on the other hand, creates a simplified version of the model around a specific prediction, enabling users to see which features were most influential in that particular decision. These tools make it possible to inspect and interpret individual predictions, which is vital for validating that models align with ethical and regulatory standards.

Interpretability, on the other hand, focuses on understanding the model's internal mechanics, such as how data flows through different layers in a neural network or how a decision tree splits at each node. Interpretability is important not only for regulatory compliance but also for risk management. For example, if a model used for monitoring is interpretable, researchers and stakeholders can understand its decision pathways, identify any biases or inconsistencies, and modify

the model to improve performance or fairness. Models that are interpretable allow organizations to audit and refine their AI systems continually, ensuring that the models are aligned with organizational objectives over time.

By enhancing both explainability and interpretability, organizations can use AI models in ways that are accountable, transparent, and fair. This level of transparency is essential, especially where the impact of AI-driven decisions often extends beyond individual organizations to society and broader stakeholders. As AI systems become increasingly embedded in sustainability-related applications, such as environmental monitoring, social equity assessments, and governance compliance, prioritizing transparency will be crucial for building public trust, aligning with ethical standards, and ensuring that AI contributes positively to sustainable and responsible business practices.

### 4.3. Glossary

You can find our Artificial Intelligence glossary here:



### 4.4. Bibliography

Literature on Artificial Intelligence evolves at a pace that makes any bibliography obsolete within months. To address this issue, we have chosen to put the bibliography online and update it regularly. You can access it here:



## 4.5. List of Figures

- Figure 1:** The four main benefits of AI applications
- Figure 2:** The rise of Artificial Intelligence capabilities
- Figure 3:** Stage of artificial intelligence adoption in the organisation
- Figure 4:** Artificial intelligence use cases in sustainable finance for asset managers and asset owners (in % of respondents)
- Figure 5:** Key challenges in artificial intelligence implementation for sustainable finance for asset managers and asset owners (in % of respondents)
- Figure 6:** AI applications in sustainable finance along the data-to-decision workflow
- Figure 7:** Quanthome AI generating a building-level ESG analysis
- Figure 8:** Quanthome AI system architecture
- Figure 9:** Clarity AI's data collection & curation workflows
- Figure 10:** Emissions reduction units trading
- Figure 11:** Combining human and artificial intelligence for trusted risk insights
- Figure 12:** Dashboard of the DIZH BioVaR demonstrator for assessing the impact of biodiversity loss and deforestation
- Figure 13:** Swiss real estate funds: average building age vs. NAV
- Figure 14:** Comparative risk & return statistics: ESG vs. Non-ESG
- Figure 15:** Cumulative return performance: ESG vs. Non-ESG
- Figure 16:** Illustrative example of the SDG profile of a thematic strategy
- Figure 17:** Illustration of AI-powered company briefs for investment research
- Figure 18:** "AI Insights" – one of Clarity AI's copilot features
- Figure 19:** A simplified timeline of AI developments

**Main Author**

Romain Leroy-Castillo, Director AI and Education, Swiss Sustainable Finance

**Supporting Authors**

Dr. Joao Carvalho, researcher at ETH Zurich, co-founder and CTO of Uthereal

**Acknowledgements**

Swiss Sustainable Finance would like to thank the following contributors for providing expert inputs:

- Clarity AI
- Conser
- EY
- Pictet
- Quanthome
- Ram AI
- RepRisk
- Universität Zürich
- ZHAW

**Cover Image**

noord.ch

**Design**

noord.ch

**Publisher**

Swiss Sustainable Finance  
Grossmünsterplatz 6  
8001 Zurich  
044 515 60 50  
sustainablefinance.ch

January 2026

The mission of Swiss Sustainable Finance (SSF) is to strengthen Switzerland's position as a leading voice and actor in sustainable finance, thereby contributing to a sustainable and prosperous economy. The association, founded in 2014, has representative offices in Zurich and Geneva. Currently, SSF unites over 250 members and network partners from financial service providers, investors, universities and business schools, public-sector entities and other interested organisations.

**Legal disclaimer**

This information material was produced by Swiss Sustainable Finance (SSF) with the greatest of care and to the best of its knowledge and belief. However, SSF provides no guarantee regarding its content and completeness, and does not accept any liability for losses that might arise from making use of this information. The opinions expressed in this information material are those of SSF at the time of writing and are subject to change at any time without notice. Unless otherwise stated, all figures are unaudited. This information material is provided for information purposes only and is for the exclusive use of the recipient. It does not constitute an offer or a recommendation to buy or sell financial instruments or services, and does not release the recipient from exercising his / her own judgment. This information material may not be reproduced either in part or in full without the written permission of SSF.