What AI Can Do for Investors in ESG

Zakaryae Boudi, Jiulin Teng

VOGOSEN

2021

Contents

Executive Summary

New and evolving challenges in ESG necessitate evidence-based ESG practice. Thankfully, recent developments in Artificial Intelligence (AI), in particular Natural Language Processing (NLP), have made a rich, data-driven ESG strategy a reality. In this White Paper, we present five AI modules that enrich an ESG Information System to underpin this strategy. These include Dynamic ESG Materiality Appraisal that identifies and prioritizes ESG factors for each individual company, ESG Sensor that alerts investors of ESG events and predict their impacts, ESG Risk Assessment that harmonizes third-party scores and facilitates proprietary scoring, ESG Opportunity Radar that identifies ESG investment opportunities, and ESG Reporting Assistance that reduces compliance burden and improves transparency. These modules and more can be developed for the need of each investor and deployed progressively such that benefits become evident quickly and without heavy investment. The progressive nature of this deployment also makes it a nature process for the organization to grow strong capabilities that bring lasting competitive advantage. Fully built, AI-Powered ESG Information System allows the investor to do more with less.

Introduction

ESG must be evidence-based and data-driven. AI with matching information system is needed to do it well in today's environment.

As ESG becomes an integral part of investing, being able to manage effectively large quantities of often unstructured data is increasingly a key competitive advantage. This is particularly true as the public grows more wary of greenwashing, and as third-party rating services use inconsistent methodologies to derive dissimilar scores. In essence, ESG must be **evidence-based** and **data-driven**.

To collect, process, analyze, and utilize big data in an actionable manner, investors can leverage artificial intelligence (AI)[1,2]: AIs are **versatile**. They can be trained for a broad range of ESG-related tasks and involve varying degrees of complexity. AIs are **fast**. They can work 24/7 at speeds that human analysts cannot match without performance degradation. AIs are **scalable**. With faster hardware (e.g. on the cloud), they can process even larger quantities of data; when demands are low, they can be scaled back easily.

In this White Paper, we demonstrate five types of AIs that investors can develop and deploy for ESG. Starting with **Dynamic ESG Materiality Appraisal**, AI is able to help investors break with static ESG materiality mapping as well as identify and rank materiality issues for each company individually. With **ESG Sensor**, investors can continually monitor events that affect material issues for their portfolio; AI is also able to forecast how these news and events will impact long-term ESG performance.

ESG Risk Assessment and **ESG Opportunity Radar** are potentially a pair of AIs that help investors manage risks and identify opportunities. On the risk side, AI can help investors harmonize third-party risk reports; moreover, it is able to support detailed, proprietary ESG risk scoring using logics that are meaningful to the decision-makers. On the opportunity side, AI is able to identify investment opportunities from the perspective of ESG at industry, company group, and company levels.

Lastly, we touch on **ESG Reporting Assistance**, which eliminates human error, increases reporting speed and consistency, and enhances transparency to not only regulators but also clients and other stakeholders.

We must emphasize two points: First, the use of AI in ESG is far from limited to these five areas. They serve as examples to illustrate that many areas of ESG either can only be done with AIs or can only be **done well** with them.

Second, a well-designed **information system** is the foundation of any data strategy and the success of AI. It affords privacy and security as well as accessibility and scalability to the investors. On its top, various AI modules can be deployed to provide meaningful services.

Indeed, AIs can be significantly more cost-effective than human analysts in many ESG tasks. It may serve as **an equalizer** in the investment industry, for smaller players armed with purposefully-designed AIs can compete, at least in ESG analytical capabilities, with big investors[3].

While the amount of data involved and the complexity of the algorithms may appear daunting to the untrained eye, the methods that we use are **proven** by years of real-world applications such as credit rating[4] and default forecasting[5]. Further, we recommend that investors deploy AI-powered ESG information system **progressively**, with close involvement of the people who will actually use it[6]. This way, investors can reap the benefits quickly, and the organization can grow with the information system, thereby building robust competitive advantage.

This is also why we advise against using SaaS or third-party scores in ESG **blindly**: With a proprietary information system powered by AI, investors have full control over their data and their decision-making. It also grows over time into the **organization's capabilities**. In fact, a good ESG manager today is one who is tech-savvy and who masters advanced information systems to help the organization beyond basic spreadsheet-based analysis.

Last but not least, the AI-powered ESG information system that we discuss in this White Paper helps **all teams in the organization work together** more smartly. It makes data sharing and presentation across the firm straightforward, while AI-based decision maps reveal reasons behind each signal, so that the decision-makers can verify before taking action together. Naturally, this information system can be built in conjunction with, or as an extension to, existing information systems that the investment teams use. Outputs from the ESG modules, thereby, can feed directly into the fundamental analysis.

AI in ESG is a trend that brings once-in-a-generation **opportunities**. We at Vogosen look forward to working with investors to bring it to fruition.

We recommend that investors deploy AI-powered ESG information system progressively, with close involvement of the people who will be using it.

1. Dynamic ESG Materiality Appraisal

Hetereogeneity among companies and the evolving ESG space makes it critical to identify and prioritize ESG factors dynamically.

Materiality refers to **all** environmental, social, and governance (ESG) aspects that, if not well managed or dealt with, can result in value destruction and other financially adverse consequences. In practice, companies and investors tend to focus on material ESG aspects with the help of the SASB standard[7] as well as multiple ESG data sources[8] in their strategic decision making.

However, two factors call into question the effectiveness of this static approach: On the one hand, each company operates in a unique set of geographic, temporal, supply chain, and market circumstances. Their exposure to each ESG aspect and the relative severity of each exposure are **heterogeneous** even for companies in the same industry sector.

On the other hand, ESG material aspects and factors are becoming increasingly mixed and **interconnected** in new combinations as a result of changes in company practices, business models, market conditions, and indeed the envelope of technology. A deeper understanding at a more granular level beyond conventional sector- and geography-based delineations can support more effective decision-making.

For example, two software companies may be exposed to dramatically different sets of material issues, depending on when, where, and how they operate. One needs to deal with data privacy and security considerations, because of local tensions and cybercriminal threats. The other needs to be more concerned with intellectual property protection and customer welfare. Material aspects, in effect, constitute the hotspot for both companies and investors willing to address ESG risks—and the multiple market dynamics might be redefining that hotspot at breakneck speed.

These factors make it critical to **identify and prioritize ESG factors** that become material on a **dynamic basis**. Materiality appraisal, likewise, needs to adapt to the unique, evolving context, while providing a granular grasp of the signals and triggers that may lead to unsuspected risks.

Hence, a rich pool of factors and parameters should displace industry and sector classes as the unit of analysis to which materiality is approached. They should also be organized and accessible in an **actionable** manner. Such deep assessment of the material issues that are currently applicable for a specific company or a given portfolio enables stakeholders and investors to spot actual ESG risks better, to ensure that their strategies are relevant and up-to-date, and to build more effective forecasts. A dynamic materiality taxonomy/ontology sits at the core of each of these key activities and is necessary to reap the full reward of materiality analysis[9]. It also underpins the **ESG Sensor**, **ESG Risk Assessment**, and **ESG Opportunity Radar** that we discuss in other sections of this White Paper.

Currently, the major barrier to this end is the limited capacity to **access and process timely information** that is fragmented, unstructured, and heterogeneous. This is where Big Data and AI can shine for Dynamic ESG Materiality Assessment[10].

Big Data and AI technologies can be

applied to harness a wide variety of data sources, based on models' ability to **identify relevant ESG information** and **capture underlying relationships**, especially with regards to financial materiality as well as risk management.

In particular, **Natural Language Processing** (NLP)[11], a subset of AI techniques, can be used to analyse data feeds (i.e., news, tweets, financial data, etc.) in real-time and to forecast aspects that may become material for a company in a given context, taking into account static and slowly evolving information. Crucially, AI can perform such an assessment in a way that human analysts cannot, in terms of speed, quantity, granularity, and consistency.

A concrete example of leveraging AI for dynamic ESG materiality appraisal is **live processing of news feeds at scale**. First, investors can start by deploying targeted crawlers governed by manually chosen rules, graph metrics, and/or machine learning-based algorithms to collect thousands of news articles, in real time, from a large set of media and web portals.

Second, data ingestion would rely on a container-based data-processing architecture (Big Data) such as the Kappa Architecture. Successive NLP-based algorithms extract and normalize all the relevant pieces of information. For instance, the **Named-Entity Recognition** (NER) module detects industry, company names, locations, and regulatory authorities, etc. **Dependency Parsing** modules with specialized **classification models**, meanwhile, isolate meaningful phrases / paragraphs and map them to the Materiality ontology. These can be fine-tuned on large transformer-based machine learning **Language Models** such BERT[12] and GPT-2. An additional **Sentiment Analysis** module can infer a sentiment score—before handing them as further inputs to the analytics block of the system[13].

Third, the analytics block generates "signals" or insights in such an information system. It combines outputs from unstructured data fed by the NLP with other available data, including metadata (e.g. publication date, portal address, and article length, etc.) and other computed indicators (e.g. topic frequency, topic growth, and average sentiment scores, etc.) from the news source as well as financial, supply chain, and market data sourced from APIs (e.g. Bloomberg).

This block can be built with a spectrum of techniques. To start with, a **rules-based** analytics system can be programmed and fine-tuned manually with the help of experts. The advantage of such a such is that it is quick to deploy and does not require AI training. Its disadvantage can be found in its lack of adaptability and its rigidity: In other words, all updates need to be made manually, and it is incapable of identifying trends that the programmers / experts are unaware of.

In conjunction, analytics can be enriched by on a range of **machine learning-based**

Dynamic ESG Materiality Appraisal AI affords investors timely access and analysis of ESG information backed by traceable reasoning.

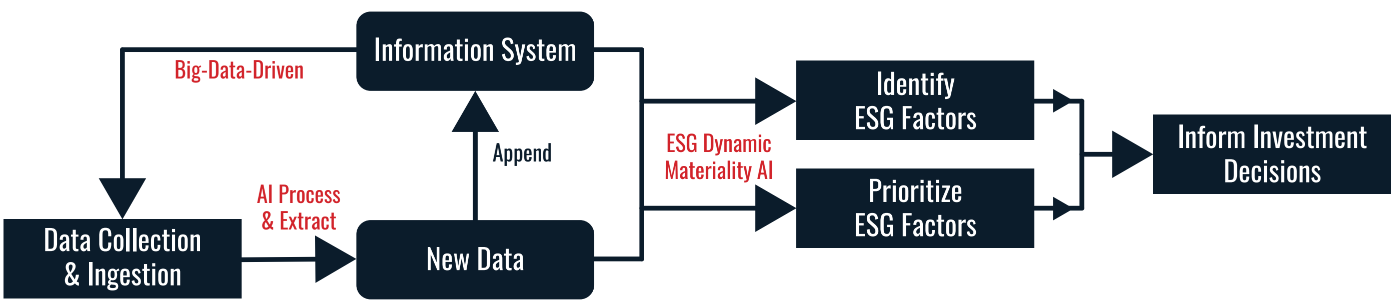

Figure 1. Overview of Dynamic ESG Materiality Appraisal

A hallmark of AI-powered ESG Information System is the ability to trace the reasoning to the data sources on demand.

(statistical) models. Depending on the amount of inputs taken into consideration and the availability of data for training, these proprietary models can explore patterns and output signals in varying specificity and granularity. They are adaptable and flexible, which means that these models, when trained with sufficient data, are more capable of identifying shifting dynamics in ESG Materiality for each specific company. They can also detect patterns that experts are unaware of or overlook.

With a well-structured analytics block that offers high-quality processing, distribution, and visualization , asset managers and institutional investors not only get the patterns and signals: they are also provided with reasoning **traceable** to the ingested data source(s)[14]. As such, decision makers can verify before taking action.

It should be noted, however, that the application of AI and NLP in the ESG space have yet to be implemented at scale. Substantial investment and particular skill sets are needed. Each element needs to be implemented, tested, and deployed progressively.

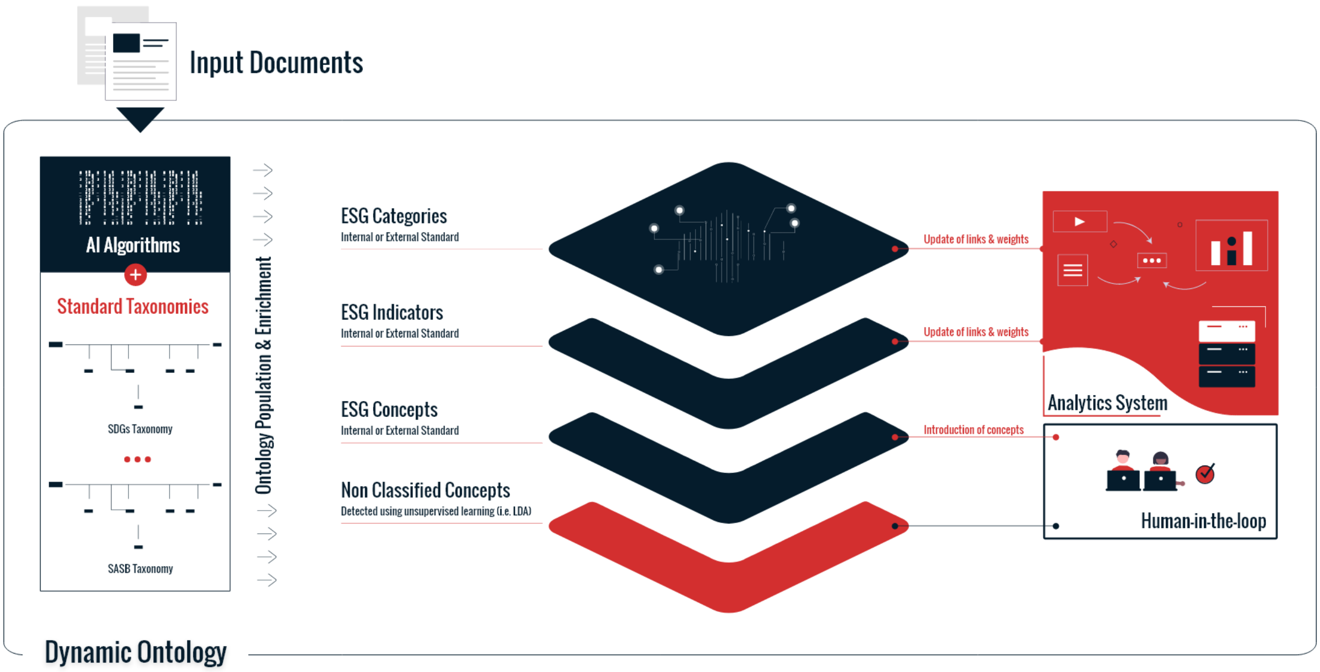

Figure 2. Dynamic Ontology is the Backbone of ESG Materiality Appraisal and ESG Sensor

2. ESG Sensor

The motivation for Dynamic ESG Materiality Appraisal also applies to the assessment of ESG performance in a portfolio. Namely, each company has its unique set of exposures, and these evolve continually and in some cases dramatically within a short period of time.

While some of the factors driving ESG performance are long-lasting, others are heavily **event-driven**. In particular, the social aspects of ESG can be quite sensitive to events, perhaps on a sector- or geography-basis. Governance issues, while changing more slowly and predictably, may also be sensitive to specific events ranging from those that impact all companies to those that affect only a single company. The same is true for environmental issues, which perhaps deliver fewer but nonetheless impactful surprises.

While decision-makers may wish to follow these changes personally, as the size of the portfolio increases it quickly becomes impossible to cover all essential sources in a timely fashion.

With AI, investors can monitor **more sources**, weighted in their preference, **with virtually no lag**, covering not only companies in their portfolio but also potential targets. Moreover, a more advanced AI can forecast ESG performance based on collected information, giving asset managers and ESG teams a head start in tackling potential risks and in assessing potential opportunities before they become public.

The **ESG Sensor-Monitor AI** can ingest, process, and analyze unstructured data similar to the Dynamic ESG Materiality Appraisal AI that we described in the last section. In short, a customized crawler selects desired data from sources such as news feeds, Tweets, press releases, and regulatory announcements. This data is stored in a document store (NoSQL database or Document Store function of modern SQL databases such as Oracle, SQL Server, PostgreSQL, or MySQL).

As the raw data is fed into the database, the NLP module leverages appropriate language models based on deep neural networks to extract pieces of information and assess the data holistically. Here, the shorter length and less formal language of Tweets juxtaposed with the longer length and more formal language of press releases illustrate the need for specialized models.

The analyzer module here, in comparison, is somewhat different than the one for Materiality Appraisal: It can similarly be either rule-based or statistical (machine learning-based). However, it needs to be programmed or trained for a different task. In this case, **alerting the investor of actionable and/or otherwise significant events**. As before, this module will provide traceable alerts, meaning that the decision maker can assess not only the source(s) but also be presented with meaningful motivations (ontology)[15].

The **ESG Sensor-Forecast AI** goes one step further. Building on the NLP module of the ESG Sensor-Monitor AI, the Forecast AI incorporates other data available in the information system to **forecast the impact of the event** on future ESG performance.

This is possible thanks to two facts: One,

ESG Sensor AI alerts investors of actionable events and forecasts their impact on future ESG performance.

ESG Sensor AI can process more data faster and identify non-obvious patterns compared to human analysts.

an AI can **process more data faster** than human analysts can. Two, a well-trained AI can **identify non-obvious patterns** in forecasting that human analysts, by nature of not being able to process data in the same order of magnitude, cannot.

It should be noted that this is a more involved AI and would require significant training and tuning. As it is the case with Dynamic ESG Materiality Appraisal, this forecast AI can be continually improved after deployment.

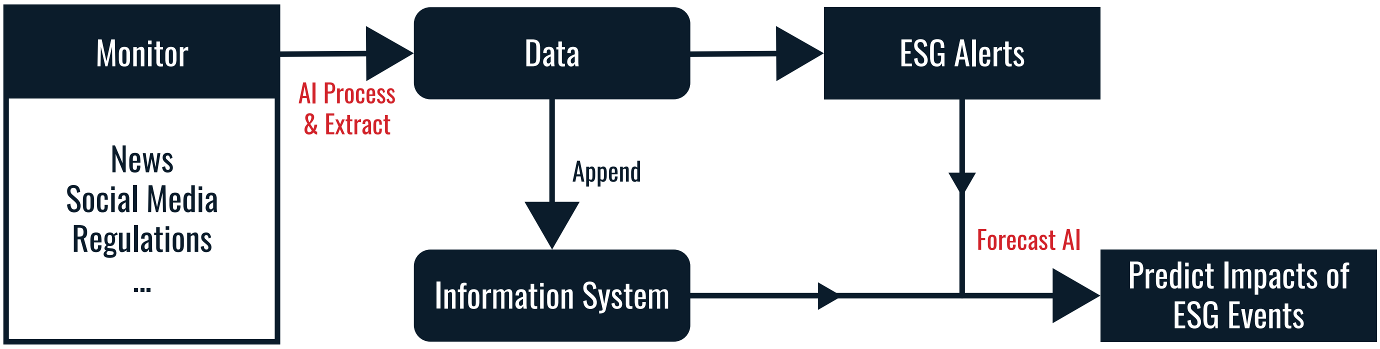

Figure 3. Overview of ESG Sensor

3. ESG Risk Assessment

Business and investment communities find themselves increasingly exposed to new types of ESG issues that make ESG risk assessment an integral part of their investment decisions.

There is not a single best way to assess a company's ESG risks, but all current methods attempt to construct a global risk score or rating in a process that progressively abstracts different combinations of ESG and financial materiality inputs.

However, for effective investment decision-making, opaque risk scores are insufficient. Investors need to be able to **explain the rationale** behind the scores, so as to avoid greenwashing or delegating decisions to unaccountable third parties.

Meanwhile, due to heterogeneity in approaches to the abstractions, third-party ESG risk scores usually vary markedly among scoring providers[16]. Directly comparing different scores without a solid grasp of the logic involved can be more misleading than ignoring ESG altogether[17,18]. Further, such divergences complicate incorporating ESG risk analysis to one's investment decisions.

With an information system backed by ESG Risk Assessment AI, however, these issues can be tackled.

ESG Risk Assessment – Harmonizer AI helps investors harmonize third-party risk reports. Using various techniques in NLP, it is able to analyze each risk report / score individually, deconstruct their abstractions, and map them to in-house standards and approaches[19]. It is also able to highlight the factors underlying various risk scores. In this process, the Harmonizer AI leaves trails of reasoning behind that can later be used to explain the reasoning on demand.

An important feature of AI in harmonizing heterogeneous external ESG scores is its ability to **identify complex patterns** that are non-obvious to human analysts. As such, the ESG Risk Assessment – Harmonizer AI can validate the consistencies of third-party scores and **perform meta analysis**.

ESG Risk Assessment – Scorer AI takes this one step further to underpin proprietary ESG risk scoring. It can build logical reasoning around identified risks and their assessment. In conjunction with ESG Sensor, which we discussed in the previous section, and the information system in general, this AI will be able to uncover data elements that would help, for example, to measure the financial magnitude each risk factor can have on a specific company or industry.

Further AI and analytics modules can be deployed to predict future risks (probability, frequency, and context, etc.) with trends in regulation and current events as inputs.They can also predict the consequences of each potential event as well as the capacity of companies to respond.

We should note here that the **internal standards and approaches** of each investor are vital for risk assessment. By codifying these standards and approaches as the foundation of the ESG Risk Assessment AI, investors benefit from **directly-actionable outputs** backed by logics that are easy to trace. This not only improves the effectiveness

ESG Risk Assessment AI harmonizes hetereogeneous 3rd-party ESG scores and performs meta analysis. It also underpins proprietary ESG risk scoring.

ESG Risk Assessment AI can help investors bolster their risk management while improving transparency and trust.

of investment decision-making but also increases **transparency and trust** to clients.

Moreover, the ESG Risk Assessment AI allows investors to weigh dynamically risk factors not only for risk assessment but also for scenario simulations.

Building on the architecture that we have discussed in the previous sections, investors can begin developing ESG Risk Assessment functionalities by creating a **knowledge database**. Such an ESG knowledge database includes information on portfolios, ESG scores, engagement meeting notes, assumptions, and analysis documents, etc.

Behind the scenes, data is passed to a **graph database** within the information system. Graph databases are purpose-built to store data as nodes, allowing us to capture and navigate relationships between these nodes flexibly. They are particularly well suited for ontology-like knowledge representations[20], where data points can be highly connected with evolving relationships[21].

With this data configuration, a **question-answer AI model** (i.e. logic-based NLP system) can be deployed over the graph database as a means to retrieve any relevant information relating to a specific question in a logical stream. Questions on ESG risks can be submitted either by ESG analysts or portfolio managers in free natural language[22]. Here, the AI uses the input questions to develop an optimal query taking into account the data available, before extracting potential answers from the knowledge base, organized as logical reasoning sequences that are validated through the ontology graph.

In this implementation example, the ontology graph can itself exploit an AI model to dynamically discover and update relationships between data elements, especially with regard to risk factors. This could be a link between a company profile, a news article, and/or a risk factor. This capacity underpins the development of an analytics block aimed at consolidating in-house ESG analytics.

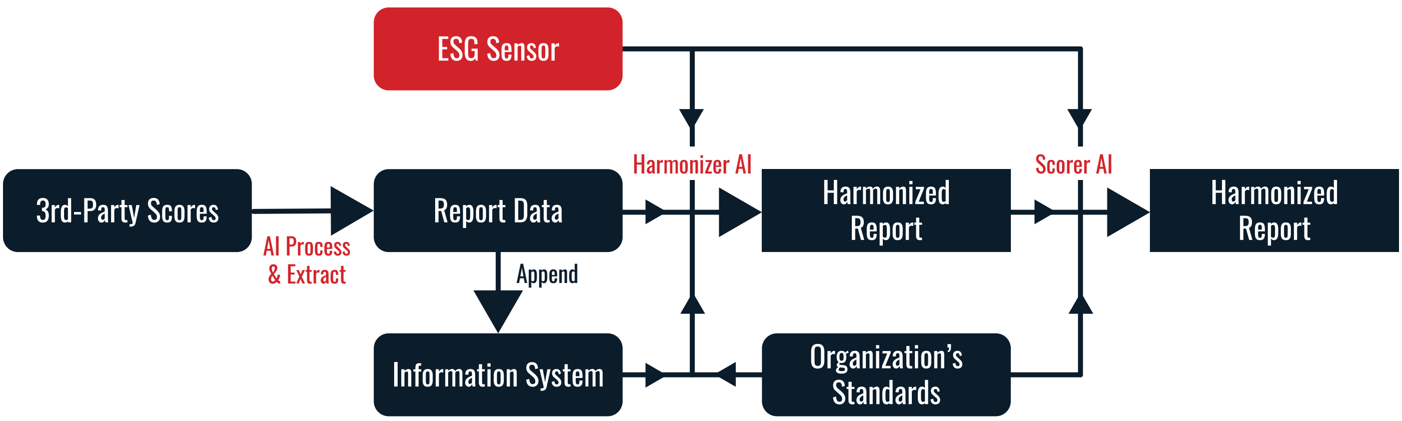

Figure 4. Overview of ESG Risk Assessment

4. ESG Opportunity Radar

Opportunities and risks often come hand in hand. In ESG, investors that can identify opportunities before they materialize reap the most rewards. At the same time, events, macroeconomic data, and company-specific data, etc. impact each investment differently at multiple levels.

Decision makers who wish to be one step ahead of the competition need an information system that can help them ingest, process, and analyze large amounts of data from diverse sources. Moreover, AI is particularly valuable here in **identifying patterns** and **forecasting upcoming opportunities**.

The ESG Opportunity Radar is the most **Big-Data**-driven aspect of AI in ESG in the sense that three levels of data in multiple formats from the broadest set of sources are leveraged simultaneously, each in its own fashion.

At the **Macro** level, investors can leverage a setup similar to the **Dynamic ESG Materiality Appraisal**, which we discuss in detail in another section. The difference is that here we collect data with industry and geography focuses and run them with other macro data through a forecasting AI that is aimed to identify non-obvious patterns that human analysts cannot.

Sources of data for Macro ESG Opportunity Radar include updates in macroeconomic policy and regulations, macroeconomic data, industry reports, and news as well as events with industry and geographic implications. The goal of ESG Opportunity Radar at this level is to help inform the investors' overall investment strategy.

Astute readers will note that, while some of the aforementioned sources are easily scoped, one cannot properly scope news and events without analyzing them. Indeed, this is why Macro ESG Opportunity Radar is optimally deployed in conjunction with other ESG AI components such as the **ESG Risk Assessment** and **Dynamic ESG Materiality Appraisal** to improve the synergy of the information system.

At the **Meso** level, investors can further delve into company groups. These are companies in the same industry with similar geographic presence (e.g. Coca Cola and PepsiCo) that share a similar ESG opportunity profile. This system can be set up similar to the **ESG Sensor**.

Sources of data for Meso ESG Opportunity Radar include the raw data as well as output of Macro ESG Opportunity Radar. It also combines them with company-specific data that range from financials to news and events.

The goal of ESG Opportunity Radar at the meso level is to help investors validate their investment strategy and inform any adjustment of their portfolio. ESG components that have particular synergy with it include **ESG Risk Assessment** and **ESG Sensor**.

At the **Micro** level, investors can implement the most elaborate AI system in ESG. Single companies are continually monitored, and detailed data that can inform the system of the company's R&D, business model, supply chain, marketing, and various other aspects of their operations are taken into account.

ESG Opportunity Radar AI can identify ESG-based patterns and investment opportunities at the industry, company-cluster, and company levels.

ESG Opportunity Radar works in concert with ESG Risk Assessment to bolster investors in ESG

This is where the information ingestion can go beyond news and press releases. Research collaborations, patent applications, shipping data, energy consumption data, personnel changes, and marketing campaigns can all be progressively included in the model.

The goal of Micro ESG Opportunity Radar is to move ahead of the market in entries and exits.

Across the three levels, we encounter perhaps the most diverse collection of data formats. For language-based data, **Natural Language Processing** (NLP) can help us extract relevant information from large amounts of texts in multiple languages. This is discussed in further detail in Dynamic ESG Materiality Appraisal. Structured data such as financials can be normalized and directly incorporated into the model. Data that go beyond texts and numbers need specific processing first.

For this system to inform investment in your organization, its outputs need to be **customized to work with your decision-making process**. In other words, the full potential of the AI can be unleashed when the outputs go beyond a single (or a single set of) score[23].

The outputs of ESG Opportunity Radar, together with that of ESG Risk Assessment, can be fed into your **Fundamental Analysis** to provide a holistic assessment of each company. For those we prefer individual indicators, the system can also be customized to output those that work with your logics.

Moreover, it is always possible to retrace the reasoning behind each output: with the click of a button, the ESG team and asset managers can trace the decision tree back to the sources (with collapsible details and links) so that ESG-related decisions can be truly informed.

Architecturally, the ESG Opportunity Radar is rather similar to that ESG Risk Assessment: just as risk and opportunity in investment, they are two sides of the same coin.

Figure 5. Overview of ESG Opportunity Radar

5. ESG Reporting Assistance

With the growing importance of ESG, most investors—especially small investment firms—must do more in less time. Information systems and tools that enable insights into ESG processes and allow smart automation of repeatable yet evolving tasks, are gaining momentum.

Instead of wasting man-hours, ESG Reporting Assistance AI **automatically** and **efficiently** takes care of tedious jobs with a level of **consistency** that human analysts cannot match. Tasks such as reading through reams of ESG documents and questionnaires, drafting responses, clustering relevant pieces of information together, and scouring petabytes of ESG data to find answers are where AI excels. In contrast, AI-less information systems and tools have difficulty in these qualitative areas, making an effective ESG integration unlikely.

A major challenge for investors in ESG reporting is information flow management. The sheer quantity of ESG data points and documents that could and should be considered compounded by their lack of uniformity—which is also the case for disclosure standards and auditory inputs—make it difficult to organise, manage, and synchronize between ESG and compliance teams. AI, and specifically Natural Language Processing (NLP) technology, has an exceptional ability to **extract information from unstructured data** and can be trained to recognise ESG concepts and/or classify information with high degree of precision. With a well-structured information system in place, this allows investors to **speed up the ESG reporting processes** and thus **contain costs** and **reduce compliance risk**.

Moreover, the entire history of ESG reporting data can be tracked and analysed. From these, ESG Reporting Assistance AI can provide crucial insights for **future** reporting endeavours, allowing reporting teams to work in a fast, smart, and **evidence-based** environment with advanced analytics support. For example, it can infer potential compliance inconsistencies between reports and specific responses and forecast new changes in ESG questionnaires based on previous evolutions and their underlying data. AI can identify patterns in regulatory reporting requirements and help reporting teams anticipate needs and prepare the information beforehand.

For example, **SFDR** is a regulation framework introduced by the EU to boost transparency and accountability in investment management and across capital markets, mainly by setting new and strict reporting rules for sustainability risk disclosure and management[24]. These reporting requirements apply in particular for investment funds.

Under this framework, investors are required to disclose a set of 18 mandatory and pre-set **Principle Adverse Impact** (PAI) ESG indicators as part of the SFDR regulation, as well as several self-chosen ones from a list of 46. Such a disclosure requirement poses many challenges, especially for the PAI indicators for which companies' reported data is unclear, fragmented, or non-standardized. Thus, it is important for ESG teams to identify existing data gaps, efficiently bridge their available information, and sometimes source other datasets to

ESG Reporting Assistance AI can speed up ESG reporting, contain compliance costs, and reduce compliance risk.

build reliable data.

Further, the regulation requires investors to report their data usage processes, including sources, any measurements being taken to ensure the best possible coverage, data alignment, and metric calculations. For example, when switching between providers to output a compound metric, investors need to explain how field names, descriptions and indicators are mapped to each other.

ESG Reporting Assistance AI improves reporting quality and transparency.

Armed with ESG Reporting Assistance AI, ESG teams can better ensure only **high quality, well aligned, and explainable data** streams feed into SFDR analysis and reporting activities.

In deploying information systems enhanced by ESG Reporting Assistance AI, investors can connect external APIs where appropriate and consolidate data into their own databases. This allows for faster response time in data access; more importantly, investors can identify, fill the gaps, and enrich the data with that from other sources. Such an approach also enables using NLP capabilities of the AI more readily to process and analyze the data and perform the aforementioned tasks. It echoes the **progressive approach to deployment** that we advocate.

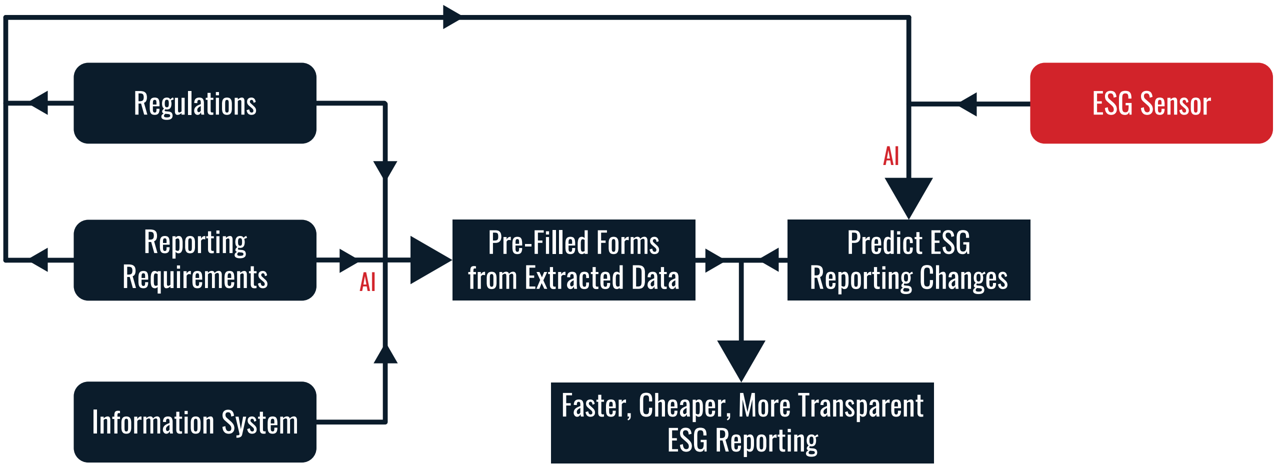

Figure 6. Overview of ESG Reporting Assistance

Vogosen's Approach to Implement AI-Powered ESG Information Systems

At this point, you may wonder how you can get started. Indeed, you may question how an information system with so many new features and applications can be built and, justifiably, doubt whether it can integrate into your investment process. We agree. This is why Vogosen builds AI-powered ESG information systems **progressively**, with **the investors' inputs and choices** taken into account at each step in accordance with our 5C framework.

Vogosen's 5C framework starts from Categorizing data sources to Centralizing data access to Clarifying analytical outputs to Creating analytics workflows to Conveying the output. It is developed under the belief that investors must have control over how their ESG information is collected, processed, analyzed, and presented in order for the information system to be directly actionable for decision-making.

Progressive Deployment

To deploy a new AI-powered ESG information system that can grow with the organization, we follow six clear steps in an Agile manner so that each step not only prepares for the next step(s) but also foresees future needs[25,26].

First, the investor defines the **Objective** of the system with us. Initially, the objective may simply be to track material issues of the portfolio more clearly or to monitor a large set of information sources in a timely fashion; over time, new objectives can be added. It is important to define the outputs clearly here: these **outputs** should be meaningful for your decision-making and/or contain actionable signals[27].

Second, we identify with the investor **Inputs** to the system. Some inputs may be readily available, though to interpret them appropriately depends on the objectives of the information system as well as other inputs that must be taken into consideration. Other inputs may not be readily available, and we need to find out how to collect them. Part of these may require subsystems, for example, outputs from other AI modules.

Third, together we iron out the **Preprocessing** of the inputs so that they are primed for producing the outputs in a clear, unbiased fashion. Indeed, here is also when the investor determines with us whether to adopt machine learning.

Fourth, we work out a solid **Data Structure** and a set of **Algorithms** with the investor. We assess how the investor currently manages investment data—that is, data beyond ESG. If the investor has yet to deploy any modern information system and still rely on spreadsheets, we first propose an easy-to-use, cheap-to-implement database system with secure GUI (like spreadsheets but much more powerful) that protects data security, ensures data consistency and integrity, and facilitates data sharing across teams in the organization. Beyond conventional relational (SQL) databases, we use other data storages that are optimized for specific tasks, e.g. Graph DB for semantic queries and ontology.

Fifth, the investor chooses with us the desired **User Interface** elements, including features of the application such as

Vogosen works closely with investors to design, develop, and deploy AI-Powered ESG Information Systems progressively.

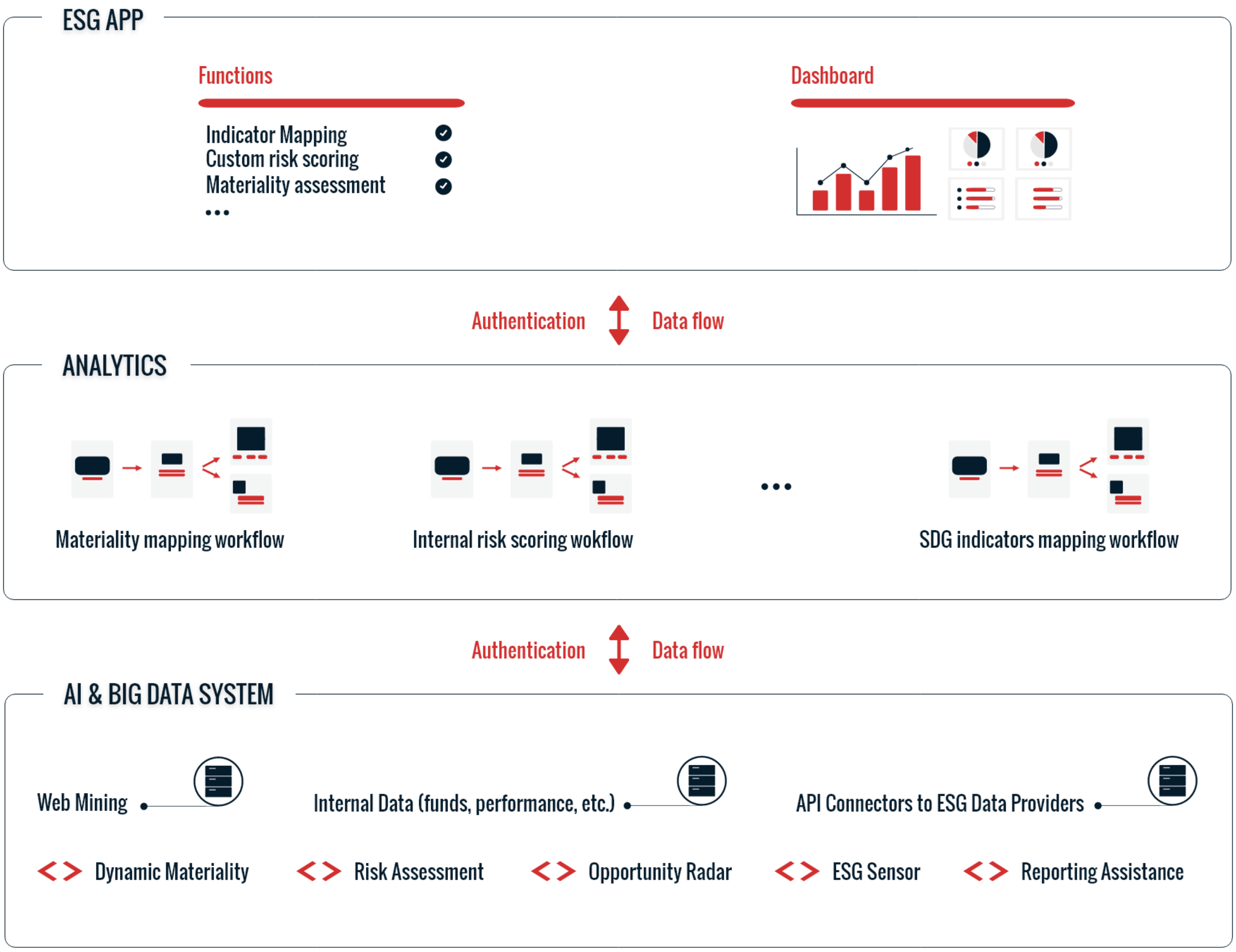

Figure 7. Overview of AI-Powered ESG Information System

visualization, that make sense to them.

Sixth, we define a **Governance and Maintenance** structure with the investor so that the information system is resilient, secure, and efficient. This involves indicators and warning systems that signal issues before they undermine the entire system.

Throughout the process, Agile is used with continual testing. Users receive training on each component of the information system, while their feedback informs continual improvement and adaptation.

Generally, we recommend following the sequence of Dynamic ESG Materiality Appraisal, ESG Sensor, and ultimately ESG Risk Assessment & ESG Opportunity Radar in expanding the objectives of the information system. Within the more advanced AI modules, namely ESG Risk Assessment and ESG Opportunity Radar, we recommend that Macro-level AI be deployed first, followed by Meso- and, eventually, Micro-level AIs.

We recommend progressive deployment of AI-powered ESG information systems not only because these modules have logical dependency on each other but, more importantly, because it is critical to grow relevant organizational capabilities and broaden organizational vision in-house in order to take full advantage of such a new technology[28]. In plain words, AI in ESG is a shiny new tool, and investors need to learn how to use it and where (beyond which they are currently doing) can it be used.

Web Mining Architecture

Efficiently aggregating a large number of data points starts with the implementation of a **scalable** web-mining system able to capture—in real time—all ESG related information publicly available on the Internet.

Such a system must operate within a Big Data architecture that is adapted to the size and dynamics of data as well as the targeted web sources. Its architect needs to take into account **security and data integrity**, in particular when combining cloud infrastructure and local data centres. This process involves guaranteeing real-time mining when needed, ensuring the optimization of the database structure, adapting web-crawlers to site structures (targeted and exploratory crawling), incorporating metadata, and pre-processing functions. The outputs can then feed the analytical systems in place (specialized analytics for ESG research, ESG internal apps, etc.). We can also use these feed data to (re)train the AI modules.

NLP-Driven Systems in ESG

ESG has its own language: The environment aspect covers topics such as biodiversity and climate change, which relate to **domain-specific** knowledge, concepts, indicators, and metrics. The same is true for the social and governance aspects of ESG.

This language is not only critical for investment stakeholders to evaluate and/or comply with and/or compare ESG practices but also influences how NLP models identify items that data systems track and analyse—**concept identification & extraction**. How things relate to one another—**clustering and classification**—makes its way into programs and data designs, influences AI training strategies and outputs, and ultimately dictates whether the data exploited is actually of value to the organization.

From straightforward knowledge man-

We ensure a scalable architecture that guarantees security and data security. This is the foundation of a powerful and versatile AI system in ESG.

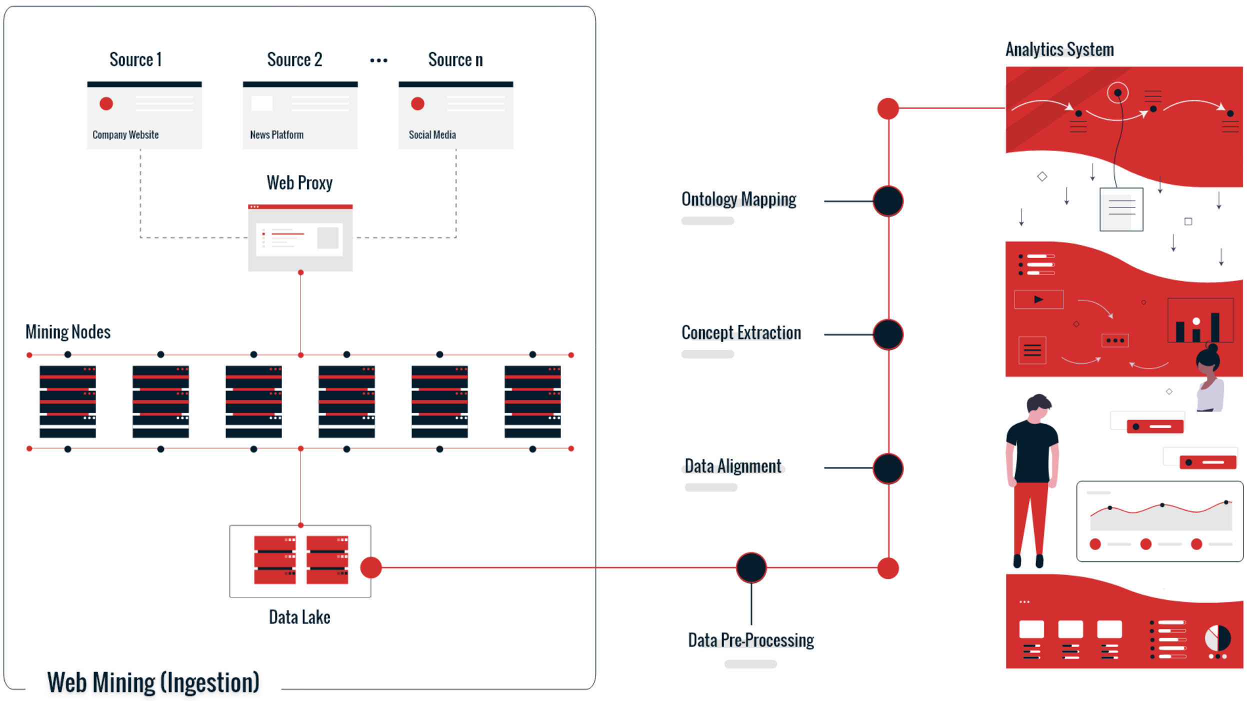

Figure 8. Web Mining Architecture

AI in ESG largely leverages NLP. In building this, we recommend harnessing open-source technologies and limit dependence on third-party services.

agement to sophisticated AI models, ontologies have proved great potential in capturing expertise while being particularly apposite to today's data abundance and digital transformation. An **ontology** is generally intended to act as a standard—sort of common language—forming a set of controlled vocabularies and concepts. Overall, these ontologies bring opportunity to ESG teams to have control on algorithms and data they use, paving the way to create resilient ESG planning models, increase productivity, and deliver effective guidance to investment strategies. Investors can thus embrace a data-driven approach, expanding the boundaries and reshaping traditional patterns.

We use **dynamic ontologies** to model evolving ESG concepts and taxonomies[29], which are aimed to facilitate the alignment and exchanging of ESG data, such as ESG material aspects, SDG indicators[30], multi-provider risk scores, etc., and the development of powerful ESG information systems. Classification techniques ranging from complex language models to simple sentiment analysis can be then used on top of these ontology-based models to enrich the data and the analytics process.

Sentiment analysis is one of the common NLP topics that investors can start implementing easily, especially because it has low complexity, requires manageable data and fits with a large number of ESG applications[31]. For example, sentiment analysis can be integrated to the risk and opportunity analytics as a way to help screen early signals. It can also support portfolio monitoring, engagement strategies and marketing initiatives.

From a technical perspective, sentiment analysis falls into the broad category of supervised learning text classification, where the model inputs a sentence and outputs a score for each sentiment class (the number of classes can vary, but it is common to use two to three simple classes: positive, neutral and negative). One popular and simple technique for developing sentiment analysis models

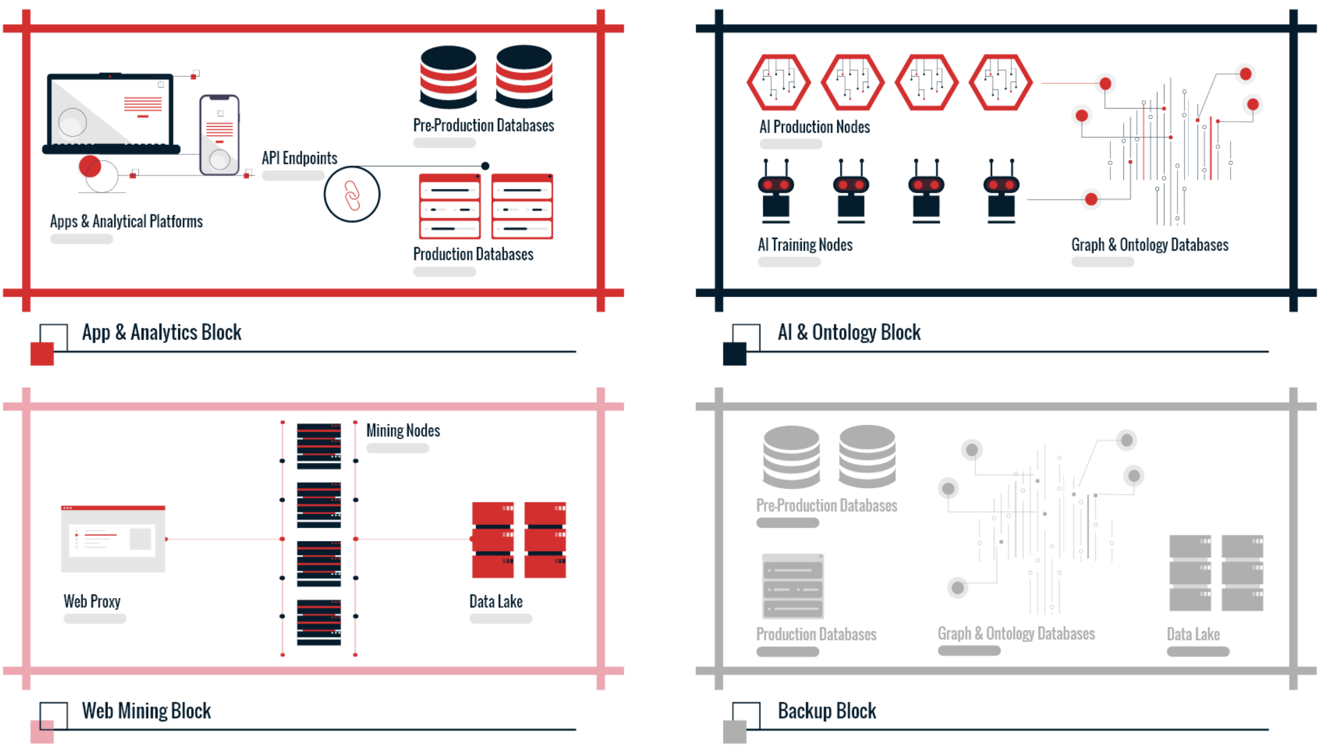

Figure 9. Architecture of an AI-Powered ESG Information System

is to use a bag-of-words representation that transforms sentences into vectors of weighted words.

Technology stacks for implementation can vary and evolve based on organizational capabilities, existing practices and technical dependencies. We recommend however, whenever it is possible, to **harness open-source technologies** (for data architectures, app deployments, etc.) so as to **limit third party dependencies** (cloud, SaaS, etc.) as well as to ensure a total oversight on one's data and technology skills, interoperability, and operating costs.

APIs, Apps, and User Interfaces

The AI-powered ESG information system can be deployed as pure APIs upon which investor's in-house development team build their proprietary interfaces. Alternatively, this information system can come with application interfaces built to the specifications of the investor on an Agile basis.

In either case, the user interface should be designed and developed progressively, in the same spirit of the progressive deployment of the entire ESG Information System. That is, while in the beginning a more structured user interface is preferred, as the system evolves it will be possible to enrich the user interface with AI, too. For example, investment managers will be able to ask questions to the information system directly, and the AI will identify the appropriate module to seek the corresponding answer[32,33].

The ESG Information System can also be integrated with software and information systems that other parts of the organization use.

.

Index

References

1 OECD (2021), *Artificial Intelligence, Machine Learning and Big Data in Finance: Opportunities, Challenges, and Implications for Policy Makers*, https://www.oecd.org/finance/artificial-intelligence-machine-learning-big-data-in-finance.htm.

2 M. Chui, and B. McCarthy (2020), "An Executive's Guide to AI", McKinsey & Company.

3 C. Stier, D. Dalton, and K. Ehrsam (2019), "Artificial Intelligence: The Next Frontier for Investment Management Firms", Deloitte.

4 M. Hurley, and J. Adebayo (2016), "Credit Scoring in the Era of Big Data", *18 Yale Journal of Law & Technology*.

5 M. Moscatelli, et al. (2019), "Corporate Default Forecasting with Machine Learning", Bank of Italy, Economic Research and International Relations Area.

6 O. Fliche, and S. Yang (2018), "Artificial Intelligence: Challenges for the Financial Sector", ACPR, Banque de France.

7 Sustainability Accounting Standards Board, *SASB Standards*, Accessed September 2021, https://www.sasb.org/standards.

8 United Nations Sustainable Stock Exchanges Initiative (2021), *ESG Guidance Database*, https://sseinitiative.org/esg-guidance-database.

9 S. Sung, et al. (2008), "Efficient Concept Clustering for Ontology Learning using an Event Life Cycle on the Web", *Proc. 2008 ACM SYmposium on Applied Computing*, ACM, Fortaleza, Brazil.

10 G. Bala et al. (2015), "Tracking Companies' Real Time Sustainability Trends: Cognitive Computing's Identification of Short-Term Materiality Indicators", Available at SSRN: https://ssrn.com/abstract=2570616.

11 R. Collobert, et al. (2011), "Natural Language Processing (Almost) from Scratch", *Journal of Machine Learning Research 12*, pp. 2493-2537.

12 Z. Wu, and D. C. Ong (2021), "Context-Guided BERT for Targeted Aspect-Based Sentiment Analysis", *arXiv*, https://arxiv.org/abs/2010.07523v2.

13 J. Xu (2019), "How NLP is Advancing Asset Management", *Synced*, https://syncedreview.com/2019/07/13/how-nlp-is-advancing-asset-management.

14 A. Holzinger (2018), "From Machine Learning to Explainable AI", *Keynote Talk at IEEE DISA 2018 Conference*, Kosice, August 23, 2018.

15 S. He, et al. (2014), "CASIA@V2: A MLN-based Question Answering System over Linked Data", In: Cappellato, L., Ferro, N., Halvey, M., and Kraaij, W., editors. *CLEF 2014 Labs and Workshops, Notebook Papers*. CEUR Workshop Proceedings.

16 F. Berg, J. Kölbel, and R. Rigobon (2020), "Aggregate Confusion: The Divergence of ESG Ratings", Available at SSRN: https://ssrn.com/abstract=3438533.

17 IOSCO Board (2021), *Environmental, Social and Governance (ESG) Ratings and Data Products Providers (July 2021)*, https://www.iosco.org/library/pubdocs/pdf/IOSCOPD681.pdf

18 R. Boffo, and R. Patalano (2020), "ESG Investing: Practices, Progress and Challenges", OECD Paris, https://www.oecd.org/finance/ESG-Investing-Practices-Progress-and-Challenges.pdf.

19 T. Jaulin, et al. (2021), "Artificial Intelligence Solutions to Support Environmental, Social, and Governance Integration in Emerging Markets", Amundi Asset Management and International Finance Corporation.

20 W. L. Chang, D. Boyd, O. Levin (2019), "NIST Big Data Interoperability Framework: Volume 6, Reference Architecture", National Institute of Standards and Technology, U.S. Department of Commerce.

21 R. Gayathri, and V. Uma (2018), "Ontology Based Knowledge Representation Technique, Domain Modeling Languages and Planners for Robotic Path Planning: A Survey", *ICT Express, vol. 4*, issue 2.

22 Z. Yang, et al. (2018), "HotpotQA: A Dataset for Diverse, Explainable Multi-hop Question Answering", *arXiv*, https://arxiv.org/abs/1809.09600.

23 A. Hughes, M. A. Urban, and D. Wójcik (2021), "Alternative ESG Ratings: How Technological Innovation Is Reshaping Sustainable Investment", *Sustainability*, vol. 13(6).

24 EU (2019), *Regulation (EU) 2019/2088 of the European Parliament and of the Council of 27 November 2019 on Sustainability-Related Disclosures in the Financial Services Sector*, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02019R2088-20200712.

25 J. Zhang, and E. Bareinboim (2020), "Can Humans Be out of the Loop?", CausalAI, https://causalai.net/r64.pdf.

26 L. Dupont, O. Fliche, and S. Yang (2020), "Governance of Artificial Intelligence in Finance", ACPR, Banque de France.

27 Barclay Hedge (2018), *BarclayHedge Survey: Majority of Hedge Fund Pros Use AI/Machine Learning in Investment Strategies*, https://www.barclayhedge.com/insider/barclayhedge-survey-majority-of-hedge-fund-pros-use-ai-machine-learning-in-investment-strategies.

28 IOSCO Board (2020), *The Use of Artificial Intelligence and Machine Learning by Market Intermediaries and Asset Managers (June 2020)*, https://www.iosco.org/library/pubdocs/pdf/IOSCOPD658.pdf.

29 D. M. Blei, A. Y. Ng, and M. I. Jordan (2003), "Latent Dirichlet Allocation", *Journal of Machine Learning Research 3*, pp. 993-1022.

30 United Nations Department of Economic and Social Affairs, *Sustainable Development Goals*, https://sdgs.un.org/goals.

31 A. Radford, R. Jozefowicz, and I. Sutskever (2017), "Learning to Generate Reviews and Discovering Sentiment", arXiv, https://arxiv.org/abs/1704.01444.

32 C. Alberti, K. Lee, and M. Collins (2019), "A BERT Baseline for the Natural Questions", *arXiv*, https://arxiv.org/abs/1901.08634.

33 E. Choi et al. (2018), "QuAC: Question Answering in Context", *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Brussels, October-November 2018.

The Authors

Zakaryae Boudi

As an entrepreneur, Zakaryae Boudi developed an in-depth expertise in multiple areas including new technologies, public policies, and international development.

Before launching one of the leading labor market intelligence startups in France, Zakaryae had the opportunity to work in international collaborative teams on diverse economic and social challenges. In particular, he led the development of the global PPP and financing centre at the International Union of Railways as part of the Sustainability Unit.

Zakaryae also teaches several courses on Big Data and analytics as part of the professoral team at Aivancity, a new technology & business school. He holds an engineering degree from the Ecole Mohammadia d'Ingénieurs (Morocco) as well as a Master's degree from Sciences Po Paris (France). Recently, he completed a PhD thesis on AI systems' control using mathematical formal methods.

Jiulin Teng

Jiulin Teng has deep understanding of key topics in management science and new technologies. He also has broad knowledge in macroeconomics, business policy, and software engineering.

Upon completing his dissertation on the application of game theory in competitive strategy, Jiulin Teng worked as a consultant at McKinsey & Company (Belgium). More recently, he has launched and participated in a number of startups in blockchain, IT, and AI. He is a proficient database architect and is a full-stack developer with good command of modern programming languages. He has also taken an interest in Natural Language Processing.

Jiulin Teng holds a PhD in Strategic Management and Business Policy from HEC Paris (France). He also has a Master's degree in Bioentrepreneurship from Karolinska Institute (Sweden) and a Bachelor's degree in Biological Sciences from Nankai University (China).

VOGOSEN
2021

Web: vogosen.com
Email: inquiry@vogosen.com