

Smart Sustainability: How Artificial Intelligence helps to enrich Sustainable Investing

Valentijn van Nieuwenhuijzen

In the dynamic domain of investment management, a discernible paradigm shift is underway, compelling practitioners to reassess traditional approaches within an evolving landscape. Central to this evolution are dynamics that extend beyond customary metrics, prompting investment managers to synthesize considerations at the intersection of technological advancement and ethical responsibility. Against the backdrop of market adaptability to transformative forces, astute managers find themselves positioned at a juncture where exploration of emerging technologies converges with a conscientious appraisal of broader societal and environment imperatives.

Yes indeed, this machine-generated introduction (OpenAI, 2023) was prompted with instructions to generate a text that leaves the reader guessing if the article is about artificial intelligence or sustainable investing.

A DISCERNIBLE PARADIGM SHIFT IS UNDERWAY

Both artificial intelligence (AI) and sustainable investing (SI) are top of mind for investors. Breakthrough technological advances in the field of Natural Language Processing (NLP), like the great success of ChatGPT and other large language models, have heightened interest of investors in AI this year. For a bit longer already the growing demand for sustainable investments has caught investors' attention with equal force. In this article we explore what is happening at the intersection of AI and SI when it comes to the work of investment professionals. Taking inspiration from recent publications on environmental, social and governance (ESG) topics will hopefully trigger further thinking on the opportunities that novel AI offers to investment professionals and asset owners.

NATURAL LANGUAGE PROCESSING

The concept of AI spans a broad universe of technologies and has a rich history that includes periods of widely held enthusiasm as well as the proverbial AI winter. Despite AI's depth and breadth, many sustainable investing use-cases specifically leverage NLP techniques. This is typical for the field of sustainable investing and is at least partially explained by the specific data challenges that this rapidly evolving domain is facing. Where

investors typically leverage tabular datasets in their research, they find themselves without those for many concepts when it comes to sustainability themes. Even when consensus grows on how to measure a specific sustainability theme and self (or regulatory) reporting starts (for example using tonnes of CO2 emissions to express climate impact), the next topics of interest present themselves already. The rapid evolution and diversity of sustainability topics make it fertile ground for investment researchers that can navigate alternative datasets using NLP.

Valentijn van Nieuwenhuijzen

Co-Head of Multi Asset Solutions (MAS) and Global Head for Sustainability for Public Investing at Goldman Sachs



The value of NLP as a technique to capitalize on alternative datasets is illustrated in recent research on sustainability. Consider Sautner et al. (2023) as a first of three examples handpicked from the 2023 vintage of papers that span both the E and S space of ESG. Sautner et al. (2023) create a novel firm-level climate change exposure measurement from earnings calls and investigate its relationship to real world impact (green hiring, green patent generation) and financial market outcomes (risk, risk-premia). To this end metrics needed to be created from things like earnings call transcripts, job postings and patent filings. Often textual sources of information lack the identifiers that practitioners typically use to navigate financial datasets. NLP can play a vital role in matching document to company identifiers. In this case attributing patents to the right organisation by connecting those organisation names to the names on patent filings is essential, but applications go beyond that, just think about social media or news articles in which companies' names are mentioned. Indicators like this can be helpful across the investment management value chain. They can aid investors in idea generation, portfolio construction and risk management.

A second illustration is on measuring biodiversity. With attention for biodiversity on the rise, it is important for investors to explore how biodiversity exposure can be quantified. Related to that: any relationship between exposures and financial returns is important to learn about. Giglio et al. (2023) demonstrate how biodiversity risks can be quantified. To this end, they first construct a news-based measure of biodiversity attention using newspaper articles and Google search activity. Thereafter, amongst others, they construct a firm specific measure of biodiversity attention based on corporate disclosures (financial reports). Studying the relationship between company and sector level biodiversity attention and the news-based biodiversity attention, they find that equity markets are already pricing biodiversity risks. Next to using classical NLP methods like keyword-based search, they apply attention-based transformer model to determine sentiment. Sentiment classification is a critical technique for differentiating between biodiversity risk and opportunity. Thanks to the recent shift to attention-based models that incorporate context, sentiment classification can be performed with improved precision. They no longer must rely on a 'bag of words' and run the risk of missing valuable nuancing in preceding or following words. Quantifying biodiversity (risk) is high on many investors' agenda; hence some level of practical ingenuity is required. NLP is likely to play an important role here.

Outside the environmental domain of climate and biodiversity, sustainable finance professionals are potentially even more challenged by the lack of availability of data in the social domain. Lohre et al. (2023) set out to measure social controversies and their impact on stock returns. Against a backdrop of earlier research that found negative ESG news to be detrimental to financial returns, they zoom in on social controversies. However, until now only a few datasets for social controversies exist, and for the widely known ones there is a high level of

disagreement (low overlap) amongst their data vendors. To solve for this the researchers trained a bespoke classification model: ControversyBERT. When constructing their model that classifies news articles about a firm as containing a social controversy or not, they utilize an attention-based model to bring the context awareness discussed in the example above. With classical NLP approaches a sentence of the form "Company XYZ launches policy to tackle gender inequality" would be classified as a social controversy due to word matching on 'gender inequality'; however, with contextual awareness added this is unlikely to be flagged as a controversy (Lohre et al., 2023, p3.). Both deepening investors' understanding of the relation between social controversies and financial returns, as well as improved accuracy of classifying sentiment are important to the development of sustainable investing. The former as it answers one of the most fundamental questions for investors, the later as improved accuracy increases the level of comfort that practitioners can have with the model output.

INVESTMENT RESEARCH AND BEYOND

By now the usage of AI already has a decent history in investment management, an industry that is hyper competitive and focused on finding new ways to outperform the market or better tailor to evolving client needs. In contrast to the long history of analysing classical financial datasets that include prices, volumes and the fundamental data originating from accounting statements, the usage of textual analyses is a nascent phenomenon. The adoption of natural language processing techniques in their tool kits gives investment professionals an ocean of new opportunities for finding unique insights that can complement classical financial analyses at a scale. Suddenly information that was not systematically available in traditional datasets can be harvested at scale and be the ingredients for an additional layer of investment analyses.

USAGE OF AI IS INCREASINGLY EMBRACED IN THE INDUSTRY

For analyses of the sustainability profile of a firm, it is increasingly important to also consider forward looking elements like intentions, commitments, targets, etc. Historically, this would take researchers a substantial amount of time as they first need to find the information and then interpret it. AI has now reached the level of sophistication at which it can support researchers in both stages. It can reduce the time spend on search, thereby freeing up capacity for the more creative and expertise-based activity of interpretation of information. In this interpretation phase the scalability of AI can help researchers again, this time by enhancing their analyses via increased objectivity and completeness. Information extracted with help of NLP can for example be shown against a larger, more complete peer group, mitigating typical human biases in the process. When it comes to news or other local content, machine translation has now

reached the level of maturity that local language articles can be incorporated into the analyses. As a result, more timely and complete insights on developments concerning investments internationally can be considered.

LOOKING AHEAD

NLP unlocks alternative datasets that contain a magnitude of relevant sustainability information for investors. It enables refined analyses via contextual awareness that can improve NLP model accuracy. It facilitates the joining of these new insights to classical finance datasets so that both real world and financial impact can be researched.

NLP UNLOCKS ALTERNATIVE DATASETS

So, does this imply all is perfect? Obviously not. Working with new tools and datasets also introduces new risks that need to be managed carefully. For tools, model suitability is an important consideration: the limited transparency of the popular context aware models might not be blocking for a researcher that is using it to speed up information gathering. However, for feeding directly into trade-signals a different degree of explainability is often required. Practical ways of navigating explainability are evolving. For example, in today's world of Large Language Models (LLMs) their powerful generative capabilities (they can 'generate' new text), summarization is a popular application. Summarizing an annual report of an investee firm on its sustainability information can be a timesaving exercise. However, the tolerance for model hallucination will be low if such summaries will be used for direct decision making. To not fully scope out summarization from investors' toolkits,

a differentiation can be made between extractive and abstractive summarization methods. Extractive NLP preserves the original text and therefore has a high degree of explainability, a summary created with extractive NLP will act as a highlighter; each bit of information is traceable to its source and no new text is created. Abstractive NLP will be more eloquently written, but at the risk of hallucination text, hence not the most suitable for direct decision making.

When it comes to datasets there are several biases at play. Reporting biases, that might make available sustainability data not representative for the overall population of investable securities, as well as size and geographical biases that might skew portfolio construction unintendedly, just to name a few. Although considerations around model suitability and dataset biases are not new to finance professionals, domain expertise in both sustainable investing as well as data science will prove critically important to avoid the pitfalls on this journey and enable investors to capitalize on opportunities along the way. So more than enough focus needed on safeguarding prudent human oversight and effective quality control, but with these aspects in play a much smarter way of navigating sustainable investing opportunities and risks can be identified.

References

- OpenAI. (2023). ChatGPT (November 2023 version) [GPT-3.5]. <https://chat.openai.com/chat>
- Giglio, S., Kuchler, T., Stroebel, J., & Zeng, X. (2023). Biodiversity Risk. National Bureau of Economic Research. <https://doi.org/10.3386/w31137>
- Lohre, H., Nolte, S., Ranganathan, A., Rother, C., & Steiner, M. (2023). ControversyBERT: Detecting Social Controversies and Their Impact on Stock Returns. In *The Journal of Impact and ESG Investing* (Vol. 4, Issue 1, pp. 87–108). With Intelligence LLC. <https://doi.org/10.3905/jesg.2023.1.076>
- SAUTNER, Z., VAN LENT, L., VILKOV, G., & ZHANG, R. (2023). Firm-Level Climate Change Exposure. In *The Journal of Finance* (Vol. 78, Issue 3, pp. 1449–1498). Wiley. <https://doi.org/10.1111/jofi.13219>