ESG & Impact Investing Text Classification and Quantification with Neural Networks

Viktor Aghajanyan (vaghaja1@jhu.edu)

Abstract—ESG and Impact Investing text classification and quantification are complicated tasks due to domain-specific language and diversity of issues. The effectiveness of general language models in predicting the topics of sustainability issues and quantifying such texts is low. We propose an end-to-end pipeline of models trained for ESG/Impact Investing tasks to convert unstructured textual data into company-level intelligence scores across ESG/Impact issues using the latest advances in deep learning. In particular, neural language models have been built to:

- Convert long ESG/Impact texts into few-sentence-summaries
- Classify and filter ESG/Impact-related texts from general news flow
- Detect ESG/Impact Investing topics across SASB and SDG framework categories
- Construct a standardized & scalable quantification model to quantify ESG/Impact texts

The final aggregated pipeline of models allows us to extract any material publicly available data about a fund or company, summarize long articles into few-sentence-summaries, filter out articles unrelated to ESG and Impact, automatically identify the sustainability categories of such articles across SASB and SDG frameworks, and quantify texts across these categories to generate exponentially-weighted company or fund-level intelligence scores. This work largely addresses the need for this technology in Private Equity and Asset Management as later described in the Introduction section. It was our intention to make this research paper as simple as possible, given the complexity of neural language models applied, to reach a wider audience in the PE space.

Index Terms—Sustainable Accounting Standards Board (SASB), Sustainable Development Goals (SDGs), Robustly Optimized BERT Pretraining Approach (RoBERTa), Bidirectional Encoder Representations from Transformers (BERT), Text-To-Text Transfer Transformer (T5)

I. INTRODUCTION

What is ESG Investing? According to Financial Times [1], ESG is a term commonly used by investors to assess corporate behavior across ESG topics and to determine the future financial performance of companies. The same source defines ESG Investing as investing which prioritizes optimal ESG factors. This is not to be confused with Impact Investing. According to the Global Impact Investing Network (GIIN) [2], impact investments are investments made with the intention to generate positive, measurable social and environmental impact alongside a financial return. However, often Impact Investing is just whatever you want it to be. As long as you control the narrative, money will flow to your fund. The reason is that there is no precise definition of what constitutes ESG/impact and how to measure it. But according to WSJ [3], continued interest in ESG issues has driven assets in sustainable funds to $2 trillion globally at the end of Q1 2021.

What is needed is a reliable and scalable technology that would track companies and funds ESG/Impact records across a trusted ESG/Impact framework real-time and help investor achieve their investment goals. This will largely address the narrative-versus-action gap issue.

THE MARKET NEED ADDRESSED:

How to measure the return on investment tied to a commitment to i.e., gender diversity?

Since the concept of ESG and its incorporation into business processes are relatively novel topics, quantitative data about companies' and funds' ESG and Impact performances lack, and the quality of existing data is low. For investors who are looking to make sustainable investments, the understanding of ESG impacts, risks, and opportunities is fairly complex due to lack of available disclosed historical data. Currently, investors of public companies largely rely on annual company-published sustainability reports and corporate social responsibility reports for such information. A few problems with only relying on such reports are that 1) the data is reported annually, 2) the data remains mostly unaudited and unreliable, and 3) companies tend to leave certain things out due to lack of standardized reporting framework.

Further to this, the opacity of ESG reporting by Private Equity firms contrasts with the increasing transparency provided by many public companies [4]. Investments in a specific e-commerce site or in machinery to improve productivity are easier to track and quantify, but an investment in ESG or impact is often hard. For example, how might an investor or GP actually measure the return on investment tied to a commitment to diversity? The business need for standardization and quantification for ESG is obvious.

Commensurate amount of unstructured textual data about companies' ESG practices are being created each day. We leverage data from over 100,000 sources, relying on only on trusted news sources to extract data to train our models. Data from third-party sources hinges not on company intentions but societal outcomes of company's or fund's actions.

The final data output would enable investors and analysts to review time series data on specific ESG/Impact issues such as GHG emissions, waste management, or employee diversity to assess progress or lack of progress on such measures.

II. RELATED LITERATURE

This part of the paper discusses previous researches conducted in the ESG/Impact text classification. ESG investing has come a long way in the past few years [3]. It is important to note that all existing papers, to the best of my knowledge, are solely focused on ESG issues and ESG frameworks of public companies and not private companies. It was, however, our intention to adapt our technology for both private and public company needs.

Nugent and Leidnera (2020) [5] from Refinitiv Labs attempted to train a deep learning model on a large financial news archive to detect controversies reported in financial and business news. They designed a multi-class ESG dataset provided by Refinitiv, consisting of 31,605 news articles each annotated across 20 ESG controversy categories with a significant class imbalance in the dataset. Their best language model achieved 84% F1-score on their data. However, if we stay away from general scores and investigate the model test results more closely, on specific topics such as Public

Health, the accuracy is around 50%. This study also was not concerned about sentiment classification or regression but deals only with detection of ESG controversies.

Guo et al. (2020) [6] attempted to measure the predictive power of ESG related financial news on stock volatility. For that they developed a pipeline of ESG news extraction, news representations, and Bayesian inference of deep learn- ing models. Their findings in different geographies confirm that ESG news flow integration may contribute to building profitable investment strategies beyond tradition methods. This research was not concerned about the ESG topic detection but only building better risk-adjusted investment strategies by integrating ESG news flow. Their models' performance showed almost 3 times smaller one-week forward volatility predicting rooted mean squared error than traditional volatility predictions.

Sokolov et al. (2021) [7] from RiskLab based in Canada demonstrated the feasibility of identifying ESG risks using social media data applying state-of-the art natural language processing (NLP) techniques. Our final training dataset con- tains 6,000 unique tweets, with 1,468 tweets covering 10 general ESG topics. This is the closest research to our research, but results are based on training and evaluating the model using only Twitter data. Their model across an example class has a precision of 60% at a 60% recall and it would need to be improved further to create a production-ready ESG scoring solution.

While there are other relevant works conducted on this subject, these three researches are the most relevant and recent to what we are trying to achieve.

III. METHODS

A. Bidirectional Encoder Representations from Transformers

BERT is a neural language model capable of learning representations from large volumes of text [8]. BERT is a state- of-the-art machine learning model used for classification and regression tasks. Jacob Devlin and his colleagues developed BERT at Google in 2018. BERT has been trained on English Wikipedia (2,500M words) and BooksCorpus (800M words) and achieved the best accuracies for many NLP tasks. As opposed to directional models, which read the text input sequentially (left-to-right or right-to-left), the BERT Transformer encoder reads the entire sequence of words at once. Therefore, it is considered bidirectional, though it would be more accurate to say that it's non-directional. This characteristic allows the model to learn the context of a word based on all of its surroundings (left and right of the word).

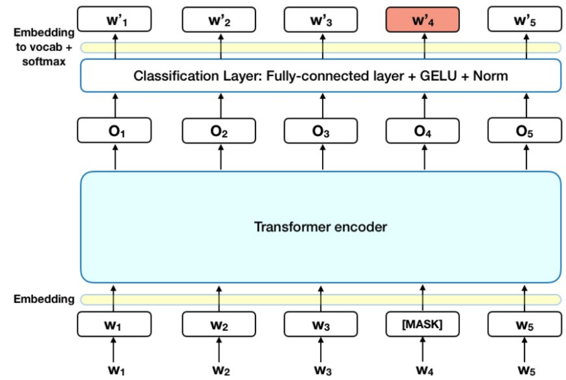
There are two general BERT variations: The base model is a 12-layer, 768-hidden, 12-heads, 110M parameter neural network architecture, whereas the large model is a 24-layer, 1024-hidden, 16-heads, 340M parameter neural network architecture. In technical terms, the prediction of the output words requires:

- Adding a classification layer on top of the encoder output.
- Multiplying the output vectors by the embedding matrix, transforming them into the vocabulary dimension.
- Calculating the probability of each word in the vocabulary with softmax.

Overall, the main ideas under BERT can be summarized as:

- Bidirectional: to understand the text you're looking you'll have to look back (at the previous words) and forward (at the next words)
- Transformers: The "Attention Is All You Need" paper presented the Transformer model. The Transformer reads entire sequences of tokens at once. In a sense, the model is non-directional, while LSTMs read sequentially (left-to-right or right-to-left).
- Pre-trained contextualized word embeddings: The ELMO paper introduced a way to encode words based on their meaning/context. For example, nails have multiple meanings - fingernails and metal nails.

FIGURE I
BERT CLASSIFICATION EXPLAINED

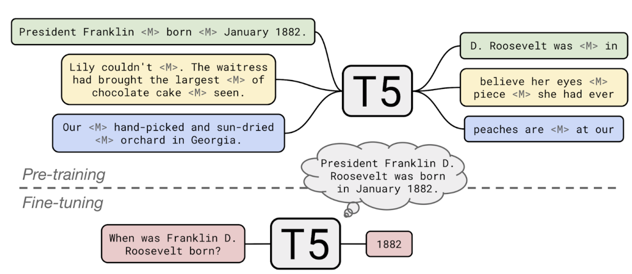


II. T5: TEXT-TO-TEXT TRANSFER TRANSFORMER

The Text-to-Text Transfer Transformer or T5 is a type of Transformer that is capable of being trained on a variety of tasks with a uniform architecture. It was created by Google AI and was published in the paper "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer". T5 model proposes all NLP tasks into a unified text-to-text format where the input and output are always text strings. [9] This formatting makes one T5 model fit for multiple tasks. T5 takes in text input from left for various NLP tasks and outputs the text for that respective task as can be seen in Figure 2 below.

The data that was used to the model is the largest by far compared to other models such as BERT. The authors have named it C4 (Colossal Clean Crawled Corpus). It's approximately 700GB in size and is the cleaned version Common Crawl. The authors have mentioned that the cleaning in sense of extracting is only English text, removing code lines, etc. It's a high-quality English language corpus that is available for download. Also, the T5 model, on C4, achieves state-of-the-art results on many NLP benchmarks while being flexible enough to be fine-tuned to a variety of important downstream tasks.

FIGURE 2
T5 SUMMARY EXPLAINED



T5 is fine-tuned on a variety of tasks from GLUE and SuperGLUE benchmarks as well as the CNN/DM for question answering, and the WMT EN-FR, EN-GE, and EN-RO for translation [10]. Some of the use-cases include:

- Sentence acceptability judgment (CoLA)
- Sentiment analysis (SST-2)
- Paraphrasing / sentence similarity (MRPC; STS-B; QQP)
- Natural language inference (MNLI, QNLI, RTE, CB)

- Sentence completion (COPA)
- Word sense disambiguation (WIC)
- Question answering (MultiRC, ReCoRD; BoolQ)

After giving some background on the models applied, the training and implementation of these models for ESG tasks will be discussed in the next section.

III. PROPOSED APPROACH

In this section, the implementation of BERT and T5 Transformer models will be engaged to build 4 models for different ESG and Impact Investing tasks. This includes collecting and hand-labeling over 26,000 textual ESG data for:

1) T5 Summary Model fine-tuned for ESG & Impact-related text
2) RoBERTa fine-tuned for ESG & Impact vs Financial text classification across 2 categories
3) BERT fine-tuned for SASB & SDG framework categories text classification across 26 and 17 labels
4) BERT fine-tuned for Sentiment Classification & Regression for quantification of texts

News Data Extraction has been one of the most important steps in the pipeline because the quality of the models can only be as good as the input, which is textual data. No formal news aggregation service has been used in the data collection process. Hundreds of thousands of news sources have been accessed to retrieve company/fund historical news articles real-time, working with raw data files.

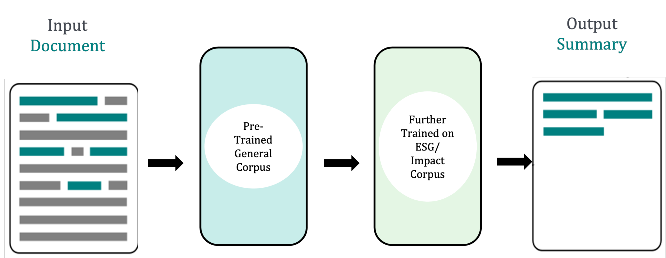
A. T5 Summary Model for Long ESG & Impact Articles

As mentioned in the previous section, in early 2020, Google introduced the Text-To-Text Transfer Transformer (T5) model. It also introduced a new open-source pre-training dataset, called the Colossal Clean Crawled Corpus (C4). The T5 model, pre-trained on C4, achieved state-of-the-art results on many NLP benchmarks while being flexible enough to be fine-tuned to a variety of important downstream tasks.

The goal with fine-tuning T5 model for ESG has been to build a text summarization model that would allow us to compress long ESG and Impact articles into two or three-sentence summary "objects" as inputs to classification and quantification models. In addition, just also being able to read such texts for labeling and analysis are powerful tasks by themselves as will be show in later sections.

To build this model, around 10,000 ESG and Impact-related texts and summaries have been generated for training. Figure 3 below presents the general idea of what has been accomplish. But there is no better way to discuss the model's accuracy other than share some of the model's summary predictions which has been done in Table 4.

FIGURE 3
T5 SUMMARY EXPLAINED



B. RoBERTa for ESG & Impact vs Financial Texts

After collecting and labeling about 10,000 ESG/Impact and Financial Texts (5000 each) and compressing the content into a few-sentence-summaries using our fine-tuned T5 model, we can now fine-tune RoBERTa for the two-label classification task. We did not separately discuss RoBERTa but RoBERTa is an improved recipe for training BERT models [11] that did better on this particular task.

First, I will address the importance of solving this classification problem as adding additional potentially lengthy process may seem unclear from the operational standpoint. Often times when collecting textual data, we gather large amount of data about a company or fund, most of which is unrelated to ESG and Impact. After this, we had to filter unrelated data out. The problem with building a rule-based approach or filtering unrelated news articles out using keywords is that we miss out many related articles, and we do not filter out unrelated articles. For example, if we want to gather textual data about "Employee Diversity" topic about a company using a rule-based approach, the keyword "diversity" will most likely be one of the words in the "if this then that" logic. The problem with this logic is that texts unrelated to "Employee Diversity" topic would be filtered in such as biodiversity. In addition, many topic related texts would be filtered out if they do not contain that specific set of words. In theory, it is possible to write cases for each of these cases but then it is neither flexible nor reliable in production.

The neural net model does not face this problem as it does not look up for specific set of words but rather looks at the meaning and intent of the text given the train data.

The results of ESG/Finance Text Classification are reported in Table 1:

TABLE I
ESG/FINANCE TEXT CLASSIFICATION ACCURACY REPORT

		precision	recall	f1-score	support
esg	0	0.99	0.99	0.99	132
financial	1	0.98	0.98	0.98	112
	accuracy			0.99	244
	macro avg	0.99	0.99	0.99	244
	weighted avg	0.99	0.99	0.99	244

C. BERT BASE FOR SASB & SDG TEXT CLASSIFICATION MODEL

After addressing the text summarization and ESG/Finance classification problems, we can now fine-tune BERT for SASB and SDG classification problem. This has been a rather challenging and comprehensive problem to address due to the number of categories/labels (26), ESG expertise required, and the amount of data collected and labeled.

Overall, around 26,000 ESG and Impact-related texts have been labeled: 1000 across each SASB label. The inputs to this model are the summarized and filtered ESG/Impact texts. The objective is to predict whether the information is related to any number of SASB or SDG categories or sub-categories, which includes "GHG Emissions", "Data Security", or "Employee Inclusion & Diversity", etc. One of the business use-cases for this classification model is assessing companies' or funds' performance across any ESG or Impact topic of choice and only addressing ESG issues that investors feel strongly about. The results of BERT BASE for SASB Text Classification are given in Table 2 in the next page:

TABLE 2
SASB CLASSIFICATION REPORT

		precision	recall	f1-score	support
Business_Ethics	0	0.93	0.92	0.92	96
Data_Security	1	0.86	0.90	0.88	87
Access_And_Affordability	2	0.88	0.87	0.87	54
Business_Model_Resilience	3	0.92	0.85	0.88	57
Competitive_Behavior	4	0.94	0.87	0.90	119
Critical_Incident_Risk_Management	5	0.87	0.87	0.87	63
Customer_Welfare	6	0.92	0.92	0.92	71
Director_Appointment_Removal	7	0.90	0.90	0.90	56
Employee_Engagement_Inclusion_And_Diversity	8	0.90	0.84	0.87	83
Employee_Health_And_Safety	9	0.89	0.92	0.90	134
Human_Rights_And_Community_Relations	10	0.88	0.91	0.89	129
Labor_Practices	11	0.90	0.84	0.87	147
Management_Of_Legal_And_Regulatory_Framework	12	0.91	0.86	0.88	123
Physical_Impacts_Of_Climate_Change	13	0.95	0.92	0.93	111
Product_Quality_And_Safety	14	0.86	0.84	0.85	38
Product_Design_And_Lifecycle_Management	15	0.95	0.91	0.93	90
Selling_Practices_And_Product_Labeling	16	0.93	0.91	0.92	48
Supply_Chain_Management	17	0.95	0.87	0.91	75
Systemic_Risk_Management	18	0.90	0.92	0.91	81
Waste_And_Hazardous_Materials_Management	19	0.88	0.87	0.87	84
Water_And_Wastewater_Management	20	0.85	0.86	0.85	75
Air_Quality	21	0.91	0.88	0.89	104
Customer_Privacy	22	0.88	0.88	0.88	86
Ecological_Impacts	23	0.93	0.88	0.90	93
Energy_Management	24	0.88	0.88	0.88	87
GHG_Emissions	25	0.90	0.92	0.91	83
	accuracy			0.9	2274
	macro avg	0.90	0.89	0.89	2274
	weighted avg	0.90	0.89	0.89	2274

The accuracy results of the classification model across are reported in the table above. Generally, there are categories in the SASB framework that are close by definition, which makes the SASB classifier to "confuse" certain categories. For example, Labor Practices and Employee Health & Safety or Critical Incident Risk Management and Data Security. These have caused the accuracy of the model to decrease slightly in this report. However, in practice, when deploying the model for production, we do not only consider one category prediction for each text, but two or more labels given the confidence score of the prediction is high enough. The categories can be adapted for any chosen ESG or Impact framework. In case we choose a framework with fewer categories, the accuracy of the prediction model will most definitely increase even more. Overall, around 90% accuracy already showcase the potential for applying these models in production. In next sections a sample dataset will be presented to demonstrate the practical significance of this model.

The model architecture parameters are given below (not everything has been shared, but these are the best parameters):

- "architectures": ["BertForSequenceClassification"],
- "attention_probs_dropout_prob": 0.1,
- "gradient_checkpointing": false,
- "hidden_act": "gelu",
- "hidden_dropout_prob": 0.1,
- "hidden_size": 768,
- "initializer_range": 0.02,
- "intermediate_size": 3072,
- "layer_norm_eps": 1e-12,
- "max_position_embeddings": 512,
- "model_type": "bert",
- "num_attention_heads": 12,
- "num_hidden_layers": 12,
- "type_vocab_size": 2,
- "vocab_size": 30522

III. BERT BASE FURTHER TRAINED FOR ESG & IMPACT TEXT QUANTIFICATION

After building the datasets for the SASB and SDG classification model, we manually labeled the datasets for sentiment classification and regression tasks with BERT. Creating a reliable quantification model that allows us to generate scores of textual data related to companies' ESG and impact practices in real-time is essential to solving any ESG/Impact Investing. The motivation to build our own model stems from the fact that the accuracy of general- purpose models in prediction the class of the sentiment is extremely low, yet alone predicting a continuous target. There are other mechanisms for quantifying texts that do not involve regression problem, which is a relatively hard model to build because datasets with continuous targets are required. One such technique is detection ESG risks, classifying these risks based on severity, calculating number of incidents and then applying some standardization technique to produce ESG risk scores. This technique, in fact, is used by many prominent ESG data providers. We believe that our quantification model is more robust and has more use-cases than others because of its accurate numeric predictions.

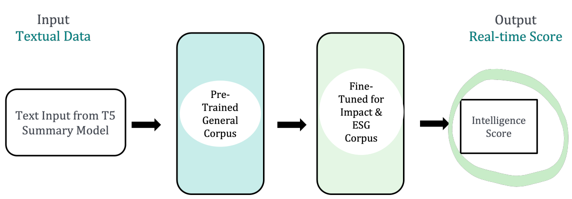
As for the accuracy of the model, on the subject matter of this paper, we will only report the results of the sentiment classification accuracy and share some prediction results of BERT regression model. To build and evaluate the accuracy of the regression model, MSELoss function of pytorch has been used. While for the classification model, the CrossEntropyLoss function has been used. Generally, this is the only difference between the models as far as this task is concerned. However, there is no similar model built in this area, to the best of my knowledge, to compare results for MSE results of regression task. An example predictions will be provided in next sections that would help understand what type of problem it is solving. The results of the BERT Sentiment Classification are reported in table 3 in next page.

TABLE 3
BERT CLASSIFICATION ACCURACY REPORT

		precision	recall	f1-score	support
positive	0	0.93	0.95	0.94	245
negative	1	0.89	0.92	0.90	254
neutral	2	0.91	0.9	0.90	289
	accuracy			0.91	788
	macro avg	0.91	0.92	0.92	788
	weighted avg	0.91	0.92	0.92	788

The report shows that the Sentiment Classification model is highly accurate in predicting ESG and Impact texts. The same could be said for the regression model.

FIGURE 4
BERT CLASSIFICATION EXPLAINED



This visual representation explains the general logic behind implementation. The input are texts from the T5 summary model that we trained for ESG/Impact tasks. The output is score from [-1,1] that we exponentially scale with weighted average techniques to arrive at the final Intelligence Score that is the output.

IV. USING FINE-TUNED MODELS FOR QUANTIFICATION AND CLASSIFICATION

In this section the results of our set of models will be presented. The table 4 below shows the score comparison among our fine-tuned models, traditional sentiment models and Amazon's AWS model on the summary texts provided. It shows that our model is best on new test data as it was realistically trained for Impact/ESG Investing texts. On the first example, BT Group has reportedly cut its emissions, committing to net-zero emission target. Our BERT model scored this as overwhelmingly positive as evidenced by 80+ score. (More details about the methodology will be provided). The same article has been rated around 50 and 51 by traditional NLTK based models and AWS Comprehend. These surely do not reflect the reality under the ESG and Impact context. On the same example, the same text, which has been generated by our fine-tuned T5 model, is detected as related to GHG Emissions and Physical Impacts of Climate Change. When deployed for production, not only one but first two predicted labels will be considered. This will overwhelmingly improve the real accuracy of the model. There are two other examples provided in Table 3, which also indicate similar logical conclusions. For example, third row/case tells us that "GitLab has acquired Peach Tech…". Again, in the ESG context this is a positive event under the Competitive Behavior and Systemic Risk Management, automatically generated by our fine-tuned model. Our regression model generated a score for this at 76. NLTK-based models would tell us that it is only at 55, and same with AWS Comprehend. This well explains that the usage of general-purpose models for ESG and Impact would be a mistake.

After gathering 1000s of textual data about a company or a fund, applying fine-tuned T5 Summary model to generate compressed versions of texts, filtering ESG/Impact texts, classifying data across ESG/Impact issues, and producing scores for individual texts, it is our intention convert summary-level model outputs into company or fund-level scores. These intelligence or insight scores, when updated daily, will then be applied to build smart beta portfolios by weighting ESG/Impact issues based on investor preferences. We are not only concerned about quantifying public company performance but also private. We first take the daily standardized scores of ESG performance changes that highlight opportunities and controversies each day and apply MiniMax model for standardization. Then, we resample data into average weekly/monthly data and apply Exponentially Weighted Moving Average to the daily ESG/Impact performance graph. If we represent the raw daily data sequence by $\{x_t\}$ at time t and note the output of the exponential smoothing algorithm as $\{s_t\}$, s_t, which is the company-level weekly/monthly ratings, will be:

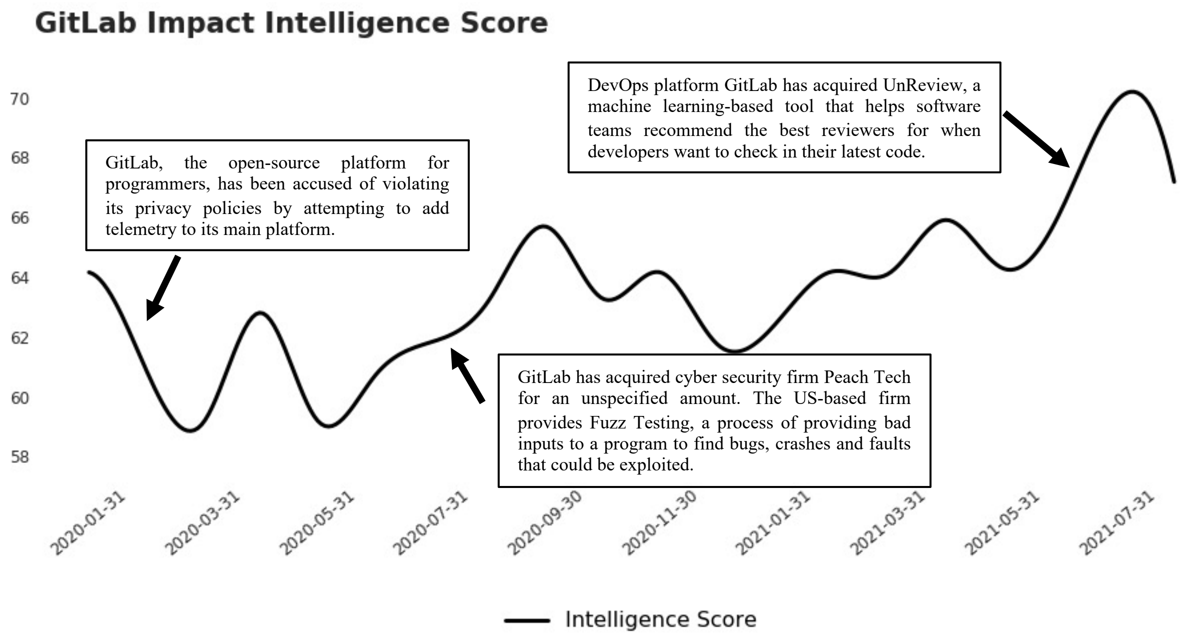
$$s_t = \alpha x_t + (1-\alpha)s_{t-1} = s_{t-1} + \alpha(x_t - s_{t-1}).$$

Where alpha is the smoothing factor $0 < \alpha < 1$.

TABLE 4
FINE-TUNED MODEL RESULTS COMPARISON TABLE

Fine-Tuned BERT Classification Model Predictions	**T5 Model Summary**	**Fine-Tuned BERT Regression Model Predictions**	**Traditional Sentiment Model Results**	**Amazon AWS Model Results**
GHG_Emissions, Physical_Impacts_Of_Climate_Change	BT cuts emissions by 14% over the last 12 months, putting the company on course to reach net-zero emissions by 2045. The 14% reduction in emissions recorded over the last financial year means that BT has reduced emissions by 57% since 2016/17. This puts the company on track to reduce emissions by 87% by 2031 and reach net-zero by 2045.	**84.39**	**50.01**	**51.35**
Customer_Privacy, Data_Security	GitLab, the open-source platform for programmers, has been accused of violating its privacy policies by attempting to add telemetry to its main platform. An email sent on October 23 informed GitLab users of the changes that would be added to the platform and the new Terms of Service the users would have to agree to. However, GitLab CEO Sid Sijbrandij admitted that adding user data monitoring was a mistake and did so without discussion with the community.	**21.31**	**57.215**	**40.28**
Competitive_Behavior Systemic_Risk_Management	GitLab has acquired Peach Tech, a security software firm specializing in protocol fuzz testing and dynamic application security testing (DAST) API testing, and Fuzzit, a continuous fuzz testing solution providing coverage-guided testing. This makes GitLab's DevSecOps offering the first security solution to offer both coverage-guided and behavioral fuzz testing techniques.	**76.21**	**55.357**	**55.06**

FIGURE 5
EXAMPLE INTELLIGENCE SCORE GRAPH FOR A PRIVATE COMPANY



Smoothed time series data are then calculated across each 26 SASB and 17 SDG labels to generate Intelligence Scores across each category for investors or analysts. An example Intelligence Score graph is demonstrated in Figure 5 above for a company called GitLab. For this company, a total of 1,000 articles have been collected over the past year and eight months related to its ESG practices. The score graph is the result of monthly sampling and exponential smoothing with equal weights.

IV. IMPLEMENTATION DETAILS

Computational Costs: Pre-processing and training BERT for ESG & Impact texts took significant computational resources. The entire model pre-processing and training procedure took roughly 30 days of computational runtime using a Nvidia Tesla P100 12gb (and significant CPU power and memory for pre-processing tasks). 30 days of continuous run-time is a significant investment and may be beyond the reach of some companies. Added to the fact that more than 1 year worth of training and testing data have been collected and labeled by ESG experts.

V. CONCLUSION AND FUTURE WORK

In this paper, we have introduced four neural language pre-trained models further trained for ESG and Impact investing tasks. Both BERT and T5, which have separately been discussed in this paper, are state-of-the-art models that essentially can handle any language-based task. We collected and hand-labeled over 26,000 ESG and Impact texts and fine-tuned highly accurate 4 models over the past roughly two years.

The pipeline of models is capable of extracting any material publicly available data about a fund or company, summarizing long articles into few-sentence-summaries, filtering out articles unrelated to ESG and Impact, automatically identifying the topics of such articles across SASB and SDG frameworks, and quantifying texts across these categories to generate exponentially weighted company or fund-level intelligence scores. To the best of our knowledge, this is the first use of ESG & Impact domain neural language models (BERT and T5) applied for Private Equity and Impact Investing.

As discussed in Literature Review section, there have been multiple implementations of Transformers for ESG tasks. However, all existing published works have only applied BERT (we apply BERT RoBERTa and T5 that has been introduced in 2020) and focus on capital markets and not the private industry, again to the best of my knowledge. Further to this, while there are research works conducted separately for sentiment analysis, topic detection, ESG stock volatility prediction, ESG portfolio integration, etc. This is also the first work to hand-label domain-specific textual data to build the whole pipeline from automatic extraction of information to prediction of scores to generation of scores.

Some of the primary areas of future work will be to collect more training data focused more on Impact Investing beyond SDGs such as IRIS (Impact Reporting and Investment Standards), expanding datasets beyond articles, improve scoring methodology (more weights for data from prominent sources), running models for several investment funds to compare fund-level intelligence score movements vs fund money flow, and, generally, focusing more on use-cases of this technology for private equities and their portfolio companies.

In the last few years, we have seen an increasing capital allocation toward ESG and Impact funds and companies and integration of ESG in portfolio management. According to WSJ and Boston Trust Walden Co [3], in both the public and private markets, there’s no shortage of products—ESG opportunities abound; investors can stick to their values-based investing ideals and still be properly allocated across asset classes. The literature on ESG integration is new and mixed. Initially, investors were skeptical about excess return generating potential of ESG data. Over time as more data has been published, it seemed that ESG integration in portfolios may not be compromised for lower returns. Most recently, literature has evolved to claim that ESG integration in portfolios may be a good risk reduction measure and may produce alpha. According to most recent Morningstar report [12], in Q1 2021, a net of $21.5 billion flowed into ESG-managed mutual funds and ETFs – a record amount and almost double the net inflows in the year-earlier quarter. In page 7, information about the number of funds across different sustainability-related themes and their assets are shown.

FIGURE 6
Where is the money going among sustainable-investment themes as of Q1 2021

Source: Mornigstar, WSJ

Going back to the question asked in the Introduction section about how might a GP actually measure the return on investment tied to a commitment to gender diversity? The answer now can be answered using this set of models.

1) Pick a framework assume either SDG "Gender Equality" or SASB "SASB Employee Engagement Diversity & Inclusion"
2) Collect and filter data textual across this category
3) Run fine-tuned classification and quantification models to quantify employee diversity performance of Palladium Equity Partners after making a hypothetical investment in March 2021

In June 2021, Palladium Equity Partners appointed Meahgan O'Grady as the Director of Business Development", according to Prnewswire [13]. In August 2021, Palladium Equity Partners promoted partner Eugenie Cesar-Fabian the head of ESG & sustainability, according to Themiddlemarket [14], building upon Palladium's commitment to ESG and diversity, equity and inclusion.

In this circumstance, investors in Palladium Equity will be able to determine if their investment has been justified. Further to this, we did not need the PE to share any ESG information at the GP or the portco level, disclose data or qualitative case studies.

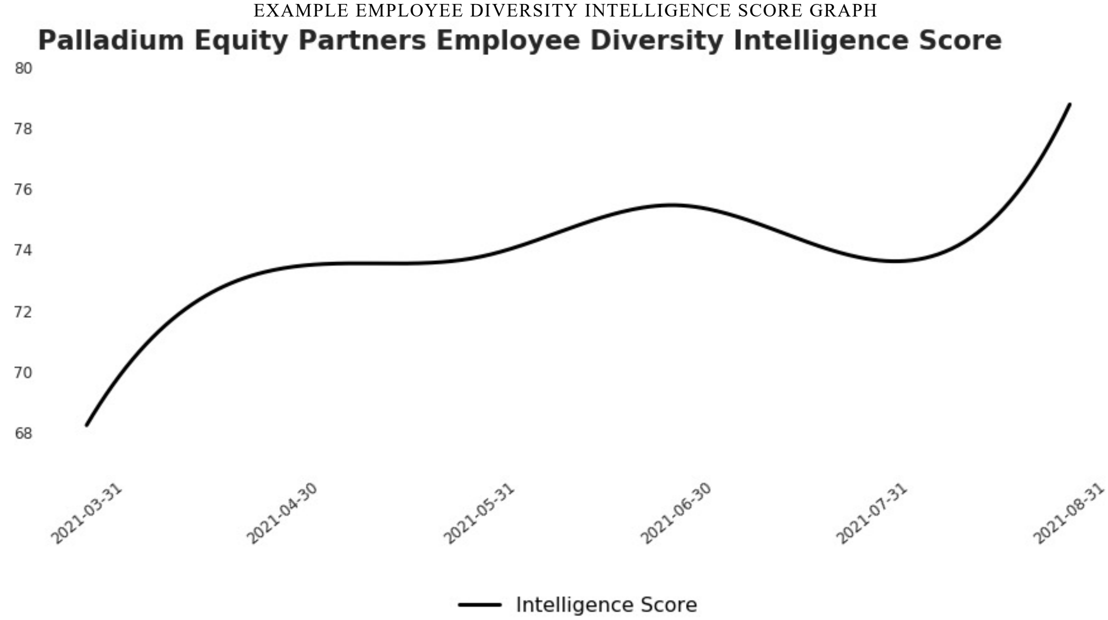
REFERENCES

[1] The Financial Times LTD. Definition of ESG. http://lexicon.ft.com/Term?term=ESG.

[2] What you need to know about impact investing. The GIIN. https://thegiin.org/impact-investing/need-to-know/#what-is-impact-investing.

[3] Munk, C. W. (2021, June 27). The new math of socially responsible investing. The Wall Street Journal. https://www.wsj.com/articles/socially-responsible-investing-11624288038.

FIGURE 7
EXAMPLE EMPLOYEE DIVERSITY INTELLIGENCE SCORE GRAPH



[4] Pucker, K., & Kotsantonis, S. (2020). Private equity makes ESG promises. but their impact is often superficial. Institutional Investor. https://www.institutionalinvestor.com/article/b1m8spzx5bp6g7/Private-Equity-Makes-ESG-Promises-But-Their-Impact-Is-Often-Superficial.

[5] Nugent, T., Stelea, N., & Leidner, J. (2020, October 16). Detecting ESG topics using domain-specific language models and data augmentation approaches. https://arxiv.org/pdf/2010.08319v1.pdf.

[6] Guo, T., Jamet, N., Betrix, V., & Piquet, L. (2020, May 5). ESG2Risk: A Deep Learning Framework from ESG News to Stock Volatility Prediction. https://arxiv.org/pdf/2005.02527v1.pdf.

[7] Sokolov, A., Mostovoy, J., Ding, J., & Seco, L. (2021). Building Machine Learning Systems for Automated ESG Scoring. https://jesg.pm-research.com/content/pmrjesg/early/2021/01/09/jesg.2021.1.010.full.pdf.

[8] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019, May 24). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. https://arxiv.org/pdf/1810.04805.pdf.

[9] Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., & Liu, P. J. (2020, July 28). Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. https://arxiv.org/pdf/1910.10683.pdf.

[10] Versloot, C. (2021, June 21). What is the T5 transformer and how does it work? MachineCurve. https://www.machinecurve.com/index.php/question/what-is-the-t5-transformer-and-how-does-it-work/.

[11] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019, July 26). Roberta: A robustly optimized Bert pretraining approach. arXiv.org. https://arxiv.org/abs/1907.11692.

[12] Morningstar Manager Research. (2021, July). Global Sustainable Fund Flows: Q2 2021 in Review. https://www.morningstar.com/lp/global-esg-flows.

[13] Melendez, J. (2021, August 17). Palladium Equity Partners Expands Team. Mergers & Acquisitions. https://www.themiddlemarket.com/latest-news/palladium-equity-partners-expands-team.

[14] Palladium Equity Partners. (2021, August 16). Palladium equity partners recruits Carlos Reyes from the World Bank. https://www.prnewswire.com/news-releases/palladium-equity-partners-recruits-carlos-reyes-from-the-world-bank-appoints-partner-eugenie-cesar-fabian-to-head-of-esg--sustainability-and-promotes-dominick-barbieri-to-general-counsel--chief-compliance-officer-301355378.html.