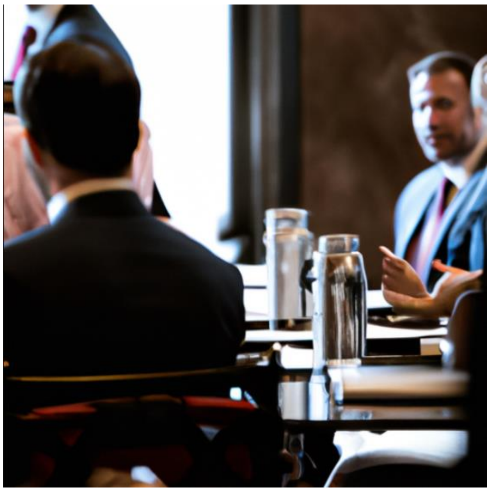
Using AI to extract sentiment from conference calls

Wang Chun Wei, Realindex Investments

For Institutional Use Only

July 2023



An AI generated image of a conference call meeting (using OpenAI DALL-E)

Extracting alpha from conference calls

At Realindex, our ongoing research is moving strongly into the area of natural language processing (NLP), and how AI tools can be used to generate alpha. In particular, we have been looking at applying NLP on conference calls for some time, and insights from our research in this area have made their way into our alpha model. We have already published twice on this topic:

- In September 2022: linguistic clarity in conference calls can hint to future stock return performance. We showed that our variation of the Gunning Fog index utilizing i) sentence length, ii) % of complex words and iii) % of numeric words, provided us with information about managers may try to obfuscate future negative financial performance. Stocks with higher obfuscation scores under-performed stocks with lower obfuscation scores.
- In December 2022: extracting linguistic tone or sentiment from conference calls. Tone scores were constructed using the Loughran McDonald (2012, 2016) word dictionary. This is a common approach used in academia and industry, for example, it is used in Brockman, Li and Price (2015).

Since launching these signals into production, we were pleased to see their performance during the banking crisis in March 2023. In particular, these signals drove much of the underweight of Credit Suisse and US regional banks like SVB Financial, Signature Bank and First Republic in all of our portfolios. For example, Figure 1 shows the signals and holdings for SVB. We are keen to continue work in this exciting area of research and believe there is more to be uncovered.

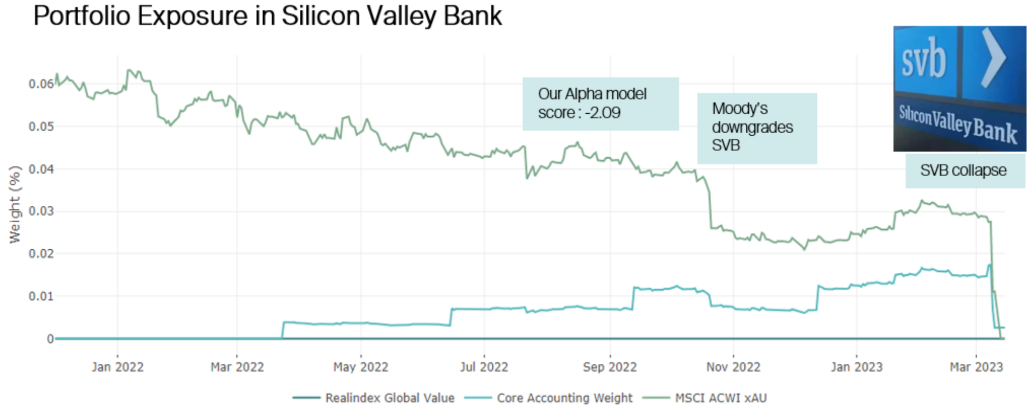
Here, we show that estimating linguistic tone from conference call transcripts can be enhanced using more advanced NLP techniques. In particular, we have focused on employing the BERT (Bidirectional Encoder Representations from Transformers) family of language models to assist us in sentiment classification. BERT was developed in 2018 by the Google AI language team. We use a fine-tuned version of BERT developed by Huang, Wang and Yang (2022), adapted for financial text.

In the following section, we will briefly explain large language models, BERT and FinBERT, and how we implement them on conference call transcripts. We also show how FinBERT can dramatically improve performance over the more traditional Loughran-McDonald word dictionary approach.

SVB Financial Group

Our alpha model signals and our portfolio construction process made sure we were never exposed to SVB

Portfolio Exposure in Silicon Valley Bank



Whilst the Core Accounting weights wanted to hold between 1-2bps in SVB,

Our enhancements process made sure the portfolio <u>**never**</u> invested in SVB.

None of our Value or DA strategies had any SVB.

SVB total alpha score was, for a long time, the worst among US banks for us

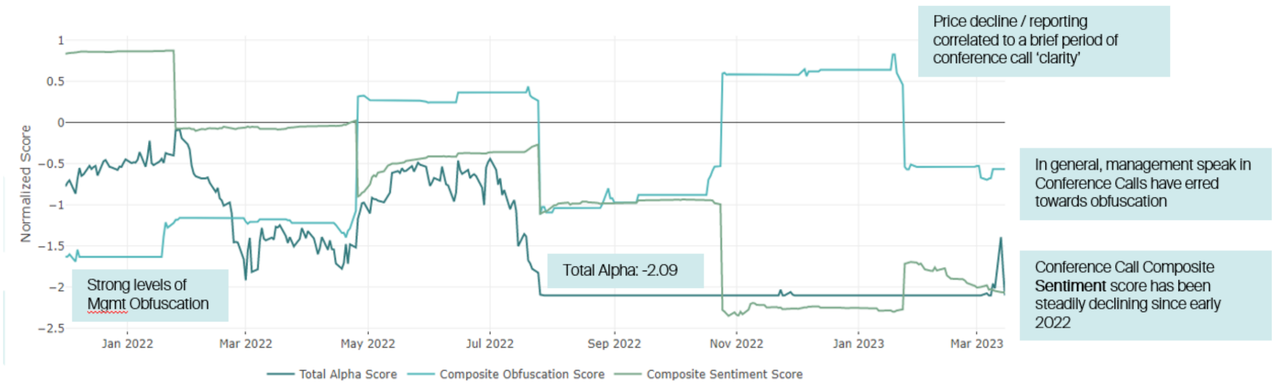


Figure 1. SVB Financial Group

Source: Realindex, Factset as at 31 March 2023

In a nutshell: Large language models, BERT and FinBERT

NLP algorithms have advanced significantly over the last decade. For instance, Word2Vec (Mikolev et al., 2013), GloVe (Pennington et al., 2014), FastText (2016), ELMo (Peters et al., 2018), BERT (Devlin et al., 2018) and GPT (Radford et al., 2018).

Despite these advancements, a lot of existing research on natural language processing (NLP) in the accounting and finance domain remains largely focused on the "bag of words" approach. For instance, see the work of Li (2010), Brockman, Li and Price (2015) and Loughran and McDonald (2016).

In this section, we will briefly explain the intuition behind these newer NLP algorithms in simple terms, and how this is relevant for us.

Word2Vec (Mikolev et al., 2013) uses word level embedding. This means words are represented as a numerical vector (hence the name word2vec). This approach cannot capture the context of the word, as the same words will always have the same vector no matter what text surrounds them. However, this approach is a step above what is known as "one-hot" encoding where words are completely independent of each other. The word2vec word vector is learnt from co-occurrences in English language text - for example, the word 'Japan' and the word 'sushi' will be closer to each other in the vector space than the words 'Japan' and 'Jamaica'.

This lack of contextualization means Word2Vec is unable to distinguish between the word 'bank' in 'a river bank' versus 'Deutsche Bank'. The inability to understand context often leads to humorous outcomes. Figure 2 below shows how AI can misrepresent words or mix up the context.

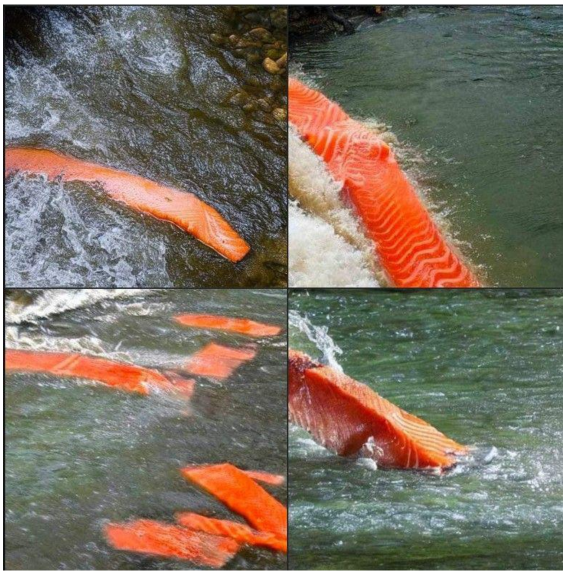


Figure 2. 'Salmon swimming in water' generated by AI

Large language models such as ELMo, BERT and GPT are improvements on this – they have *contextualized* word embeddings. This means words can have different vectors depending on their context. In order to gain contextual awareness, these models need to operate at the sentence level (rather than at the word level in Word2Vec).

Here, we focus on BERT. BERT word embeddings are based on three components:

- Token Embedding: where each word piece token is converted into a 768-dimensional vector representation
- Segment Embedding: this determines which sentence the word belongs to (for instance 0 for the first sentence, 1 for the second sentence and so on.)
- Position Embedding: the index position of the word in the sentence

These three components in the word embedding gives BERT the ability to see words in the context of a sentence and of a paragraph of text. BERT is then trained using a two-stage process: an unsupervised stage and a supervised fine-tuning stage.

- In the unsupervised stage, it learns from a large corpus of text including Wikipedia (circa 2.5bn words) and Google's BooksCorpus (circa 800mil words). BERT performs two training tasks; masked language model (MLM) where it learns the surrounding words within each sentence and next sentence prediction (NSP) where it learns the relationship between sentences by trying to predict if one sentence follows another.
- A supervised stage then fine-tunes the BERT model for specific use cases; for example (in our case), the sentiment classification.

Our work heavily relies on the work of Huang, Wang and Yang (2022). They develop a BERT model for financial data (FinBERT). In the unsupervised stage, FinBERT learns semantic and syntactic information from unlabelled financial text including corporate filings, analyst reports and conference calls. In the supervised learning stage for sentiment classification, Huang, Wang and Yang (2022) use 10,000 sentences from analyst reports for training. They show that having finance vocabulary, despite the smaller training sample, outperforms general models such as BERT. They also show that FinBERT outperforms existing NLP techniques such as Loughran-McDonald, Naive Bayes, SVMs, CNNs, LSTMs and random forests for sentiment classification.[1]

Below we show a FinBERT example using the 2021 Q3 conference call for Tesla. We show three different chunks of text from the management section that yield different sentiment classifications from FinBERT – the sentiment (Neutral, Positive, Negative) plus a score indicating confidence in the classification.

[1] A cluster of abbreviations: SVMs are Support Vector Machines, CNNs are Convolutional Neural Networks, LSTMs are Long Short-Term Memory networks (a type of neural network). The entire field is known as Machine Learning (ML)

Text	Sentiment
Good afternoon, everyone, and welcome to Tesla's Third Quarter 2021 Q&A Webcast. My name is Martin Viecha, Senior Director of Investor Relations. And I'm joined today by our CFO, Zachary Kirkhorn, and our Senior VP, Drew Baglino, as well as other executives. Our Q3 results were announced at about 3 PM Central Time in the updated deck we published at the same link as this webcast. During this call, we will discuss our business outlook and make forward-looking statements. These comments are based on our predictions and expectations as of today. Actual events or results could differ materially due to a number of risks and uncertainties, including those mentioned in our most recent filings with the SEC.During the question-and-answer portion of today's call, please limit yourself to one question and one follow-up. Please use the Raise Hand button to join the question queue.	Neutral (score 0.99)
Additionally, we have made great progress increasing production volumes of Model S and have recently started the ramp and deliveries of Model X. It will take a bit more time to get this program back to prior volumes, but based on demand, we are targeting to exceed historical production levels. We have also completed the transition of our Shanghai factory as our main export hub. This has enabled us to supply more vehicles to the North America market and to introduce Model Y to Europe. Due to part shortages and logistics variability, we have not been able to run our factories at full capacity. It's important to note that while we have roughly doubled deliveries year-to-date, this has been exceptionally difficult to achieve. I want to thank our supply chain team for their incredible work and our production teams for showing impressive flexibility as we make adjustments real time. This team's expertise in the chip industry across all tiers has made a huge difference in managing through these challenges.	Positive (score 0.99)
The Model S has now returned to positive gross margin and we expect this to increase with higher production and the ramp of Model X.As was the case in Q2, there was some net benefit from pricing actions. However, this remains small in the context of other contributors. Please keep in mind that given backlog it will take time for the impact of recent changes to flow through our financials. Note that we are also not yet recognizing additional revenue from the FSD beta program. Supply chain challenges, including expedites, continue to provide cost headwinds, as was also the case with FX this quarter. While we are seeing an impact from the rising commodity and labor costs, we have also been adjusting pricing which should help to compensate. Overall, as I mentioned in our last call, our P&L continues to benefit from the marginal profitability of each incremental unit with higher fixed cost absorption.	Negative (score 0.85)

Table 1. Sample FinBERT Sentiment classification based on 2021 Q3 Conference call for Tesla

Empirical results showing AI potency

Our tone signals are generated by cutting up or "chunking" larger transcripts into smaller and more manageable sections for FinBERT. We employ a process where score weights these sections to generate our final score. In Table 2 we provide some quick summary performance results of our FinBERT signals over a period of 10 years in weekly frequency.[2] The signals are broken into three sections: the analyst questions, the management answers and the prepared presentation section. The results are very strong, with positive information ratios (IRs[3]) for each signal

[2] The returns are quintile spreads – that is, quintile 1 return less quintile 5 return

[3] Active return divided by active risk

Signal	Mean	Stdev	IR	Turnover	T-stat	Hit Rate
North American Tone						
Question	1.52%	5.28%	0.29	208%	1.01	51.7%
Answer	5.09%	7.02%	0.73	173%	2.55	57.0%
Management Prepared Remarks	3.45%	6.73%	0.51	168%	1.80	57.0%
Question YoY	3.58%	4.34%	0.83	246%	2.78	54.9%
Answer YoY	4.00%	4.47%	0.90	248%	3.02	55.9%
Management Prepared Remarks YoY	4.02%	5.80%	0.69	234%	2.33	55.4%
Europe Tone						
Question	5.02%	5.56%	0.90	211%	3.17	56.4%
Answer	2.91%	7.43%	0.39	181%	1.37	54.4%
Management Prepared Remarks	5.30%	7.92%	0.67	175%	2.35	54.8%
Question YoY	6.86%	5.03%	1.37	236%	4.60	58.3%
Answer YoY	3.91%	4.78%	0.82	237%	2.76	51.7%
Management Prepared Remarks YoY	4.68%	6.47%	0.72	228%	2.44	55.4%

Table 2. Conference Call FinBERT Tone Return Spread Performance

Data date range: 2010-06-08 to 2022-09-27

In Charts 1 & 2, we plot the cumulative long-short return spread performance across time. In Europe, analyst FinBERT sentiment seems particularly strong. The cumulative returns are strong, plus steady and consistent over time.

Chart 1a. North American FinBERT Long/Short Spread Performance

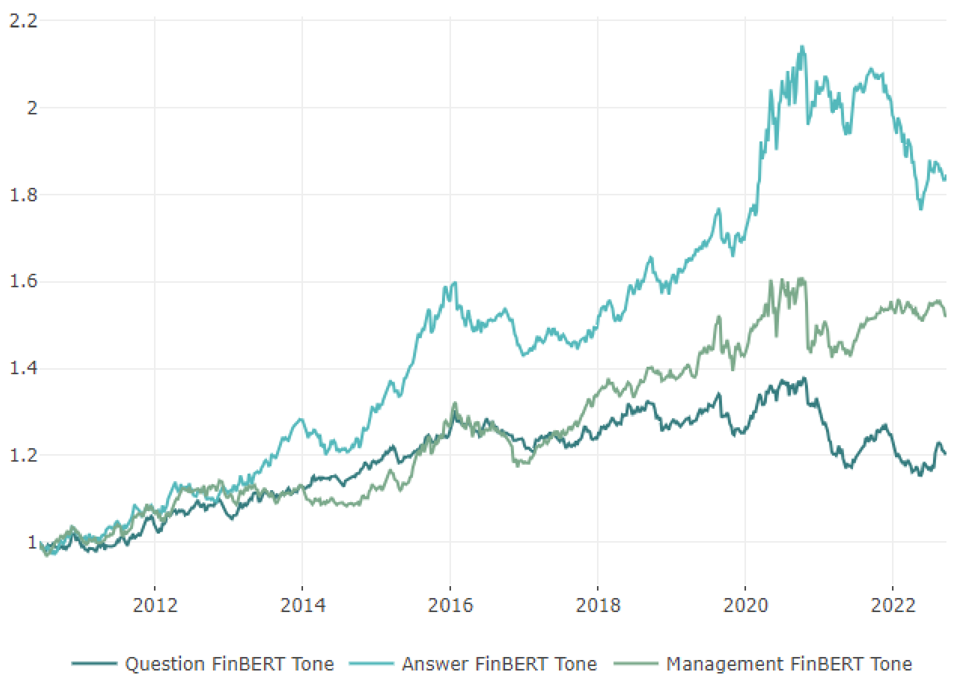


Chart 1b. North American FinBERT YoY Change Long/Short Spread Performance

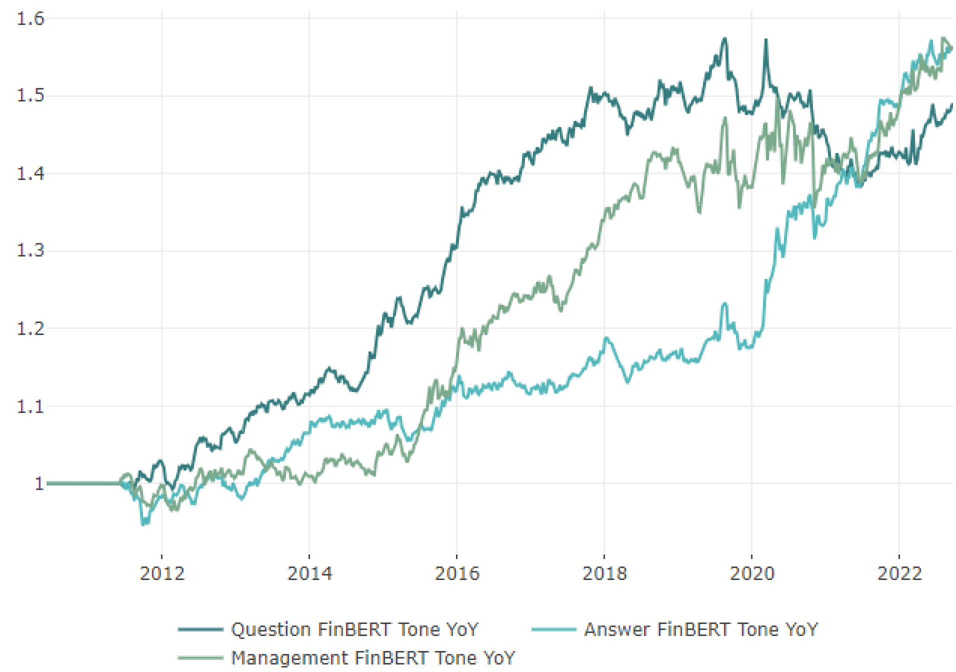


Chart 1a & 1b North American FinBERT Performance (equally weighted)

Date range: Jun 2010 to Sep 2022

Chart 2a. Europe FinBERT Long/Short Spread Performance

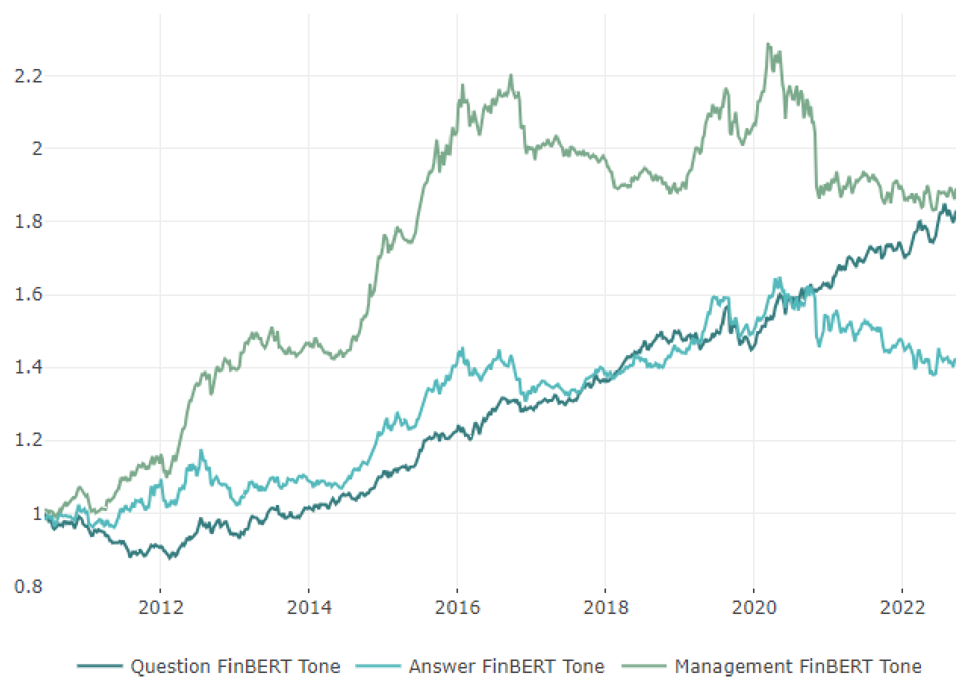


Chart 2b. Europe FinBERT YoY Change Long/Short Spread Performance

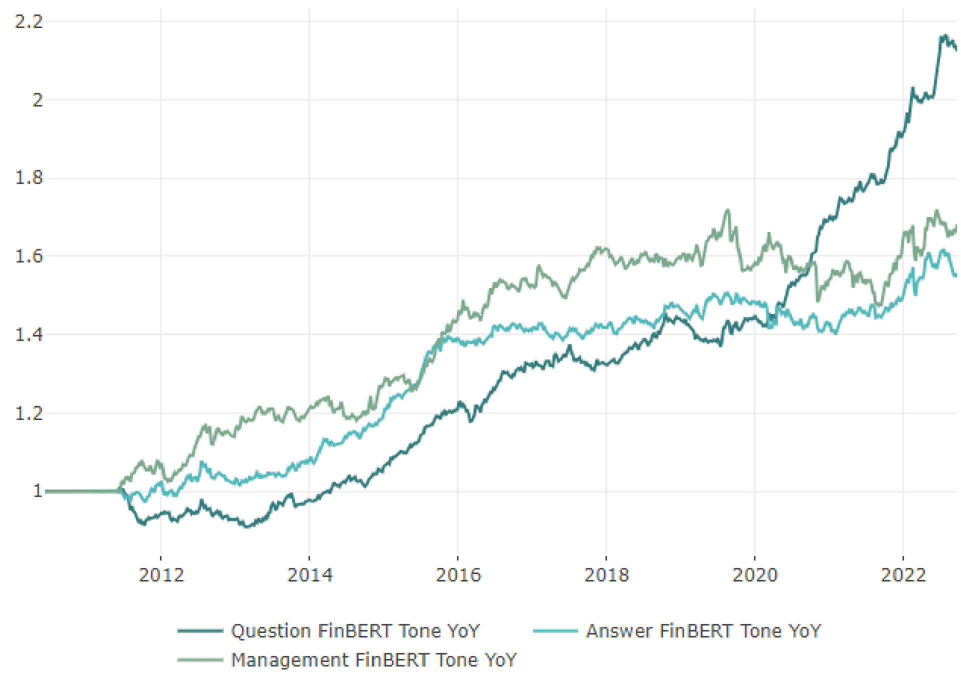


Chart 2a & 2b European FinBERT Performance (equally weighted)

Date range: Jun 2010 to Sep 2022

What we're doing with this at Realindex

In this short note, we show that implementing FinBERT or a variation of BERT pre-trained on financial data can provide meaningful information on future cross-sectional stock price returns. We are looking to implement a variation of this sentiment score for our alpha model as it is both additive and significant in alpha generation.

References

Devlin, Jacob; Chang, Ming-Wei; Lee, Kenton; Toutanova, Kristina (2018). “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”

Huang, A.H., Wang, H. and Y. Yang (2023) FinBERT: A Large Language Model for Extracting Information from Financial Text, Contemporary Accounting Research, Vol 40, Issue 2, 806-841

Jeffrey Pennington, Richard Socher, and Christopher Manning (2014) GloVe: Global Vectors for Word Representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 1532–1543, Doha, Qatar. Association for Computational Linguistics.

Peters ME, Neumann M, Iyyer M, Gardner M, Clark C, Lee K, Zettlemoyer L (2018). “Deep contextualized word representations”. arXiv

Radford, Alec; Narasimhan, Karthik; Salimans, Tim; Sutskever, Ilya (2018). “Improving Language Understanding by Generative Pre-Training”

Disclaimer

This material has been prepared and issued by First Sentier Investors Realindex Pty Ltd (ABN 24 133 312 017, AFSL 335381) (**Realindex**), which forms part of First Sentier Investors, a global asset management business. First Sentier Investors is ultimately owned by Mitsubishi UFJ Financial Group, Inc (**MUFG**), a global financial group.

This material is directed at persons who are 'wholesale clients' (as defined under the Corporations Act 2001 (Cth) (**Corporations Act**)) and has not been prepared for and is not intended for persons who are 'retail clients' (as defined under the Corporations Act). This material contains general information only. It is not intended to provide you with financial product advice and does not take into account your objectives, financial situation or needs. Before making an investment decision you should consider, with a financial advisor, whether this information is appropriate in light of your investment needs, objectives and financial situation.

Any opinions expressed in this material are the opinions of the individual author at the time of publication only and are subject to change without notice. Such opinions: (i) are not a recommendation to hold, purchase or sell a particular financial product; (ii) may not include all of the information needed to make an investment decision in relation to such a financial product; and (iii) may substantially differ from other individual authors within First Sentier Investors.

We have taken reasonable care to ensure that this material is accurate, current, and complete and fit for its intended purpose and audience as at the date of publication. No part of this material may be reproduced or transmitted in any form or by any means without the prior written consent of Realindex.