Marketing Material
For Professional Investors Only

Financial Sentiment Analysis with Large Language Models:

An Introductory & Comparative Study on News Flow

Systematic Equity Team - Emmanuel Hauptmann, Valentin Betrix, Nicolas Jamet, Tian Guo, Louis-Alexandre Piquet

Convinction in Fundamentals.
Powered by Technology.

Introduction.

Sentiment Analysis is the process by which a positive, negative or neutral tone is assigned to text. In finance, the role of sentiment analysis is to extract a numerical signal from financial text.[2,4,10,13] In the last years, the sentiment analysis process has transitioned from traditional dictionary-based approaches[3] to Machine Learning and Natural Language Processing (NLP) based methods leveraging text embeddings and (large) language models (LLMs).[7,8,12] Text embedding is an NLP technique that transforms words, phrases or documents into numerical representations in the high-dimensional vector space, where semantically similar entities are located closer to each other.[1] This technique has significantly improved NLP research, but cannot consistently adapt to the variability of text meaning in different contexts.

The advent of LLMs, such as BERT (Bidirectional Encoder Representations from Transformers),[6] GPT (Generative Pre-trained Transformer) and their variants has pushed NLP beyond traditional text embedding methods. LLMs' success is mainly driven by their Transformer architecture, a neural network specialised for contextualised text embedding.[5] Figure 1 illustrates the timeline of some important language model techniques.

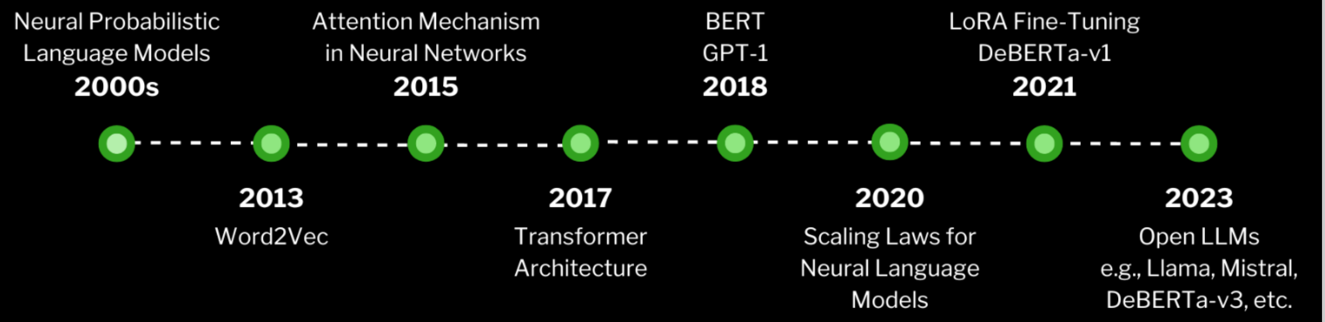

Figure 1: Some Impactful Techniques in the History of Language Models.

Financial Sentiment Analysis with LLMs

In this paper, we will introduce two techniques for undertaking financial sentiment analysis using LLMs. Firstly, we present the pre-training and fine-tuning process of adapting an LLM for financial sentiment analysis, i.e., from BERT to FinBERT.[6,7] Then, we describe a Transformer variant DeBERTa (Decoding-enhanced BERT with disentangled attention)[9,11] and FinDeBERTa featuring the technique of disentangled representations (or embeddings). As shown in Table 1, DeBERTa presents a competitive performance on several NLP benchmarks.

Model #Train	CoLA Mcc 8.5k	QQP Acc 364k	MNLI-m/mm Acc 393k	SST-2 Acc 67k	STS-B Corr 7k	QNLI Acc 108k	RTE Acc 2.5k	MRPC Acc 3.7k	Avg.
$\text{BERT}_{\text{large}}$	60.6	91.3	86.6/-	93.2	90.0	92.3	70.4	88.0	84.05
$\text{RoBERTa}_{\text{large}}$	68.0	92.2	90.2/90.2	96.4	92.4	93.9	86.6	90.9	88.82
$\text{XLNet}_{\text{large}}$	69.0	92.3	90.8/90.8	**97.0**	92.5	94.9	85.9	90.8	89.15
$\text{ELECTRA}_{\text{large}}$	69.1	92.4	90.9/-	96.9	92.6	95.0	88.0	90.8	89.46
$\text{DeBERTa}_{\text{large}}$	70.5	92.3	91.1/91.1	96.8	92.8	95.3	88.3	91.9	90.00
$\text{DeBERTaV3}_{\text{large}}$	**75.3**	**93.0**	**91.8/91.9**	96.9	**93.0**	**96.0**	**92.7**	**92.2**	**91.37**

Table 1: Comparison of (encoder) LLMs on GLUE benchmark (the higher the value, the better the model)[11]

Experimentally, we compare FinBERT and FinDeBERTa by backtesting the sentiment factors obtained when applying them to financial news flow. The result shows that FinBERT and FinDeBERTa notably outperform the dictionary methods.

From BERT to FinBERT through Fine-tuning

BERT is among the first generation of LLMs built with the Transformer architecture, which implements a self-attention mechanism that considers both left and right contexts for each word in a text sequence, as shown in Figure 2 (a).

Before training BERT on real text data, it has no language knowledge. The pre-training step is to let BERT encode the general language pattern in a large corpus, e.g., Wikipedia, BookCorpus, as illustrated by Figure 2 (a).

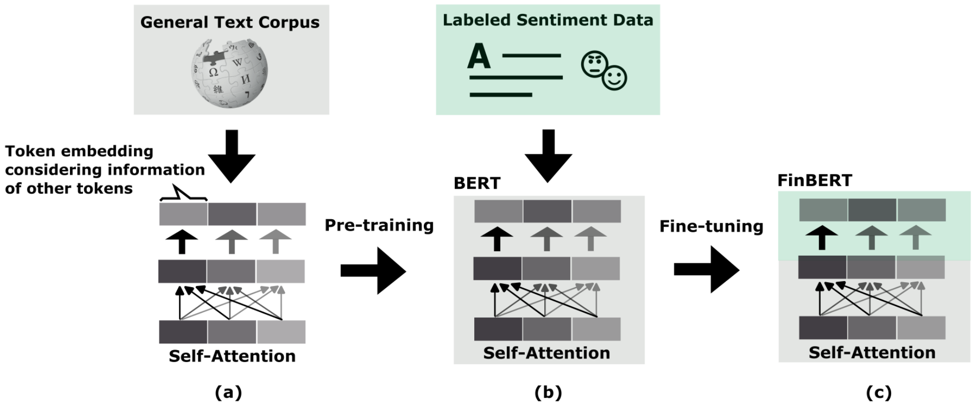

Figure 2: Pre-training and fine-tuning Paradigm. The small gray blocks represent individual tokens of a text sequence. (a) the self-attention first computes pair-wise similarities of tokens and then represents each token with a similarity-weighted combination of all tokens' embeddings. (b) after the pre-training, BERT encodes the general language knowledge. Then, the labelled sentiment data is used to fine-tune BERT. (c) after fine-tuning, financial sentiment analysis obtains the tailored model FinBERT.

However, the pre-trained BERT might be sub-optimal for specific tasks, such as financial sentiment analysis. This is because some domain-specific vocabulary is less representative in the general corpus,[7] as financial terms like 'dividend yield,' 'earnings per share,' or 'price-to-earnings ratio,' etc. As a result, the pre-trained BERT might be less capable of determining these terms' sentiments.

Figure 2 (b) shows the fine-tuning phase typically uses a small and labelled sentiment dataset, such as the financial PhraseBank, to adapt BERT to the sentiment analysis.[4] The parameters of the pre-trained BERT are marginally updated with better-fitting financial text and associated sentiment. The outcome is the fine-tuned BERT named FinBERT in Figure 1 (c), which contains both knowledge of general languages and financial text sentiments.

From FinBERT to FinDeBERTa through Disentangled Representations

Enabling disentangled representations in the self-attention mechanism can help improve BERT-like models in the literature. We take DeBERTa as an example to explain this idea.[9,11]

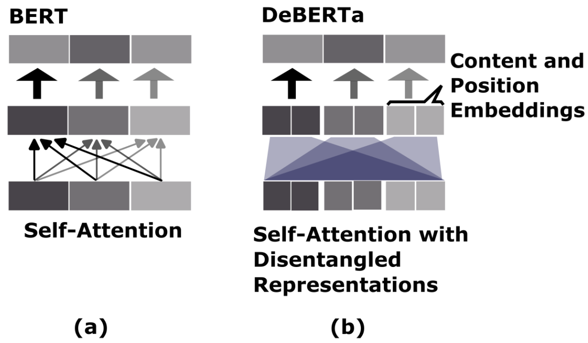

Figure 3: Disentangled representations in DeBERTa. (a) BERT self-attention mechanism. (b) the self-attention in DeBERTa uses separate content and position embeddings to represent one token. Unlike BERT in (a), the pair-wise similarity calculation is based both on the content and position embeddings.

As shown in Figure 3 (b), the disentangled self-attention in DeBERTa has two representations that account for each token's positional and content information, respectively. This is motivated by the observation that the meaning of words or phrases depends not only on the content but also on the position. For example, the words 'dividend' and 'yield' refer to the financial ratio (dividend/price) if they are adjacent while having a different meaning if they are apart. Then, similarly, through the pre-training and fine-tuning paradigm, we can build FinDeBERTa for financial sentiment tasks.

Results on News Flow

In this part, we report the backtest results to demonstrate the predictive power of the sentiment factors for the stock forward return. To compare beyond FinBERT and FinDeBERTa, we add a baseline, FinVader, a dictionary-based method adapted to add a financial sentiment lexicon.[12] The sentiment factor is constructed by applying these sentiment models to financial news (provided by StreetAccount) and then aggregating the sentiment values as the factors of the corresponding stocks.

Europe Coverage	Europe #news	North America Coverage	North America #news	Global Coverage	Global #news
56%	5.9million	68%	17.6million	52%	26.5million

Table 2: Statistics of News Flow Data in Different Universes.
The coverage refers to the percentage of stocks in the universe that have associated news.

Total Number	Positive Instances	Negative Instances	Neutral Instances
14780	3988	1841	8951

Table 3: Statistics of Labelled Financial Sentiment Data PhraseBank

Backtest Results

The backtest is performed on three different investment universes: Europe (Figure 4), North America (Figure 5) and Global, including both developed and emerging markets (Figure 6). The rebalancing frequency is monthly. Details of the news flow and labelled financial sentiment data are given in Table 2 and 3.

The cumulative performance charts below show that FinDeBERTa mostly outperforms FinVader and FinBERT. When comparing the charts of the all cap segment to those of large and small and mid (SMID) caps, it shows that the all cap performance is highly attributed to the SMID cap, suggesting that SMID companies might be more affected beyond intraday movements by news events. Large cap stocks tend to be more stable on this monthly horizon and the impact of news is likely to be evidenced at higher frequencies, e.g., intraday patterns. We leave this for future work.

As shown in the decile return bar plot in Figures 4, 5 and 6, the decile return of stocks is positively correlated to the factor value. Therefore, this shows a certain monotonicity which is favourable for quantitative investing.

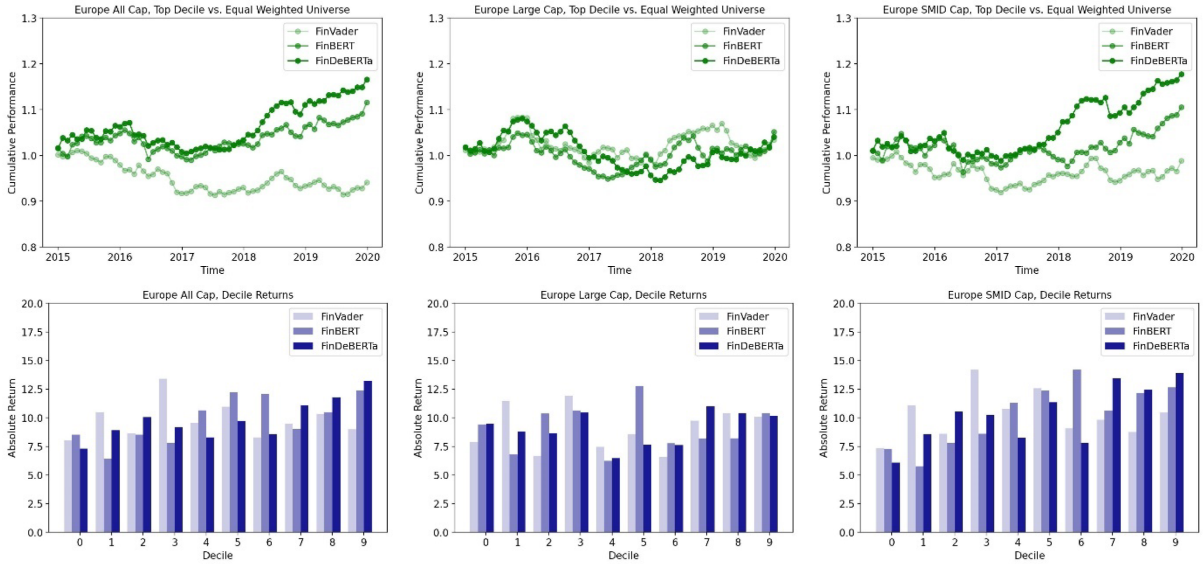

Figure 4: Europe Investment Universe. Top row: the cumulative return of stocks in the top decile of the sentiment factor relative to the universe average of all cap, large cap, and small and mid (SMID) cap. Bottom row: the absolute return of the stocks falling in each decile of the sentiment factor.

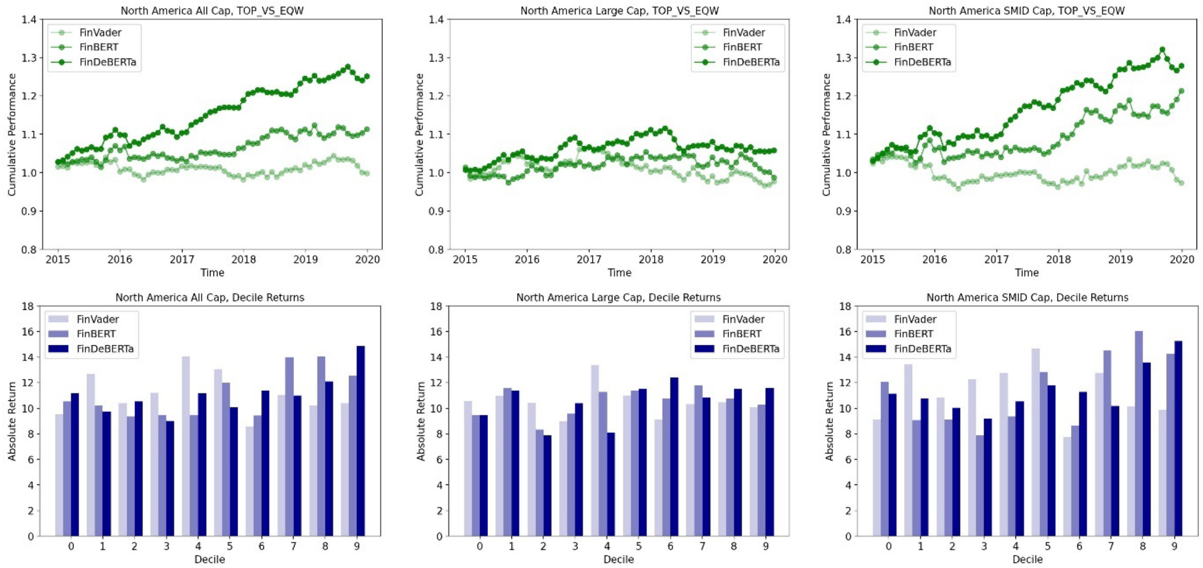

Figure 5: North America Investment Universe. Top row: the cumulative return of stocks in the top decile of the sentiment factor relative to the universe average of all cap, large cap and small and mid (SMID) cap. Bottom row: the absolute return of the stocks falling in each decile of the sentiment factor.

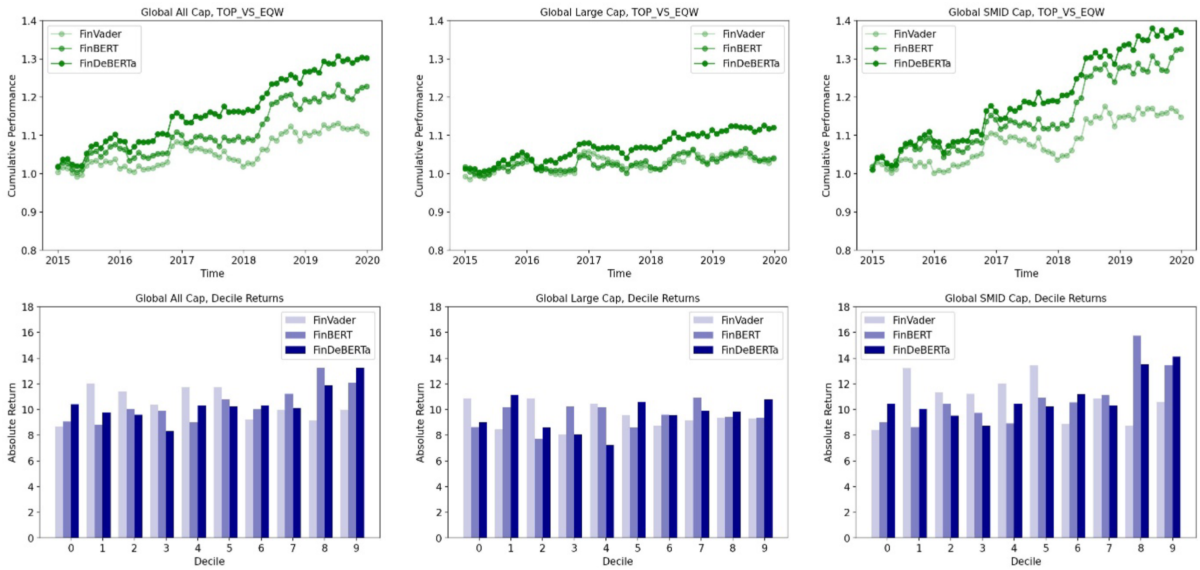

Figure 6: Global Investment Universe. Top row: the cumulative return of stocks in the top decile of the sentiment factor relative to the universe average of all cap, large cap, and small and mid (SMID) cap. Bottom row: the absolute return of the stocks falling in each decile of the sentiment factor.

Qualitative Analysis

Next, we provide a qualitative analysis using news examples to get a sense of these models' different capabilities.

Ground truth		FinVader	FinBERT	FinDeBERTa
Positive News	Company upgraded to hold from sell at … Investment Bank.	Neutral	Positive	Positive
	The earnings have increased by 10% in the last quarter.	Positive	Positive	Positive
	Company reports Q4 income CHF1.39B vs consensus CHF1.34B.	Neutral	Neutral	Positive
	Company reports FY EPS CHF2.89 vs consensus CHF1.96.	Neutral	Neutral	Positive
Negative News	Company downgraded to sell from hold at … Investment Bank.	Neutral	Negative	Negative
	The operating costs have increased by 10% in Q3.	Positive	Positive	Negative
	Company reports FY net income CHF1.66B vs consensus CHF1.89B.	Neutral	Neutral	Negative
	Company reports FY EPS $0.39 vs consensus $0.46.	Neutral	Neutral	Negative

Figure 7: News Examples and Sentiment Labels by Models. The green (or red) underlines indicate the determinative words of the positive (or negative) sentiments.

In Figure 7, we present two news groups: the positive and negative. Each group has four news pieces. The company or entity names are anonymised to avoid entity bias. The green (or red) underlines indicate the determinative words of the positive (or negative) sentiments.

In each group, the first two news items are relatively easy, as the underlined words' sentiment is evident to determine. The last two are challenging, as they require the sentiment model to recognise the finance-report-specific comparison pattern (e.g., income CHF 1.39B vs consensus CHF 1.34B).

In both groups, FinDeBERTa gives correct labels. FinBERT fails on the news, which includes the comparison pattern and the one where the words combined reverse the sentiment, i.e., 'operating costs' and 'increased.' This implies the FinBERT's context-understanding ability is still weak.

Conclusion

In this paper, we introduce the techniques to adapt LLMs to financial sentiment analysis tasks through the pre-training and fine-tuning paradigm. We applied two fine-tuned LLMs, FinBERT and FinDeBERTa to financial news flow and backtested the derived sentiment factors. The results demonstrate that the news sentiment captured by LLMs-based models has the desired predictive power for stock forward returns. Moreover, the effect of news sentiment varies across market segments, i.e., the sentiment of SMID cap stocks is more predictive for the forward return on the monthly rebalancing horizon, while large cap stocks tend to be more stable on this horizon. In our future work, we will be exploring the impact of news at higher rebalancing frequencies.

References

1. Mikolov, Tomas, et al. ‘Distributed representations of words and phrases and their compositionality.’ Advances in neural information processing systems 26 (2013).
2. Wang, Chuan-Ju, et al. ‘Financial sentiment analysis for risk prediction.’ Proceedings of the Sixth International Joint Conference on Natural Language Processing. (2013).
3. Hutto, Clayton, and Eric Gilbert. ‘Vader: A parsimonious rule-based model for sentiment analysis of social media text.’ Proceedings of the international AAAI conference on web and social media. Vol. 8. No. 1. (2014).
4. Malo, Pekka, et al. ‘Good debt or bad debt: Detecting semantic orientations in economic texts.’ Journal of the Association for Information Science and Technology 65.4 (2014): 782-796.
5. Vaswani, Ashish, et al. ‘Attention is all you need.’ Advances in neural information processing systems 30 (2017).
6. Devlin, Jacob, et al. ‘Bert: Pre-training of deep bidirectional transformers for language understanding.’ arXiv preprint arXiv:1810.04805 (2018).
7. Araci, Dogu. ‘Finbert: Financial sentiment analysis with pre-trained language models.’ arXiv preprint arXiv:1908.10063 (2019).
8. Guo, Tian, et al. ‘Esg2risk: A deep learning framework from esg news to stock volatility prediction.’ arXiv preprint arXiv:2005.02527 (2020).
9. He, Pengcheng, et al. ‘DeBERTa: Decoding-enhanced BERT with Disentangled Attention.’ International Conference on Learning Representations. (2020).
10. Guo, Tian, et al. “A deep learning framework for climate responsible investment.” https://www.ram-ai.com/sites/default/files/2021-02/a-deep-learning-framework-for-climate-responsible-investment.pdf (2021).
11. He, Pengcheng, Jianfeng Gao, and Weizhu Chen. ‘DeBERTaV3: Improving DeBERTa using ELECTRA-Style Pre-Training with Gradient-Disentangled Embedding Sharing.’ The Eleventh International Conference on Learning Representations (2023).
12. Petr, Korab. FinVADER: VADER sentiment classifier updated with financial lexicons. https://github.com/PetrKorab/FinVADER?tab=readme-ov-file (2023).
13. Zhang, Boyu, et al. ‘Enhancing financial sentiment analysis via retrieval augmented large language models.’ Proceedings of the Fourth ACM International Conference on AI in Finance. 2023.

Disclaimer

Important Information:

The information and analyses contained in this document are based on sources deemed to be reliable. However, RAM Active Investments S.A. cannot guarantee that said information and analyses are up-to-date, accurate or exhaustive, and accepts no liability for any loss or damage that may result from their use. All information and assessments are subject to change without notice.

This document has been drawn up for information purposes only. It is neither an offer nor an invitation to buy or sell the investment products mentioned herein and may not be interpreted as an investment advisory service. It is not intended to be distributed, published or used in a jurisdiction where such distribution, publication or use is forbidden, and is not intended for any person or entity to whom or to which it would be illegal to address such a document. In particular, the investment products are not offered for sale in the United States or its territories and possessions, nor to any US person (citizens or residents of the United States of America). The opinions expressed herein do not take into account each customer's individual situation, objectives or needs. Customers should form their own opinion about any security or financial instrument mentioned in this document. Prior to any transaction, customers should check whether it is suited to their personal situation, and analyse the specific risks incurred, especially financial, legal and tax risks, and consult professional advisers if necessary.

This document is strictly confidential and addressed solely to its intended recipient; its reproduction and distribution are prohibited. It has not been approved by any financial Authority. Issued in Switzerland by RAM Active Investments S.A. ((Rue du Rhône 8 CH-1204 Geneva)) which is authorised and regulated in Switzerland by the Swiss Financial Market Supervisory Authority (FINMA). Issued in the European Union and the EEA by the Management Company RAM Active Investments (Europe) S.A., 51 av. John F. Kennedy L-1855 Luxembourg, Grand Duchy of Luxembourg. Distributed in the United Kingdom to professional investors by RAM AI Advisory, a trading name of Sturgeon Ventures LLP utilised under exclusive license. Sturgeon Ventures LLP (FRN: 452811) is authorised and regulated by the Financial Conduct Authority (FCA).

No part of this document may be copied, stored electronically or transferred in any way, whether manually or electronically, without the prior agreement of RAM Active Investments S.A.

RAM Active Investments SA

Geneva

Rue du Rhône 8
1204 Geneva – SWITZERLAND

Tel : +41 58 726 87 00