Short Interest Trend Prediction with Large Language Models

Zhaomin Xiao
zhaominxiao@my.unt.edu
University of North Texas
USA

Yachen Cui
yachen.cui@unt.edu
University of North Texas
USA

Zhelu Mai
zhelumai@my.unt.edu
I-66 Express Mobility Partners
USA

Zhuoer Xu
patrick.xu.work@gmail.com
Hewlett Packard Enterprise
USA

Jiancheng Li
jianchengli@my.unt.edu
University of North Texas
USA

ABSTRACT

This paper studies the problem of short interest trend prediction using large language models. To do so, we provide a formal task definition and create a dataset for this task. We conduct extensive experiments with various types of large language models in different settings of in-context learning. Our results show that large language models have gained knowledge pertaining to short sale interest trends, and providing examples is beneficial.

CCS CONCEPTS

• **Computing methodologies** → **Information extraction**; • **Social and professional topics** → **Computing and business**.

KEYWORDS

Short Interest Trend, Natural Language Processing, Large Language Model

ACM Reference Format:
Zhaomin Xiao, Yachen Cui, Zhelu Mai, Zhuoer Xu, and Jiancheng Li. 2024. Short Interest Trend Prediction with Large Language Models. In *2024 the 8th International Conference on Innovation in Artificial Intelligence (ICIAI 2024), March 16–18, 2024, Tokyo, Japan.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3655497.3655500

News title	Label
Cobalt International Energy *Hits 52-Week High* of $20.68	increase
Polaris CEO *Bullish* About Sales to Military (PII, HOG)	increase
Gabelli & Co. *Downgrades* Lindsay Corporation to Hold	decrease
Credit Suisse *Maintains* Underperform on Patterson Companies, Inc., Raises PT to $45.00	decrease

Figure 1: An example of our dataset.

1 INTRODUCTION

Information is the most vital asset in capital market investment and trading. Traders make decisions based on the information they gather, especially when it comes to short selling. Unlike traditional trading, which involves buying low and selling high, short selling entails borrowing and selling stocks with the anticipation that their prices will decrease before repurchasing and returning the borrowed stocks. Short sellers' information collection and distribution significantly impact the real economy, stock market efficiency, financial decisions, and auditing [5, 17]. It is revealed that public information is extremely important to short sellers because they have advanced skills in processing public information. Other than traditional quantitative financial reporting, [18] short sellers prefer news and initiate more trading on such days. Knowing how this public information. Understanding how short sellers process information from the news is both interesting and crucial for research.

On a daily basis, the financial realm experiences a deluge of information, offering investors a wealth of data to navigate. From the venerable insights of the Wall Street Journal to the rapid-fire updates on social media from market experts, a multitude of sources inundate screens with analytical reports, algorithmic predictions, and real-time alerts. Recognizing the pivotal role of succinct communication, news organizations meticulously curate titles to encapsulate the essence of complex stories. These titles act as windows into the main points, enabling readers to glean quick and meaningful insights amid global finance's dynamic and swiftly evolving landscape without the need for in-depth exploration of entire articles.

In our dataset, words and phrases associated with a potential change in future stock prices include *Hit*, *Bullish*, *Downgrades*, and *Maintains* in the news titles. These terms are more likely to be linked to shifts in short interest. Figure 1 illustrates the information extracted from news titles for predicting short-interest trends. For example, *Hits 52-Week High* and *Bullish* explicitly indicate that the stock price will bump up in the future. Sophisticated short sellers perceive this information as a sign of market overreaction, leading to an overvaluation of the stocks. Consequently, they engage in short selling, causing a rise in short interest. This pattern allows us to observe an increase in short interest. Likewise, the *Downgrades* and *Maintains* directly indicate that the stock price will decrease in the future. Likewise, instead of betting on stock price will continue to go down, the sophisticated short sellers identify that as a good opportunity to cover the short position; therefore, short interest is likely to increase in such instances.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
ICIAI 2024, March 16–18, 2024, Tokyo, Japan
© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 979-8-4007-0930-2/24/03
https://doi.org/10.1145/3655497.3655500

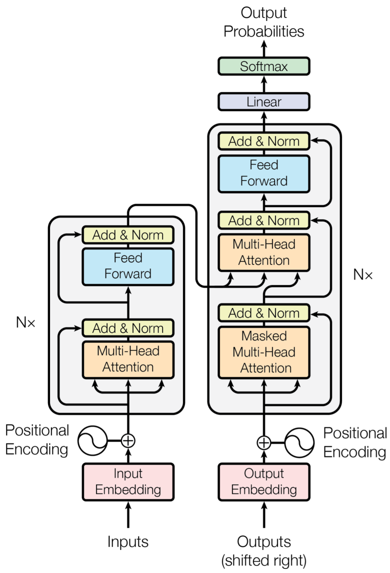

Figure 2: Architecture of the Transformer.

The research on short-interest trend prediction opens the door to many applications. One of them is in portfolio management, where accurate predictions of short interest trends can inform investment strategies and risk management. Financial institutions and hedge funds can leverage these predictions to optimize their portfolios, enhancing overall performance. Second, regulatory bodies may benefit from this research in monitoring and regulating short selling activities, contributing to the stability and integrity of financial markets. Third, traders and investors can use short interest trend predictions to make informed decisions, adjusting their positions based on anticipated market movements.

The main contributions of this paper are: 1) a novel task of short interest trend prediction, 2) a dataset for this task created by scraping the financial news related to public companies and their daily short sell data, 3) and experimental results with various LLMs in different settings of in-context learning showing that this challenging task can be automated by LLMs and providing examples before conducting inference is generally beneficial.

2 BACKGROUND

2.1 Short Interest Trend Prediction

Short-sellers are highly skilled and informed traders, and their activities affect stock market performance and efficiency[10]. Traditionally, finance researchers employ short interest, defined as *the number of shares shorted* over *total number of shares outstanding*, to measure the portion of shares to be shorted. The short interest serves as a predictor to predict future stock price underperformance ([11]). This phenomenon contributes to increased market efficiency and enhances the price discovery process for overvalued stocks. Detecting the change in short interest would help investors recognize the timely movement of short sellers and prepare for future investment and asset allocation.

Model input:

Review: Delicious food! Sentiment: Positive
Review: Terrible environment! Sentiment: Negative
Review: DIRTY PLACE!!! Sentiment: Negative
...
Review: Good meal and REALLY ENJOY THEM! Sentiment:

Model output:

Negative

Figure 3: An example of in-context learning.

2.2 Large Language Models

The objective of language model is to predict the next token based on the preceding tokens. In the pretraining phase, tokens following the current one are masked, and the model utilizes the prior tokens to predict the current token. Large language models (LLMs), are extremely scaled-up versions of original language models. They have at least 1 billion parameters and are typically pretrained on 0.1 trillion to 1 trillion tokens. LLMs can be categorized into two branches: one based on an encoder-decoder architecture and the other on a decoder-only architecture. In the context of LLMs, the encoder and decoder correspond to components in the Transformer architecture, as shown in 2. Each transformer block consists of multi-head self-attention layers, feed-forward networks, and skip connections. Representative works on encoder-decoder-based LLMs are T5 [9] and UL2 [14]. Representative works on decoder-based LLMs are LLaMA [15], LLaMA 2 [16], OPT [26], PaLM2 [1], and BLOOM [12]. Previous studies indicate that naive LLMs struggle to follow instructions and generate precise predictions. To address this, researchers have introduced instruction-tuning, a process where the parameters of naive LLMs are adjusted using carefully designed instructions for downstream tasks. FLAN is an influential work in instruction-tuning, featuring a collection of datasets consisting of 15M instances for 1,836 NLP tasks. After instruction-tuning on FLAN Collection [7], LLMs demonstrate their capability to comprehend instructions and produce accurate predictions.

2.3 In-Context Learning

In the context of LLM, in-context learning involves providing domain-specific or task-specific examples to LLMs before conducting inference [4]. Although it is not very clear, most researchers conclude that the improvement brought by in-context learning is from the ability to grasp the intricacies and nuances of particular subjects or industries. This adaptability ensures that the model remains relevant and accurate within specific contexts, as it can dynamically adjust its language generation patterns based on the input it receives. That said, if we provide a few examples before the question, LLMs can be "taught" to think about the problem more reasonably.

Figure 3 shows an example of in-context learning. To better elicit LLMs' ability to understand language, we can provide a few examples [1] before asking for the sentiment of the last review ("*Good meal and REALLY ENJOY THEM!*"). The presence of examples can "teach" LLMs to understand the review from the aspect of sentiment analysis.

[1]Note that the number of examples is typically determined empirically.

```
Read the news title below and determine how
the short interest of {company} changes.

{news title}

OPTIONS:
1. increase
2. decrease

ANSWER:
```

Figure 4: Our prompt. *company* is the name of the company we ask for the short interest trend.

3 DATASET CREATION

We choose to collect data from the Financial Industry Regulatory Authority (FINRA) [2] and Benzinga [3]. Specifically, we retrieve daily short sale data from FINRA, focusing solely on publicly traded companies listed on NASDAQ, NYSE, and AMEX. We work with these markets because they are US-based and significantly impact global trading markets. Our dataset spans from 2009 to 2020, excluding news published before 2009 due to scraping challenges and post-2020 trading data unavailability. For news titles, we employ an open web-scraping tool [4] to extract titles from Benzinga, excluding titles with over 50 tokens or any non-ASCII characters. Applying these filters could manage computational costs and align with most text processing APIs. Finally, our dataset comprises 136,192 news titles, each paired with the short interest trend of the associated company.

We create our dataset by matching the scraped news titles and daily short selling data. Specifically, we use the stock ticker of each publicly traded company and the timestamp of news as the primary keys to match the news title and the company's short selling data at the time when the news is published. We link the news title with the company's short selling data at the time of news publication, denoting the short interest on that day as S_t^c, where c and t represent the company and news timestamp, respectively. The short short interest trend of company c at time t is `increase` if $S_{t+1}^c - S_{t-1}^c > 0$. The short interest trend of company c at time t is `decrease` if $S_{t+1}^c - S_{t-1}^c < 0$. Note that we do not consider the case when $S_{t+1}^c - S_{t-1}^c = 0$ because it is very rare in our pilot study (less than 0.5% of all data).

4 DATASET ANALYSIS

54.7% of the 136,192 news titles in our dataset are labeled as `increase` and the remaining 45.3% labeled as `decrease`. We create stratified train and test splits (80% / 20%). The lengths of news titles in the training and test sets are similar (16.53 tokens vs. 16.48 tokens), but their vocabulary sizes differ significantly. The training data has 66,624 unique tokens, while the test data has 30,787 unique tokens. This difference in vocabulary usage allows the model to capture various aspects of financial knowledge.

5 EXPERIMENTS

We establish the majority label as the baseline. Among LLMs, we experiment with FLAN-T5 [2], FLAN-UL2 [13], and LLaMA 2 [16]. FLAN-T5 and FLAN-UL2 belong to the category of instruction-tuned encoder-decoder LLMs. Both models undergo instruction tuning on the FLAN Collection. As for LLaMA 2, we opt for LLaMA 2 Chat (7B), which is instruction-tuned on chat datasets specifically designed for dialogue-related tasks.

Figure 4 shows the prompt we use in our experiments with LLMs. We explicitly inform LLMs that we need them to determine how the short interest of the associated company changes. Following this, we provide the options from which we expect LLMs to select. Lastly, we use a flag ("ANSWER: ") to enable LLMs to make predictions by completing the prompts. We conduct experiments in different in-context settings by presenting 0/1/5/10 examples before providing our question. We do not experiment with more examples due to the maximum input lengths of these LLMs, ranging from 2k to 4k. Providing excessive examples would result in input lengths exceeding the models' limits, leading to truncation.

We use the 11B version of FLAN-T5 [5], which was released by Google. We use the 20B checkpoint of FLAN-UL2. [6] The LLaMA 2 Chat we use is released by Meta. [7] All LLMs are hosted by Huggingface [19]. To manage the computational cost, we set `truncation` as `True`, and the maximum input length is set as 2,048. [8] When loading the model, we use their 8-bit versions to control the GPU memory usage. Previous work shows that truncating the model parameter does not lead to performance degradation. [3] We conduct all experiments using Pytorch [8] and one NVIDIA GeForce RTX 3090.

6 RESULTS

Table 1 shows the weighted average F1 scores of LLMs in different settings of in-context learning. All LLMs outperform the majority baseline, indicating that this challenging task can be automated by LLMs. The best result (average F1: 0.47) is achieved by FLAN-UL2. The results reveal several insights.

- There exists a substantial disparity in performance between `increase` and `decrease`, suggesting that the financial knowledge acquired by LLMs during their pretraining process is imbalanced.
- Presenting more examples proves beneficial in most instances. For both FLAN-T5 and FLAN-UL2, the weighted average F1 scores show an improvement as the number of examples increases. Taking `increase` as an example, FLAN-T5's performance rises from 0.03 in the 0-shot setting to 0.31 in the 10-shot setting, and FLAN-UL2's performance increases from 0.18 in the 0-shot setting to 0.39 in the 10-shot setting. In contrast, LLaMA 2 does not gain from additional examples, as its F1 score drops from 0.17 in the 0-shot setting to 0.00 in the 10-shot setting. We attribute this to LLaMA 2's struggle with processing lengthy inputs effectively.
- Even the latest LLM, such as FLAN-UL2, struggles with this problem, achieving only an F1 score of 0.47. This contradicts the belief that LLMs can seamlessly handle any NLP task. To address this domain-specific challenge, we propose

[2]https://www.finra.org/finra-data/browse-catalog/short-sale-volume-data/daily-short-sale-volume-files
[3]https://www.benzinga.com/
[4]https://github.com/miguelaenlle/Scraping-Tools-Benzinga

[5]https://huggingface.co/google/flan-t5-xxl
[6]https://huggingface.co/google/flan-ul2
[7]https://huggingface.co/meta-llama/Llama-2-7b-chat-hf
[8]Recall that in the setting of 10-shot in-context learning, the model input could be very long.

	increase				decrease				Average			
	0-shot	1-shot	5-shot	10-shot	0-shot	1-shot	5-shot	10-shot	0-shot	1-shot	5-shot	10-shot
Majority	0.71				0.00				0.35			
FLAN-T5	0.03	0.23	0.30	0.31	0.62	0.59	0.60	0.59	0.30	0.40	0.38	0.44
LLaMA 2	0.17	0.06	0.00	0.00	0.40	0.17	0.14	0.31	0.28	0.08	0.10	0.39
FLAN-UL2	0.18	0.16	0.34	0.39	0.61	0.61	0.58	0.56	0.37	0.36	0.45	**0.47**

Table 1: Weighted average F1 scores of various LLMs in the different settings of in-context learning.

that LLMs should acquire more domain-specific knowledge during their pretraining phase. Furthermore, relying solely on pretraining LLMs for the task of next token prediction may be inadequate, as a comprehensive understanding of a company requires a broader foundation of commonsense knowledge in the finance domain.

7 CONCLUSIONS AND FUTURE WORKS

This paper introduces a new task—short interest trend prediction—and evaluates the performance of several large language models (LLMs) in this domain. To create the dataset, we collect financial news titles from Benzinga and daily short sell data from FINRA. We choose to work with FLAN-T5, FLAN-UL2, and LLaMA 2 since they represent both encoder-decoder and decoder-only architectures. We investigate the impact of in-context learning by varying the number of examples provided to each LLM. The findings reveal that LLMs can automate the task, providing more examples generally improves performance, and further refinements in pretraining are necessary to tackle our task.

In the future, we plan to investigate the role of instruction-finetuning in our task. Specifically, we plan to conduct experiments using Parameter-Effificient Finetuning Methods (PEFT), such as LoRA [6], to further finetune the pretrained LLMs so that they can learn more domain-specific knowledge during the finetuning process. Additionally, we also plan to investigate whether and how the information carried by different modalities and other contextual content contributes to our task. Incorporating more than a single modality has been proven useful in many tasks, such as spatial information extraction [20], sentiment analysis [25], named entity recognition [23], and corporate event prediction [22]. Regarding the contextual content, it is also used in [21] and [24]. But to the best of our knowledge, taking into account multiple modalities and contextual content in short interest trend prediction still remains raw.

REFERENCES

[1] Rohan Anil, Andrew M. Dai, Orhan Firat, Melvin Johnson, Dmitry Lepikhin, Alexandre Tachard Passos, Siamak Shakeri, Emanuel Taropa, Paige Bailey, Z. Chen, Eric Chu, J. Clark, Laurent El Shafey, Yanping Huang, Kathleen S. Meier-Hellstern, Gaurav Mishra, Erica Moreira, Mark Omernick, Kevin Robinson, Sebastian Ruder, Yi Tay, Kefan Xiao, Yuanzhong Xu, Yujing Zhang, Gustavo Hernandez Abrego, Junwhan Ahn, Jacob Austin, Paul Barham, Jan A. Botha, James Bradbury, Siddhartha Brahma, Kevin Michael Brooks, Michele Catasta, Yongzhou Cheng, Colin Cherry, Christopher A. Choquette-Choo, Aakanksha Chowdhery, C Crépy, Shachi Dave, Mostafa Dehghani, Sunipa Dev, Jacob Devlin, M. C. D'iaz, Nan Du, Ethan Dyer, Vladimir Feinberg, Fan Feng, Vlad Fienber, Markus Freitag, Xavier García, Sebastian Gehrmann, Lucas González, Guy Gur-Ari, Steven Hand, Hadi Hashemi, Le Hou, Joshua Howland, An Ren Hu, Jeffrey Hui, Jeremy Hurwitz, Michael Isard, Abe Ittycheriah, Matthew Jagielski, Wen Hao Jia, Kathleen Kenealy, Maxim Krikun, Sneha Kudugunta, Chang Lan, Katherine Lee, Benjamin Lee, Eric Li, Mu-Li Li, Wei Li, Yaguang Li, Jun Yu Li, Hyeontaek Lim, Han Lin, Zhong-Zhong Liu, Frederick Liu, Marcello Maggioni, Aroma Mahendru, Joshua Maynez, Vedant Misra, Maysam Moussalem, Zachary Nado, John Nham, Eric Ni, Andrew Nystrom, Alicia Parrish, Marie Pellat, Martin Polacek, Oleksandr Polozov, Reiner Pope, Siyuan Qiao, Emily Reif, Bryan Richter, Parker Riley, Alexandra Ros, Aurko Roy, Brennan Saeta, Rajkumar Samuel, Renee Marie Shelby, Ambrose Slone, Daniel Smilkov, David R. So, Daniela Sohn, Simon Tokumine, Dasha Valter, Vijay Vasudevan, Kiran Vodrahalli, Xuezhi Wang, Pidong Wang, Zirui Wang, Tao Wang, John Wieting, Yuhuai Wu, Ke Xu, Yunhan Xu, Lin Wu Xue, Pengcheng Yin, Jiahui Yu, Qiaoling Zhang, Steven Zheng, Ce Zheng, Wei Zhou, Denny Zhou, Slav Petrov, and Yonghui Wu. 2023. PaLM 2 Technical Report. *ArXiv* abs/2305.10403 (2023). https://api.semanticscholar.org/CorpusID:258740735

[2] Hyung Won Chung, Le Hou, Shayne Longpre, Barret Zoph, Yi Tay, William Fedus, Eric Li, Xuezhi Wang, Mostafa Dehghani, Siddhartha Brahma, Albert Webson, Shixiang Shane Gu, Zhuyun Dai, Mirac Suzgun, Xinyun Chen, Aakanksha Chowdhery, Sharan Narang, Gaurav Mishra, Adams Yu, Vincent Y. Zhao, Yanping Huang, Andrew M. Dai, Hongkun Yu, Slav Petrov, Ed H. Chi, Jeff Dean, Jacob Devlin, Adam Roberts, Denny Zhou, Quoc V. Le, and Jason Wei. 2022. Scaling Instruction-Finetuned Language Models. *CoRR* abs/2210.11416 (2022). https://doi.org/10.48550/ARXIV.2210.11416 arXiv:2210.11416

[3] Tim Dettmers, Mike Lewis, Younes Belkada, and Luke Zettlemoyer. 2022. LLM.int8(): 8-bit Matrix Multiplication for Transformers at Scale. *CoRR* abs/2208.07339 (2022). https://doi.org/10.48550/ARXIV.2208.07339 arXiv:2208.07339

[4] Qingxiu Dong, Lei Li, Damai Dai, Ce Zheng, Zhiyong Wu, Baobao Chang, Xu Sun, Jingjing Xu, Lei Li, and Zhifang Sui. 2023. A Survey on In-context Learning. arXiv:2301.00234 [cs.CL]

[5] Gustavo Grullon, Sébastien Michenaud, and James P Weston. 2015. The real effects of short-selling constraints. *The Review of Financial Studies* 28, 6 (2015), 1737–1767.

[6] Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, and Weizhu Chen. 2021. LoRA: Low-Rank Adaptation of Large Language Models. *CoRR* abs/2106.09685 (2021). arXiv:2106.09685 https://arxiv.org/abs/2106.09685

[7] Shayne Longpre, Le Hou, Tu Vu, Albert Webson, Hyung Won Chung, Yi Tay, Denny Zhou, Quoc V. Le, Barret Zoph, Jason Wei, and Adam Roberts. 2023. The Flan Collection: Designing Data and Methods for Effective Instruction Tuning. In *International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA (Proceedings of Machine Learning Research, Vol. 202)*, Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (Eds.). PMLR, 22631–22648. https://proceedings.mlr.press/v202/longpre23a.html

[8] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems.* Curran Associates Inc., Red Hook, NY, USA, Article 721, 12 pages.

[9] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *J. Mach. Learn. Res.* 21 (2020), 140:1–140:67. http://jmlr.org/papers/v21/20-074.html

[10] David Rapach, Matthew C Ringgenberg, and Guofu Zhou. 2016. Appendix to'Short Interest and Aggregate Stock Returns'. *Available at SSRN 2675298* (2016).

[11] David E Rapach, Matthew C Ringgenberg, and Guofu Zhou. 2016. Short interest and aggregate stock returns. *Journal of Financial Economics* 121, 1 (2016), 46–65.

[12] Teven Le Scao, Angela Fan, Christopher Akiki, Ellie Pavlick, Suzana Ili'c, Daniel Hesslow, Roman Castagn'e, Alexandra Sasha Luccioni, Franccois Yvon, Matthias

Gallé, Jonathan Tow, Alexander M. Rush, Stella Biderman, Albert Webson, Pawan Sasanka Ammanamanchi, Thomas Wang, Benoît Sagot, Niklas Muennighoff, Albert Villanova del Moral, Olatunji Ruwase, Rachel Bawden, Stas Bekman, Angelina McMillan-Major, Iz Beltagy, Huu Nguyen, Lucile Saulnier, Samson Tan, Pedro Ortiz Suarez, Victor Sanh, Hugo Laurenccon, Yacine Jernite, Julien Launay, Margaret Mitchell, Colin Raffel, Aaron Gokaslan, Adi Simhi, Aitor Soroa Etxabe, Alham Fikri Aji, Amit Alfassy, Anna Rogers, Ariel Kreisberg Nitzav, Canwen Xu, Chenghao Mou, Chris C. Emezue, Christopher Klamm, Colin Leong, Daniel Alexander van Strien, David Ifeoluwa Adelani, Dragomir R. Radev, Eduardo Gonz'alez Ponferrada, Efrat Levkovizh, Ethan Kim, Eyal Natan, Francesco De Toni, Gérard Dupont, Germán Kruszewski, Giada Pistilli, Hady ElSahar, Hamza Benyamina, Hieu Trung Tran, Ian Yu, Idris Abdulmumin, Isaac Johnson, Itziar Gonzalez-Dios, Javier de la Rosa, Jenny Chim, Jesse Dodge, Jian Zhu, Jonathan Chang, Jorg Frohberg, Josephine L. Tobing, Joydeep Bhattacharjee, Khalid Almubarak, Kimbo Chen, Kyle Lo, Leandro von Werra, Leon Weber, Long Phan, Loubna Ben Allal, Ludovic Tanguy, Manan Dey, Manuel Romero Muñoz, Maraim Masoud, María Grandury, Mario vSavsko, Max Huang, Maximin Coavoux, Mayank Singh, Mike Tian-Jian Jiang, Minh Chien Vu, Mohammad Ali Jauhar, Mustafa Ghaleb, Nishant Subramani, Nora Kassner, Nurulaqilla Khamis, Olivier Nguyen, Omar Espejel, Ona de Gibert, Paulo Villegas, Peter Henderson, Pierre Colombo, Priscilla Amuok, Quentin Lhoest, Rheza Harliman, Rishi Bommasani, Roberto L'opez, Rui Ribeiro, Salomey Osei, Sampo Pyysalo, Sebastian Nagel, Shamik Bose, Shamsuddeen Hassan Muhammad, Shanya Sharma Sharma, S. Longpre, So maieh Nikpoor, S. Silberberg, Suhas Pai, Sydney Zink, Tiago Timponi Torrent, Timo Schick, Tristan Thrush, Valentin Danchev, Vassilina Nikoulina, Veronika Laippala, Violette Lepercq, Vrinda Prabhu, Zaid Alyafeai, Zeerak Talat, Arun Raja, Benjamin Heinzerling, Chenglei Si, Elizabeth Salesky, Sabrina J. Mielke, Wilson Y. Lee, Abheesht Sharma, Andrea Santilli, Antoine Chaffin, Arnaud Stiegler, Debajyoti Datta, Eliza Szczechla, Gunjan Chhablani, Han Wang, Harshit Pandey, Hendrik Strobelt, Jason Alan Fries, Jos Rozen, Leo Gao, Lintang Sutawika, M Saiful Bari, Maged S. Al-Shaibani, Matteo Manica, Nihal V. Nayak, Ryan Teehan, Samuel Albanie, Sheng Shen, Srulik Ben-David, Stephen H. Bach, Taewoon Kim, Tali Bers, Thibault Févry, Trishala Neeraj, Urmish Thakker, Vikas Raunak, Xiang Tang, Zheng-Xin Yong, Zhiqing Sun, Shaked Brody, Y Uri, Hadar Tojarieh, Adam Roberts, Hyung Won Chung, Jaesung Tae, Jason Phang, Ofir Press, Conglong Li, Deepak Narayanan, Hatim Bourfoune, Jared Casper, Jeff Rasley, Max Ryabinin, Mayank Mishra, Minjia Zhang, Mohammad Shoeybi, Myriam Peyrounette, Nicolas Patry, Nouamane Tazi, Omar Sanseviero, Patrick von Platen, Pierre Cornette, Pierre Franccois Lavall'ee, Rémi Lacroix, Samyam Rajbhandari, Sanchit Gandhi, Shaden Smith, Stéphane Requena, Suraj Patil, Tim Dettmers, Ahmed Baruwa, Amanpreet Singh, Anastasia Cheveleva, Anne-Laure Ligozat, Arjun Subramonian, Aur'elie N'ev'eol, Charles Lovering, Daniel H Garrette, Deepak R. Tunuguntla, Ehud Reiter, Ekaterina Taktasheva, Ekaterina Voloshina, Eli Bogdanov, Genta Indra Winata, Hailey Schoelkopf, Jan-Christoph Kalo, Jekaterina Novikova, Jessica Zosa Forde, Xiangru Tang, Jungo Kasai, Ken Kawamura, Liam Hazan, Marine Carpuat, Miruna Clinciu, Najoung Kim, Newton Cheng, Oleg Serikov, Omer Antverg, Oskar van der Wal, Rui Zhang, Ruochen Zhang, Sebastian Gehrmann, Shachar Mirkin, S. Osher Pais, Tatiana Shavrina, Thomas Scialom, Tian Yun, Tomasz Limisiewicz, Verena Rieser, Vitaly Protasov, Vladislav Mikhailov, Yada Pruksachatkun, Yonatan Belinkov, Zachary Bamberger, Zdenvek Kasner, Zdeněk Kasner, Amanda Pestana, Amir Feizpour, Ammar Khan, Amy Faranak, Ananda Santa Rosa Santos, Anthony Hevia, Antigona Unldreaj, Arash Aghagol, Arezoo Abdollahi, Aycha Tammour, Azadeh HajiHosseini, Bahareh Behroozi, Benjamin Ayoade Ajibade, Bharat Kumar Saxena, Carlos Muñoz Ferrandis, Danish Contractor, David M. Lansky, Davis David, Douwe Kiela, Duong Anh Nguyen, Edward Tan, Emi Baylor, Ezinwanne Ozoani, Fatim Tahirah Mirza, Frankline Ononiwu, Habib Rezanejad, H.A. Jones, Indrani Bhattacharya, Irene Solaiman, Irina Sedenko, Isar Nejadgholi, Jan Passmore, Joshua Seltzer, Julio Bonis Sanz, Karen Fort, Lívia Dutra, Mairon Samagaio, Maraim Elbadri, Margot Mieskes, Marissa Gerchick, Martha Akinlolu, Michael McKenna, Mike Qiu, Muhammed Ghauri, Mykola Burynok, Nafis Abrar, Nazneen Rajani, Nour Elkott, Nourhan Fahmy, Olanrewaju Samuel, Ran An, R. P. Kromann, Ryan Hao, Samira Alizadeh, Sarmad Shubber, Silas L. Wang, Sourav Roy, Sylvain Viguier, Thanh-Cong Le, Tobi Oyebade, Trieu Nguyen Hai Le, Yoyo Yang, Zachary Kyle Nguyen, Abhinav Ramesh Kashyap, Alfredo Palasciano, Alison Callahan, Anima Shukla, Antonio Miranda-Escalada, Ayush Kumar Singh, Benjamin Beilharz, Bo Wang, Caio Matheus Fonseca de Brito, Chenxi Zhou, Chirag Jain, Chuxin Xu, Clémentine Fourrier, Daniel Le'on Perin'an, Daniel Molano, Dian Yu, Enrique Manjavacas, Fabio Barth, Florian Fuhrimann, Gabriel Altay, Giyaseddin Bayrak, Gully Burns, Helena U. Vrabec, Iman I.B. Bello, Isha Dash, Ji Soo Kang, John Giorgi, Jonas Golde, Jose David Posada, Karthi Sivaraman, Lokesh Bulchandani, Lu Liu, Luisa Shinzato, Madeleine Hahn de Bykhovetz, Maiko Takeuchi, Marc Pàmies, María Andrea Castillo, Marianna Nezhurina, Mario Sanger, Matthias Samwald, Michael Cullan, Michael Weinberg, M Wolf, Mina Mihaljcic, Minna Liu, Moritz Freidank, Myungsun Kang, Natasha Seelam, Nathan Dahlberg, Nicholas Michio Broad, Nikolaus Muellner, Pascale Fung, Patricia Haller, R. Chandrasekhar, Renata Eisenberg, Robert Martin, Rodrigo L. Canalli, Rosaline Su, Ruisi Su, Samuel Cahyawijaya, Samuele Garda, Shlok S Deshmukh, Shubhanshu Mishra, Sid Kiblawi, Simon Ott, Sinee Sang-aroonsiri, Srishti Kumar, Stefan Schweter, Sushil Pratap Bharati, Tanmay Laud, Théo Gigant, Tomoya Kainuma, Wojciech Kusa, Yanis Labrak, Yashasvi Bajaj, Y. Venkatraman, Yifan Xu, Ying Xu, Yu Xu, Zhee Xao Tan, Zhongli Xie, Zifan Ye, Mathilde Bras, Younes Belkada, and Thomas Wolf. 2022. BLOOM: A 176B-Parameter Open-Access Multilingual Language Model. *ArXiv* abs/2211.05100 (2022). https://api.semanticscholar.org/CorpusID:253420279

[13] Yi Tay. 2023. A new open source Flan 20B with UL2. https://www.yitay.net/blog/flan-ul2-20b

[14] Yi Tay, Mostafa Dehghani, Vinh Q. Tran, Xavier Garcia, Jason Wei, Xuezhi Wang, Hyung Won Chung, Dara Bahri, Tal Schuster, Huaixiu Steven Zheng, Denny Zhou, Neil Houlsby, and Donald Metzler. 2023. UL2: Unifying Language Learning Paradigms. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net. https://openreview.net/pdf?id=6ruVLB727MC

[15] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurélien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023. LLaMA: Open and Efficient Foundation Language Models. *CoRR* abs/2302.13971 (2023). https://doi.org/10.48550/ARXIV.2302.13971 arXiv:2302.13971

[16] Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, Dan Bikel, Lukas Blecher, Cristian Canton-Ferrer, Moya Chen, Guillem Cucurull, David Esiobu, Jude Fernandes, Jeremy Fu, Wenyin Fu, Brian Fuller, Cynthia Gao, Vedanuj Goswami, Naman Goyal, Anthony Hartshorn, Saghar Hosseini, Rui Hou, Hakan Inan, Marcin Kardas, Viktor Kerkez, Madian Khabsa, Isabel Kloumann, Artem Korenev, Punit Singh Koura, Marie-Anne Lachaux, Thibaut Lavril, Jenya Lee, Diana Liskovich, Yinghai Lu, Yuning Mao, Xavier Martinet, Todor Mihaylov, Pushkar Mishra, Igor Molybog, Yixin Nie, Andrew Poulton, Jeremy Reizenstein, Rashi Rungta, Kalyan Saladi, Alan Schelten, Ruan Silva, Eric Michael Smith, Ranjan Subramanian, Xiaoqing Ellen Tan, Binh Tang, Ross Taylor, Adina Williams, Jian Xiang Kuan, Puxin Xu, Zheng Yan, Iliyan Zarov, Yuchen Zhang, Angela Fan, Melanie Kambadur, Sharan Narang, Aurélien Rodriguez, Robert Stojnic, Sergey Edunov, and Thomas Scialom. 2023. Llama 2: Open Foundation and Fine-Tuned Chat Models. *CoRR* abs/2307.09288 (2023). https://doi.org/10.48550/ARXIV.2307.09288 arXiv:2307.09288

[17] David Tranfield, David Denyer, and Palminder Smart. 2003. Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British journal of management* 14, 3 (2003), 207–222.

[18] Bastian Von Beschwitz, Oleg Chuprinin, and Massimo Massa. 2017. Why do short sellers like qualitative news? *Journal of Financial and Quantitative Analysis* 52, 2 (2017), 645–675.

[19] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, and Jamie Brew. 2019. HuggingFace's Transformers: State-of-the-art Natural Language Processing. *CoRR* abs/1910.03771 (2019). arXiv:1910.03771 http://arxiv.org/abs/1910.03771

[20] Zhaomin Xiao and Eduardo Blanco. 2022. Are People Located in the Places They Mention in Their Tweets? A Multimodal Approach. In *Proceedings of the 29th International Conference on Computational Linguistics*, Nicoletta Calzolari, Chu-Ren Huang, Hansaem Kim, James Pustejovsky, Leo Wanner, Key-Sun Choi, Pum-Mo Ryu, Hsin-Hsi Chen, Lucia Donatelli, Heng Ji, Sadao Kurohashi, Patrizia Paggio, Nianwen Xue, Seokhwan Kim, Younggyun Hahm, Zhong He, Tony Kyungil Lee, Enrico Santus, Francis Bond, and Seung-Hoon Na (Eds.). International Committee on Computational Linguistics, Gyeongju, Republic of Korea, 2561–2571. https://aclanthology.org/2022.coling-1.226

[21] Zhaomin Xiao, Yan Huang, and Eduardo Blanco. 2023. Context Helps Determine Spatial Knowledge from Tweets. In *Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, Nusa Dua, Bali, 149–160. https://aclanthology.org/2023.findings-ijcnlp.13

[22] Zhaomin Xiao, Zhelu Mai, Zhuoer Xu, Yachen Cui, and Jiancheng Li. 2023. Corporate Event Predictions Using Large Language Models. In *2023 10th International Conference on Soft Computing & Machine Intelligence (ISCMI)*. 193–197. https://doi.org/10.1109/ISCMI59957.2023.10458651

[23] Jianfei Yu, Jing Jiang, Li Yang, and Rui Xia. 2020. Improving Multimodal Named Entity Recognition via Entity Span Detection with Unified Multimodal Transformer. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault (Eds.). Association for Computational Linguistics, Online, 3342–3352. https://doi.org/10.18653/v1/2020.acl-main.306

[24] Xinchen Yu, Eduardo Blanco, and Lingzi Hong. 2022. Hate Speech and Counter Speech Detection: Conversational Context Does Matter. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Marine Carpuat, Marie-Catherine de Marneffe, and Ivan Vladimir Meza Ruiz (Eds.). Association for Computational Linguistics, Seattle, United States, 5918–5930. https://doi.org/10.18653/v1/2022.

naacl-main.433
[25] Amir Zadeh, Minghai Chen, Soujanya Poria, Erik Cambria, and Louis-Philippe Morency. 2017. Tensor Fusion Network for Multimodal Sentiment Analysis. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, Martha Palmer, Rebecca Hwa, and Sebastian Riedel (Eds.). Association for Computational Linguistics, Copenhagen, Denmark, 1103–1114. https://doi.org/10.18653/v1/D17-1115
[26] Susan Zhang, Stephen Roller, Naman Goyal, Mikel Artetxe, Moya Chen, Shuohui Chen, Christopher Dewan, Mona T. Diab, Xian Li, Xi Victoria Lin, Todor Mihaylov, Myle Ott, Sam Shleifer, Kurt Shuster, Daniel Simig, Punit Singh Koura, Anjali Sridhar, Tianlu Wang, and Luke Zettlemoyer. 2022. OPT: Open Pre-trained Transformer Language Models. *ArXiv* abs/2205.01068 (2022). https://api.semanticscholar.org/CorpusID:248496292