



Robust NLP for Finance (RobustFin)

Sameena Shah
JPMorgan AI Research
New York, USA
sameena.shah@jpmchase.com

Xiaodan Zhu
Queen's University
Kingston, Canada
xiaodan.zhu@queensu.ca

Gerard de Melo
HPI / University of Potsdam
Potsdam, Germany
gerard.demelo@hpi.de

Armineh Nourbakhsh
JPMorgan AI Research
New York, USA
armineh.nourbakhsh@jpmchase.com

Xiaomo Liu
JPMorgan AI Research
New York, USA
xiaomo.liu@jpmchase.com

Zhiqiang Ma
JPMorgan AI Research
New York, USA
zhiqiang.ma@jpmchase.com

Charese Smiley
JPMorgan AI Research
Chicago, USA
charese.h.smiley@jpmchase.com

Zhiyu Chen
Meta
Redmond, USA
zhiyuchen@meta.com

ABSTRACT

Natural language processing (NLP) technologies have been widely applied in business domains such as e-commerce and customer service, but their adoption in the financial sector has been constrained by industry-specific performance standards and regulatory restrictions. This challenge has created new opportunities for core research in related areas. Recent advancements in NLP, such as the advent of large language models, has encouraged adoption in the finance sector. However, compared to other domains, finance has stricter requirements for robustness, explainability, and generalizability. Given this background, we propose to organize the first Robust NLP for Finance (RobustFin) workshop at KDD '23 to encourage the study of and research on robustness and explainability technologies with regard to financial NLP. The goal of the workshop is to extend the applications of NLP in finance, while motivating further research in robust NLP.

CCS CONCEPTS

- **Computing methodologies** → **Natural language processing**;
- **Information systems** → *Enterprise information systems*.

KEYWORDS

model robustness, natural language processing, financial services

ACM Reference Format:

Sameena Shah, Xiaodan Zhu, Gerard de Melo, Armineh Nourbakhsh, Xiaomo Liu, Zhiqiang Ma, Charese Smiley, and Zhiyu Chen. 2023. Robust NLP for Finance (RobustFin). In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '23)*, August 6–10, 2023, Long Beach, CA, USA. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3580305.3599211>

Permission to make digital or hard copies of part or all 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 third-party components of this work must be honored. For all other uses, contact the owner/author(s).
KDD '23, August 6–10, 2023, Long Beach, CA, USA
© 2023 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0103-0/23/08.
<https://doi.org/10.1145/3580305.3599211>

1 WORKSHOP MOTIVATIONS AND TOPICS

Financial services is a document-rich industry. This has led to great demand for advanced NLP techniques to process financial corpora in order to curate knowledge, augment human intelligence, and support industry professionals in making informed business decisions [1, 6]. Recent advances in the NLP domain have motivated the emergence of key applications such as the automation of due diligence protocols, financial question answering (QA) and reasoning, conversational AI in finance, summarizing business entity earnings calls, and analyzing investment sentiment of social media posts [8]. Due to the unique industry standards and regulations [5], robustness is a fundamental consideration underlying nearly all financial NLP applications. The concept of *robustness* may connote many specific definitions and real-life considerations. In finance, the key concerns of robustness refer to better generalizability, higher explainability, lower risk of catastrophic failure in unforeseen circumstances, and diligent fairness considerations. Common NLP models like large language models trained on generic corpora usually deteriorate when processing financial texts despite having achieved state-of-the-art results in other applications [7, 10, 11]. The business and regulatory requirements for explainability and the reasoning behind decision makings for the financial industry are often uncompromising, while many NLP modeling techniques still lack this characteristic [3].

Based on years of research experience in this domain, the organizers recapitulate the key aspects of these challenges: massive volume of data, large variation in the data format, low signal-to-noise ratio, scarcity of expert annotated datasets, task ambiguity, challenges regarding data integrity and privacy [4], domain shift [9], and high performance requirements set by industry and regulatory standards. With that, we envision a multi-year workshop series aiming to advance the state of the art for robustness and explainability technologies with regard to financial NLP. While these challenges have been studied in many generic contexts, we believe working on this vertical high-stakes domain, i.e. finance, will help consolidate the development of solutions. The workshop will be composed of three components: invited talks, paper presentations and posters, along with a shared task competition. The workshop welcomes

original research submissions related to robustness and explainability research on NLP, with a focus on the financial domain and/or financial corpora, which will include, but not be limited to, the topics listed below. The workshop will welcome both position and regular research papers.

- Adversarial attacks and defense to improve NLP model robustness and privacy protection.
- Transfer learning applications and their robustness and explainability studies.
- Robustness research against concept drift, domain shift, and noisy datasets; NLP model generalization.
- Language modeling on financial corpora including tabular and numerical data, and multi-modal modeling; numeracy and quantitative reasoning, fact verification, and QA over tabular data.
- Reconciling structured and unstructured knowledge to improve model interpretability; financial knowledge graph construction.
- Few-shot learning and prompt methods.
- Synthetic or genuine financial textual datasets and benchmarks.
- Reasoning and textual entailment in financial applications.
- Empirical studies on robust NLP models.

2 SHARED TASK

There will be one shared task associated with the workshop, Conversational Finance Question Answering (ConvFinQA) [2]. We believe this representative task can encourage NLP researchers and practitioners to contemplate the unique requirements of the domain-specific tasks when proposing new solutions and using transfer learning for the finance domain.

Financial report understanding is a very common and fundamental exercise in finance. This exercise usually involves multiple steps such as quantity/metric extraction, symbolic/arithmetic manipulation, and multi-hop numerical reasoning. The recent advances in large pre-trained language models are far from satisfactory when addressing complex reasoning in financial reports. We propose the ConvFINQA shared task to the community. The task dataset contains 3,892 conversations comprising 14,115 questions carefully composed by domain annotators. More details about the dataset, contest tasks, and resources will be available on our workshop website and the leaderboard will be hosted on CodaLab. We invite participants to explore robust modeling approaches for this challenge.

3 WORKSHOP FORMAT AND ATTENDANCE

The workshop will include invited talks, paper presentations, and shared task demonstrations. Selected high-quality submissions will be presented in the workshop. Shared task contestants will be required to report their approaches and results, and top-ranked teams will be invited to present their work during the workshop.

The target audiences are mainly 1) academic researchers interested in applying relevant research to industrial problems, and 2) industrial practitioners from financial institutions such as applied research scientists within financial services companies and labs.

4 COMPARABLE WORKSHOPS

The organizers have organized three Knowledge Discovery workshops at AAAI. We believe robustness is vital to text analytics for financial applications, and it is significantly understudied, which has motivated us to organize this workshop at KDD. Another comparable workshop is Financial Technology and Natural Language Processing (FinNLP), held at IJCAI 2022. The scope of FinNLP is general applications of NLP to the FinTech domain. The NeurIPS 2018 Workshop on Challenges and Opportunities for AI in Financial Services and NeurIPS 2019 Workshop on Robust AI in Financial Services discussed four key aspects of AI in finance – fairness, explainability, accuracy, and privacy, while ours emphasizes the robustness research on financial NLP.

Disclaimer. This paper was prepared for informational purposes by the Artificial Intelligence Research group of JPMorgan Chase & Co. and its affiliates ("JP Morgan"), and is not a product of the Research Department of JP Morgan. JP Morgan makes no representation and warranty whatsoever and disclaims all liability, for the completeness, accuracy or reliability of the information contained herein. This document is not intended as investment research or investment advice, or a recommendation, offer or solicitation for the purchase or sale of any security, financial instrument, financial product or service, or to be used in any way for evaluating the merits of participating in any transaction, and shall not constitute a solicitation under any jurisdiction or to any person, if such solicitation under such jurisdiction or to such person would be unlawful.

REFERENCES

- [1] Longbing Cao. 2022. AI in Finance: Challenges, Techniques, and Opportunities. *ACM Comput. Surv.* 55, 3, Article 64 (feb 2022), 38 pages.
- [2] Zhiyu Chen, Shiyang Li, Charese Smiley, Zhiqiang Ma, Sameena Shah, and William Yang Wang. 2022. ConvFinQA: Exploring the Chain of Numerical Reasoning in Conversational Finance Question Answering. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*. 6279–6292.
- [3] Marina Danilevsky, Shipi Dhanorkar, Yunyao Li, Lucian Popa, Kun Qian, and Anbang Xu. 2021. Explainability for Natural Language Processing. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. 4033–4034.
- [4] Oluwaseyi Feyisetan, Borja Balle, Thomas Drake, and Tom Dieth. 2020. Privacy- and Utility-Preserving Textual Analysis via Calibrated Multivariate Perturbations. In *Proceedings of the 13th International Conference on Web Search and Data Mining (Houston, TX, USA) (WSDM '20)*. Association for Computing Machinery, New York, NY, USA, 178–186.
- [5] Eren Kurshan, Hongda Shen, and Jiahao Chen. 2021. Towards Self-Regulating AI: Challenges and Opportunities of AI Model Governance in Financial Services. In *Proceedings of the First ACM International Conference on AI in Finance (New York, New York) (ICAIF '20)*. Association for Computing Machinery, New York, NY, USA, Article 49, 8 pages. <https://doi.org/10.1145/3383455.3422564>
- [6] Pascal Muam Mah, Iwona Skalna, John Muzam, and Lilian Song. 2022. Analysis of Natural Language Processing in the FinTech Models of Mid-21st Century. *Journal of Information Technology and Digital World* 4 (09 2022), 183–211. <https://doi.org/10.36548/jitdw.2022.3.005>
- [7] Raj Shah, Kunal Chawla, Dheeraj Eidnani, Agam Shah, Wendi Du, Sudheer Chava, Natraj Raman, Charese Smiley, Jiaao Chen, and Diyi Yang. 2022. When FLUE Meets FLANG: Benchmarks and Large Pretrained Language Model for Financial Domain. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*. 2322–2335.
- [8] Narges Tabari, Piyusha Biswas, Bhanu Praneeth, Armin Seyeditabari, Mirsad Hadzikadic, and Wlodek Zadrozny. 2018. Causality Analysis of Twitter Sentiments and Stock Market Returns. In *Proceedings of the First Workshop on Economics and Natural Language Processing*. Association for Computational Linguistics, Melbourne, Australia, 11–19. <https://doi.org/10.18653/v1/W18-3102>
- [9] Tianlu Wang, Rohit Sridhar, Diyi Yang, and Xuezhi Wang. 2021. Identifying and Mitigating Spurious Correlations for Improving Robustness in NLP Models. <https://doi.org/10.48550/ARXIV.2110.07736>
- [10] Xuezhi Wang, Haohan Wang, and Diyi Yang. 2022. Measure and Improve Robustness in NLP Models: A Survey. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, Seattle, United States, 4569–4586. <https://doi.org/10.18653/v1/2022.naacl-main.339>
- [11] Yi Yang, Mark Christopher Siy UY, and Allen Huang. 2020. FinBERT: A Pretrained Language Model for Financial Communications. arXiv:2006.08097 [cs.CL]