

YIXUAN SU

Cohere, 51 Great Marlborough St, London W1F 7JT, UK

ys484@outlook.com ◇ (+44)07579873639 ◇ Personal Page ◇ Google Scholar ◇ Semantic Scholar ◇ Github

I am a senior research scientist and modelling lead on agentic reasoning at Cohere. I led the modelling development of Command-A-Reasoning, and also contributed to Command-A/R7B/R+/R.

EXPERIENCE

Cohere , London, United Kingdom Senior Research Scientist Modelling Lead, Agentic Reasoning	May 2023 - Present
Google DeepMind , London, United Kingdom Student Collaborator <i>Advised by Dani Yogatama</i>	Feb 2022 - June 2022
Amazon AWS AI , Seattle, United States Research Scientist Intern <i>Advised by Lei Shu, Yi Zhang</i>	May 2021 - Oct 2021
Apple Research , Cambridge, United Kingdom Research Scientist Intern <i>Advised by David Vandyke</i>	Sep 2020 - Dec 2020
Tencent AI Lab , Shenzhen, China Research Intern <i>Advised by Yan Wang</i>	Sep 2018 - Aug 2019
Microsoft Research Asia , Beijing, China Research Intern <i>Advised by Lei Ji</i>	Feb 2017 - July 2017

EDUCATION

University of Cambridge, United Kingdom PhD in Computation, Cognition and Language Language Technology Lab Advisor: Prof. Nigel Collier Thesis Committee: Prof. Andreas Vlachos, Prof. Mirella Lapata	Sep 2019 - June 2023
University of Cambridge, United Kingdom MPhil in Machine Learning, Speech and Language Technology Department of Engineering Advisor: Prof. Anna Korhonen and Dr. Simon Baker Adwarded with Distinction Degree (Top 5%)	Sep 2017 - Sep 2018
Beijing Institute of Technology, China Bachelor of Science Department of Engineering	Sep 2013 - July 2017

PUBLICATIONS

- [1] **Prompt Compression for Large Language Models: A Survey**
Zongqian Li, Yinhong Liu, Yixuan Su, Nigel Collier

In Proceedings of the North American Chapter of Association for Computational Linguistics (NAACL 2025).

- [2] **To code, or Not to Code? Exploring Impact of Code in Pre-training**
Viraat Aryabumi, [Yixuan Su](#), Raymond Ma, Adrien Morisot, Ivan Zhang, Acyr Locatelli, Marzieh Fadaee, Ahmet Üstün, Sara Hooker
In Proceedings of The Thirteenth International Conference on Learning Representations (ICLR 2025)
- [3] **Replacing Judges with Juries: Evaluating LLM Generations with a Panel of Diverse Models**
Pat Verga, Sebastian Hofstatter, Sophia Althammer, [Yixuan Su](#), Aleksandra Piktus, Arkady Arkhangorodsky, Minjie Xu, Naomi White, Patrick Lewis.
On Arxiv Preprint (arXiv:2404.18796)
- [4] **StarCoder 2 and The Stack v2: The Next Generation**
StarCoder Team.
On Arxiv Preprint (arXiv:2402.19173)
- [5] **Unlocking Structure Measuring: Introducing PDD, an Automatic Metric for Positional Discourse Coherence**
Yinhong Liu, [Yixuan Su](#), Ehsan Shareghi, Nigel Collier.
In Proceedings of the North American Chapter of Association for Computational Linguistics (NAACL 2024).
- [6] **Instruct-SCTG: Guiding Sequential Controlled Text Generation through Instructions**
Yinhong Liu, [Yixuan Su](#), Ehsan Shareghi, Nigel Collier.
On Arxiv Preprint (arXiv:2312.12299)
- [7] **Specialist or Generalist? Instruction Tuning for Specific NLP Tasks**
Chufan Shi, [Yixuan Su](#), Cheng Yang, Yujiu Yang, Deng Cai.
In Proceedings of Empirical Methods in Natural Language Processing (EMNLP 2023).
- [8] **Biomedical Named Entity Recognition via Dictionary-based Synonym Generalization**
Zihao Fu, [Yixuan Su](#), Zaiqiao Meng, Nigel Collier.
In Proceedings of Empirical Methods in Natural Language Processing (EMNLP 2023).
- [9] **Repetition In Repetition Out: Towards Understanding Neural Text Degeneration from the Data Perspective**
Huayang Li, Tian Lan, Zihao Fu, Deng Cai, Lemao Liu, Nigel Collier, Taro Watanabe, [Yixuan Su](#).
In Advances in Neural Information Processing Systems (NeurIPS 2023).
- [10] **PandaGPT: One Model To Instruction-Follow Them All**
[Yixuan Su](#), Tian Lan, Huayang Li, Jialu Xu, Yan Wang, Deng Cai.
In Proceedings of the 1st Workshop on Taming Large Language Models (TLLM) at SIGDIAL 2023
- [11] **OpenAlpaca: A Fully Open-Source Instruction-Following Model Based On OpenLLaMA**
[Yixuan Su](#), Tian Lan, Deng Cai.
- [12] **Contrastive Search Is What You Need For Neural Text Generation**
[Yixuan Su](#), Nigel Collier.
In Transactions on Machine Learning Research (TMLR 2023).
- [13] **Plug-and-Play Recipe Generation with Content Planning**
Yinhong Liu, [Yixuan Su](#), Ehsan Shareghi, Nigel Collier.
In Proceedings of the 2nd Workshop on Natural Language Generation, Evaluation, and Metrics (GEM 2022)

- [14] **Momentum Decoding: Open-ended Text Generation As Graph Exploration**
Tian Lan*, Yixuan Su*, Shuhang Liu, Heyan Huang, Xian-Ling Mao.
On Arxiv Preprint (arXiv:2212.02175)
- [15] **An Empirical Study On Contrastive Search And Contrastive Decoding For Open-ended Text Generation**
Yixuan Su, Jialu Xu.
On Arxiv Preprint (arXiv:2211.10797)
- [16] **Language Models Can See: Plugging Visual Controls in Text Generation**
Yixuan Su, Tian Lan, Yahui Liu, Fangyu Liu, Dani Yogatama, Yan Wang, Lingpeng Kong, Nigel Collier.
On Arxiv Preprint (arXiv:2205.02655)
- [17] **A Contrastive Framework for Neural Text Generation**
Yixuan Su, Tian Lan, Yan Wang, Dani Yogatama, Lingpeng Kong, Nigel Collier.
In Advances in Neural Information Processing Systems (NeurIPS 2022 **Spotlight**).
- [18] **Measuring and Reducing Model Update Regression in Structured Prediction for NLP**
Deng Cai, Elman Mansimov, Yi-An Lai, Yixuan Su, Lei Shu, Yi Zhang.
In Advances in Neural Information Processing Systems (NeurIPS 2022).
- [19] **From Easy to Hard: A Dual Curriculum Learning Framework for Context-Aware Document Ranking**
Yutao Zhu, Jian-Yun Nie, Yixuan Su, Haonan Chen, Xinyu Zhang, Zhicheng Dou.
In Proceedings of the Conference on Information and Knowledge Management (CIKM 2022).
- [20] **TaCL: Improving BERT Pre-training with Token-aware Contrastive Learning**
Yixuan Su, Fangyu Liu, Zaiqiao Meng, Tian Lan, Lei Shu, Ehsan Shareghi, Nigel Collier.
In Findings of the North American Chapter of Association for Computational Linguistics (NAACL 2022).
- [21] **Multi-Task Pre-Training for Plug-and-Play Task-Oriented Dialogue System**
Yixuan Su, Lei Shu, Elman Mansimov, Arshit Gupta, Deng Cai, Yi-An Lai, Yi Zhang.
In Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL 2022).
- [22] **Rewire-then-Probe: A Contrastive Recipe for Probing Biomedical Knowledge of Pre-trained Language Models**
Zaiqiao Meng, Fangyu Liu, Ehsan Shareghi, Yixuan Su, Charlotte Collins, Nigel Collier.
In Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL 2022).
- [23] **A Survey on Retrieval-Augmented Text Generation**
Huayang Li*, Yixuan Su*, Deng Cai*, Yan Wang*, Lemao Liu*.
On Arxiv Preprint (arXiv:2202.01110)
- [24] **Exploring Dense Retrieval for Dialogue Response Selection**
Tian Lan, Deng Cai, Yan Wang, Yixuan Su, Xian-Ling Mao, Heyan Huang.
On Arxiv Preprint (arXiv:2110.06612)
- [25] **Plan-then-Generate: Controlled Data-to-Text Generation via Planning**
Yixuan Su, David Vandyke, Sihui Wang, Yimai Fang, Nigel Collier.
In Findings of Empirical Methods in Natural Language Processing (EMNLP 2021).
- [26] **Few-Shot Table-to-Text Generation with Prototype Memory**
Yixuan Su, Zaiqiao Meng, Simon Baker, Nigel Collier.
In Findings of Empirical Methods in Natural Language Processing (EMNLP 2021).
- [27] **Dialogue Response Selection with Hierarchical Curriculum Learning**

Yixuan Su*, Deng Cai*, Qingyu Zhou, Zibo Lin, Simon Baker, Yunbo Cao, Shuming Shi, Nigel Collier, Yan Wang.

In Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL 2021).

- [28] **Keep the Primary, Rewrite the Secondary: A Two-Stage Approach for Paraphrase Generation**

Yixuan Su, David Vandyke, Simon Baker, Yan Wang, Nigel Collier.

In Findings of the Annual Meeting of the Association for Computational Linguistics (ACL 2021).

- [29] **Non-Autoregressive Text Generation with Pre-trained Language Models**

Yixuan Su, Deng Cai, Yan Wang, David Vandyke, Simon Baker, Piji Li, Nigel Collier.

In Proceedings of the European Chapter of the Association for Computational Linguistics (EACL 2021).

- [30] **Prototype-to-Style: Dialogue Generation with Style-Aware Editing on Retrieval Memory**

Yixuan Su, Yan Wang, Deng Cai, Simon Baker, Anna Korhonen, Nigel Collier.

In IEEE/ACM Transactions on Audio Speech and Language Processing (TASLP 2021).

- [31] **Stylistic Dialogue Generation via Information-Guided Reinforcement Learning**

Yixuan Su, Deng Cai, Yan Wang, Simon Baker, Anna Korhonen, Xiaojiang Liu.

On Arxiv Preprint (arXiv:2004.02202)

INVITED TALKS

- January 2023, IR Group, University of Glasgow
- October 2022, NLG Student Webinar, Chinese Information Processing Society of China
- August 2022, MLNLP Webinar
- August 2022, NLP Group, Princeton University
- June 2022, Language Technology Lab, University of Cambridge
- April 2022, NLP Group, University of Washington
- April 2022, Language Team, DeepMind, London, UK
- March 2022, NLP Group, University of Oxford
- March 2022, NLP Group, Nara Institute of Science and Technology
- February 2022, NLP Group, Tencent AI Lab, Shenzhen, China

PROFESSIONAL SERVICE

Program Committee/Reviewer:

- IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)
- The International Conference on Machine Learning (ICML)
- The Conference on Neural Information Processing Systems (NeurIPS)
- International Conference on Learning Representations (ICLR)
- The Annual Meeting of the Association for Computational Linguistics (ACL)
- The Conference on Empirical Methods in Natural Language Processing (EMNLP)
- The European Chapter of the Association for Computational Linguistics (EACL)
- ACL Rolling Review (ARR)