

# Cheng Tang

E-mail: tangch30@outlook.com    Homepage: <https://github.com/tangch30/tangch30.github.io>

## RESEARCH SHORT RECAP

---

During my PhD, I spent time studying unsupervised learning methods, including the k-means method and principal component analysis, from a theoretical perspective. At Amazon, I applied large language models to product-inspired problems in natural language processing and information retrieval.

## GAP YEARS

---

Due to complicated situation, I was away from academia and the job market for several years. I took this opportunity to catch up on research topics in ML theory and LLM's. I am currently especially interested in research at the intersection of ML theory and applications, including but not limited to:

- Domain adaptation, causal inference, and emerging learning paradigms such as meta learning, self learning, and multi-modal learning.
- LLM's, including its training and evaluation, and ways to improve its reasoning and comprehension capabilities.
- Revisiting topics such as deep matrix factorization and low-rank methods.
- Time-series and its applications in finance
- Legal AI

## EMPLOYMENT

---

**Postdoc**, (Incomplete for personal reasons)  
Anonymous Institution, 2021

**Applied Scientist II**, New York, NY USA  
AI Labs, Amazon Web Services, Jul. 2018 – Jan. 2021  
Managers: Bing Xiang & Andrew Arnold

- Experience: I explored neural embedding models, transformer-based models (BERT and BART), and their applications to domain adaptation, reading comprehension, recommendation, and ranking.

## RESEARCH INTERNSHIPS

---

**Research Intern**, Cambridge, UK  
Machine Intelligence group, Microsoft Research, Jun. 2017– Aug. 2017  
Mentors: Alex Gaunt & Ryota Tomioka, Jul., 2017– Sep., 2017

- Project: Explored the application of a neural-logical model to the problem of understanding the hidden rules that govern the gene expression process.

**Research Visitor**, Tübingen, Germany

Theory of Machine Learning group, University of Tübingen, Germany, Mar. 2017– Jun. 2017

Mentor: Ulrike von Luxburg, Mar., 2017– Jun., 2017

- Project: Theoretical exploration of the role of sub-sampling in random forests

## EDUCATION

---

**The George Washington University**, Washington, DC USA

Ph.D., Computer Science, May 2018

- Advisor: Claire Monteleoni

B.S., Mathematics, Magna Cum Laude, GPA 3.77/4.0, May 2012

## PREPRINT AND ONGOING WORKS

---

- C. Tang, D. Garreau, C. Monteleoni, “Multiple-source adaptation with the Hedge algorithm” (Preprint link: [https://tangch30.github.io/assets/files/msa\\_preprint.pdf](https://tangch30.github.io/assets/files/msa_preprint.pdf)), 2026.
- C. Tang, Foundation of auto-differentiation and computational graphs, in-preparation.
- C. Tang, On the stability of hyperparameters during data scaling, in-preparation.
- C. Tang, “On the tightness of linear relaxation based robustness certification methods”, 2022.
- C. Tang, Andrew Arnold, “Neural document expansion for ad-hoc information retrieval” (arXiv: <https://arxiv.org/pdf/2012.14005>), 2020.

## CONFERENCE PROCEEDINGS

---

- C. Tang, “Exponentially convergent stochastic k-PCA without variance reduction”, NeurIPS, 2019.
- C. Tang, D. Garreau, U. von Luxburg, “When do random forests fail?”, Proceedings of 32nd Conference on Neural Information Processing Systems (NeurIPS), 2018.
- C. Tang and C. Monteleoni, “Convergence rate of stochastic k-means”, AISTATS, 2017.
- C. Tang and C. Monteleoni, “On Lloyd’s algorithm: new theoretical insights for clustering in practice,” Proceedings of the 19th International Conference on Artificial Intelligence and Statistics (AISTATS), pp. 1280-1289, 2016.

## ACADEMIC REVIEWING EXPERIENCE

---

AISTATS (2017, 2019, 2020, 2021), ICML (2015, 2018, 2019, 2020), NeurIPS (2016, 2017, 2018), ICLR (2017, 2019, 2020, 2021), AAAI (2019, 2020, 2021), TPAMI (2019, 2020)

## SELECTED HONORS AND AWARDS

---

- Engineer Alumni Association Scholarship, SEAS, GWU, 2015– 2016
- Louis P. Wagman Endowment Fellowship, GWU, 2013
- Presidential Academic Scholarship, GWU, 2008– 2012
- Ranked 165/230,000, National College Entrance Exam (admitted to Fudan University), Sichuan province, China, 2008.

## **ACADEMIC TEACHING EXPERIENCE**

---

- Discrete Structures II, Graduate Teaching Assistant, Department of Computer Science, GWU, Fall, 2016
- Machine Learning, Undergraduate Teaching Assistant, Department of Computer Science, GWU, Spring 2012
- Math and Politics, Undergraduate Teaching Assistant, Department of Mathematics, GWU, Fall, 2011

## **COMPUTER LANGUAGES**

---

Python (including deep-learning frameworks such as TensorFlow/PyTorch/MXNet), MATLAB, Unix/Linux