# Yuntian DENG

Google Scholar◇ Personal Website
yuntian@uwaterloo.ca

## Appointment

**University of Waterloo** — Waterloo, Canada
Incoming Assistant Professor — Starting Fall 2024

**Vector Institute** — Toronto, Canada
Incoming Faculty Affiliate — Starting Fall 2024

**Allen Institute for AI** — Seattle, WA
Young Investigator (Postdoc) — Jul 2023

## Education

**Harvard University** — Cambridge, MA
Ph.D. in Computer Science — May 2023
Advisors: Alexander Rush and Stuart Shieber

**Carnegie Mellon University** — Pittsburgh, PA
Master in Language Technologies — Aug 2016
Advisor: Eric Xing

**Tsinghua University** — Beijing, China
Bachelor of Engineering, Department of Automation — Jul 2014

## Research Interests

My research interests center on the intersection of natural language processing, machine learning, and multi-agent systems. Specifically, I am interested in exploring how large language models (LLMs) can communicate and collaborate to solve complex tasks together, and how they can be trained to specialize in different domains for a division of labor. My key focus areas include:

- Inducing Latent Language for Inter-LLM Communication: Developing methods to induce a specialized language for LLM communication, thereby enabling LLMs to leverage each other s expertise.
- Communication for Models Across Modalities: Extending Inter-LLM communication methods to enable collaboration among models that specialize in different modalities, such as language, image, and sensory data.
- Collaborative Training for Division of Labor among Models: Exploring ways to foster a division of labor among models, using communication as a tool to distribute knowledge among them during the training process.

I also work on open-source projects such as OpenNMT, Im2LaTeX, LaTeX2Im, and Steganography to make my research efforts more readily available for developers and researchers.

## Awards

| | |
|---|---|
| Argonne National Lab Impact Award | Jan 2023 |
| University of Chicago Rising Stars in Data Science | Nov 2022 |
| ACM Gordon Bell Prize | Nov 2022 |
| **Nvidia Fellowship** (10 awardees) | Dec 2021 |
| Harvard Certificates of Distinction in Teaching | Spring 2019, Fall 2020, Fall 2021 |
| Twitch Fellowship Finalist | Jan 2021 |
| DAC 2020 Best Paper Award | Jul 2020 |
| **Baidu Fellowship** (10 awardees) | Feb 2019 |
| French American Doctoral Exchange Program (French Embassy, 10 US laureates) | Jul 2018 |
| ACL 2017 Best Demo Paper Award Runner-Up | Aug 2017 |
| Silver Medal in the 26th Chinese Physics Olympiad | Nov 2009 |

## Publications

*Equal Contribution.

### Selected Papers

26 **Yuntian Deng**, Kiran Prasad, Roland Fernandez, Paul Smolensky, Vishrav Chaudhary, Stuart Shieber. Implicit Chain of Thought Reasoning via Knowledge Distillation. *arXiv 2023*

25 John X. Morris, Chandan Singh, Alexander M. Rush, Jianfeng Gao, **Yuntian Deng**. Tree Prompting: Efficient Task Adaptation without Fine-Tuning. *EMNLP 2023*

24 **Yuntian Deng**, Volodymyr Kuleshov, Alexander Rush. Model Criticism for Long-Form Text Generation. *EMNLP 2022*

23 **Yuntian Deng**, Anton Bakhtin, Myle Ott, Arthur Szlam, Marc'Aurelio Ranzato. Residual Energy-Based Models for Text Generation. *ICLR 2020*

22 **Yuntian Deng**, Alexander Rush. Cascaded Text Generation with Markov Transformers. *NeurIPS 2020*

21 Sebastian Gehrmann, **Yuntian Deng**, Alexander M. Rush. Bottom-Up Abstractive Summarization. *EMNLP 2018*

20 Zachary Ziegler*, **Yuntian Deng***, Alexander Rush. Neural Linguistic Steganography. *EMNLP 2019* ***oral*** (**Demo**: steganography.live)

19 **Yuntian Deng**, Noriyuki Kojima, Alexander Rush. Markup-to-Image Diffusion Models with Scheduled Sampling. *ICLR 2023* (Demo: huggingface.co/spaces/yuntian-deng/latex2im)

18 **Yuntian Deng***, Yoon Kim*, Justin Chiu, Demi Guo, Alexander M. Rush. Latent Alignment and Variational Attention. *NeurIPS 2018*

17 **Yuntian Deng**, Anssi Kanervisto, Jeffrey Ling, Alexander M. Rush. Image-to-Markup Generation with Coarse-to-Fine Attention. *ICML 2017* (Demo: im2markup.yuntiandeng.com)

**Journal Papers**

16 Anton Bakhtin*, **Yuntian Deng***, Sam Gross, Myle Ott, Marc'Aurelio Ranzato, Arthur Szlam. Residual Energy-Based Models for Text. *JMLR 2021*

**Conference Papers**

15 Richa Rastogi, Yair Schiff, Alon Hacohen, Zhaozhi Li, Ian Lee, **Yuntian Deng**, Mert R. Sabuncu, Volodymyr Kuleshov. Semi-Parametric Inducing Point Networks and Neural Processes. *ICLR 2023*

14 Maxim Zvyagin*, Alexander Brace*, Kyle Hippe*, **Yuntian Deng***, Bin Zhang, Cindy Orozco Bohorquez, Austin Clyde, Bharat Kale, Danilo Perez-Rivera, Heng Ma, Carla M. Mann, Michael Irvin, J. Gregory Pauloski, Logan Ward, Valerie Hayot, Murali Emani, Sam Foreman, Zhen Xie, Diangen Lin, Maulik Shukla, Weili Nie, Josh Romero, Christian Dallago, Arash Vahdat, Chaowei Xiao, Thomas Gibbs, Ian Foster, James J. Davis, Michael E. Papka, Thomas Brettin, Rick Stevens, Anima Anandkumar, Venkatram Vishwanath, Arvind Ramanathan, GenSLMs: Genome-scale language models reveal SARS-CoV-2 evolutionary dynamics. **ACM Gordon Bell Special Covid Prize**

13 Justin Chiu*, **Yuntian Deng***, Alexander Rush. Low-Rank Constraints for Fast Inference in Structured Models. *NeurIPS 2021*

12 **Yuntian Deng**, Alexander Rush. Sequence-to-Lattice Models for Fast Translation. *EMNLP 2021 Findings*

11 Keyon Vafa, **Yuntian Deng**, David Blei, Alexander Rush. Rationales for Sequential Predictions. *EMNLP 2021* ***oral***

10 Anton Bakhtin, Sam Gross, Myle Ott, **Yuntian Deng**, Marc'Aurelio Ranzato, Arthur Szlam. Real or Fake? Learning to Discriminate Machine from Human Generated Text. *arXiv 1906.03351*

9 Thierry Tambe, En-Yu Yang, Zishen Wan, **Yuntian Deng**, Vijay Janapa Reddi, Alexander Rush, David Brooks, Gu-Yeon Wei. Algorithm-Hardware Co-Design of Adaptive Floating-Point Encodings for Resilient Deep Learning Inference. *Design Automation Conference (DAC) 2020* ***Best Paper Award***

8 **Yuntian Deng**, David Rosenberg, Gideon Mann. Challenges in End-to-End Neural Scientific Table Recognition. *ICDAR 2019*

7 Pengtao Xie, **Yuntian Deng**, Yi Zhou, Abhimanu Kumar, Yaoliang Yu, James Zou, Eric P Xing. Learning Latent Space Models with Angular Constraints. *ICML 2017*

6 Guillaume Klein, Yoon Kim, **Yuntian Deng**, Jean Senellart, Alexander M Rush. OpenNMT: Open-Source Toolkit for Neural Machine Translation. *ACL 2017* ***Best Demo Paper Award Runner-Up***

5 Zichao Yang, Zhiting Hu, **Yuntian Deng**, Chris Dyer, Alex Smola. Neural Machine Translation with Recurrent Attention Modeling. *EACL 2017*

4 Xuezhe Ma, Yingkai Gao, Zhiting Hu, Yaoliang Yu, **Yuntian Deng**, Eduard Hovy. Dropout with Expectation-linear Regularization. *ICLR 2017*

3 Hao Zhang, Zhiting Hu, **Yuntian Deng**, Mrinmaya Sachan, Zhicheng Yan, Eric P. Xing. Learning Concept Taxonomies from Multi-modal Data. *ACL 2016*

2 Pengtao Xie, **Yuntian Deng**, Eric P. Xing. Diversifying Restricted Boltzmann Machine for Document Modeling. *KDD 2015*

1 Zhiting Hu, Poyao Huang, **Yuntian Deng**, Yingkai Gao, Eric P. Xing. Entity Hierarchy Embedding. *ACL 2015*

## Teaching Experience

**Harvard University**

| | |
|---|---|
| Head TA, Intro to Computational Linguistics and NLP (Stuart Shieber) | Aug 2021 - Dec 2021 |
| Head TA, Intro to Computational Linguistics and NLP (Stuart Shieber) | Aug 2020 - Dec 2020 |
| TA, Natural Language Processing (Alexander Rush) | Jan 2019 - May 2019 |
| Lecture, Natural Language Processing - Translation (Alexander Rush) | Feb 13, 2019 |
| TA, Advanced Machine Learning (Alexander Rush) | Aug 2017 - Dec 2017 |

**Cornell University**

| | |
|---|---|
| Lab Material Preparation, Break Through Tech AI (Alexander Rush) | Aug 2021 |

**Carnegie Mellon University**

| | |
|---|---|
| TA, Graduate Probabilistic Graphical Models (Matthew Gormley, Eric Xing) | Jan 2016 - May 2016 |
| TA, Graduate Machine Learning (Ziv Bar Joseph, Eric Xing) | Aug 2015 - Dec 2015 |

**Revere High School**

| | |
|---|---|
| Volunteer Instructor of AP CS (Microsoft Philanthropies) | Sep 2020 - May 2021 |

## Professional Service

2024: Area Chair of ICLR

2023: Area Chair of EMNLP GEM Workshop, Reviewer of TACL, ICML, ACL, ACL ARR, ACL Demo, AACL, EMNLP, Foundations and Trends in Signal Processing

2022: Reviewer of ACL ARR, EMNLP, ICML, NeurIPS, ICLR, EACL, ML Reproducibility Challenge, ACL Demo, NAACL Demo, and EMNLP GEM Workshop.

2021: Expert Reviewer of ICML. Oral Session Volunteer of ACL. Reviewer of NeurIPS, ACL, ACL ARR, ICLR, AAAI, EACL, NAACL, ML Reproducibility Challenge, ICLR EBM Workshop, IEEE TIIS, and NEJLT. Volunteer of EMNLP.

2020: Top 10% Reviewer of NeurIPS. Reviewer of ICML, ACL, EMNLP, AACL, AAAI, ACL Demo, COLING, IEEE TIFS, Transactions on Information Systems, and Journal of Computer Science and Technology. Volunteer of ICML. Volunteer moderator of ACL.

2019: Reviewer of NeurIPS, EMNLP, ICLR, NAACL, AAAI, IEEE TNNLS, ACM Computing Surveys, NeurIPS LIRE Workshop, NeurIPS Reproducibility Challenge, EMNLP Summarization Workshop, ICML GraphReason Workshop, and ICLR DeepGenStruct Workshop.

2018: Reviewer of ICLR.

2016: Volunteer of ICML.

## Community Service

| | |
|---|---|
| ACL Mentorship Session on CV and Statement of Purpose | Oct 10, 2023 |
| Childcare Chair at NAACL 2022 D&I Committee | Apr - Jul 2022 |
| Organizer and Mentor of Harvard Women in CS NLP Reading Group | Sep 2020 - Present |
| Volunteer Instructor of AP CS at Revere High School | Sep 2020 - May 2021 |
| Leader of Harvard English Language Table | Sep 2021 - Jan 2023 |

## Internships

| | |
|---|---|
| Nvidia / Argonne National Lab (Anima Anandkumar / Arvind Ramanathan) | May - Dec 2022 |
| Facebook AI Research (Marc'Aurelio Ranzato) | May - Dec 2019 |
| Bloomberg CTO Office (David Rosenberg, Gideon Mann) | Jan - Aug 2017 |
| UCSD (Charles Elkan) | Jul - Sep 2013 |

## Talks

| | |
|---|---|
| AI2 Semantic Scholars: Structure Modeling in Language Models | Aug 2023, Seattle, Washington |

| | |
|---|---|
| NYU AI School 2023: Natural Language Processing | Jun 2023, New York, NY |
| AI2 Mosaic: Structure Modeling in Language Models | Jun 2023, Seattle, Washington |
| UMass Amherst: Structural Coherence in Text Generation | May 2023, remote |
| Georgia Tech: Structural Coherence in Text Generation | Apr 2023, Atlanta, Georgia |
| Harvard Kempner Institute: Structural Coherence in Text Generation | Mar 2023, remote |
| U Alberta: Structural Coherence in Text Generation | Mar 2023, Edmonton, Canada |
| TTIC: Structural Coherence in Text Generation | Feb 2023, Chicago |
| U Waterloo ECE: Structural Coherence in Text Generation | Feb 2023, remote |
| Cornell Seminar in NLU: Structural Coherence in Text Generation | Feb 2023, New York |
| U Waterloo CS: Structural Coherence in Text Generation | Jan 2023, Waterloo, Canada |
| U Chicago Rising Star: Model Criticism for Long-Form Text Generation | Nov 2022, Chicago |
| Princeton NLP: Model Criticism for Long-Form Text Generation | Nov 2022, Princeton |
| EMNLP 2021 Findings: Sequence-to-Lattice Models for Fast Translation | Oct 2021, remote |
| OpenAI: Residual Energy-based Models for Text Generation | Apr 2021, remote |
| NeurIPS 2020: Cascaded Text Generation with Markov Transformers | Oct 2020, remote |
| Baidu: Cascaded Text Generation with Markov Transformers | Jun 2020, remote |
| ICLR 2020: Residual Energy-based Models for Text Generation | Apr 2020, remote |
| LinkedIn: Residual Energy-based Models for Text Generation | Mar 2020, remote |
| Wayfair: Neural Encoder-Decoder Models | Nov 2019, Boston |
| EMNLP 2019: Neural Linguistic Steganography | Nov 2019, Hong Kong |
| FAIR NLP Meeting: Energy-Based Models for Text Generation | Aug 2019, New York City |
| NeurIPS 2018: Latent Alignment and Variational Attention | Dec 2018, Montreal, Canada |
| The French National Center for Scientific Research: Variational Attention | Jul 2018, France |
| Association for Machine Translation in the Americas: OpenNMT | Mar 2018, Boston |
| Open-Source Neural Machine Translation Workshop: Image-to-Text | Mar 2018, Paris, France |
| Tencent Social Network Group TSAIC: OpenNMT | Dec 2017, Shenzhen, China |
| ICML: Image-to-Markup Generation | Aug 2017, Sydney, Australia |

**Open Source Projects**

| | |
|---|---|
| Cascaded Generation (125 Github stars) | PyTorch |
| Neural Lingustic Steganography (170 Github stars) | PyTorch |
| Variational Attention (325 Github stars) | PyTorch |
| Image-to-Markup Generation (1.1k Github stars) | LuaTorch |
| OpenNMT-py (6.3k Github stars) | PyTorch |
| OpenNMT (2.4k Github stars) | LuaTorch |
| Attention OCR (1.1k Github stars) | Tensorflow |