Peng (Billy) Xu
Canada: (780)6558396
Canada Permanent Resident

April 5, 2022
pxu4@ualberta.ca
Researcher and Software Engineer

Career Highlights

- Google Cloud Conversational AI (2022-present):
  - Lead IC of Synopsis Project, the next generation dialogue manager for Google Dialogflow
  - Oversee the overall design of Synopsis and make critical judgement calls
- Borealis AI (2019-2021):
  - Lead IC of Turing Text2SQL project, which was successfully applied in banking business.
  - Grew from Research Engineer to Senior Researcher in three years (two promotions)
- Published 9 (8 first-authored) peer-reviewed publications at top AI conferences (200+ citations).
- One US patent granted, four US patent applications pending.

Professional and Research Experience

- **Google** Remote in Ontario
  Software Engineer, Machine Learning *2022.1-Present*
  - Design and implement Synopsis, the next generation dialogue manager of Google Dialogflow.
  - Lead the collaboration with Google Research to improve Synopsis with the latest techniques.
- **Borealis AI (RBC Institute for Research)** Toronto
  Senior Machine Learning Researcher *2018.7-2022.1*
  - Implement and improve the state-of-the-art semantic parsing models to build a Text-to-SQL system, which has been successfully applied to different banking business, including direct investing and commercial banking. Demo: https://turing-app.borealisai.com.
  - Publish five first-authored papers at ACL 2019, AISTATS 2020, ICML 2020, ACL 2021 on various topics in natural language processing and machine learning.
  - Mentor three internship research projects on semantic parsing.
  - File five patent applications based on the above-mentioned research work.
- **Shannon AI** Beijing
  NLP Algorithm Intern *2018.5*
  - Contribute to the development of an intelligent financial question answering system
  - Build the pipeline to automatically extract information from company announcements
- **DBpedia** Remote
  Active Contributor and Mentor for Google Summer of Code (GSoC) *2016.4-2018.9*
  - Complete a project on inferring infobox template class mappings from Wikipedia and WikiData in 2016
  - Mentor two GSoC projects about entity embeddings based on DBpedia in 2017
  - Mentor one GSoC project about embeddings for Out-Of-Vocabulary resources in 2018

- **University of Alberta** Edmonton
  Research Assistant under Prof. Denilson Barbosa *2016.5-2018.7*
  - Research on fine-grained entity type classification (FETC) with deep neural models
  - Research on relation extraction (RE) with deep neural models
  - Research on incorporating knowledge base information to facilitate neural information extraction, mainly on task of FETC and RE
  - Publish two first-authored papers at NAACL 2018 and NAACL 2019 on FETC and RE
  - Be awarded the runner up for the Outstanding Master Thesis Award
- **Tsinghua University** Beijing
  Research Assistant under Prof. Jie Tang *2014.9-2015.6*
  - Develop AMiner which is a website that offers comprehensive search and mining services for academic community in `python` and `scala`
- **Tsinghua University** Beijing
  Research Assistant under Prof. Zhihu Du *2014 Summer*
  - Research on efficient partical-mesh spreading in `C` and `CUDA`

Skills

- Demonstrated technical leadership by delivering multiple business products
- Strong communication skills and ample cross-teams cross-functions collaboration experience
- Extensive knowledge on Natural Language Processing (NLP) and Machine Learning (ML)
  - **Conversational AI**: Task-Oriented Dialogue, Semantic Parsing and Text Generation
  - **Foundation Models**: Large Language Models and Transformer Optimization
  - **Knowledge Acquisition**: Knowledge Graphs and Information Extraction
- Professional in Python for ML research and experimentation
  - Deep Learning: PyTorch, Tensorflow, Keras, JAX
  - Machine Learning and Data Science : Scikit-learn, XGBoost, Pandas, Numpy
  - Data Visualization: Matplotlib, Altair, Graphviz
  - Experimentation: Tensorboard, Gin-config
- Professional in Kotlin for ML product development and deployment
- Professional in software engineering
  - Concurrency, OOD, Memory, Typing, Functional Programming and beyond
  - Version Control (Git, Mercurial)
  - Unit Testing (unittest.mock, JUnit, Mockito)
  - Coding Productivity (Vim, Tmux, Bash)
  - Code Review and Technical Writing
- Familiar with Java, C/C++, LaTeX

Peer-Reviewed Publications

(* indicates equal contribution)

1. *Optimizing Deeper Transformers on Small Datasets*
   The 59th Annual Meeting of the Association for Computational Linguistics (ACL), 2021, *long paper*
   Peng Xu, Dhruv Kumar, Wei Yang, Wenjie Zi, Keyi Tang, Chenyang Huang, Jackie Chi Kit Cheung, Simon J.D. Prince, Yanshuai Cao

2. *TURING: an Accurate and Interpretable Multi-Hypothesis Cross-Domain Natural Language Database Interface*
   The 59th Annual Meeting of the Association for Computational Linguistics (ACL), 2021, *demonstration track*
   Peng Xu*, Wenjie Zi*, Hamidreza Shahidi, Akos Kadar, Keyi Tang, Wei Yang, Jawad Ateeq, Harsh Barot, Meidan Alon, Yanshuai Cao

3. *On Variational Learning of Controllable Representations for Text without Supervision*
   The 37th International Conference on Machine Learning (ICML), 2020
   Peng Xu, Yanshuai Cao, Jackie Chi Kit Cheung

4. *Better Long-Range Dependency By Bootstrapping A Mutual Information Regularizer*
   The 23rd International Conference on Artificial Intelligence and Statistics (AISTATS), 2020
   Yanshuai Cao* and **Peng Xu***

5. *A Cross-Domain Transferable Neural Coherence Model*
   The 57th Annual Meeting of the Association for Computational Linguistics (ACL), 2019, *long paper*
   Peng Xu, Hamidreza Saghir, Jin Sung Kang, Teng Long, Avishek Joey Bose, Yanshuai Cao, Jackie Chi Kit Cheung

6. *Connecting Language and Knowledge with Heterogeneous Representations for Neural Relation Extraction*
   The 17th Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2019, *short paper*
   Peng Xu, Denilson Barbosa

7. *Neural Fine-Grained Entity Type Classification with Hierarchy-Aware Loss*
   The 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2018, *long oral paper*
   Peng Xu, Denilson Barbosa

8. *Matching Resumes to Job Descriptions with Stacked Models*
   The 31st Canadian Conference on Artificial Intelligence, 2018.
   Peng Xu, Denilson Barbosa

9. *Efficient Particle-Mesh Spreading on GPUs*
   The 15th International Conference on Computational Science (ICCS), 2015
   Xiangyu Guo, Xing Liu, **Peng Xu**, Zhihu Du, Edmond Chow

Technical Reports

1. *Shortcut Learning Hypothesis of Modern Language Models*
   Blog post, 2022
   Peng Xu

2. *Transformers: Introduction, Extensions and Training*
   Tutorials, 2021
   Simon Prince, **Peng Xu**

3. *Hierarchical Neural Data Synthesis for Semantic Parsing*
   arXiv preprint, 2021
   Wei Yang, **Peng Xu**, Yanshuai Cao

4. *Investigations on Knowledge Base Embedding for Relation Prediction and Extraction*
   arXiv preprint, 2018
   Peng Xu, Denilson Barbosa

Thesis

- *Towards Neural Information Extraction without Manual Annotated Data*
  M.Sc. Thesis, **runner up for the Outstanding Master Thesis Award**

Patents Granted

1. *System and Method for Cross-domain Transferable Neural Coherence Model*
   US Patent 11,270,072.

Patents Filed

1. *System and Method for Machine Learning with Long-Range Dependency*
   US Patent App. 16/809,267.

2. *System and Method for Controllable Machine Text Generation Architecture*
   US Patent App. 16/881,843.

Peer-Review Services

Conference Reviewers:

- *ACL Rolling Review*: 2021, 2022
- *NAACL-HLT*: 2021

Journal Reviewers:

- *Journal of Intelligent Information Systems*: 2022
- *Computational Intelligence*: 2022
- *Natural Language Engineering*: 2022
- *PeerJ Computer Science*: 2022
- *SCIENCE CHINA Information Sciences*: 2022
- *Distributed and Parallel Databases*: 2022

Education

- **University of Alberta** Edmonton, AB
  M.Sc. Computer Science *2016.1 - 2018.7*
  - GPA: 4.0/4.0
  - Supervisor: Prof. Denilson Barbosa
- **Beijing University of Posts and Telecommunications** Beijing
  B.Eng. Electronic Information Engineering *2011.9 - 2015.6*
  - Major: 89.62/100, Overall: 87.06/100
  - Thesis Advisor: Prof. Jie Tang