# Honglei Liu

liuhonglei@gmail.com
https://honglei.ai/
https://www.linkedin.com/in/honglei-liu-a53a435a/

## Education

2012 – 2017 **Ph.D., Computer Science**
University of California, Santa Barbara (UCSB)
Advisor: *Prof. Xifeng Yan*
Thesis: *Mining Patterns and Networks from Sequence Data*

2010 – 2012 **M.S., Computer Science** *(Graduate with Honors, top 1%)*
Northeastern University (NEU), China
Advisors: *Prof. Xiaochun Yang and Prof. Bin Wang*
Thesis: *Approximate Text Alignment Techniques and Optimization Approaches*
*(Outstanding Graduate Dissertation Award)*

2006 – 2010 **B.S., Computer Science** *(Honors Class, top 3%)*
Northeastern University (NEU), China
Thesis: *Design and Implementation of an Approximate Querying Approach Based on Berkeley DB*
*(Outstanding Undergraduate Dissertation Award)*

## Experience

09/2021 – Present **Head of Machine Learning and Data Platform**, Fast
Built and led the ML & Data org to support all the data and machine learning needs at Fast, with three sub-teams
- Data Platform
- Recommendation
- Search

The ML & Data group's mission is to empower Fast to easily embed machine learning and data in our strategy, decision making, and product experiences so that we can bring value to our customers. On the platform side, this group is responsible for: scaling our online infrastructure to track, measure, and quantify user touchpoints, building a data and machine learning platform that supports batch and streaming data processing and prediction at scale, enhancing our product experiences with experimentation infrastructure. On the product side, it is also this group's core responsibility to build recommendations, search, and other machine learning powered products that can increase engagement with our customers and supercharge Fast sellers with increased revenue.

04/2020 – 09/2021 **Research Scientist Manager**, Facebook
Built and led a team of research scientists and engineers working on conversational AI, ranking, and assistant technologies

02/2020 – 04/2020 **Staff Research Scientist**, Facebook

## Experience (continued)

| | |
|---|---|
| 01/2019 – 02/2020 | **Senior Research Scientist**, Facebook<br>Tech lead at Assistant team<br>- Led a highly xfn workstream on Conversational Recommendation/Search that involved >7 sub-teams and >10 engineers<br>- Delivered many 0 to 1 efforts end-to-end, including shipping multiple assistant domains from scratch<br>- Single-handedly designed and built a recommender system which now has >10 engineers contributing/maintaining it<br>- Open-sourced ReAgent: a modular, end-to-end platform for building reasoning systems, driving tens of billions of decisions per day<br>- Published 2 research papers on Federated Learning and 1 research paper on Conversational Recommendation |
| 07/2017 – 12/2018 | **Research Scientist**, Facebook<br>Focus on Natural Language Understanding, Dialog, Personalization and Deep Learning<br>- Multiple NLU and ranking model improvements that led to significant metrics gains for M Suggestion (FB Assistant on Messenger), including a 12% absolute ctr gain<br>- Shipped the first Reinforcement Learning (RL) model for Messenger (second successful use case of RL at Facebook), serving billions of users<br>- Shipped the first on-device NLP model for Messenger by solving big challenges such as vocabulary compression, knowledge distillation, multitask learning, and on-device deployment<br>- Open-sourced PyText: a deep-learning based NLP modeling framework<br>- Published 2 research papers on online learning and RL |
| 06/2016 – 09/2016 | **Intern**, Facebook<br>Topic: *Indexing and Mining Billions of Time Series* |
| 07/2015 – 09/2015 | **Bioinformatics Intern**, Illumina<br>Topic: *Fast Specificity Checking for Multiplex PCR Primer Design* |

## Publications

1. **Adding Chit-Chats to Enhance Task-Oriented Dialogues**

   *Kai Sun, Seungwhan Moon, Paul Crook, Stephen Roller, Becka Silvert, Bing Liu, Zhiguang Wang, Honglei Liu, Eunjoon Cho, Claire Cardie. Proc. of the Annual Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2021).*

2. **NUANCED: Natural Utterance Annotation for Nuanced Conversation with Estimated Distributions**

   *Zhiyu Chen, Honglei Liu, Hu Xu, Seungwhan Moon, Hao Zhou, Bing Liu. arXiv preprint arXiv:2010.12758 (2020).*

3. **User Memory Reasoning for Conversational Recommendation**

   *Hu Xu, Seungwhan Moon, Honglei Liu, Bing Liu, Pararth Shah, Bing Liu, Philip S. Yu. Proc. of Int. Conf. on Computational Linguisticsto (COLING 2020).*

4. **Federated User Representation Learning**

   *Duc Bui, Kshitiz Malik, Jack Goetz, Honglei Liu, Seungwhan Moon, Anuj Kumar, Kang G. Shin. arXiv preprint arXiv:1909.12535 (2019).*

5. **Active Federated Learning**

*Jack Goetz, Kshitiz Malik, Duc Bui, Seungwhan Moon, Honglei Liu, Anuj Kumar. Workshop on Federated Learning for Data Privacy and Confidentiality at Neural Information Processing Systems (NeurIPS 2019).*

6. **Global Textual Relation Embedding for Relational Understanding**

   *Zhiyu Chen, Hanwen Zha, Honglei Liu, Wenhu Chen, Xifeng Yan, Yu Su. Proc. of the Annual Meeting of the Association for Computational Linguistics (ACL 2019). (Short Paper)*

7. **Interpretability of Deep Reinforcement Learning Models in a Conversational System**

   *Honglei Liu, Parath Shah, Wenxuan Li, Wenhai Yang, Anuj Kumar. in submission*

8. **Explore-Exploit: A Framework for Interactive and Online Learning**

   *Honglei Liu, Anuj Kumar, Wenhai Yang, Benoit Dumoulin. Systems for Machine Learning Workshop at Neural Information Processing Systems (NeurIPS 2018).*

   Open-sourced as **ReAgent**, a modular, end-to-end platform for building reasoning systems

9. **PoQaa: Text Mining and Knowledge Sharing for Scientific Publications**

   *Keqian Li, Ping Zhang, Honglei Liu, Hanwen Zha, Xifeng Yan. Proc. of Int. Conf. on Knowledge Discovery and Data Mining (KDD 2018). (demo)*

10. **In Vitro Validation of in Silico Identified Inhibitory Interactions**

    *Honglei Liu, Daniel Bridges, Connor Randall, Sara A. Solla, Bian Wu, Paul Hansma, Xifeng Yan, Kenneth S. Kosik, Kristofer Bouchard. Journal of Neuroscience Methods 321 (2019): 39-48.*

11. **Global Relation Embedding for Relation Extraction**

    *Yu Su*, Honglei Liu*, Semih Yavuz, Izzeddin Gur, Huan Sun, Xifeng Yan. Proc. of the Annual Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2018). (*: Equal Contribution)*

12. **Active Learning of Functional Networks from Spike Trains**

    *Honglei Liu, Bian Wu. SIAM Int. Conf. on Data Mining (SDM 2017).*

13. **Fast Motif Discovery in Short Sequences**

    *Honglei Liu, Fangqiu Han, Hongjun Zhou, Xifeng Yan, Kenneth S. Kosik. Proc. of Int. Conf. on Data Engineering (ICDE 2016).*

    **Software** licensed to **SerImmune Inc.** to produce real world value

14. **ALAE: Accelerating Local Alignment with Affine Gap Exactly in Biosequence Databases**

    *Xiaochun Yang, Honglei Liu, Bin Wang. Proc. of Int. Conf. on Very Large Data Bases (VLDB 2012).*

15. **Approximate Substring Query Algorithms Supporting Local Optimal Matching**

    *Honglei Liu, Xiaochun Yang, Bin Wang, Rong Jin. Journal of Frontiers of Computer Science and Technology, 2011.*

## Open Source

**ReAgent** A modular, end-to-end platform for building reasoning systems. It closes the loop of turning actions into feedback, and feedback into training data for RL and online learning. ReAgent is used at Facebook to drive tens of billions of decisions per day.

**PyText** A deep-learning based NLP modeling framework built on PyTorch. PyText is used at Facebook to iterate quickly on new modeling ideas and then seamlessly ship them at scale.

## Open Source (continued)

**ASC** A fast motif discovery tool that is 10,000 times faster than MEME while preserving the same accuracy. ASC+MEME reduces the running time of MEME from weeks to a few minutes with even better accuracy. ASC was licensed to SerImmune Inc. funded by NIH, illumina, Merck, etc. to find motifs from massive protein sequences generated by modern sequencing techniques.