

Sia, Xin Yun Suzanna

Web Page: <https://suzyahyah.github.io>
Email: ssia1@jhu.edu
Github: <https://github.com/suzyahyah>
Contact: (+1)4434499249

RESEARCH INTERESTS	Inductive Biases in Transformer self-attention models Previously worked on Multimodal Fusion, Topic Modeling, Multilingual IR, Deep Latent Variable Modeling, Non-parametric Grammar Induction, Argument Mining, Chatbots, Sentiment Analysis
EDUCATION	PhD Computer Science (NLP) , Johns Hopkins University 2018- ▪ Primary Advisor: Kevin Duh M.Tech Knowledge Engineering , National University of Singapore Jan 2014 – Dec 2016 ▪ Awarded Honors (Distinction) ▪ Thesis: An Expert System for Energy Efficient Resource Management. Advisor: Zhu Fangming BSc. A.I & Psychology , University of Edinburgh Sep 2009 – Jul 2013 ▪ Awarded First Class Honors ▪ Awarded Best CS Thesis, Video Meeting Search Interfaces. Advisor: Steve Renals
AWARDS / SCHOLARSHIPS	Best Poster , NYAS Speech and Dialogue Symposium , (1/60) 2019 PhD Scholarship , DSO National Labs (declined for JHU RAship) 2018 2nd Place , NUS/NUHS-MIT Datathon 2018 3rd Place , 26th Association for Computing Machinery CIKM Analyticup 2017 Kinetic Award, GPA, GPA , DSO National Labs 2016, 2017, 2018 JASSO Scholarship , Japan Government 2016, 2017 Best Computer Science Final Year Project , University of Edinburgh 2013 Best Poster , British Computing Society Lovelace Colloquium 2013 Overseas Undergraduate Scholarship , Defence Science Technology Agency 2009 – 2013
PUBLICATIONS	Sia, S., Duh, K, Adaptive Mixture LDA for Low-resource Topic Modeling <i>Proceedings of the 2021 Conference European Association of Computational Linguistics</i> . [paper] 2021 Sia, S., Dalmia, A., Mielke, S., Tired of Topic Models? Clusters of Pretrained Word Embeddings Make for Fast and Good Topics too! <i>Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing</i> . [paper] 2020 Sun, S., Sia, S., Duh, K., CLIReval: Evaluating Machine Translation as a Cross-Lingual Information Retrieval Task, <i>Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations</i> [paper] 2020 Xie, J., Sia, S., Garcia, P., Povey, D., Khundanpur, S. Mixture of Speaker-type PLDAs for Children's Speech Diarization [paper] <i>arXiv:2008.13213</i> 2020 Sia, S., Jaidka, K., Duh, K., A semi-supervised hierarchical generative model of argumentation. <i>Natural Language, Dialog and Speech Symposium, New York Academy of Sciences</i> . 2019 Sia, S., Li L.J.A., Hierarchical Module Classification in Mixed Initiative Conversational Agent System, <i>Proceedings of the 24th ACM International on Conference on Information and Knowledge Management</i> . ACM, 553–562. 12 2017
SERVICE	Reviewer EMNLP 2019 (Outstanding Reviewer); Journal of Computational Linguistics, ACL 2020
TEACHING	Teaching Assistant (Reviews), CS465/665 NLP, CS462/682 Deep Learning, CS475/675 Machine Learning, Johns Hopkins 2019 Student Mentor , Shibaura Institute of Technology, Saitama, Japan, [Report] [Program] 2016

WORK EXPERIENCE	Research Intern , Facebook AI Applied Research Jun 2020 – Oct 2020
	<ul style="list-style-type: none"> ▪ Multimodal Hatespeech detection, multimodal indexing and retrieval. Advisor: Lambert Mathias ▪ Made several code contributions to Facebook’s open source Multimodal Fusion Library
	Research Engineer , DSO National Laboratories, Singapore Jun 2015 – Aug 2018
	<ul style="list-style-type: none"> ▪ Defence related NLP projects: Information Retrieval, conversational agents, recommender systems and social computing. Advisor: Chieu Hai Leong
	Research Assistant , Singapore Management University Aug 2016 – Dec 2016
	<ul style="list-style-type: none"> ▪ Clustered user sub-groups via collaborative filtering and Variational Inference for Probabilistic Matrix Factorization. Advisor: Jiang Jing
	Psychologist , DSO National Laboratories, Singapore Jul 2013 – Jun 2015
	<ul style="list-style-type: none"> ▪ Conducted experiments to evaluate various cognitive and social sensing systems including wearable technology, Microsoft Kinect, and cognitive test batteries. Human factors and usability studies.
	Visiting Researcher , Stanford University Mar 2012 – Jul 2012
	<ul style="list-style-type: none"> ▪ Designed metric and implemented the algorithm for automatically scoring a sequence of decisions in an online choice based task. Advisor: Dan Schwartz
PROJECTS	Infinite PCFGs for social grammars 2019
	<ul style="list-style-type: none"> ▪ Hierarchical Dirichlet Process for learning non-parametric grammars. Grammars are evaluated on sentiment classification using easy-first parsing tree LSTMs. [Interim write-up]
	Patient Conversation Simulator 2017
	<ul style="list-style-type: none"> ▪ Simulated patient conversation for medical education, demonstrated system with explainable response selection at the 26th ACM Conference of International Knowledge Management.
	Undergraduate Admissions QA System , National University of Singapore 2017
	<ul style="list-style-type: none"> ▪ Programmed a live chat system for the Undergraduate Computing admissions cycle. Deployed on NUS School of Computing Website and Facebook page. [Twitter]
	Computational Propaganda Project Feb 2016 – May 2016
	<ul style="list-style-type: none"> ▪ Developed anomaly detection algorithms for Taiwan case study in the Computational Propaganda Project by Oxford Internet Institute. [Guardian-article]
SKILLS/TOOLS	<p><u>Programming</u>: Python, Java, Cython, bash, Octave/Matlab, R, HTML5/Javascript/CSS, Haskell</p> <p><u>Research</u>: Git, Scikit-learn, SciPy, NumPy, Pandas, Gensim, Stanford NLP, wordnet/nltk, pymc3, NetworkX, Thulac, Protege, DBPedia, SPARQL, Mallet (Java), EJML (Java), \LaTeX</p> <p><u>Deep Learning Frameworks</u>: Tensorflow, PyTorch, AllenNLP, Keras</p> <p><u>Web</u>: Amazon Web Services, Heroku, NodeJS, ExpressJS, Flask, Scrappy, Selenium, Django</p> <p><u>Databases</u>: MongoDB, PostgreSQL, Elasticsearch, Neo4j</p> <p><u>Psychology</u>: Amazon Turk, experimental design, participant recruitment, interface design</p>
GRADUATE COURSEWORK	Natural Language Processing, Advanced Bayesian Statistics (Non-parametrics), Linguistic and Sequence Modelling, Distributed Systems, Introduction to Statistics, Information Theory, Randomized Algorithms, Object-Oriented Architecture and Design
ADDITIONAL COURSES / CERTIFICATIONS	Deep Learning Specialization (16 weeks), Coursera/DeepLearning.AI, [Cert] 2018
	Machine Learning Specialization (25 weeks), Coursera/University of Washington [Cert] 2017
	Machine Learning Summer School, Max Planck Institute of Intelligent Systems, Germany 2017
	ITIL Foundation in IT Service Management, AXELOS, UK, [Cert] 2015