

# Sia, Xin Yun Suzanna

Web Page: <https://suzyahyah.github.io>

Email: [ssia1@jhu.edu](mailto:ssia1@jhu.edu)

Github: <https://github.com/suzyahyah>

RESEARCH AREAS	Prior experience with Multimodal Fusion, Topic Modeling, Multilingual IR, Deep Latent Variable Modeling, Non-parametric Grammar Induction, Argument Mining, Chatbots	
EDUCATION	<b>PhD Computer Science (NLP)</b> , Johns Hopkins University ▪ Primary Advisor: Kevin Duh ▪ Thesis: In-context Learning for Machine Translation (In-progress) <b>M.Tech Knowledge Engineering</b> , National University of Singapore ▪ Awarded Honors Distinction, completed part-time while working full-time at DSO Labs ▪ Thesis: An Expert System for Energy Efficient Resource Management. Advisor: Zhu Fangming <b>BSc. A.I &amp; Psychology</b> , University of Edinburgh ▪ Full Scholarship from Singapore government ▪ Awarded First Class Honors ▪ Best Undergrad CS Thesis, Video Meeting Search Interfaces. Advisor: Steve Renals	2018-   Jan 2014 – Dec 2016  Sep 2009 – Jul 2013
AWARDS / SCHOLARSHIPS	<b>Best Poster</b> , <a href="#">NYAS Speech and Dialogue Symposium</a> , (1/60) <b>PhD Scholarship</b> , DSO National Labs (declined for JHU RAship) <b>2nd Place</b> , NUS/NUHS-MIT Datathon <b>3rd Place</b> , 26th Association for Computing Machinery CIKM Analyticup <b>JASSO Scholarship</b> , Japan Government	2019 2018 2018 2017 2016, 2017
PUBLICATIONS (PEER REVIEWED)	Sia, S., Duh, K., <a href="#">Prefix Embeddings for In-context Machine Translation</a> . <i>AMTAS 2022</i> Dalmia, A. Sia, S. <a href="#">Clustering with UMAP: Why and How Connectivity matters</a> . <i>Workshop on Graphs and more Complex structures for Learning and Reasoning @AAAI 2022</i> <a href="#">Sia, S.*, Jaidka, K*, Ahuja, H., Chhaya, K., Duh, K. Offer a Different Perspective: Modeling the Belief Alignment of Arguments in Multi-party Debates</a> <i>EMNLP 2022</i> <a href="#">Sia, S., Duh, K. Adaptive Mixture LDA for Low-resource Topic Modeling</a> <i>EACL 2021</i> <a href="#">Sia, S., Dalmia, A., Mielke, S., Tired of Topic Models? Clusters of Pretrained Word Embeddings Make for Fast and Good Topics too!</a> <i>EMNLP 2020</i> <a href="#">Sun, S., Sia, S., Duh, K., CLIReval: Evaluating Machine Translation as a Cross-Lingual Information Retrieval Task</a> <i>ACL 2020 Demo</i> <a href="#">Sia, S., Jaidka, K., Duh, K., A semi-supervised hierarchical generative model of argumentation</a> . <i>Natural Language, Dialog and Speech Symposium, New York Academy of Sciences</i> . <a href="#">Sia, S., Li L.J.A., Hierarchical Module Classification in Mixed Initiative Conversational Agent System</a> , <i>CIKM Demo 2017</i>	2022 2022 2022 2021 2020 2020 2019 2017
UNDER REVIEW/ARXIV	<a href="#">Sia, S., Duh, K., Improving In-Context Learning for Machine Translation via PMI-based Separator Initialisation</a> . <a href="#">Sia, S., Belyy, A., Khabsa, M., Amjad, A., Zettlemoyer, L., Mathias, L., On the Logical Satisfiability of Counterfactuals for Evaluating Faithful Explanations in NLI Tasks</a> . <a href="#">Xie, J., Sia, S., Garcia, P., Povey, D., Khundampur, S. Mixture of Speaker-type PLDAs for Children's Speech Diarization</a> <i>arXiv:2008.13213</i>	2022 2022 2020
TEACHING	<b>Teaching Assistant</b> NLP(Reviews), DL, ML, Intro Human Language Technology <b>Student Mentor</b> , Shibaura Institute of Technology, Saitama, Japan, <a href="#">[Report]</a> <a href="#">[Program]</a>	2019-2021 2016

<b>WORK EXPERIENCE</b>	Research Intern, Facebook AI Applied Research (2 Summers)	Jun 2020/1 – Oct 2020/1
	<ul style="list-style-type: none"> <li>▪ Multimodal Explainability for Logical Inference</li> <li>▪ Multimodal Hatespeech detection, multimodal indexing and retrieval.</li> <li>▪ Made several code contributions to Facebook's open source Multimodal Fusion Library. <a href="#">[PRs]</a></li> </ul>	
	Engineer, DSO National Laboratories, Singapore	Jun 2015 – Aug 2018
	<ul style="list-style-type: none"> <li>▪ Defence related NLP projects: Information Retrieval, conversational agents, recommender systems and social computing. Advisor: Chieu Hai Leong</li> </ul>	
	Research Assistant, Singapore Management University	Aug 2016 – Dec 2016
	<ul style="list-style-type: none"> <li>▪ Clustered user sub-groups via collaborative filtering and Variational Inference for Probabilistic Matrix Factorization. Advisor: Jiang Jing</li> </ul>	
	In-house Consultant, Ministry of Defence, Singapore	2015 – 2016
	<ul style="list-style-type: none"> <li>▪ Provided technical recommendations in Strategic Policy paper on Fake News Detection</li> <li>▪ Workflow Analysis and technology recommendations for Intelligence Analysts</li> <li>▪ Review team decision-making processes for Chemical-Toxin-Radiological-Nuclear Safety Council.</li> </ul>	
	Psychologist, DSO National Laboratories, Singapore	Jul 2013 – Jun 2015
	<ul style="list-style-type: none"> <li>▪ Conducted experiments to evaluate various cognitive and social sensing systems including wearable technology, Microsoft Kinect, and cognitive test batteries. Human factors and usability studies.</li> </ul>	
	Visiting Researcher, Stanford University	Mar 2012 – Jul 2012
<b>OTHER PROJECTS</b>	<b>Infinite PCFGs for social grammars</b>	2019
	<ul style="list-style-type: none"> <li>▪ Hierarchical Dirichlet Process for learning non-parametric grammars. Grammars are evaluated on sentiment classification using easy-first parsing tree LSTMs. <a href="#">[Interim write-up]</a></li> </ul>	
	<b>Patient Conversation Simulator</b>	2017
	<ul style="list-style-type: none"> <li>▪ Simulated patient conversation for medical education, demonstrated system with explainable response selection at the 26th ACM Conference of International Knowledge Management.</li> </ul>	
	<b>Undergraduate Admissions QA System</b> , National University of Singapore	2017
<b>SKILLS/TOOLS</b>	<ul style="list-style-type: none"> <li>▪ Programmed a live chat system for the Undergraduate Computing admissions cycle. Deployed on NUS School of Computing Website and Facebook page. <a href="#">[Twitter]</a></li> </ul>	
	<b>Computational Propaganda Project</b>	Feb 2016 – May 2016
	<ul style="list-style-type: none"> <li>▪ Developed anomaly detection algorithms for Taiwan case study in the Computational Propaganda Project by Oxford Internet Institute. <a href="#">[Guardian-article]</a></li> </ul>	
	<u>Programming:</u> Python, bash, Java, Cython, R, HTML5/Javascript/CSS <u>Libraries:</u> Scikit-learn, SciPy, NumPy, Pandas, HuggingFace, Gensim, Stanford NLP, wordnet/nltk, pymc3, NetworkX, Thulac, Protege, DBPedia, Mallet (Java), EJML (Java) <u>Deep Learning Frameworks:</u> PyTorch, Tensorflow AllenNLP, Keras <u>Web:</u> Amazon Web Services, Heroku, NodeJS, ExpressJS, Flask, Scrapy, Selenium, Django <u>Databases:</u> MongoDB, PostgreSQL, Elasticsearch, Neo4j <u>Psychology:</u> Amazon Turk, experimental design, participant recruitment, interface design	
<b>GRADUATE COURSEWORK</b>	Information Theory, Advanced Bayesian Statistics (Non-parametrics), Distributed Systems, Introduction to Statistics, Randomized Algorithms, Object-Oriented Architecture and Design	
<b>ADDITIONAL COURSES / CERTIFICATIONS</b>	Deep Learning Specialization (16 weeks), Coursera/DeepLearning.AI, <a href="#">[Cert]</a>	2018
	Machine Learning Specialization (25 weeks), Coursera/University of Washington <a href="#">[Cert]</a>	2017
	Machine Learning Summer School, Max Planck Institute of Intelligent Systems, Germany	2017
	ITIL Foundation in IT Service Management, AXELOS, UK, <a href="#">[Cert]</a>	2015