Dr. Khyathi Raghavi Chandu

Research Scientist & Visiting Professor

@ khythiraghavi@gmail.com 🌐 khyathiraghavi.github.io/ 🎓 Google Scholar

🎓 Education

2021	**Carnegie Mellon University, PhD in Language Technologies**
2018	**Carnegie Mellon University, Masters in Language Technologies**
2016	**IIIT-H, Bachelors in Computer Science**

💼 Work Experience

AI2	**Research Scientist,**
	❯ Multimodality, LLM training, LLM Evaluation, Distillation, Instruction-tuning, Uncertainty modeling, Agents
Jio Institute	**Visiting Professor,**
	❯ Designed and taught a post-graduate course on LLMs and Beyond for Deep Learning for NLP
	❯ 🔗 Course Curriculum
Meta AI	**Research Scientist, Feb 2022 - Jan 2023**
Meta AI	**Research Intern, May 2021 - August 2021**
	❯ Multimodal Conversational AI
Google Research	**Research Intern, May 2020 - August 2020**
	❯ Weak Supervision of content selection for better Controllability in Image Captioning
Apple	**Research Intern, May 2019 - August 2019**
	❯ Transfer Learning for Code-Switched Named Entity Recognition
Univ. of Pennsylvania	**Research Intern, May 2015 - August 2015**
	❯ Converting unstructured medical abstracts to structured abstracts.

📢 Talks

Netflix	**Invited Talk, July 2024**
Allen Institute for AI	**Invited Talk, Aug 2022**
	❯ Topic : Anchoring Multiview Narrative Generation
Univ of Utah	**Colloquium Talk, April 2022**
	❯ Topic : Multiview Narrative Generation 🔗 Link
Univ of Pittsburgh	**Guest Lecture in Introduction to NLP, Dec 2021**
	❯ Topic : Grounding and Multimodality 🔗 Link
Microsoft Research	**Talk, May 2021**
	❯ Topic : Multimodal Narrative Intelligence
Google	**PhD Intern Research Conference, July 2020**
	❯ Topic : Towards effective controllability of multimodal narrative generation
WiNLP at ACL'20	**Closing Remarks from organizers, July 2020**
	❯ Topic : Progress of Widening NLP and lessons learnt from virtual workshop
CMU	**LTI Summer Seminar Series, July 2020**
	❯ Topic : Components of Neural Text Generation 🔗 Details
CMU	**Guest Lecture in Neural Networks for NLP, April 2020**
	❯ Topic : Natural Language Generation 🔗 Youtube Link
CMU	**Guest Lecture in Introduction to NLP, April 2020 & May 2021**
	❯ Topic : Natural Language Generation
Facebook	**Invited Research Talk, October 2019**
	❯ Topic : Multimodal Narrative Intelligence
Apple	**Guest Talk, August 2019**
	❯ Topic : Visual Story and Storyboard Generation
Univ of West Scotland	**Invited Talk, April 2021**
	❯ Topic : Content Anchoring In Multimodal Texts : A Prequel with PhD advice

👊 RESEARCH SERVICE

Organizer

> Workshop on Narrative Understanding at EMNLP 2024
> Workshop on Novel Ideas in Learning-to-Learn through Interaction at EMNLP 2023 🔗 *[link]*
> Workshop on 'Generation, Evaluation, and Metrics' at EMNLP 2023 🔗 *[link]*
> Workshop on Narrative Understanding at NAACL 2023 🔗 *[link]*
> Workshop on 'Generation, Evaluation, and Metrics' at EMNLP 2022 🔗 *[link]*
> Workshop on Narrative Understanding at NAACL 2022 🔗 *[link]*
> Widening NLP Workshop at EMNLP 2021 🔗 *[link]*
> Widening NLP at ACL 2020 🔗 *[link]*
> CMU LTI Open House Poster Session 2019

Co-chair

> Doctoral Symposium at AIMLSystems 🔗 *[link]*
> Academic Inclusion, Diversity & Inclusion at ACL 2020, ACL 2021, NAACL 2022, 🔗 *[link]*

Area Chair

> Neurips 2024, COLM 2024, ACL ARR Feb 2024, Apr 2024, June 2024,
> Generation and Summarization track at AACL 2023

Panel Moderator

> EACL 2021 (Topic : Low resource languages) 🔗 *[link]*
> WiNLP Satellite event at AACL 2020 (Topic : Social Biases and Future of NLP) 🔗 *[link]*
> Academic Inclusion at ACL 2020

PC Member

> Humanizing AI at IJCAI 2019 and 2018, 🔗 *[link]*

Conference Reviewer

> Neurips 2023, AAAI 2023, EMNLP 2022, COLING 2022, NAACL 2022, EMNLP 2021, EMNLP 2020, **ACL 2020 (Outstanding reviewer)**, EACL 2021, COLING 2020, AACL 2020, EACL 2020, AAAI 2020, ACL 2019, AAAI 2019, NAACL 2018 ; GenChal Proposals at INLG 2022

Workshop Reviewer

> GEM 2021, CALCS-NAACL'21, ALVR-ACL 2020, SRW-ACL 2020, WiNLP-ACL 2020, CALCS-ACL 2018, Code-Switching Workshop-Interspeech 2020

Journal Reviewer

> JAIR 2022, TALLIP 2020

Editor

> ML @ CMU Blog 🔗 *[link]*

🏆 ACADEMIC HONORS & AWARDS

> **Amazon Simbot Challenge** : Our team was among the 10 university teams to be selected to take part in the Simbot Challenge.
> **Rising Stars EECS 2020** : Selected and invited to Rising Stars 2020 at UC Berkeley 🔗 [link]
> **Shortlisted** and Nominated for **Apple Fellowship** 2020
> **Winners of BioAsq Challenge :** in the sixth edition of BioAsq Challenge 🔗 [link]
> **CMU LTI Graduate Fellowship :** Fully funded covering tuition, travel expenses & monthly stipend.
> **Amazon Alexa Prize Sponsorship :** Our team was amongst 12 others around the world to be awarded 100,000$ stipend by Amazon to compete in the Alexa Prize.
> **ISCA Grant** : The grant includes 1-year ISCA membership and cash amount to support travel costs.
> **S.N. Bose Scholarship 2015** : Awarded to top 1.5% students at national level to carry out summer research internship in the US. 🔗 [link]
> **Dean's Research Award -IIIT-H 2015** : Awarded for undergraduate students actively pursuing research.
> **Deans' Academic Merit List** : Received in Fall 2011, Spring 2012, Fall 2013, Spring 2014 in IIIT-H.
> **Best All Rounder Student IIIT-H 2015** : Awarded for consistent involvement and all round contributions in academics and extra-curricular activities.
> **Best Project in IISC Summer School** (2nd place)
> **Best Sports Woman - IIIT-H :** For 3 consecutive academic years 2012, 2013, 2014.
> **Best Outgoing Student** : Awarded in Sanghamitra School for all round performance.

Conferences

- *Certainly Uncertain : A Benchmark and Metric for Multimodal Epistemic and Aleatoric Awareness* (under review) 🔗 [paper]
- *DataComp-LM : In search of the next generation of training sets for language models* (under review) 🔗 [paper]
- *The Art of Saying No : Contextual Noncompliance in Language Models* (under review) 🔗 [paper]
- *WildBench : Benchmarking LLMs with Challenging Tasks from Real Users in the Wild* (under review) 🔗 [paper]
- *WildHallucinations : Evaluating Long-form Factuality in LLMs with Real-World Entity Queries* (under review)
- *Rewardbench : Evaluating reward models for language modeling* (under review) 🔗 [paper]
- *L3GO : Language Agents with Chain-of-3D-Thoughts for Generating Unconventional Objects* (under review) 🔗 [paper]
- *OLMO : : Accelerating the Science of Language Models* (**ACL 2024**) 🔗 [paper]
- *Dolma : An open corpus of three trillion tokens for language model pretraining research* (**ACL 2024**) 🔗 [paper]
- *The unlocking spell on base llms : Rethinking alignment via in-context learning* (**ICLR 2024**) 🔗 [paper]
- *The Generative AI Paradox:"What It Can Create, It May Not Understand"* (**ICLR 2024**) 🔗 [paper]
- *Lumos : Learning agents with unified data, modular design, and open-source llms* (ACL Findings 2024) 🔗 [paper]
- *Deal, or no deal (or who knows)? Forecasting Uncertainty in Conversations using Large Language Models* (ACL Findings 2024) 🔗 [paper]
- *Selective" Selective Prediction" : Reducing Unnecessary Abstention in Vision-Language Reasoning* (ACL Findings 2024) 🔗 [paper]
- *On the Role of Summary Content Units in Text Summarization Evaluation* (**NAACL 2024**) 🔗 [paper]
- *How Far Can Camels Go? Exploring the State of Instruction Tuning on Open Resources* (**NeurIPS 2023**) 🔗 [paper]
- *NovaCOMET : Open Commonsense Foundation Models with Symbolic Knowledge Distillation* (EMNLP Findings 2023) 🔗 [paper]
- *Localized Symbolic Knowledge Distillation for Visual Commonsense Models* (NeurIPS 2023) 🔗 [paper]
- *LOWRECORP : the Low-Resource NLG Corpus Building Challenge* (**INLG 2023**)
- *Maximizing Task Consistency Improves Continual Learning for Dialogue State Tracking* (**EMNLP 2023**)
- *A Needle in a Haystack : An Analysis of High-Agreement Workers on MTurk for Summarization* (**ACL 2023**)
- *Inference-Time Policy Adapters (IPA) : Tailoring Extreme-Scale LMs without Fine-tuning* (**EMNLP 2023**)
- *NovaCOMET : Open Commonsense Foundation Models with Symbolic Knowledge Distillation* (**EMNLP 2023**)
- *Curriculum Script Distillation for Multilingual Visual Question Answering* (arxiv)

📖 PUBLICATIONS

Conferences

› *"Denoising Large-Scale Image Captioning from Alt-text Data using Content Selection Models"* (**COLING 2022**)

› *Multilingual Multimodality : A Taxonomical Survey of Datasets, Techniques, Challenges and Opportunities* (under review)

› *Curriculum Script Distillation for Multilingual Visual Question Answering* (arxiv)📑 [paper]

› *GEMv2 : Multilingual NLG Benchmarking in a Single Line of Code* (EMNLP 2022 Demos) 📑 [paper]

› *Switch Point biased Self-Training : Re-purposing Pretrained Models for Code-Switching* (**EMNLP 2021**)

› *Grounding 'Grounding' in NLP* (**ACL 2021 Findings**) 📑 [paper]

› *The GEM Benchmark : Natural Language Generation, its Evaluation and Metrics* 📑 [paper]

› *"Positioning yourself in the maze of Neural Text Generation"* (*submitted, under review*)

› "*Reading Between the Lines : Exploring Infilling in Visual Narratives*" (**EMNLP 2020**)📑 [paper]

› *"Style Transfer as Vantage Point for Code-Switching"* (**InterSpeech 2020**) 📑 [paper]

› *"Storyboarding of Recipes : Grounded Contextual Generation"* at **ACL 2019.** 📑 [paper]

› *"Speech Synthesis for Mixed-Language Navigation Instructions"* **InterSpeech 2017**. 📑 [paper]

› *"Answer ka type kya he? Learning to classify questions in code-mixed language".* in International Conference on World Wide Web **WWW 2015** . 📑 [paper]

› *"WebShodh : A Code Mixed Factoid Question Answering System for Web"* in **CLEF 2017** 📑 [paper]

› *"Nee Intention enti?" Towards Dialog Act Recognition in Code-Mixed Conversations.* **IALP 2017**. 📑 [paper]

› *"Domain Adaptation in Morphological Analysis".* in International Journal of Languages, Literature and Linguistics 2015. 📑 [paper]

Journals

› *A Survey of Code-switched Speech and Language Processing* (under review) 📑 [paper]

Proceedings

› Proceedings of the Workshop on Novel Ideas in Learning-to-Learn through Interaction (NILLI 2023) 📑 [paper]

› Proceedings of the Third Workshop on Natural Language Generation, Evaluation, and Metrics (GEM) 📑 [paper]

› Proceedings of the The Fourth Workshop on Narrative Understanding

› Proceedings of the The Fifth Widening Natural Language Processing Workshop

› Proceedings of the The Fourth Widening Natural Language Processing Workshop 📑 [paper]

Theses

› *PhD Thesis Proposal : Effective Anchoring of Multimodal Narrative Generation* (CMU) 📑 [paper]

› *Code-Mixed Question Answering for Indian Languages* (**Research thesis at IIIT-H**)

📖 PUBLICATIONS

Workshops

- *CodemixedNLP : An Extensible and Open NLP Toolkit for Code-Mixing* in CALCS at NAACL 2021. ☑ [paper]
- *Entity Skeletons for Visual Storytelling* in LVVU at CVPR 2020, ALVR at ACL 2020 . ☑ [paper]
- *"My Way of Telling a Story" : Persona based Grounded Story Generation* in StoryNLP at ACL 2019. ☑ [paper]
- *"Comparative Analysis of Neural QA models on SQuAD"* in MRQA at ACL 2018. ☑ [paper]
- *"Code-Mixed Question Answering Challenge : Crowd-sourcing Data and Techniques"* in CALCS at ACL 2018. ☑ [paper]
- *"Language Informed Modeling of Code-Switched Text."* in CALCS at ACL 2018. ☑ [paper]
- *"Tackling Code-Switched NER : Participation of CMU."* in CALCS at ACL 2018. ☑ [paper]
- *"Tackling Biomedical Text Summarization : OAQA at BioASQ 5B."* in BioNLP at ACL 2017. ☑ [paper]
- *Ontology-Based Retrieval & Neural Approaches for BioAsq Ideal Answer Generation* in Bio-Asq at EMNLP 2018. [paper]
- *Extraction Meets Abstraction : Ideal Answer Generation for Biomedical Questions* in BioAsq at EMNLP 2018. ☑ [paper]

Technical Reports

- Amazon Alexa team *"Building CMU Magnus from User Feedback".* in Amazon Alexa Prize Summit 2017. ☑ [Report]
- *"Textually Enriched Neural Module Networks for Visual Question Answering".* ☑ [Arxiv]
- *"Entity Skeletons as Intermediate Representations for Visual Storytelling".* at BayLearn 2020 ☑ [link]

Book

- *"Language Technologies for Code-Switching : an account of sequence, semantic and generation tasks".* online soon

✏️ TEACHING

Deep Learning for NLP (LLMs and beyond) – Visiting Professor	Jio Institute, 2024
Question Answering (TA)	CMU, Spring 2018
Design & Engineering of Intelligent Information Systems (TA)	CMU, Fall 2017
Graphics (TA)	IIIT-H, Spring 2016 & Spring 2014
Natural Language Processing (TA)	IIIT-H, Fall 2015
Natural Language Applications (TA)	IIIT-H, Spring 2015
Structured Systems Analysis and Design (TA)	IIIT-H, Fall 2014
Science (TA)	IIIT-H, Fall 2013