

Shehzaad Dhuliawala

BIO My research focuses on building reasoning systems that are factual, well-calibrated, and reliable. I aim to develop models that can perform effectively and generalize across several tasks.

Areas of research: Factuality, Reasoning, Alignment, Calibration

✉ shehzaadzd at gmail.com 🏠 shehzaadzd.github.io 🌐 @shehzaadzd

🔍 Google Scholar (Citations: 1400+) 🌐 @shehzaadzd

EDUCATION

ETH Zürich

PhD., Computer Science, 2020-2025 (expected)

Advisors: Prof. Mrinmaya Sachan, Prof. Thomas Hofmann

University of Massachusetts Amherst

M.S., Computer Science, 2016-2018

Advisor: Prof. Andrew McCallum

Visvesvaraya National Institute of Technology,

B.Tech., Computer Science and Engineering

AWARDS

IBM PhD Fellowship Award 2021

RESEARCH & PROFESSIONAL EXPERIENCE

Research Intern

May 2024 - Sept 2024

FAIR, Meta AI, New York

Mentor: Jason Weston

-Developed an [adaptive decoding](#) approach that learns how to choose the best decoding hyperparameters for the LLM based on the input improving performance compared to fixed hyperparameters.

Research Intern

June 2023 - Sept 2023

FAIR, Meta AI, New York

Mentor: Jason Weston

-Developed the [Chain-of-Verification](#) approach where the LLM uses verification questions to mitigate hallucinations and improve factuality for long-form generations.

Research Engineer

June 2018 - Oct 2020

Microsoft Research Montréal

-Worked on adapting question-answering systems to specialized domains such as product manuals

-Researched novel methods to train a system to answer user queries against arbitrary documents using limited amounts of annotated data

SKILLS & FRAMEWORKS

Python, Pytorch, Huggingface Transformers, Torchtune, Git

RECENT AND
RELEVANT
PUBLICATIONS

**Adaptive Decoding via Latent Preference Optimization
(Preprint 2024)**

Shehzaad Dhuliawala, Ilya Kulikov, Ping Yu, Asli Celikyilmaz, Jason Weston, Sainbayar Sukhbaatar, Jack Lanchantin

**Towards Aligning Language Models with Textual Feedback
(EMNLP 2024)**

Sac Abadal Lloret, Shehzaad Dhuliawala*, Keerthiram Murugesan, Mrinmaya Sachan.*

**Chain-of-Verification Reduces Hallucinations in Large Language Models
(ACL 2024 Findings)**

Shehzaad Dhuliawala, Mojtaba Komeili, Jing Xu, Roberta Raileanu, Xian Li, Asli Celikyilmaz, Jason Weston.

**A Diachronic Perspective on User Trust in AI under Uncertainty
(EMNLP 2023 Oral Top 6%)**

Shehzaad Dhuliawala, Vilém Zouhar, Mennatallah El-Assady, and Mrinmaya Sachan.

**Variational Classification: A Probabilistic Generalization of the Softmax Classifier
(TMLR 2023)**

Shehzaad Dhuliawala, Mrinmaya Sachan, Carl Allen

**Calibration of Machine Reading systems at Scale
(ACL 2022 findings)**

Shehzaad Dhuliawala, Leonard Adolphs, Rajarshi Das, and Mrinmaya Sachan.

**Case-based Reasoning for Better Generalization in Text-Adventure Games
(ICLR 2022)**

Mattia Atzeni, Shehzaad Dhuliawala, Keerthi Muruganesan, Mrinmaya Sachan

**TopiOCQA: Open-domain Conversational Question Answering with Topic Switching
(TACL 2021)**

Vaibhav Adhlaka, Shehzaad Dhuliawala, Kaheer Suleman, Harm de Vries, Siva Reddy

OTHER
PUBLICATIONS

**Implicit Personalization in Language Models: A Systematic Study
(EMNLP 2024 Findings)**

Zhijing Jin, Nils Heil, Jiarui Liu, Shehzaad Dhuliawala, Yahang Qi, Bernhard Scholkopf, Rada Mihalcea, Mrinmaya Sachan.

How to Query Language Models?

Leonard Adolphs, Shehzaad Dhuliawala, and Thomas Hofmann

Preprint 2021

A Simple Approach to Case-Based Reasoning in Knowledge Bases

(Accepted at **AKBC 2020**, Best paper runner-up)

Rajarshi Das, Ameya Godbole, Shehzaad Dhuliawala, Manzil Zaheer, Andrew McCallum

Multi-step Retriever-Reader Interaction for Scalable Open-domain Question Answering

(**ICLR 2019**)

Rajarshi Das, Shehzaad Dhuliawala, Manzil Zaheer, Andrew McCallum

Towards Domain Adaptation from Limited Data for Question Answering (ASRU 2019 Life-long Learning Workshop)

T.J Hazen, Shehzaad Dhuliawala, Daniel Boies

Go for a Walk and Arrive at the Answer: Reasoning Over Paths in Knowledge Bases using Reinforcement Learning

(Accepted at **ICLR 2018**, Won best paper at **AKBC** at **Neurips 2017**)

Shehzaad Dhuliawala^{}, Rajarshi Das^{*}, Manzil Zaheer, Luke Vilnis, Ishan Durugkar, Akshay Krishnamurthy, Alex Smola, Andrew McCallum*