ELLIE Y. CHENG

ellieyhc@csail.mit.edu
elliecheng.com
github.com/ellieyhcheng

EDUCATION

MASSACHUSETTS INSTITUTE OF TECHNOLOGY — 2022 - PRESENT
PH.D. IN COMPUTER SCIENCE
RESEARCH FOCUS: PROGRAMMING SYSTEMS WITH AND FOR LLMS
ADVISED BY MICHAEL CARBIN

MASSACHUSETTS INSTITUTE OF TECHNOLOGY — 2024
S.M IN ELECTRICAL ENGINEERING AND COMPUTER SCIENCE
Thesis: Inference Plans for Hybrid Probabilistic Inference
ADVISED BY MICHAEL CARBIN

UNIVERSITY OF CALIFORNIA, LOS ANGELES — 2018 - 2022
B.S. IN COMPUTER SCIENCE AND ENGINEERING

HONORS AND AWARDS

COMPUTING RESEARCH ASSOCIATION (CRA) OUTSTANDING UNDERGRADUATE RESEARCHER AWARD HONORABLE MENTION — 2022

INVITED TALKS

FLIP-HOISTING: A PROBABILISTIC PROGRAM OPTIMIZATION FOR EXACT INFERENCE — 2021
THE INTERNATIONAL CONFERENCE ON PROBABILISTIC PROGRAMMING (PROBPROG)

CONFERENCE PUBLICATIONS

SHARING STATE BETWEEN PROMPTS AND PROGRAMS — 2026
ELLIE Y. CHENG, LOGAN WEBER, TIAN JIN, MICHAEL CARBIN
INTERNATIONAL CONFERENCE ON LEARNING REPRESENTATIONS (ICLR)
https://arxiv.org/abs/2512.14805

PLANNED DIFFUSION — 2026
DANIEL ISRAEL*, TIAN JIN*, ELLIE CHENG, GUY VAN DEN BROECK, ADITYA GROVER, SUVINAY SUBRAMANIAN, MICHAEL CARBIN
INTERNATIONAL CONFERENCE ON LEARNING REPRESENTATIONS (ICLR)
https://arxiv.org/abs/2510.18087

LEARNING TO KEEP A PROMISE: SCALING LANGUAGE MODEL DECODING PARALLELISM WITH LEARNED ASYNCHRONOUS DECODING 2025

TIAN JIN*, ELLIE Y. CHENG*, ZACK ANKNER, NIKUNJ SAUNSHI, BLAKE M. ELIAS, AMIR YAZDANBAKHSH, JONATHAN RAGAN-KELLEY, SUVINAY SUBRAMANIAN, MICHAEL CARBIN

INTERNATIONAL CONFERENCE ON MACHINE LEARNING (ICML)

https://arxiv.org/abs/2502.11517

INFERENCE PLANS FOR HYBRID PARTICLE FILTERING 2025

ELLIE Y. CHENG, ERIC ATKINSON, GUILLAUME BAUDART, LOUIS MANDEL, MICHAEL CARBIN

PRINCIPLES OF PROGRAMMING LANGUAGES (POPL)

https://arxiv.org/abs/2408.11283

HOW CAN I EXPLAIN THIS TO YOU? AN EMPIRICAL STUDY OF DEEP NEURAL NETWORK EXPLANATION METHODS 2020

JEYA VIKRANTH JEYAKUMAR, JOSEPH NOOR, YU-HSI CHENG, LUIS GARCIA, AND MANI SRIVASTAVA

ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS (NEURIPS)

NON-ARCHIVAL PUBLICATIONS

EXPRESSING AND EXPLOITING PARALLELISM IN LANGUAGE MODEL DECODING 2024

TIAN JIN*, ELLIE Y. CHENG*, MICHAEL CARBIN

WORKSHOP ON LARGE LANGUAGE MODEL (LLM) AGENTS, ICLR

VERIFYING PERFORMANCE PROPERTIES OF PROBABILISTIC INFERENCE 2023

ERIC ATKINSON, ELLIE Y. CHENG, GUILLAUME BAUDART, LOUIS MANDEL, MICHAEL CARBIN

THE WORKSHOP ON VERIFICATION OF PROBABILISTIC PROGRAMS (VERIPROP)

https://arxiv.org/abs/2307.07355

FLIP-HOISTING: A PROBABILISTIC PROGRAM OPTIMIZATION FOR EXACT INFERENCE 2021

YU-HSI CHENG, STEVEN HOLTZEN, GUY VAN DEN BROECK, TODD MILLSTEIN

THE INTERNATIONAL CONFERENCE ON PROBABILISTIC PROGRAMMING (PROBPROG)

https://elliecheng.com/publications/ChengPROBPROG21.pdf

Extended draft: https://arxiv.org/abs/2110.10284

EXPERIENCE

BASIS AI JUN - AUG 2024

RESEARCH INTERN

STRIPE JUN - SEP 2022

SOFTWARE ENGINEERING INTERN

META PLATFORMS SEP - DEC 2021

SOFTWARE ENGINEERING INTERN

META PLATFORMS JUN - SEP 2020

SOFTWARE ENGINEERING INTERN

STATISTICAL AND RELATIONAL ARTIFICIAL INTELLIGENCE LAB, UCLA — JAN 2020 - JUN 2022

UNDERGRADUATE RESEARCH ASSISTANT

NETWORKED & EMBEDDED SYSTEMS LAB, UCLA — OCT 2019 - SEP 2021

UNDERGRADUATE RESEARCH ASSISTANT