Eugene **Choi**

530 East 13th Street, #9, New York, NY 10009, USA

(+1) 929-288-7648 | eugene.choi@nyu.edu | eugene-choi.github.io | eugene-choi | choi-eugene | Google Scholar

Summary & Technical Skills

Machine Learning engineer/researcher specializing in natural language processing (NLP) and generative models at NYU CILVR & ML^2 lab.

- **Languages**: Python, Java, C, C++, SQL, Matlab
- **Machine Learning/Data Analysis**: PyTorch, Hugging Face, fairseq, sklearn, XGBoost, NumPy, TensorFlow, JAX, pandas, Matplotlib
- **Cloud Services/Others**: AWS, GCP, Linux, Unix, Bash, Git, Vim, CUDA, OpenMP, MPI, LaTeX, Docker, Kubernetes

Education

New York University
New York, NY, USA

M.S. In Computer Science — *Jan. 2022 - Dec. 2022*

- GPA: 3.88/4.00

New York University
New York, NY, USA

B.A. In Computer Science with a minor in Mathematics — *Sep. 2014 - Dec. 2021*

- Last 2 years of GPA: 3.86/4.00
- Dean's List for Academic Year (2019-2020, 2020-2021)
- Leave of absence 2016-2019 due to mandatory military service (Sergeant in Republic of Korea Air Force at Seongnam base)

Work & Research Experience

NYU CILVR Lab
New York, NY, USA

Deep Learning Research Assistant — *Nov. 2021 - Present*

- Deep generative model research advised by Prof. Kyunghyun Cho.
- Improved representation learning using *multi-vector representation* in image segmentation, sequence completion, and machine translation.

Machine Learning for Language (ML^2) Lab
New York, NY, USA

Natural Language Processing (NLP) Research Assistant — *May 2022 - Dec. 2022*

- Efficient adaptation of pre-trained language models research advised by Prof. Sam Bowman.
- Improved current parameter-efficient fine-tuning methods (e.g. adapters, LoRA, and prefix/prompt tuning) to adapt large, pre-trained language models to downstream tasks more efficiently, using T5 and GPT-3 models with PyTorch, Hugging Face and OpenAI libraries.

Dream Security
Seoul, South Korea

Software Engineering Intern — *Jul. 2020 – Aug. 2020*

- Assisted in the development of an authentication product using Decentralized Identifiers (DIDs) for BC Cards using the Spring Framework.

Publication

A Non-monotonic Self-terminating Language Model
In Proceedings of ICLR 2023

(with Dr. Cheolhyoung Lee and Prof. Kyunghyun Cho)

- Investigated why autoregressive neural sequence models produce infinite-length sequences when such models are paired with approximate decoding algorithms, and presented the necessary conditions to mitigate such issues.
- Proposed the "non-monotonic self-terminating language model," a new method which reformulates the softmax function to guarantee the *consistency* of any language model using *incomplete probable decoding algorithms*.
- Conducted experiments in *sequence completion* with WikiText dataset using RNN, LSTM and GPT-2 to validate the effectiveness of our method.

Teaching Experience

African Master's Program in Machine Intelligence (AMMI) 2022 - Deep Learning for NLP
Senegal (Virtual)

Teaching Assistant (Lecturer: Prof. Kyunghyun Cho) — *May 2022*

- Led the decoding algorithms fundamentals lecture.
- Prepared greedy/beam search and ancestral/top-k/nucleus sampling implementations on Google Colab.

DS-UA 203 / LING-UA 52: Machine Learning for Language Understanding
New York, NY, USA

Section Leader / Teaching Assistant / Grader (Lecturer: Prof. Sam Bowman) — *Jan. 2022 – May 2022*

- Led the weekly lab session focusing on NLP algorithm implementations and advised students on their semester research projects.