

# MAHA ALKHAIRY

[www.linkedin.com/in/malkhairy](https://www.linkedin.com/in/malkhairy) ◊ <https://github.com/mahaalkh>

## INTERESTS

Research in modeling linguistic rules as computational linguistics algorithms using a combination of modern and classical approaches.

Research in machine learning and deep learning, natural language processing, computational linguistics and computational psycholinguistics; with applications to education and health domains

## RESEARCH EXPERIENCE AND PUBLICATIONS

### Natural Language Logic Inference System for Basic LSAT

*September 2022 - May 2023*

*Supervisor:* Professor Brendan O'Connor (CS, UMASS Amherst) and Professor Vincent Homer (Linguistics, UMASS Amherst)

*Other Advisors:* Professor Neil Immerman (CS, UMASS Amherst) and Professor Katrin Erk (CS, UT Austin)

*Objective:* Encode formal semantics rules to get the semantic parse of a sentence, and use logic inference tools to infer answers to LSAT logic games using the semantics parses

*Tools Used:* Python

*Publication:* TBA

### Lexicon Refinement, Expansion, and Contextualization Utilizing BERT

*September 2021 - May 2022*

*Supervisor:* Professor Brendan O'Connor and Professor Laure Thompson, CS, UMASS Amherst

*Objective:* Expand, refine and contextualize a keyword lexicon using a corpus and BERT embeddings

*Tools Used:* Python

*Publication:* TBA

### An Integrated, Bidirectional Pronunciation, Morphology, and Diacritics System

*January - September 2020*

*Supervisor:* Professor David A. Smith, CS, NEU, Boston, MA ; Professor Adam Cooper, Linguistics, NEU; Boston, MA

*Objective:* Construct an integrated pipeline with bidirectional finite state transducers that can be used as stand alone to

diacritize and morphologically analyze words; and bidirectionally transform orthography to IPA and from IPA to orthography

*Tools Used:* Python, Foma

*Publication:* Preprint is on ssrn [\[LINK\]](#)

### Clinical Note Segments Coloring and Document Multi-Labeling

*October 2018 - December 2019*

*Objective:* Develop a model using recurrent neural networks which can identify salient segments in a medical note and label them according to ICD9 categories

*Tools Used:* Pytorch, Python, Google Colab

*Publication:* Preprint is on arXiv [\[LINK\]](#)

### Linear and Nonlinear Morphological Analyzer

*September 2016 - November 2019*

*Supervisor:* Professor David A. Smith, Computer Science, Northeastern University, Boston, MA

*Objective:* Develop a concatenative and templatic finite state machine morphologizer that analyzes a Semitic word into its morphemes and composes a word from its patterns, roots, and affixes

*Tools Used:* Python, Foma

*Publication:* Proceedings of the Language Resources and Evaluation Conference (LREC), 2020 [\[LINK\]](#)

### Rule based Text to Phonetic Transcription System

*June 2015 - September 2016*

*Supervisor:* Professor Adam Cooper; Linguistics, Northeastern University; Boston, MA

*Objective:* Construct pronunciation rules using regular expressions and engineer a reversible automata to translate between text and phonetic transcription

*Tools Used:* Python, Foma

## EMPLOYMENT

---

### Computational Research Assistant

Unnamed Startup

May 2023 - August 2023

### Computational Research Assistant

Unnamed Startup

May 2022 - August 2022

### Language Consultant

Linguistics, University of Massachusetts, Amherst, MA,

Language consultant for Fields Methods class in which Arabic phrases are elicited from me. Assist in designing the orthography used for data collection.

January 2022 - May 2022

### Research Assistant

Natural Language processing Group, University of Massachusetts, Amherst, MA,

<https://nlp.cs.umass.edu/members/>

Under Professor Laure Thompson with collaboration with Brendan O'Connor

September 2021 - May 2022

### Research Assistant

Digital Scholarship Group, Northeastern University, Boston, MA,

<https://dsg.northeastern.edu/staff/staffbio/?smid=4684>

Cherokee Language Project

Project is aimed towards contributing to a learning library for Cherokee

Responsibilities include:

September 2018 - June 2020

- Create and test finite state transducers that produce the correct phonological surface-form of the word
- Create and test finite state transducers that map from the ambiguous Cherokee Syllabary to the phonological form of the word
- Assist in creating, analyzing, and annotating a lexical database of Cherokee

### Tutor

College of Computer and Information Science, Northeastern University, Boston, MA

*Class:* Fundamentals for Computer Science

*Term:* 2015 Summer I session

*Responsibilities:* Lead labs in absence of head TA, Hold office hours, Grade homework

May - June 2015

### Tutor

College of Computer and Information Science, Northeastern University, Boston, MA

*Class:* Theory of Computation

*Term:* 2015 Spring Semester

*Responsibilities:* Respond to Piazza questions, One-on-one tutoring sessions online or in person

January - May 2015

## SKILLS AND CERTIFICATIONS

---

**Programming Languages:** Python, Foma, R, Prolog-SWI

**Tools:** Tensorflow, PyTorch, Keras, sklearn, numpy, pandas, jupyter lab, google colab

**Certifications:** Neural Networks and Deep Learning, Prolog

## EDUCATION

---

### University of Massachusetts - Amherst

Boston, MA, USA  
College of Information and Computer Sciences

MS/PhD track

Expected to receive Masters of Science in Computer Science in Spring 2024

Synthesis Project : Fall 2022 - Spring 2023 "Natural Language Logic Inference System for Basic LSAT"

GPA : 3.9 / 4.0

Relevant Coursework:

September 2021 - present

Computer Science	Advanced Logic in Computer Science, Advanced Natural Language Processing, Advanced Information Assurance, Advanced Algorithms, Natural Language Processing Seminar
Linguistics	Introduction to Semantics, Computational Linguistics: Use and Meaning, Semantics and Generative Grammar, Introduction to Transformational Grammar

**Northeastern University** Boston, MA, USA  
College of Computer and Information Science (CCIS)  
*Bachelor of Science degree in Computer Science with a minor in Mathematics*  
Honors: Cum Laude | Dean's List (Fall 2014, Spring 2016, Fall 2016, Spring 2017)  
Clubs: NUWiT, NUACM, Linguistics club, Math club

*September 2013 - May 2017*

Relevant Coursework:

Computer Science	Natural Language Processing, Machine Learning, Artificial Intelligence, Algorithms and Data, Theory of Computation, Software Development, Human Computer Interaction, Data Mining Techniques
Mathematics	Statistics and Stochastic Processes, Probability and Statistics, Linear Algebra, Multivariate Calculus, Calculus
Other	Introduction to Language and Linguistics, Foundations of Psychology

## VOLUNTEER

**Vice President of a Graduate Student Organization**  
UMASS Amherst, Amherst, MA

*September 2022-Present*

**PhD Peer Mentor**  
College of Computer and Information Science, UMASS Amherst, Amherst, MA

*September 2022-Present*

**Young Alumni Advisory Board Member**  
Northeastern University, Boston, MA

*August 2021 - August 2023*

**NLP reading group organizer**  
College of Computer and Information Science, UMASS Amherst, Amherst, MA

*January 2022 - May 2022*

## SEMINARS AND PROGRAMS

**Computational Psycholinguistics Course**

Type: Course (attended as a listener)

Location: Massachusetts Institute of Technology, Cambridge, MA

Website: <https://learning-modules.mit.edu/class/index.html?uuid=/course/9/sp20/9.19#info>

*February 2020 - June 2020*

- Methods ranging from finite state automata to context-free and mildly context-sensitive grammars to model and analyze phonetics, morphology, syntax, semantics, and pragmatics
- Language models such as n-grams and probabilistic modeling with Bayes Net
- Apply word embedding, use logistic regression, and deep neural networks to model the syntax, semantics, and pragmatics of language and use it for classification and parsing of text
- Model human language acquisition

**Machine Learning Seminar at EECE Department**

Type: Seminars

Location: Northeastern University, Boston, MA

*May 2017 - August 2018*

- Critical reading of literature in deep learning, optimization, clustering and other fields within machine learning
- Discuss applications of methods such as Variational Inference, Variational Autoencoding, and Generative Adversarial Networks

**Probabilistic Programming for Advancing Machine Learning**

Type: Summer School

Location: Arlington, VA

*July 24 – August 4, 2017*

- Learnt Probabilistic Programming, inference and its applications by Gamalon and MIT teams
- Designed probabilistic models to model data sets and infer properties about the data
- Utilized BayesDB to probabilistically extract information about childrens' speech based on demographic or features