

# Yang Bai, Ph.D.

tony.bai.yang@gmail.com | +1 (352)284-2346 | Website | LinkedIn

## SUMMARY

Researcher and engineer designing Machine Learning algorithms for multi-model applications in Information Retrieval and Natural Language Processing.

## EDUCATION

### UNIVERSITY OF FLORIDA

Gainesville, FL, USA

2019 - 2024

PH.D. COMPUTER SCIENCE

GPA: 3.96/4.0

2016-2018

MSC. ELECTRICAL AND COMPUTER  
ENG.

GPA: 3.5/4.0

### SICHUAN UNIVERSITY

2012-2016 | Chengdu, China

BSC. MICRO ELECTRONICS

## SKILLS

### LANGUAGES

Python: *Expert*

Java: *Expert*

SQL/SPARQL: *Expert*

C/C++: *Intermediate*

JavaScript: *Intermediate*

F#: *Intermediate*

### TOOLS

PyTorch TensorFlow

Keras Scikit-learn

Transformers NLTK

SciPy Pillow

OpenCV OpenE

Matlab NumPy

Pandas Oracle DB

REST API Flask

Docker Akka.NET

Git Linux

Google Test JUnit

## COURSEWORKS

Elements of Machine Intelligence

Deep Learning for Computer-

Graphics

Applied Machine Learning

Trustworthy Machine Learning

Distributed Operating System

Programming Language Principles

Database Management System

Database System Implementation

Analysis of Algorithms

Advanced Data Structures

Computer Networks

## WORK EXPERIENCE

### Meta

Research Scientist | 2025 Feb - Present

- Develop search retrieval and ranking models using state-of-the-art machine learning technologies to empower the Meta family of applications.

### Meta

ML SDE Intern | Summer 2024

- Developed advanced search retrieval and ranking models using state-of-the-art ML technologies as part of the Threads Search and Ranking Engineering team, enhancing the app's functionality.

### Nokia Bell Labs

Machine Learning Intern | Summer 2022

- Proposed and implemented a retrieval-based framework to streamline ticket root cause analysis by extracting relevant log lines (10-100M/ticket). Developed a dense log retrieval system with contrastive learning, achieving a 16.1% improvement over the BM25 baseline.

### University of Florida

Graduate Student Researcher | 2019-2024

- Developed a top-ranked recursive multi-hop dense retrieval system for FEVER, introducing a novel dense sentence representation learning method for open-domain fact verification. [\[Link\]](#)
- Led the creation of a benchmark dataset for multi-answer open-domain QA and stance mining. Developed an annotation tool and baseline system to evaluate key components, significantly improving performance. [\[Link\]](#)
- Led and contributed to DARPA's AIDA project, developing a top-performing event-centric knowledge graph search engine, ranked first in NIST TAC SM-KBP2020. [\[Link\]](#)

## SELECTED PUBLICATIONS [Google Scholar](#)

RAMQA: A UNIFIED FRAMEWORK FOR RETRIEVAL-AUGMENTED MULTI-MODAL QUESTION ANSWERING

Yang Bai, C.Grant, and D.Wang | NAACL Findings 2025

M3: A MULTI-TASK MIXED-OBJECTIVE LEARNING FRAMEWORK FOR OPEN-DOMAIN MULTI-HOP DENSE SENTENCE RETRIEVAL

Yang Bai, A.Colas, and D.Wang | COLING 2024

MYTHQA: QUERY-BASED LARGE-SCALE CHECK-WORTHY CLAIM DETECTION THROUGH MULTI-ANSWER OPEN-DOMAIN QUESTION ANSWERING

Yang Bai, A.Colas, and D.Wang | SIGIR 2023

CAN KNOWLEDGE GRAPHS SIMPLIFY TEXT?

A.Colas, H.Ma, X.He, Yang Bai, and D.Wang | CIKM 2023

MORE THAN READING COMPREHENSION: A SURVEY ON DATASETS AND METRICS OF TEXTUAL QUESTION ANSWERING

Yang Bai, D.Wang | arXiv 2021