Yang Bai

baiyang94@ufl.edu | 352 284 2346 | Website | Linkedin

SUMMARY

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

EDUCATION

UNIVERSITY OF FLORIDA

Gainesville, FL, USA
2019 - 2024(Expected)
PhD. 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	OpenIE
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

University of Florida — **Graduate Student Researcher | Sep. 2019-present**

- Developed an innovative recursive multi-hop dense sentence retrieval system with a new approach for dense sentence representation learning, which underpinned the creation of an open-domain fact verification system that attained the top ranking on the FEVER leaderboard. [Link]]
- Led the creation of a benchmark dataset for a novel open-domain question-answering task, featuring multi-answer options and controversial stance mining. Crafted a user-friendly annotation tool, along with a baseline system for the new task. This system played a crucial role in assessing the impact of various components, such as information retrieval, machine reading comprehension, distinct answer selection, and stance detection, leading to a significant enhancement in the project's overall performance. [Link]]
- Individual contributor and team lead in the DARPA-sponsored project "Active Interpretation of Disparate Alternatives(AIDA)", an alternative hypotheses search engine over event-centric knowledge graphs. Our system achieved top performance at the NIST TAC SM-KBP2020 evaluation. [Link]]

Nokia Bell Labs — **Machine Learning Intern | Jun. 2022-Aug. 2022**

- Proposed and implemented a retrieval-based framework to ease ticket root cause analysis by retrieving the most relevant log lines from the attached log files (10-100M log lines/ticket) given ticket information.
  - Conducted data cleaning, processing, visualization, and analysis on massive time-series semi-structured system-level log corpus.
  - Developed a dense log retrieval system that finetunes self-pretrained tickets and log encoders through a contrastive learning framework.
  - The best model outperforms a BM25 baseline model by 16.1%.

SELECTED PUBLICATIONS **Google Scholar**

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

M3: A Multi-Task Mixed-Objective Learning Framework for Open-Domain Multi-Hop Dense Sentence Retrieval
Yang Bai, A.Colas, and D.Wang | 2023 under submission

More Than Reading Comprehension: A Survey on Datasets and Metrics of Textual Question Answering
Yang Bai,D.Wang | arXiv 2021

GAIA at SM-KBP 2020 - A Dockerized Multi-media Multi-lingual Knowledge Extraction, Clustering, Temporal Tracking and Hypothesis Generation System
M.Li,...,**Yang Bai**,..., D.Wang | TAC 2020