

Edward Kempa

+1 (813) 526-6335 | 2306 Briana Dr, Brandon, FL 33511
 kempaedward@outlook.com | <https://www.linkedin.com/in/edward-kempa/>

EDUCATION

Master of Science in Artificial Intelligence in Biomedical and Health Sciences Dec 2027
 University of Florida, Gainesville, FL

Bachelor of Science in Computer Science May 2026
 University of Florida, Gainesville, FL
Relevant Coursework: Natural Language Processing; Mathematics for Intelligent Systems;
 Machine Learning

Bachelor of Arts in Economics May 2026
 University of Florida, Gainesville, FL

Minor: Business Administration, German
Cumulative GPA: 4.00/4.00

EXPERIENCE

Research Intern at Forschungszentrum Jülich (Jülich Research Center) May 2026 – Jul 2026

- *Ontology Powered Knowledge Graph RAG for Clinical Report Generation.*
- Jülich, North Rhine-Westphalia, Germany — Supervisor: Dr. Ravichandran Rajkumar

Student Research Assistant at University of Florida NLP and Culture Lab October 2025 – May 2026

- *Ambiguity Detection and Classification in Natural Language Inference.*
- Gainesville, Florida, United States — Supervisor: Dr. Bonnie Dorr

Summer Research Intern at Forschungszentrum Jülich May 2025 – Jul 2025

- *Interpretable MRI Based Classification and Report Generation.*
- Jülich, North Rhine-Westphalia, Germany — Supervisor: Dr. Ravichandran Rajkumar

Summer Research Intern at Exaia Technologies May 2024 – Jul 2024

- *Mental Health Detection via Social Media Data.* (https://youtu.be/K_ypETOMeA)
- Aachen, North Rhine-Westphalia, Germany — Supervisor: Dr. Daniel Wiechmann

Undergraduate Student Researcher in Computational Reasoning Group Jan 2023 – May 2026

- *Deconstructing engineering students' reasoning through a consequentialist AI Ethics-Framework*
- *Exploring the Career Trajectories of Engineering and Computer Science Students*
- Gainesville, Florida, United States — Supervisor: Ashish Aggarwal

RESEARCH

Ontology Powered Knowledge Graph RAG for Clinical Report Generation May 2026 – Present

- **Extending** RAG-based clinical report generation pipeline through knowledge graph integration.
- **Developed** a task specific ontology powering patient level graphs improving abstract retrieval.
- **Prepared** a large corpus of NLI data for human annotation under an updated taxonomy.

Mechanistic Interpretability for Natural Language Inference September 2025 – December 2025

- **Adapted** LAPE algorithm to NLI, identifying neurons for entailment, contradiction, and neutral.
- **Evaluated** additive neuron boosting and ablation interventions to steer model predictions.

Mental Health Detection via Social Media Data May 2024 – October 2025

- **Constructed** a robust database of Reddit data with over 44K users among ADHD, Autism, Anxiety, Bipolar, Depression, OCD, PTSD, and over 108K Control users.
- **Implemented** updated pattern matching/exclusion criteria to identify diagnosed/control users.
- **Conducted** thorough data preprocessing removing NSFW, non-English, and other artifacts.
- **Trained** shallow ML models, BiLSTM, Pretrained Language Models, and LLMs for classification.

Interpretable MRI Based Classification and Report Generation May 2025 – July 2025

- **Evaluated** the predictive capacity of different brain regions in a three-way classification task working with MRI data from UK Biobank for three condition groups (K50+K51, K58, Control).
- **Developed** stacked ML Model architecture leveraging volume, surface, and demographic data.

- **Constructed** a Retrieval Augmented Generation (RAG) Model for taking two time points of patient MRI scans and clinical findings reports with data collected locally, ultimately generating an interpretable report explaining changes with consideration to relevant research journal abstracts.

National Institutes of Health PREPARE Challenge - Phase 2: Acoustic Track December 2024

- Goal to improve early prediction of Alzheimer's disease and related dementias (AD/ADRD).
- **Trained** BERT, RoBERTa, and Mental RoBERTa models with acoustic and speech biomarkers.
- **Generated** final predictions leveraging a Stacking Model of our best RoBERTa Model and a Tuned Fusion Model.

Modeling Engineering Graduates Career Trajectories Oct 2023, Jan – Nov 2024, Feb – Mar 2025

- **Served** a request from the Dean of the University of Florida's (UF) College of Engineering to offer insight into the creation of a program at UF's future Jacksonville location.
- **Classified** 3,400+ individuals based on their major, degrees attained, and current careers.
- **Conducted** a qualitative analysis of the career outcomes of UF engineering graduates.
- **Writing** a paper on our work to be published to the Journal of Engineering Education.

AI Ethics-Framework Research January 2023 – May 2023, March 2024

- **Investigated** human decision-making's potential to inform algorithmic decisions in AI systems.
- **Conducted** a qualitative research analysis of survey data gathered from two semesters of an introductory programming course at the University of Florida.
- **Assessed** respondents' perception of the implementation of an algorithm in self-driving cars and how their choices/reasoning changed in response to minor changes to the consequences presented.

PUBLICATIONS

- Wiechmann, D., Kerz, E., **Kempa, E.**, & Qiao, Y. (2026). *ADHD-Lang: A large-scale social media dataset for verbal behavior and digital phenotyping in adult ADHD*. In Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026) (pp. 7279–7291). European Language Resources Association (ELRA). <https://doi.org/10.63317/3uw9c6ux3f7j>
- Mankarious, S., **Kempa, E.**, Wiechmann, D., Kerz, E., Qiao, Y., & Zirikly, A. (2026). *MindSET: Advancing mental health benchmarking through large-scale social media data*. In Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026) (pp. 11241–11251). European Language Resources Association (ELRA). <https://doi.org/10.63317/4cdunjg3bziz>
- Kempa, E.**, & Aggarwal, A. (2025). *Deconstructing engineering students' ethical reasoning: A consequentialist framework for algorithmic choices*. In Proceedings of the 30th ACM Conference on Innovation and Technology in Computer Science Education (ITiCSE 2025) (Vol. 1, pp. 597–603). Association for Computing Machinery. <https://doi.org/10.1145/3724363.3729048>
- Wiechmann, D., **Kempa, E.**, Kerz, E., & Qiao, Y. (2025). *Exploring the verbal behavior of ADHD in adults on social media through NLP-derived biomarkers and machine learning*. Published in the Abstract Book of the 10th World Congress on ADHD, Prague, Czech Republic. (Poster abstract)
- Kempa, E.**, & Aggarwal, A. (2024). *Understanding engineering students' ethical and algorithmic decision preferences through a consequentialist framework*. In Proceedings of the 55th ACM Technical Symposium on Computer Science Education (SIGCSE 2024), pp. 1883. <https://doi.org/10.1145/3626253.3635421>

PRESENTATIONS

- LREC-COLING 2026 (Palma, Mallorca, Spain) May 2026
- **Presented** a 15-minute slideshow presentation on my paper entitled “ADHD-Lang: A large-scale social media dataset for verbal behavior and digital phenotyping in adult ADHD”.
- ACM ITiCSE 2025 (Nijmegen, The Netherlands) July 2025
- **Presented** a 15-minute slideshow presentation on my paper entitled “Deconstructing engineering students' ethical reasoning: A consequentialist framework for algorithmic choices”.
- RWTH Aachen University UROP Symposium (Aachen, Germany)
- **Presented** a poster on my research on “Interpretable MRI Based Report Generation”. July 2025
 - **Presented** a poster and 20-minute slideshow presentation, on my research on “Mental Health Detection via Social Media Data” from that summer. July 2024
- University of Florida Spring Symposium (Gainesville, Florida)
- **Presented** a poster on an ADHD dataset from my “Mental Health Detection” research. April 2025
 - **Presented** my poster from the 2024 ACM SIGCSE TS SRC. April 2024
 - **Presented** a poster on the initial progress of the “AI-Ethics Framework” research. April 2023

ACM Student Research Competition (Portland, Oregon)

March 2024

- **Presented** a poster in the first round and a 15-minute slideshow presentation in the final round at the annual ACM SRC of the SIGCSE TS on my “*AI-Ethics Framework*” research.

AWARDS AND RECOGNITIONFBLA Collegiate State Leadership Conference, Orlando, FL

- 2nd Foundations of Digital Design March 2026
- 1st Computer Applications, 3rd CS and Tech Case Competition March 2025
- 2nd Foundations of Management March 2024
- 1st in Microeconomics, 1st in Macroeconomics, 3rd in Foundations of Accounting March 2023

1st Place Winner, Student Research Competition

March 2024

ACM SIGCSE Technical Symposium 2024, Portland, OR

LEADERSHIP AND INVOLVEMENTDeutschklub (German Club), University of Florida

- *Vice President* August 2024 – May 2026
- *Secretary/Historian* August 2023 – May 2024

Future Business Leaders of America Collegiate, University of Florida

- *President* May 2024 – May 2026
- *Vice President* August 2024 – May 2025
- *Treasurer* August 2023 – May 2024
- *Webmaster* September 2022 – May 2023

SKILLS/CERTIFICATIONS

- Cambridge Centre for AI in Medicine AI and Machine Learning Summer School, Sep 1-5, 2025
- BCI & Neurotech Masterclass: Python for Neuroscience (g.tec medical engineering), July 16, 2025
- Datathon: Large Language Modeling in Life Sciences — Center for Biohybrid Medical Systems, RWTH Aachen University, June 5, 2025
- Building RAG Apps Using MongoDB — MongoDB Day Munich, May 27, 2025
- MR Safety Level 1 (Forschungszentrum Jülich) — MRI scanner room access, May 20, 2025
- Microsoft Office Specialist (Word, PowerPoint, Excel, Outlook)
- *Languages*: Native Proficiency in German, Speaking Proficiency in Russian
- *Programming skills*: Python (Advanced), C++ (Intermediate – Advanced), R (Intermediate)

RESEARCH OUTPUT PORTFOLIO**Link:**

https://drive.google.com/drive/folders/1unsR1ITV8TnQGIFUmwIcPevUj1FPPFMv_?usp=drive_link