Implementing Human Feedback in Reinforcement Learning for ClaraNP: Hallucination-Mitigated AI Interface Tailored for Nursing Education

Ixchel Peralta-Martinez – iyp4754@uncw.edu; Maria Duran - mfd6863@uncw.edu; Henry Salkever - hss2235@uncw.edu
Dr. Gulustan Dogan - dogang@uncw.edu; Dr. Julie Hinkle - hinklej@uncw.edu; Dr. Crystal Dodson - dodsonc@uncw.edu

Objective

Present our ClaraNP interface and the envisioned integration of Reinforcement Learning with Human Feedback (RLHF)

Implementation RLHF

Overview

- Clara NP: A web interface leveraging Falcon-7b-instruct generative AI to enhance nursing education by mitigating hallucinations and biases.

Upcoming Showcase

- Set for presentation at the AAAI Conference on Artificial Intelligence 2024, highlighting our methodology, interface testing, and hallucination reduction techniques.

Latest Developments

- Focus on the ongoing work, particularly the application of Reinforcement Learning with Human Feedback (RLHF).
- Aim to improve user interaction capabilities through state-of-the-art research.

Integration of RLHF

- RLHF is central to our research agenda, aimed at enhancing large language model capabilities within ClaraNP.
- Process involves leveraging pre-trained models and specialized datasets for question and answer generation, fine-tuned with RLHF.

Role of Expert Input

- Input from nursing education professors at UNCW crucially guides the learning direction, ensuring AI-generated outputs align with human expectations and academic literature.

Enhancements through RLHF

- RLHF seeks to align AI behavior more closely with human intentions and ethical standards, enhancing model reliability and safety.
- Aims to provide nursing students with relevant, comprehensible content, deepening their understanding.

Impact on AI Interfaces

- The RLHF approach not only enhances AI system reliability and safety but also paves the way for more intuitive and user-friendly AI interfaces.

ClaraNP development

Language Model Preparation

- **Initialization & Libraries:** Imported libraries for PDF processing, text manipulation, and AI model deployment, including document parsing and text segmentation tools.

- **Model Configuration:** Set up pre-trained models such as Instructor XL, SBERT MPNet base, and FLAN T5 base for language model preparation.
  - SBERT MPNet was selected for its utilization of Siamese and transformer architectures to produce deeply semantic sentence embeddings, enhancing comprehension.
  - In conjunction, Falcon-7B-instruct was chosen for its advanced language processing paired with SBERT's embeddings to maximize the accuracy and utility of content in nursing education.

- **PDF-Based QA System Framework**: This architecture encompasses the initial configuration of models and embeddings, the creation of vector databases from PDF content, and the setup of retrieval QA mechanisms.

- **Output Refinement and Iterative Querying:** Cleans language model outputs, eliminating extraneous tags and spaces for clearer answers. Implements a pre-prompt and enables iterative querying with a loop, assessing ten variations of each query for improved accuracy.

Hallucination Mitigation

- **PDF Processing and Text Extraction**: Utilized PDF handling and tokenizing libraries such as pdfminer, scikit-learn, and TensorFlow.
  - Defined functions for reading, tokenizing, and truncating PDF text, with an optical character recognition(OCR) fallback using pytesseract for text extraction from photocopies.

- **Advanced Tokenization and Encoding:** Implemented the AllenAI's Longformer Encoder-Decoder (LED) model for tokenizing and encoding, specifically using the 'led-large-16384-pubmed' checkpoint.

- **Semantic Similarity and Keyword Ranking:** Established a Siamese neural network (SNN) and a keyword ranking model employing a Jaccard similarity algorithm.

$$S_{weighted} = \frac{(w_{SNN} \cdot S_{SNN} + w_{keyword} \cdot S_{keyword})}{(w_{SNN} + w_{keyword})} \quad (1)$$

Where S_{SSN} represent the SNN score and $S_{keyword}$ represent the Jaccard keyword similarity. The weighted average is $S_{weighted}$.

Experiment and Preliminary Results

- **Experiment Setup:** Generated series of 10 outputs by prompting a model to make quiz questions 10% less specific with each iteration, evaluating with ClaraNP using scores ($W\_SNN = 0.6$, $W\_keyword = 0.4$).
- **Early Findings**: ClaraNP preferred original answers, aligning somewhat with human evaluators. This indicates a promising direction for prioritizing content relevance and authenticity.
- **Limitations**: Potential bias towards longer answers by the keyword similarity algorithm needs addressing, as it might favor length over correctness.

# of Fallacies	Cosine Similarity Index	SNN Similarity Index	Accuracy Index
0	1	0.50923615694046	0.85277085
1	0.92977880422880	0.542960226535797	0.81373323
2	0.860045393911641	0.542955815792083	0.76491852
3	0.720578573277321	0.51461923122406	0.65879077
4	0.581111752643	0.514575839042663	0.56115098
5	0.41840046190296	0.510691344738006	0.44608773
6	0.3486670515858	0.510729908943176	0.39728591
7	0.2324447010572	0.443137675523757	0.29565259
8	0.1162223505286	0.444635421037673	0.21474627
9	0	0.450810253620147	0.13524308

Table 1 : Accuracy Module Output for Incrementally Augmented Model Responses

RLHF Framework

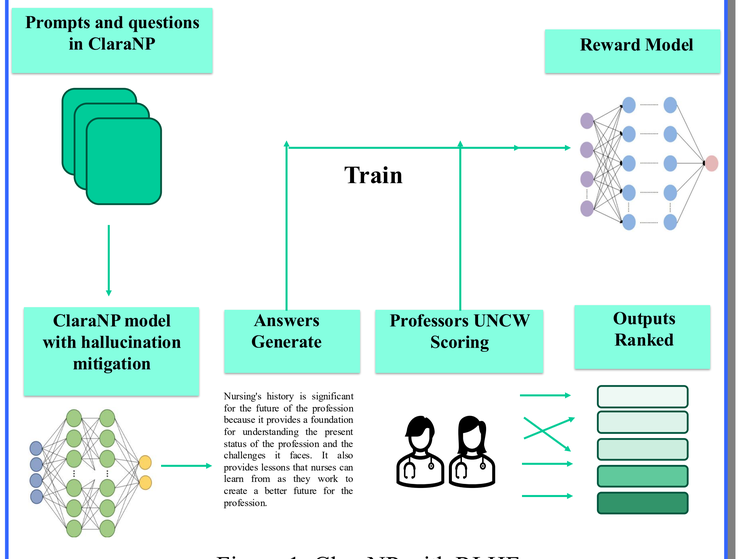

Figure 1: ClaraNP with RLHF
Adapted from Labellerr. 2023. Reinforcement Learning from Human Feedback. https://www.labellerr.com/blog/reinforcement-learning-fromhuman-feedback/.