Personal statements are a crucial part of the medical school application process, offering applicants the opportunity to showcase their personality, experiences, and motivation for pursuing a career in medicine [1]. However, many students struggle to draft and revise these essays while juggling pre-med commitments. To address this, we present the Semantic and Contextual Rubric-Based Intelligence for Biomedical Essays (SCRIBE), a novel offline tool for the automated evaluation of medical school personal statements, which fine-tunes the state-of-the-art e5-large-v2 [2]. SCRIBE provides automatic, structured feedback, lowering barriers for students who may lack access to mentoring or editing support. Our tool segments essays into semantically coherent sections, classifies each into rubric categories (Spark, Healthcare Experience, Showing Doctor Qualities, Spin), and assigns a score from 1 to 4. The SCRIBE tool was trained on manually annotated text by subject experts. The novel contributions of our research work include: (1) Development of an automated tool named SCRIBE for practical evaluation of medical personal statements by fine-tuning the state-of-the-art e5-large-v2 model, which is publicly available.

SCRIBE Tool: In our SCRIBE tool, we utilize semantic segmentation, transformer embeddings (using e5-large-v2), and an ensemble of classifiers and regressors. Personal statements are first divided into semantically coherent segments to mirror how human reviewers naturally break down an essay. We utilize a hybrid strategy combining newline-based and paragraph-based splitting with semantic similarity thresholds computed from a pre-trained sentence transformer model [3]. We compare multiple architectures, including BAAI/bge-large-en-v1.5 [4], intfloat/e5-large-v2 [2], thenlper/gte-large [5], all-mpnet-base-v2 [3], and all-MiniLM-L6-v2 [3], with these models being numbered 1 through 5, respectively. Each segment is evaluated across four rubric-based categories: Spark, Healthcare Experience, Showing Doctor Qualities, and Spin. Classified segments are passed to a corresponding regression model that assigns a rubric-aligned score from 1 to 4. These models are trained using XGBoost regressors. For each prediction, SCRIBE also reports a confidence score derived from classifier probability outputs. Segment predictions are aggregated into category-level summaries reporting average scores, rubric coverage, and matched segments. Explanations are compiled into a PDF report with feedback.

Results and Discussion: To evaluate the impact of sentence transformer choice on SCRIBE's accuracy, we tested on BAAI/bge-large-en-v1.5 [4], intfloat/e5-large-v2 [2], thenlper/gte-large [5], all-mpnet-base-v2 [3], and all-MiniLM-L6-v2 [3]. Each model generated segment embeddings, while downstream components were held constant. Results suggest that the intfloat/e5-large-v2 model offers the highest classification accuracy and rubric alignment with an overall classification accuracy of 91.7%. The strong performance of intfloat/e5-large-v2 stems from its contrastive training on large text-pair datasets, enabling fine-grained semantic similarity [2].

References:


1. L. Chandran, A.S. Chandran, and J.E. Fischel. Crafting compelling personal statements. Acad Psychiatry, 44:785–788, 2020.
2. Liang Wang, Nan Yang, Xiaolong Huang, Binxing Jiao, Linjun Yang, Daxin Jiang, Rangan Majumder, and Furu Wei. Text embeddings by weakly-supervised contrastive pretraining, 2024
3. Nils Reimers and Iryna Gurevych. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks, 2019.
4. Shitao Xiao, Zheng Liu, Peitian Zhang, and Niklas Muennighoff. C-pack: Packaged resources to advance general chinese embedding, 2023.
5. Zehan Li, Xin Zhang, Yanzhao Zhang, Dingkun Long, Pengjun Xie, and Meishan Zhang. Towards general text embeddings with multi-stage contrastive learning, 2023.