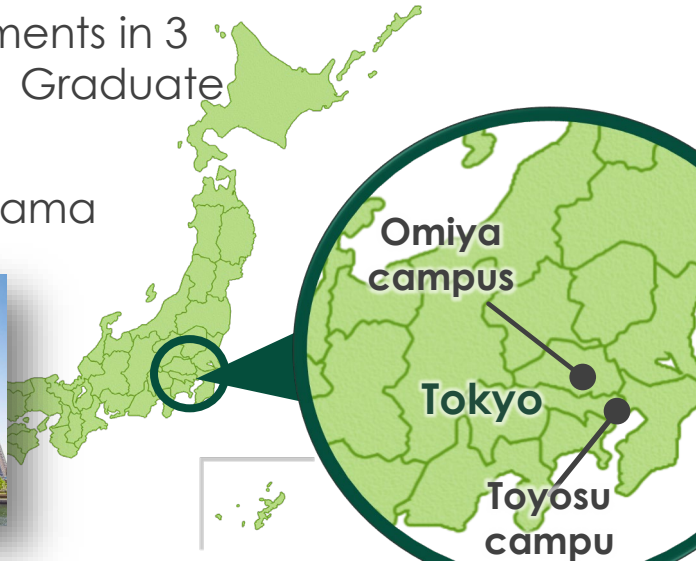
# DEEP LEARNING BASED ASSESSMENT IN HIGHER EDUCATION: CASE STUDIES

MASAOMI KIMURA, PH.D.
DIRECTOR, INTERNATIONAL PROGRAMS CENTER,
PROFESSOR, DATA SCIENCE/ENGINEERING LAB,
SHIBAURA INSTITUTE OF TECHNOLOGY

# Shibaura Institute of Technology (SIT)

- Private school for engineering education and research
- 7 Programs and 7 Departments in 3 Colleges and 1 School & 1 Graduate School
- Located in Tokyo and Saitama



# Challenges of assessment based on deep learning

- Assessment on students' essays
  - Content evaluation (essay scoring)
  - Author identification
- Program assessment for active learning programs (PBL etc.)
  - Understanding students' discussion

etc.

Hereafter, let me introduce our group's researches related to assessment

# A Method for Distinguishing Model Generated Text and Human Written Text

Hinari Shimada, Masaomi Kimura

Data Science/Engineering Laboratory

Shibaura Institute of Technology

Presented at ICIIT2024, Ho Chi Minh city, Vietnum

# Procedure of Proposed Method

1. Input each $i$-th text $x_i$ from dataset $X$ into $n$ GPT-based LLMs, and compute $n$ log likelihood $\log p_1(x_i), \ldots, \log p_n(x_i)$
2. Input $X$ into $m$ BERT-based LLMs to obtain 768-dimensional vectors, followed by dimensionality reduction to 2 dimensions using t-SNE
3. Input 1 and 2 into a feedforward neural network (FFNN) for classification

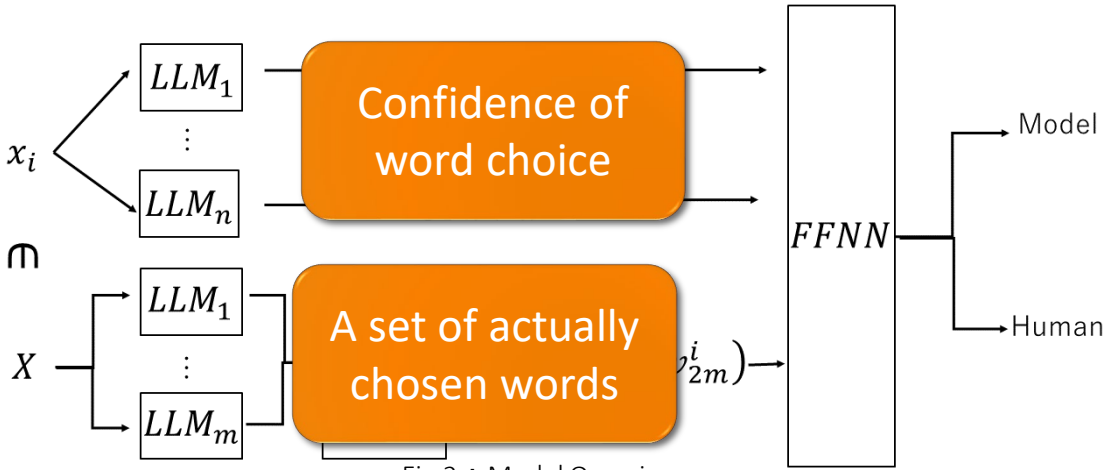



Fig.2：Model Overview

# Settings and Results

- Selection criteria of LLMs
  - Log likelihood calculation : Four GPT-based LLMs
    - Based on previous research[5] suggesting that more advanced GPT-based models tend to use sophisticated vocabulary and syntax
  - Embedding vector calculation : Four BERT-based LLMs
    - Selecting four standard base version models
- Result
  - Best accuracy compared to prior research

Table 1 : Result

| Model | Accuracy [%] |
|---|---|
| DetectGPT | 70.48 |
| Roberta-HPPT | 86.60 |
| CheckGPT | 97.95 |
| Proposed Model | **98.49** |

[5]Yikang Liu et al., "ArguGPT: evaluating, understanding and identifying argumentative essays generated by GPT models," arXiv preprint, arXiv:2304.07666, 2023.

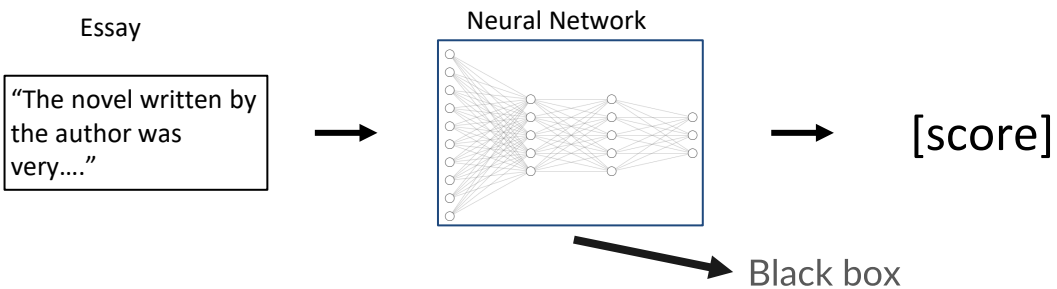
# Enhancing Automated Essay Scoring Model's Interpretability Using Triplet Loss with Teacher-Defined Rubric as Anchors

The 6th Joint International Conference on AI, Big Data and Blockchain
Istanbul, Türkiye

**Joseph Saido Gabriel and Masaomi Kimura**
Shibaura Institute of Technology, Tokyo, Japan

# **Introduction** | What is Automated Essay Scoring?

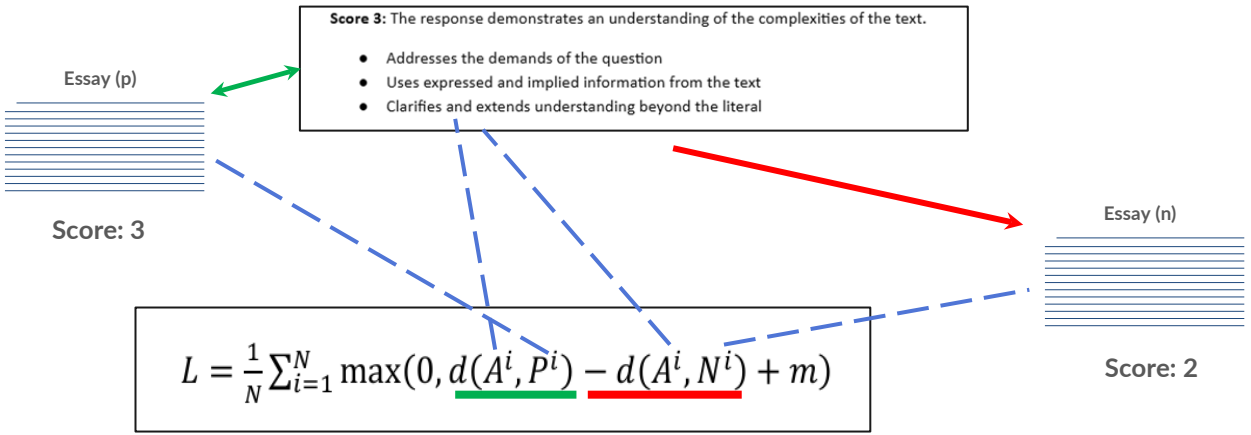
Automated Essay Scoring (AES) is a task that utilizes Natural Language Processing (NLP) techniques to assign scores to essays written by students.



# Methodology

- We propose a novel rubric-aligned AES model using triplet loss and dual-anchor triplet loss
  - Triplet loss is a key technique in metric learning, aiming to create an embedding space where similar items are closer and dissimilar ones are farther apart.

Since this essay has score three, the model will learn to structure the **essay's embedding** similarly to the corresponding rubric's embedding.



# **Experiment Result** | QWK Scores

| Models | Set 1 | Set 2 | Set 3 | Set 4 | Set 5 | Set 6 | Set 7 | Set 8 | Avg. | Explainability |
|---|---|---|---|---|---|---|---|---|---|---|
| R²BERT [6] | 0.817 | 0.719 | 0.698 | 0.845 | 0.841 | 0.847 | 0.839 | 0.744 | 0.794 | ✗ |
| NPCR [11] | 0.856 | 0.750 | 0.756 | 0.851 | 0.847 | 0.858 | 0.838 | 0.779 | 0.817 | ✗ |
| GPT-4, zero-shot, w/o rubrics [13] | 0.042 | 0.402 | 0.281 | 0.557 | 0.366 | 0.502 | 0.081 | 0.419 | 0.331 | ✓ |
| GPT-4, zero-shot, with rubrics [13] | 0.072 | 0.300 | 0.366 | 0.627 | 0.523 | 0.345 | 0.110 | 0.407 | 0.344 | ✓ |
| GPT-4, few-shot, with rubrics [13] | 0.280 | 0.338 | 0.331 | 0.784 | 0.623 | 0.728 | 0.257 | 0.454 | 0.474 | ✓ |
| Fine-tuned GPT-3.5 [13] | 0.741 | 0.618 | 0.704 | 0.859 | 0.796 | 0.848 | 0.727 | 0.614 | 0.738 | ✓ |
| Holistic rubric model | 0.810 | 0.710 | 0.735 | 0.836 | 0.822 | 0.836 | - | - | 0.770 | ✓ |
| Multiple trait rubric model | - | - | - | - | - | - | 0.730 | 0.677 | | ✓ |

# Deep Learning Based Text Mining on Global PBL Description Data

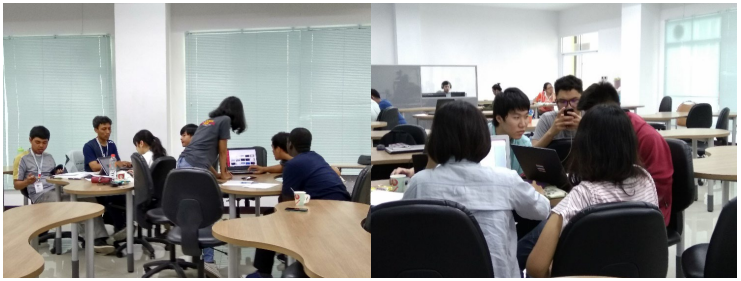
Masaomi Kimura, Hiroyuki Ishizaki, Hitoshi Nakamura, Tomoko Iwata
Shibaura Institute of Technology, Japan

Maria Anityasari, Imam Mukhlash, Mohammad Iqbal
Institut Teknologi Sepuluh Nopember, Indonesia

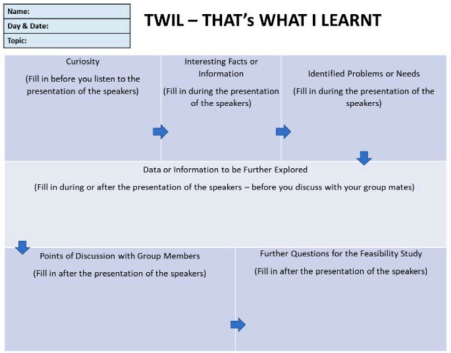


Presented at ICIIT 2025, Hanoi, Vietnam

# Global PBL



- Global PBL is a student-centered learning activity program
- These programs often involve international and interdisciplinary collaboration
- However, assessing these programs to facilitate continuous improvement remains challenging.
  - Necessary to analyze daily reports by participants

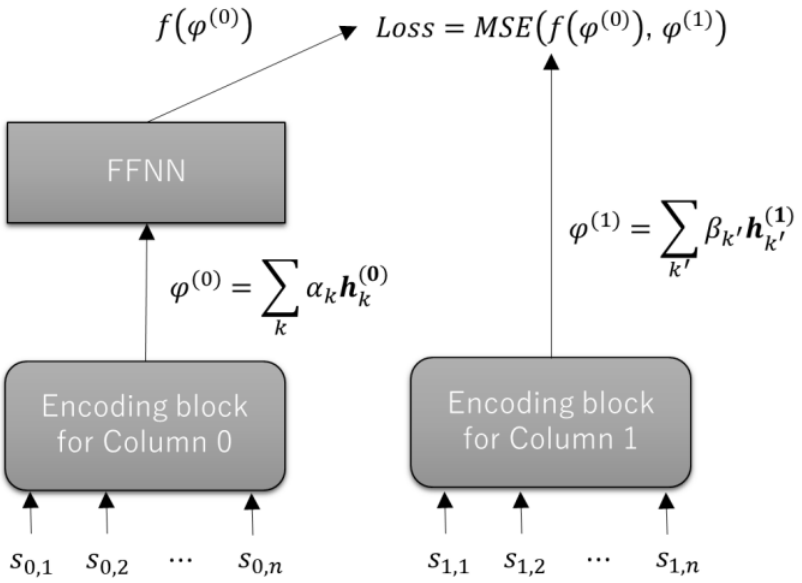


# Our proposed model

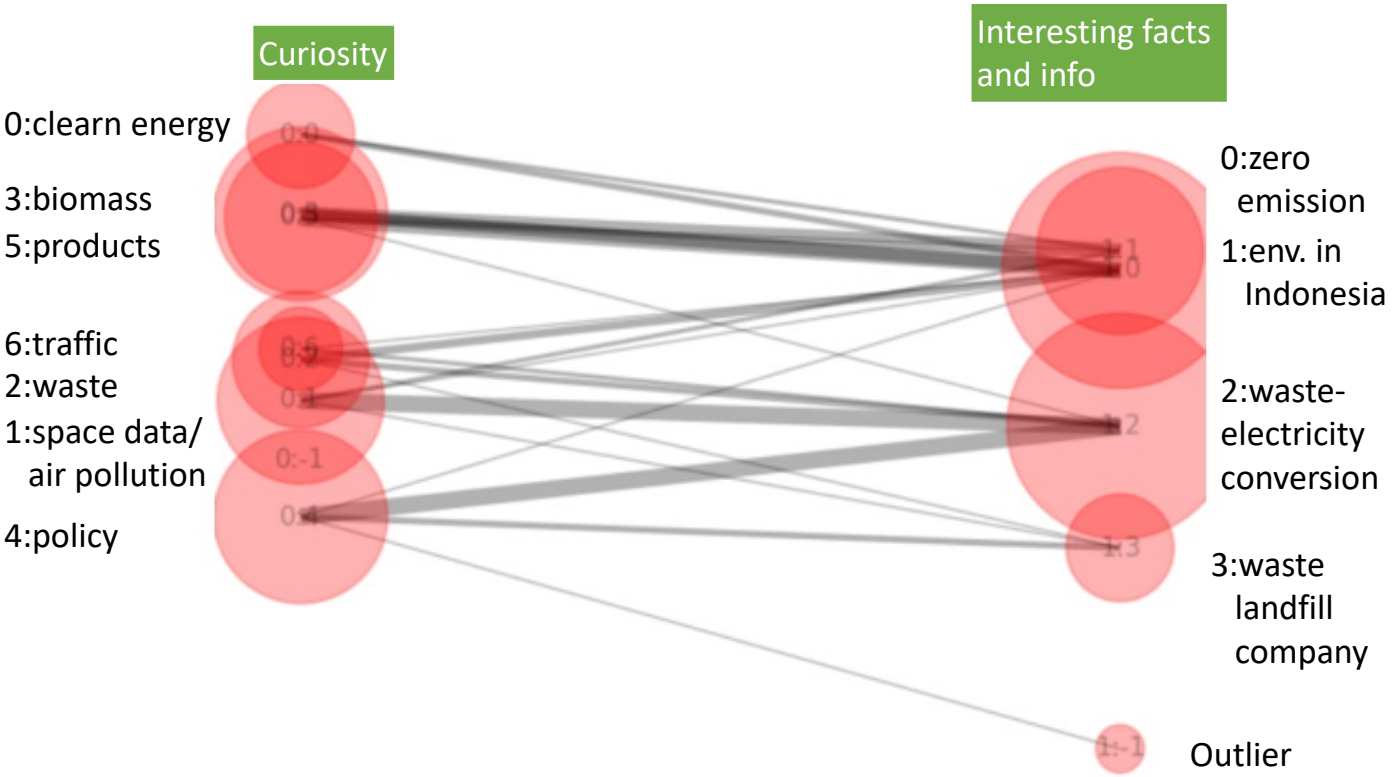
Adjustment of attention weights $\alpha$

based on the relationships of description

Text encoding by Sentence BERT



# An analysis example:
# for gPBL whose topic was environment eng.



# Thank you for your attention

**Masaomi KIMURA, Ph.D.**

Director, International Programs Center,

Professor, Data Science/Engineering Lab

Shibaura Institute of Technology

masaomi@shibaura-it.ac.jp