# 2023_"ShuWei Cup"

# Problem C：Intelligent Recognition and Detection of AI-Generated Text

In recent years, with the rapid development of information technology, various applications of artificial intelligence have emerged. Typical applications include robot navigation, speech recognition, image recognition, natural language processing, and intelligent recommendation, among others. Among these applications, large language models (Large Language Models, LLMs) led by ChatGPT have gained popularity worldwide and have been widely promoted and used. While we fully recognize the rich, intelligent, and convenient experiences that these models bring to people. It is also important to be aware of the many risks associated with tools such as AI text generation.

First, these large language models are trained based on text. And different types of languages and cultural backgrounds of different domains can have a significant impact on the generated results. Second, AI-generated results based on data may have semantic biases, lack logical coherence, and lack creativity. Finally, issues such as privacy protection, copyright protection, and the definition of related academic misconduct resulting from students using AI to generate papers pose significant difficulties and challenges to the teaching and training process for undergraduate and graduate students. In order to prevent the misuse of

AI-generated text, ensure the quality of generated content, and discuss how to address the problems caused by AI-generated papers, it is necessary to identify and detect the patterns of AI-generated text according to the topic requirements, including fields, models, images, and formulas.

To determine whether the text is AI-generated, in addition to considering factors such as meeting the word count requirement, the number of times generated, and whether it is Chinese-English translation. It is also important to note that AI currently lacks human emotion and judgment. This can lead to phenomena or styles in text generation such as "more phrases lack of examples, lack of emotion, structure, such as a single".

Please use mathematical modeling to solve the following four problems:

**Problem one:** Please use AI to rewrite some parts of the articles according to the links of twenty blogs on the Web of Science provided in Appendix I. And look for the basic rules of AI text generation, which can be inferred statistically inferred from the requirements of the number of words to be generated by AI (e.g. 200 words, 500 words, etc.), the number of times to be generated (the first time to be generated and then click on the button of "Regenerate"), whether it is a translation of the Chinese and English, and the style of the generated text.

**Problem two:** Please carefully consider whether each paragraph in the articles is generated by AI based on factors. This includes different generated languages, whether it is a translation, number of times generated, and whether there are restrictions on the number of output words, etc. for the theory and method of AI generation. Then mark the results of whether each paragraph in the ten articles provided in Appendix III is generated by AI.

**Problem three:** For the theory and method of AI generation, please carefully consider whether each paragraph in the articles is generated by AI based on factors. It includes different generated languages, whether it is a translation, number of times generated, and whether there are constraints on the output word count, etc. for the theory and method of AI generation. Then, mark the results of whether each paragraph in the ten articles provided in Appendix III is generated by AI.

**Problem four:** Please establish relevant theories and methods for further determining whether mathematical models, images, and formulas in the articles are plagiarized content. And use the examples in Appendix IV to demonstrate this and evaluate the established model.