AI Writing AND AI Paraphrasing Detection in Turnitin

Capabilities, Workflow, and Interpretation

Presented by: Chirawat Bhrombhorn

Date: 07 September 2025

Introduction to AI Paraphrasing Detection

- AI writing and AI paraphrasing detection is part of Turnitin's **AI writing indicator**
- Applies only to **English-language submissions**
- Automatically runs for institutions with AI writing detection enabled

What Can Be Detected?

- Turnitin detects:
  - AI-generated content
  - AI-bypassed content (e.g., changing sentence structure, replacing words to "bypass" detection algorithms.)
  - AI-paraphrased content (via word spinners or paraphrasing tools) “a **word spinner** refers to a **tool or technique used to automatically rewrite or paraphrase text**”
- **Color-coded highlights** in the report:
  - **Cyan:** AI-generated or bypassed
- **Purple:** AI-generated and paraphrased

AI Writing AND AI Paraphrasing report

51% detected as AI

The percentage indicates the combined amount of likely AI-generated text as well as likely AI-generated text that was also likely AI-paraphrased.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

Detection Groups

84 AI-generated only **50%**
Likely AI-generated text from a large-language model.

2 AI-generated text that was AI-paraphrased **1%**
Likely AI-generated text that was likely revised using an AI-paraphrase tool or word spinner.

Cyan: AI generated or bypassed

3.2 Participants and Sampling

The target population is Vietnamese English speakers divided into two groups (N = 320) based on their sociolinguistic environment. Group 1 (n = 160) consists of Vietnamese English speakers residing in Vietnam, while Group 2 (n = 160) comprises Vietnamese English speakers living in an English-speaking context (the United States). For group 2, participants' length of stay in the United States will be collected through the demographic survey. Since residence duration alone cannot fully capture the full extent of SLE, these data will be incorporated into the SES. The researcher will use SES scores to determine participant exposure levels, which results in three exposure groups instead of using residence duration as the sole exposure indicator. Participants in both groups are adults (18+). Previous pragmatics studies employing comparable designs have successfully yielded significant findings with relatively modest sample sizes (e.g., N = 75, Lai, 2009; N = 56, Mansour & Maryam, 2015). This study's target sample

Purple: AI-generated and paraphrased

CHAPTER 3: METHODOLOGY

The research methodology chapter describes the methods that the researcher used to conduct this study. It investigates how different sociolinguistic exposure levels affect Vietnamese English speakers' responses to compliments and criticisms. The chapter explains the sequential mixed-methods design, which combines a quantitative survey and DCT phase, followed by qualitative interviews. The chapter is divided into three main sections: (a) the research design and participants, (b) the data collection instruments and procedures (c) the data analysis and quality assurance measures.

3.1 Research Design

This study will adopt a sequential explanatory mixed-methods design (QUAN→QUAL) to investigate how sociolinguistic exposure influences Vietnamese English speakers' responses to compliments and criticisms. Scientists can achieve a complete understanding of complex research questions through the

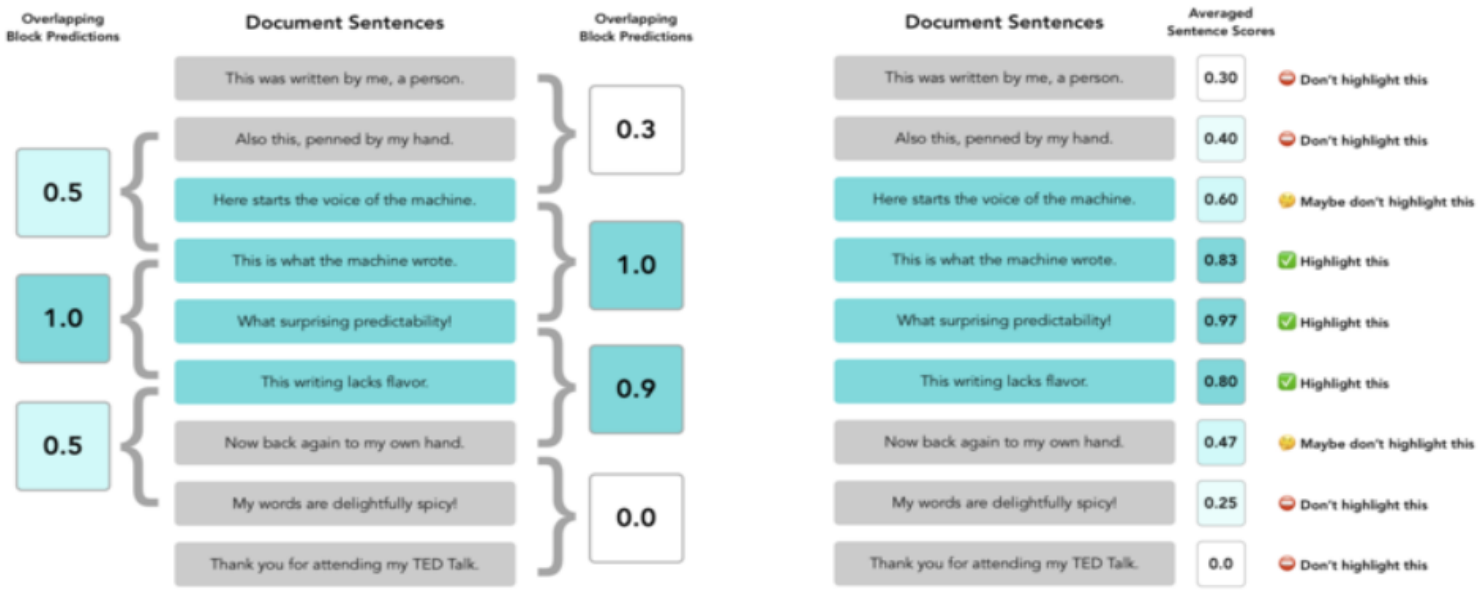
How Detection Works

- Submission is broken into overlapping segments (~5-10 sentences)
- Each sentence is scored (0 = human, 1 = AI)
- If marked as AI-generated, the **paraphrasing model** runs
- Paraphrasing model also scores each sentence (0 = not paraphrased, 1 = paraphrased)
- Final report shows:
  - AI-generated only
  - AI-generated + AI-paraphrased

How does it work?

How does it work?

When a paper is submitted to Turnitin, the submission is first broken into segments of text that are roughly a few hundred words (about five to ten sentences). Those segments are then overlapped with each other to capture each sentence in context.



Overall Prediction

- Average scores across segments produce:
  - **AI writing percentage**
  - **AI paraphrasing percentage**
- Only **prose text** is analyzed (not lists, bullet points, or code)

Technical Requirements

Requirement	Details
File size	< 100 MB
Word count	300–30,000 words
Language	English only
File types	.docx, .pdf, .txt, .rtf

Workflow Impact

- No change to user workflow
- Detection is **fully integrated** into the AI writing report
- Past submissions can be checked if **re-submitted** after release

Tools Detected

- Turnitin detects paraphrasing from:
  - **Quillbot**
  - **Grammarly (free paraphraser)**
  - **Scribbr**
  - Other open-source word spinners

Grammarly Use Clarified

- **Grammar/spelling checks:** *Not flagged*
- **Paraphrasing features:** *Likely flagged* if used to modify AI-generated text

Summary

- Turnitin detects AI-paraphrased content with high precision
- Integrated seamlessly into existing workflows
- Supports academic integrity by identifying bypassing techniques