



Laidlaw Cohort 2025  
Summer 1 Reflective Report – Research Project and Personal  
Development

# Developing a Behavioural and System-Level Tool to Detect AI Generated Content in Academic Writing

Author:  
Rachel Ranjith

Supervisor:  
Dr. Hitesh Tewari, Dr. David Lewis

September 2025



## Table of Contents

<b>Table of Contents</b>	<b>2</b>
<b>Project Overview</b>	<b>3</b>
Introduction	3
Project Outline	3
Methodology	3
Technical Results	4
Broader Results	4
Next Steps	5
<b>Personal and Professional Development</b>	<b>6</b>
Engagement with Peers, Supervisors, and Other Experts	6
Personal Development	6
Leadership	6
<b>Reflection</b>	<b>8</b>
What Went to Plan	8
What Didn't Go to Plan	8
Conclusion	9



## Project Overview

### Introduction

During the first summer of the Laidlaw Programme, my research project explored the rapidly evolving intersection of artificial intelligence and academic integrity.

The project was motivated by the increasing adoption of AI tools like ChatGPT in higher education, which are reshaping how students write, study, and learn. Early surveys suggest that a majority of students are already using AI in academic contexts, indicating that these technologies are becoming normalized. Additionally, instructors have begun adopting tools like GPTZero and Turnitin, which rely on syntactic features such as unusually complex vocabulary, sentence length, or specific punctuation patterns. However, these approaches are not sustainable, as advanced students and academics naturally produce such language.

Considering both the widespread student use of AI and the limitations of these detection tools, it becomes particularly important to critically examine the landscape of AI integration in academic contexts. Against this backdrop, I designed a project with dual objectives: first, to create a technical solution capable of detecting AI-assisted academic work, and second, to critically evaluate the broader implications of integrating AI in education. The project thus combined experimental development, empirical testing, and reflective analysis on pedagogical, ethical, and policy dimensions.

### Project Outline

The first part of the project involved developing a prototype system that could monitor student interactions and flag suspicious academic behavior. This system included keystroke tracking to detect pasting from external sources or verbatim AI output, and an integrated chatbot to simulate interactions and gauge engagement. The technical challenge lay in building a platform that was accurate yet minimally invasive, providing meaningful detection without compromising student privacy or trust.

The second part of the project focused on the theoretical dimension, examining why purely technical solutions are insufficient and exploring how AI can be responsibly integrated in academic settings without undermining learning or ethical standards. This dual approach reflected an evolution in my thinking: what began as a primarily technical investigation gradually expanded to include broader intellectual and ethical considerations, highlighting the value of integrating practical experimentation with reflective analysis.

### Methodology

The methodology was structured across four interlinked phases. The first phase was a survey of existing approaches and current state-of-the-art. I reviewed current detection tools, institutional policies, and prior research to understand the landscape of AI in academic spaces. This highlighted



significant gaps, particularly in balancing detection accuracy with ethical concerns around surveillance, as well as in keeping pace with rapidly evolving AI capabilities.

The second phase was prototype development. I designed and implemented a platform that combined keystroke tracking with a chatbot interface. The goal was to create a system capable of flagging suspicious behavior, such as copy-paste inputs from external sources or verbatim chatbot text.

The third phase focused on testing and analysis. The platform was evaluated against a range of input types, including copy-paste from external sources, verbatim retyping, chatbot-generated text, and lightly edited outputs. I then examined detection accuracy and error patterns, paying close attention to the trade-offs between false positives and false negatives.

The final phase involved critical evaluation. Here I reflected on the broader implications of AI in education, considering ethical, pedagogical, and policy dimensions. This phase emphasized that while technical tools can offer insights, they cannot fully resolve the challenges posed by AI in academic integrity.

By combining technical development with critical reflection, the methodology provided a comprehensive assessment of both the platform's capabilities and its wider significance in higher education.

## Technical Results

The empirical findings revealed that the system was highly effective at detecting simple pasted text but struggled with retyped or AI-generated content. Detection accuracy across input types is summarized as follows: copy-paste from external sources achieved 91% accuracy, while verbatim typing of external sources dropped to 22%. Copy-paste from the internal chatbot achieved near-perfect detection at 98%, whereas verbatim chatbot text was detected with only 58% accuracy, and edited verbatim outputs were detected at 31%. These results highlight the limits of keystroke tracking: while straightforward copying is reliably detected, rephrasing or AI-generated text often evades the system.

A complementary analysis of false positives and false negatives underscores this challenge, showing that most errors were false negatives, indicating the system often failed to flag AI or external inputs. False positives were minimal but carry significant ethical implications: over-tightening detection could wrongly question student credibility, creating a surveillance “witch-hunt” and undermining trust between students and institutions. Conversely, excessive leniency allows AI misuse to go undetected, potentially eroding learning outcomes and critical thinking skills. These findings demonstrate that technical solutions alone cannot fully address the complexities of academic integrity in the AI era, and that responsible integration requires a combination of technological tools, policy frameworks, and pedagogical strategies.

## Broader Results

Reflecting on these findings, I realized that technical accuracy alone cannot define the success of AI detection systems. A high rate of false negatives leaves misconduct unchecked, while reducing errors



by tightening controls risks escalating false positives, undermining trust and placing innocent students under suspicion. This tension highlights the limits of surveillance-based solutions, which risk eroding the very integrity they aim to protect by bypassing syntactic AI-detectors.

A more sustainable path lies in hybrid approaches. Detection tools can play a role, but they must be paired with thoughtful assessment design that reduces incentives for misuse, and with AI literacy initiatives that teach students to critically evaluate and responsibly apply these technologies. In this way, technical interventions support learning rather than displace it.

Ethical considerations are central. Over-regulation risks creating an environment of suspicion, where students feel monitored rather than supported. Equally, under-regulation risks normalizing uncritical reliance on AI, diminishing genuine skill development. Striking this balance requires policy frameworks that foreground fairness, transparency, and proportionality, ensuring that interventions respect student rights while safeguarding academic standards.

Looking forward, user studies are essential to test not only system usability but also student perceptions of fairness and trust. Embedding ethical safeguards into any technical solution must be treated as seriously as improving detection accuracy itself. Only then can interventions claim to be both effective and defensible in higher education.

## Next Steps

The next stage of this research will focus on developing hybrid solutions that integrate technical detection tools with redesigned assessments that reduce opportunities for misuse. Alongside this, I plan to embed AI literacy programs into the student experience, equipping learners to critically evaluate and responsibly apply these technologies rather than rely on them uncritically.

Policy refinement is another priority. Future work must examine how institutional frameworks can safeguard academic integrity without resorting to punitive surveillance, ensuring fairness, transparency, and proportionality.

Finally, user studies will be expanded to capture not only technical performance but also student perceptions of trust, credibility, and ethical acceptability. This dual focus on accuracy and lived experience will be key to creating interventions that are both effective and defensible.



## Personal and Professional Development

### Engagement with Peers, Supervisors, and Other Experts

Interactions with peers, professors, and participants in AI workshops significantly enriched the research experience. My supervisors provided an effective balance of autonomy and mentorship, facilitating independent exploration while offering support as needed. This balance suited me well, as it allowed space for independent experimentation without leaving me isolated.

Beyond this, engagement with peers and other Laidlaw Scholars proved invaluable. Conversations with them revealed nuances in AI applications across disciplines that I would not have uncovered alone, broadening both the technical and ethical scope of my work.

A particularly formative experience came from facilitating AI workshops run by a professor, where I engaged directly with senior professionals from academia, healthcare, and related fields. These exchanges were not one-way: while I learned from their perspectives on AI integration, I was also able to share insights from my own research. That reciprocity made the workshops both intellectually challenging and deeply rewarding.

### Personal Development

Working across both technical and theoretical dimensions taught me to approach problems from multiple angles. I began to see methodological limitations less as setbacks and more as prompts for critical reflection and refinement. This shift helped me grow more comfortable with uncertainty, while also sharpening my ability to construct nuanced arguments that connect empirical findings to ethical and pedagogical debates.

The technical challenges of building a detection system, balancing accuracy with privacy, tested my resilience. They required sustained problem-solving, adaptability, and creativity in the face of setbacks. Rather than being discouraged, I learned to iterate quickly, seek alternative solutions, and remain focused on the broader research aims.

Ultimately, the project gave me a clearer picture of my research style. I work best when combining detail-oriented technical rigor with reflective, big-picture thinking. This balance has strengthened both my confidence as a researcher and my resilience in navigating complex challenges.

### Leadership

This project underscored for me that leadership is not about control but about balance, between autonomy and collaboration, decisiveness and receptiveness. Working with peers on related themes, I gained perspective from their experiences, and the exchange of ideas in AI workshops broadened my own thinking. These moments highlighted how dialogue and constructive challenge can strengthen outcomes rather than dilute them.



My supervisors modeled a collaborative style of leadership, offering guidance without micromanagement. This created the conditions for me to step into ownership of the project, manage timelines, and make key methodological decisions. In doing so, I learned the importance of clarity, accountability, and persistence in sustaining progress over time.

I also became more conscious of my own leadership style: I lead by structuring work clearly, staying detail-focused, and fostering open exchange. Seeing alternative approaches in workshops helped me refine this style, particularly in valuing constructive disagreement as a catalyst for better solutions.



## Reflection

### What Went to Plan

The technical core of the project largely unfolded as anticipated. Building the prototype, conducting early copy-paste detection tests, and performing initial analyses progressed smoothly, reflecting the effectiveness of careful planning and methodical execution. Each step built naturally on the previous one, allowing me to translate abstract project goals into tangible results. The literature review provided a solid foundation for understanding AI detection strategies, which in turn informed both the design and evaluation of the system.

Engagement with my supervisors contributed to this smooth progression. Their guidance was timely and targeted, enabling me to troubleshoot challenges without stifling independent exploration. This autonomy aligned well with my PDP goals of developing detail-oriented project management, proactive problem-solving, and strategic foresight. I particularly enjoyed observing how planning and structure translated into measurable outcomes, reinforcing my confidence in taking complex ideas from conception to implementation. Overall, these successes validated my strengths in organization, analytical thinking, and communication, and demonstrated that careful preparation and strategic oversight can meaningfully shape research progress.

### What Didn't Go to Plan

Despite initial expectations, the project evolved from a primarily technical investigation into a broader inquiry encompassing theoretical, ethical, and pedagogical dimensions. Early limitations of keystroke tracking highlighted both the technical constraints of surveillance and its ethical implications, prompting me to reconsider assumptions about efficacy, privacy, and student trust. These findings naturally led to questions about pedagogy and responsible leadership in AI integration, shifting the project's focus beyond detection alone.

This evolution coincided with logistical hurdles, such as limited access to external input samples and the challenges of conducting meaningful user testing. Navigating these constraints demanded creative problem-solving and iterative adaptation. At the same time, balancing academic, extracurricular, and research responsibilities tested my resilience and time-management abilities, highlighting the importance of prioritization and sustained focus under pressure. Engaging with these overlapping challenges reinforced my understanding of the interplay between technical rigor, ethical judgment, and empathetic decision-making. By confronting these tensions, I strengthened my capacity for reflective practice, adaptability, and transparent leadership, learning to navigate complex problems while maintaining integrity and responsiveness to the human dimensions of research.



## Conclusion

Technical solutions such as keystroke tracking can be effective for certain input types but are fundamentally limited. Ethical considerations, pedagogical strategies, and policy frameworks are essential to ensure that interventions are fair, effective, and sustainable. The balance between enforcement and trust is delicate: over-surveillance risks undermining credibility and morale, while excessive leniency can erode learning and critical engagement.

Summer 1 of the Laidlaw Programme provided a transformative experience that extended beyond technical achievements. The research combined empirical investigation, critical reflection, and leadership development, offering a rich environment to grow as both a scholar and a leader. I learned to navigate complexity, integrate technical and ethical reasoning, and balance independent initiative with collaborative engagement. These insights have strengthened my confidence, enhanced my skills, and provided a foundation for future work and continued growth.