UNIVERSITY OF ÉVORA

Plagiarism Detection: Report

Javier Lamar Leon , Paulo Quaresma, Vitor Nogueira

jlamarleon@uevora.pt, pq@uevora.pt, vbn@uevora.pt

October 2024

Introduction

Plagiarism detection is a vital component of maintaining academic integrity Bretag 2014, ensuring that research papers, essays, and scholarly works are original and properly cited. With the increasing availability of online resources and the proliferation of AI writing tools, the risk of both accidental and intentional plagiarism has risen significantly. Consequently, plagiarism detection has become an indispensable step in the academic writing process, helping educators, researchers, and students avoid the ethical and professional consequences of unintentional duplication.

Detection methods for plagiarism can vary significantly based on several criteria. These methods can be monolingual or cross-lingual, depending on the number of languages involved. They can also be classified as intrinsic or extrinsic; intrinsic detection analyzes the text itself, while extrinsic detection compares the text to external sources. Additionally, detection techniques may utilize different approaches, including N-gram-based, vector-based, syntax-based, semantic-based, fuzzy-based, structural-based, and stylometric-based methods.

As the internet continues to expand, increasing the volume of textual information and introducing new tools for detecting plagiarism, such as large language models (LLMs), the urgency surrounding effective plagiarism detection has intensified.

Academic institutions and publishers heavily depend on plagiarism detection tools to scrutinize submissions and manuscripts, ensuring that the work meets the required standards of originality. While plagiarism detection is essential for preserving the credibility of academic work, the market offers a variety of tools, each with its own unique features designed to streamline this process.

The following are some of the most popular tools used in academic settings for detecting plagiarism Pudasaini et al. 2024:

1. **Turnitin** (2000) – Paid. Supports both intra-corporal (within a private database) and extra-corporal (across the web) plagiarism detection (Turnitin 2000).

2. **iThenticate** (1996) – Paid. Supports multiple document formats and is widely used for academic and research plagiarism detection (iThenticate 1996).

3. **CopyLeaks** (2015) – Paid. Supports multiple file formats, including rich text and PDFs. Suitable for detecting plagiarism across web content and educational materials (CopyLeaks 2015).

4. **Duplichecker** (2020) – Free. Allows users to copy-paste text or upload documents for basic plagiarism checks. Useful for quick and easy detection (Duplichecker 2020).

5. **Grammarly** (2016) – Paid. Provides detailed plagiarism reports that highlight the source, along with writing suggestions for improvement (Grammarly 2016).

6. **PlagScan** (2009) – Free. Handles multiple file types and offers side-by-side comparison for a thorough plagiarism review (PlagScan 2009).

Turnitin and Duplichecker stand out as prominent examples of paid and free tools, respectively. Turnitin is widely recognized in academic institutions for its comprehensive and advanced plagiarism detection capabilities. It offers in-depth comparisons across an extensive database of academic works and web sources, ensuring thorough and accurate results. Its paid nature reflects the high level of reliability and advanced features it offers, making it a preferred choice for universities and publishers committed to maintaining academic integrity.

On the other hand, Duplichecker offers a reliable alternative for those seeking free plagiarism detection. While not as feature-rich as Turnitin, it provides a convenient and accessible option for quick plagiarism checks. Users can either copy-paste text or upload documents, making it popular among individuals or smaller educational institutions with limited access to premium tools. Together, these two tools represent the spectrum of plagiarism detection solutions available to meet various needs and budgets in academia.

Turnitin

Among the tools listed, Turnitin stands out as the most widely used plagiarism detection tool in academia. The software is exclusively available to universities and other higher education institutions. Only university administrators or authorized representatives can obtain it, usually by coordinating with Turnitin's sales team for purchase or licensing.

Figure 1: Turnitin plagiarism detection logo.

Turnitin is a plagiarism detection (see logo Fig. 1) and prevention service that reviews written work for improper copying and citation errors. When a document is submitted, Turnitin compares it to a vast database of student papers, books, and websites to identify any similarities.

Today, the software is used by more than 12 million students globally. Turnitin has also partnered with Elsevier to create an open-source document comparison tool.

Recently, Turnitin introduced Turnitin Connect, an online tool that allows organizations to quickly compare submissions against its extensive repository. To enhance the accuracy of its similarity score technology, Turnitin employs machine learning techniques.

Turnitin offers a range of tools and services to a diverse audience. Each document submitted

is assigned a similarity score between 0 and 100, indicating the level of originality compared to its database.

This score helps users assess how much of the content is original. The Turnitin algorithm is continuously updated with new data to ensure accurate scoring. The threshold for what constitutes a plagiarized document may vary depending on the institution or organization using Turnitin.

A good Turnitin similarity score typically ranges between 10% to 20%, but up to 30% is generally considered acceptable. While Turnitin helps identify areas for improvement, it's important to review flagged sections carefully, as the software may sometimes misidentify or miss instances of plagiarism.

In Turnitin, the percentage refers to the similarity score, which indicates how much of a submitted document matches content in Turnitin's database. This score reflects the proportion of text in the document that matches other sources, including academic papers, websites, and other students' submissions.

Turnitin uses a color-coded system (see Fig. 2) to highlight similarities between the submitted work and preexisting sources. Each color indicates the type and degree of similarity, helping both students and instructors easily identify potential issues.

- **Blue**: No matching text found in Turnitin's database.

- **Green**: Low similarity, usually 1-24% matching text.

- **Yellow**: Medium similarity, typically 25-49%.

- **Orange**: High similarity, around 50-74%.

- **Red**: Very high similarity, 75-100% matched text.

These colors correspond to different similarity levels, with the matched text highlighted and underlined accordingly. The **red** and **orange** colors indicate significant overlap with existing sources, while **green** and **yellow** suggest lower, but still notable, matches. **Blue** indicates no matches found.

Its prominence is not merely coincidental but is underpinned by several compelling reasons:

1. **Extensive Database**: Turnitin boasts an unparalleled database comprising billions of

Color	Color Code	Indicator
Blue		0% - no matching text
Green		1-24% of matching text
Yellow		25-49% of matching text
Orange		50-74% of matching text
Red		75-100% of matching text

Figure 2: Turnitin Colors

web pages, millions of student papers, and a vast collection of academic publications. This comprehensive repository enables it to detect plagiarism from a wide array of sources, ensuring that even the most obscure instances of copying are identified.

2. **Advanced Detection Algorithms**: Turnitin employs sophisticated algorithms that go beyond simple text matching. These algorithms can detect paraphrasing, synonym substitution, and other techniques often used to disguise plagiarism. This advanced capability ensures a higher level of accuracy and reliability in identifying unoriginal work.

3. **Global Adoption**: Turnitin is a standard choice for universities and educational institutions worldwide. Its widespread adoption is a testament to its effectiveness and the trust it has garnered within the academic community. This global usage also contributes to the continuous expansion of its database, further enhancing its detection capabilities.

4. **User-Friendly Interface**: Turnitin offers an intuitive and user-friendly interface that makes it easy for educators and students to use. This accessibility ensures that the tool is not only effective but also practical for everyday use in academic settings.

5. **Integration Capabilities**: Turnitin can be seamlessly integrated with various learning management systems (LMS) such as Blackboard, Canvas, and Moodle. This integration allows for streamlined workflows, making it easier for educators to incorporate plagiarism detection into their teaching practices.

6. **Feedback and Educational Tools**: Beyond mere detection, Turnitin provides valuable feedback and educational resources. It offers tools that help students understand the importance of academic integrity and improve their writing skills, thereby fostering a culture of originality and ethical conduct.

7. **Continuous Improvement**: Turnitin is committed to continuous improvement, reg-

ularly updating its algorithms and expanding its database to keep pace with evolving plagiarism techniques. This proactive approach ensures that the tool remains effective in an ever-changing academic landscape.

8. **Legal and Ethical Compliance**: Many schools, colleges, and even publishers rely on Turnitin for maintaining academic integrity. Its use helps institutions comply with ethical standards and legal requirements, safeguarding their reputations and the value of their academic credentials.

Steps to Check for Plagiarism Using Turnitin

1. Access Turnitin

- Log in to your **Turnitin** account through your institution's portal or directly from the Turnitin website, if you have a personal account.

2. Create or Access an Assignment

- If you are a student, your instructor may have created an assignment for submission. Navigate to that assignment on your course's Turnitin dashboard.

- If you are an instructor, you need to create an assignment first:

  - Click on the **"Create Assignment"** button in the course page.

  - Set the assignment title, start date, due date, and other relevant options.

3. Submit a Document

- **Upload the document** you want to check for plagiarism. You can do this by:

  - Dragging and dropping the file.

  - Uploading the file from your computer or cloud storage services (e.g., Google Drive, Dropbox).

- Supported file types include Word, PDF, plain text, and others.

- Ensure your document is under the allowed file size (usually 100MB or 800 pages, depending on the institution).

4. Complete the Submission

- After uploading, click on **Submit** to send the file.

- Wait for the confirmation screen, which shows that the submission is complete. A digital receipt may also be generated.

5. Wait for Similarity Report

- Once submitted, Turnitin will process the document and generate a **Similarity Report**. This usually takes a few minutes, but it could take longer depending on the file size and Turnitin's current load.

- You will be notified once the report is ready.

6. View the Similarity Report

- Access the report from your Turnitin dashboard. The report will provide:

  - A **Similarity Index**: A percentage indicating how much of your text matches other sources.

  - A detailed breakdown of the **matched sources** (websites, papers, or publications).

- Click on the **colored percentage** next to your submission to open the report in the Turnitin **Document Viewer**.

7. Analyze the Similarity Matches

- In the viewer, Turnitin will highlight text from your submission that matches sources in its database.

- You can click on each match to see the source and decide whether it constitutes plagiarism or is properly cited.

- For legitimate matches (proper citations), instructors or students can ignore the flagged sections.

8. Download the Report (Optional)

- If needed, you can **download** the similarity report in PDF format for record-keeping or further review.

9. Revise and Resubmit (If Necessary)

- If you're a student and allowed multiple submissions, review the flagged sections, make any necessary revisions to address the similarities, and resubmit the document to reduce the plagiarism percentage.

Tips:

- **Instructors**: You can configure the assignment to allow or disallow students from seeing the similarity report (see Fig. 3) and control resubmissions.
- **Students**: Always check your institution's policy on resubmissions and plagiarism limits.

These steps ensure an efficient process for plagiarism detection and provide detailed feedback on document originality.
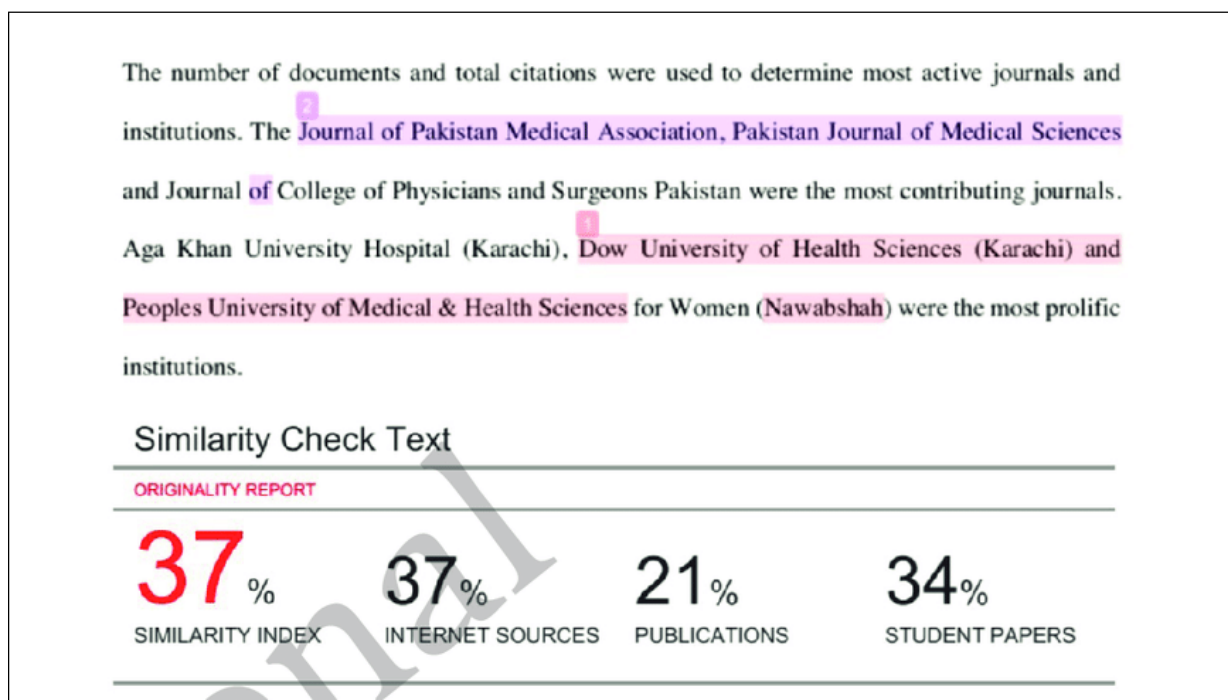

Figure 3: Demonstrative Example of a Turnitin Similarity Report

Duplichecker

DupliChecker is a robust, AI-powered plagiarism detection tool designed to help users ensure content originality while enhancing the quality of their writing. With deep scanning capabilities, support for multiple languages, a wide range of file formats, and a privacy-conscious design, DupliChecker caters to a broad spectrum of users, including writers, students, teachers, bloggers, and webmasters. Beyond plagiarism detection, it offers several value-added features like grammar checking, keyword analysis, and writing improvement suggestions, making it a one-stop shop for content integrity and quality.

Key Features

1. **Advanced Plagiarism Detection** DupliChecker employs AI to conduct in-depth scans of text against billions of web pages, academic papers, and digital content. Its detection algorithms review individual sentences for duplication and provide a detailed report showing any matched sources. The tool is effective in identifying content duplication in documents up to 25,000 words, which distinguishes it from many other plagiarism detection platforms. It provides both a plagiarism percentage and highlights duplicate sections, ensuring a thorough review.

2. **Multilingual Support and File Compatibility** A notable feature of DupliChecker is its support for multiple languages, allowing users to check content in English, French, Spanish, Russian, Portuguese, Dutch, Indonesian, Italian, and Arabic, among others. This makes it a valuable tool for global users. Additionally, DupliChecker handles multiple file formats, such as .doc, .docx, .pdf, .txt, .odt, and .rtf, providing flexibility for users who need to upload different types of documents.

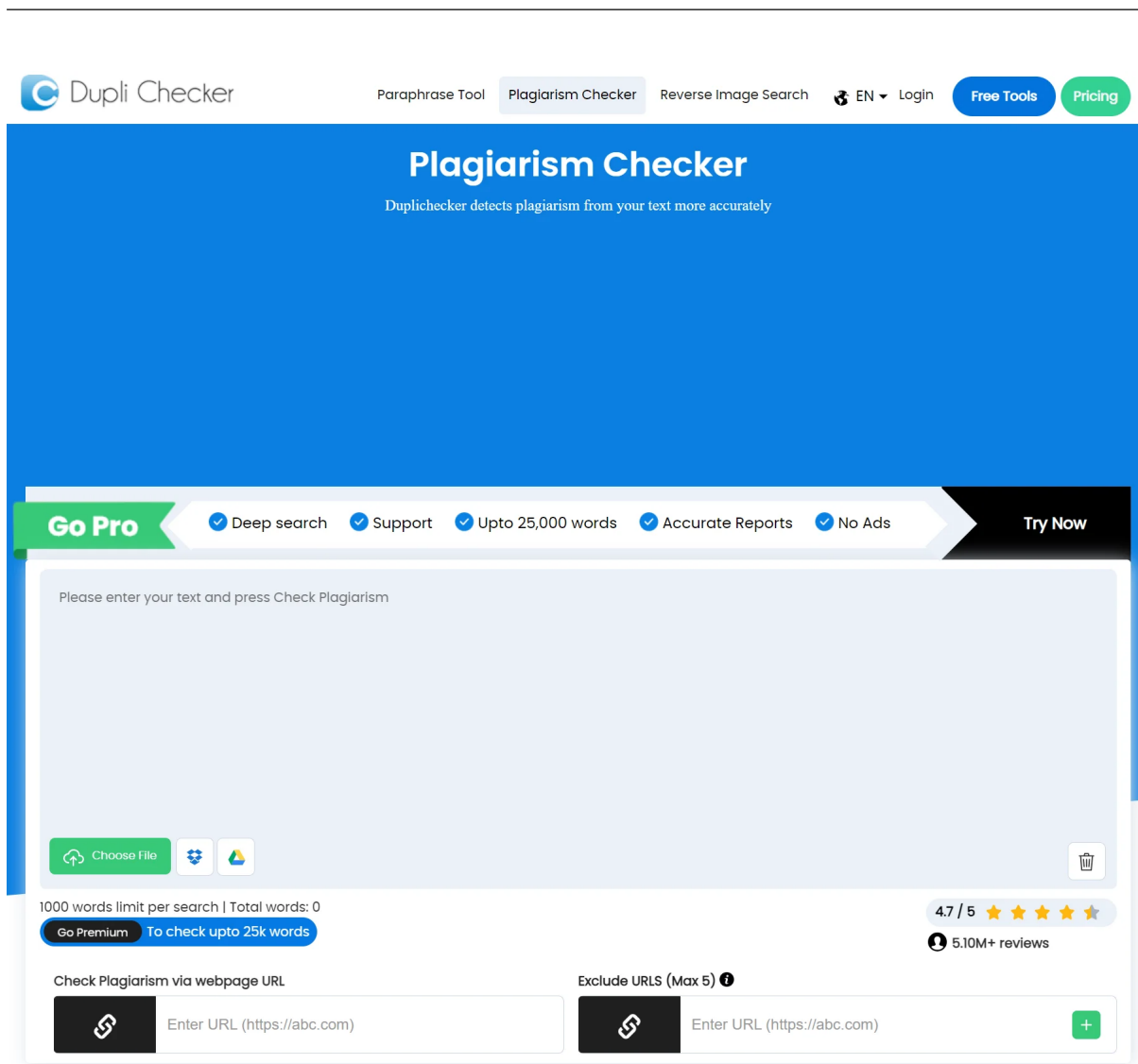3. **Grammar and Spell Checking** In addition to plagiarism detection, DupliChecker helps

Figure 4: DupliChecker.

users improve their writing by identifying grammatical errors, spelling mistakes, and punctuation issues. The built-in grammar checker offers suggestions for correcting these errors, ensuring the overall quality of the content is enhanced alongside plagiarism prevention.

4. **Keyword Analysis for SEO** Content creators looking to optimize their work for search engines can benefit from DupliChecker's keyword analysis tools. These features help identify relevant keywords, analyze keyword density, and suggest improvements to ensure the content ranks well in search engines.

5. **Privacy and Security** DupliChecker places a strong emphasis on privacy and security. It ensures that user data is not stored or shared, maintaining complete confidentiality throughout the content scanning process. This guarantees that content scanned through DupliChecker is safe from being indexed or reused elsewhere, which is particularly impor-

tant for professionals dealing with sensitive material.

6. **Writing Enhancement Suggestions** Beyond detecting duplication, DupliChecker acts as a writing assistant by providing personalized suggestions for improving the clarity, structure, and flow of content. This feature guides users in refining their text, allowing them to produce higher-quality and more engaging material.

7. **Additional Tools** DupliChecker offers more than just plagiarism detection. It includes a paraphrasing tool for rewriting plagiarized content, a reverse image search tool for verifying image authenticity, and a suite of SEO tools such as backlink checkers and malware trackers. These additional tools make it a versatile platform for a variety of digital content needs.

How to Use DupliChecker

Using DupliChecker is straightforward:

1. **Step 1**: Visit the DupliChecker homepage and either paste your text into the designated box or upload your document in one of the supported formats.

2. **Step 2**: Click the "Check Plagiarism" button.

3. **Step 3**: After processing, DupliChecker will display the plagiarism results, showing the percentage of unique content alongside the plagiarized sections, which are highlighted for easy review.

4. **Step 4**: If plagiarism is detected, users can employ the paraphrasing tool to rewrite the duplicated content and ensure it becomes original.

DupliChecker also provides users with the option to check spelling and grammar, and to use its SEO and paraphrasing tools.

Advantages and Disadvantages

Advantages

- **Comprehensive Plagiarism Detection**: DupliChecker scans text across billions of sources, providing highly accurate plagiarism detection.

- **User-Friendly**: The platform's simple interface is easy to navigate, even for beginners.

- **Privacy Guaranteed**: DupliChecker ensures that scanned content remains confidential and is not stored.

- **Multilingual and Multi-Format Support**: Its support for various languages and file formats makes it accessible to a global audience.

- **Writing Assistance**: Beyond plagiarism, the tool provides suggestions for writing improvements, grammar, and keyword analysis, adding significant value to content creators.

Disadvantages

- **Word Limit in Free Version**: The free version allows checks for only 1,000 words at a time, which limits its utility for longer documents unless users upgrade to the paid version.

- **Ads in Free Version**: Free users may experience distractions due to ads during the scanning process.

- **Accuracy Issues for Complex Texts**: While effective for simpler texts, some users have reported less accuracy in detecting plagiarism in more complex or highly technical documents.

- **Limited Free Features**: The free version is restricted in terms of word count and the number of checks allowed, which could inconvenience users with larger workloads.

DupliChecker offers both free and paid plans. The free version limits users to checking up to 1,000 words per session and may include ads. The paid version starts at $10 per month and removes these restrictions, allowing for larger word counts and more advanced features, including enhanced plagiarism detection and writing improvement tools.

DupliChecker is a comprehensive online plagiarism detection tool that goes beyond simply identifying duplicated content. Its AI-powered scanning, multilingual support, and additional features like grammar checking, keyword analysis, and writing improvement make it a valuable resource for writers, students, teachers, and professionals. While the free version has some limitations, the tool's versatility and privacy guarantees make it a leading option in the market. For those requiring more advanced functionality, the paid plans offer excellent value with access to deeper scans, more features, and the ability to handle larger texts.
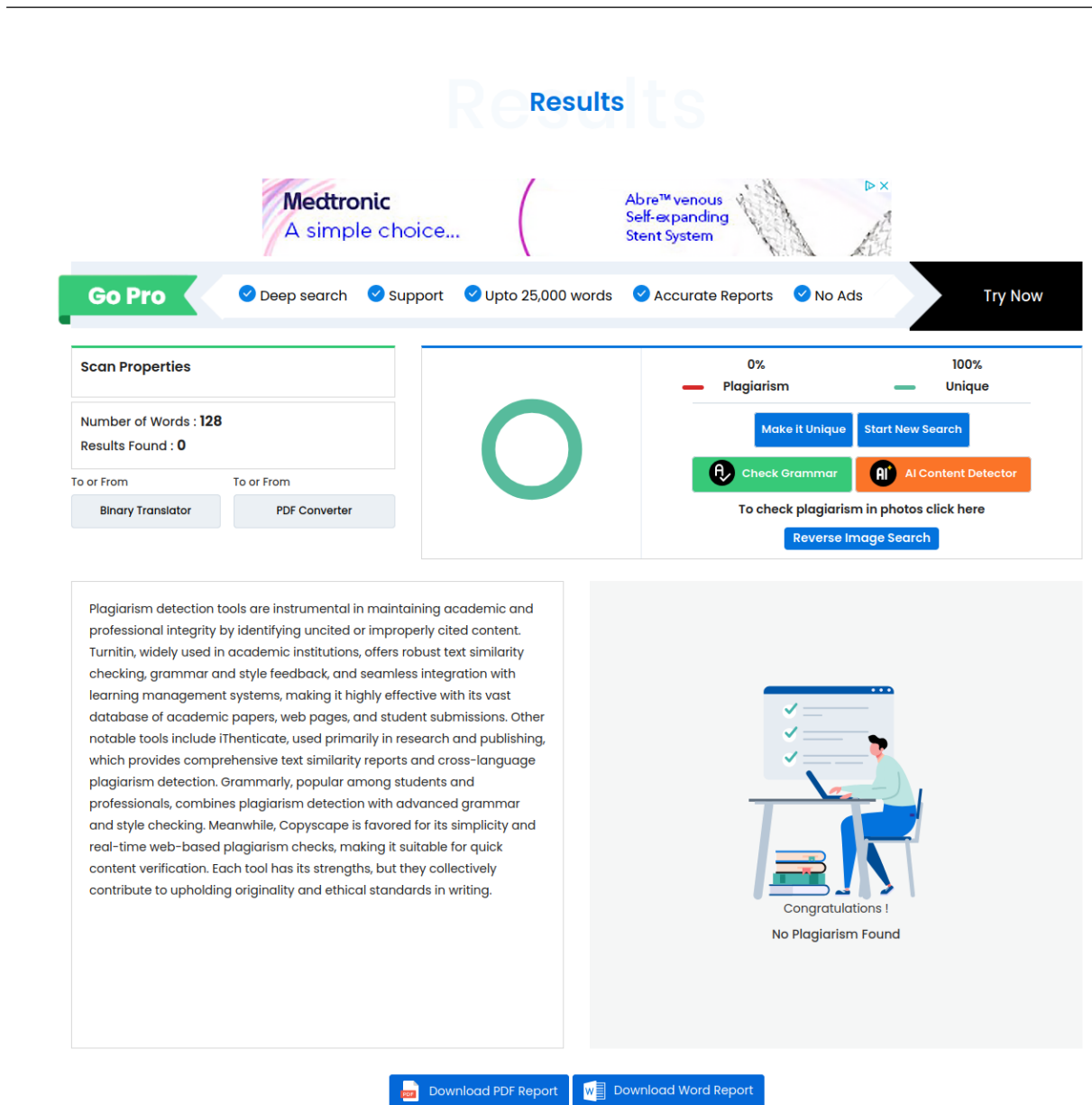
Figure 5: Demonstrative Example of a DupliChecker Plagiarism Detection Report for Text Generated by the Mistral Model.

The following two figures provide example reports generated by the Duplichecker tool, using generated text from Mistral and ChatGPT models, respectively (see 5 and 6).

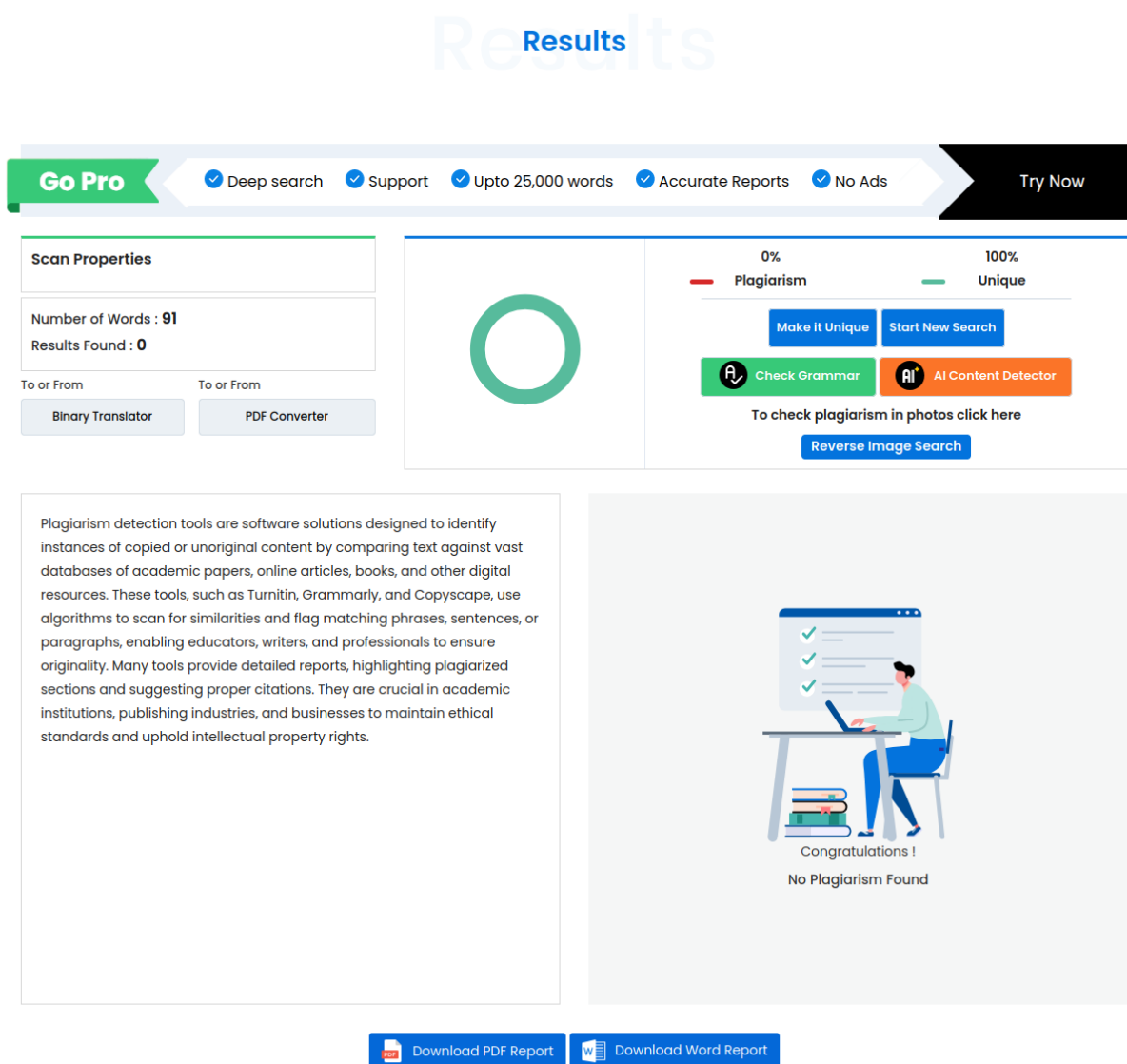We can observe that both models successfully generate completely unique text.

Figure 6: Demonstrative Example of a DupliChecker Plagiarism Detection Report for Text Generated by the ChatGPT Model.

Plagiarism and LLMs

Plagiarism detection has significantly evolved with the rise of large language models (LLMs) Pudasaini et al. 2024. While tools like Wordtune and Quillbot[1] have long been used for paraphrasing, students now also use LLMs like ChatGPT to rephrase text, attempting to evade plagiarism detection. On the other hand, LLMs themselves can be used as powerful plagiarism detection tools. Although there are several options available, such as Turnitin and DupliChecker, students can use ChatGPT for free to check if a text is plagiarized. This allows them to assess the likelihood of being caught and make adjustments to escape detection.

Biörck and Eriksson 2023 conducted experiments with prompt engineering to evaluate ChatGPT's efficiency as a plagiarism detection tool, with one prompt proving particularly effective. In another study by Khalil and Er 2023, 50 essays generated by ChatGPT were tested, and ChatGPT outperformed traditional detection tools like iThenticate in identifying plagiarism.

Moreover, students can exploit a combination of tools such as ChatGPT, Quillbot, and paraphrasing software to bypass both plagiarism and AI-generated content (AIGC) detection systems, including Turnitin, GPTZero[2], and DupliChecker. By generating content with ChatGPT and then paraphrasing it, students can make the text appear both plagiarism-free and undetectable by AIGC detectors. This demonstrates how easily academic dishonesty can occur without being flagged by standard detection methods, posing challenges for educators.

Following the example in Figure 1, the text generated by ChatGPT is identified as fully AI-generated when evaluated with GPTZero (see 7).

---

[1]Wordtune (`https://www.wordtune.com`) and Quillbot (`https://www.quillbot.com`) are AI writing assistants that help improve writing quality and efficiency.

[2]GPTZero (`https://gptzero.me`) is a tool designed to detect AI-generated text and assess the likelihood of a given text being written by an AI.
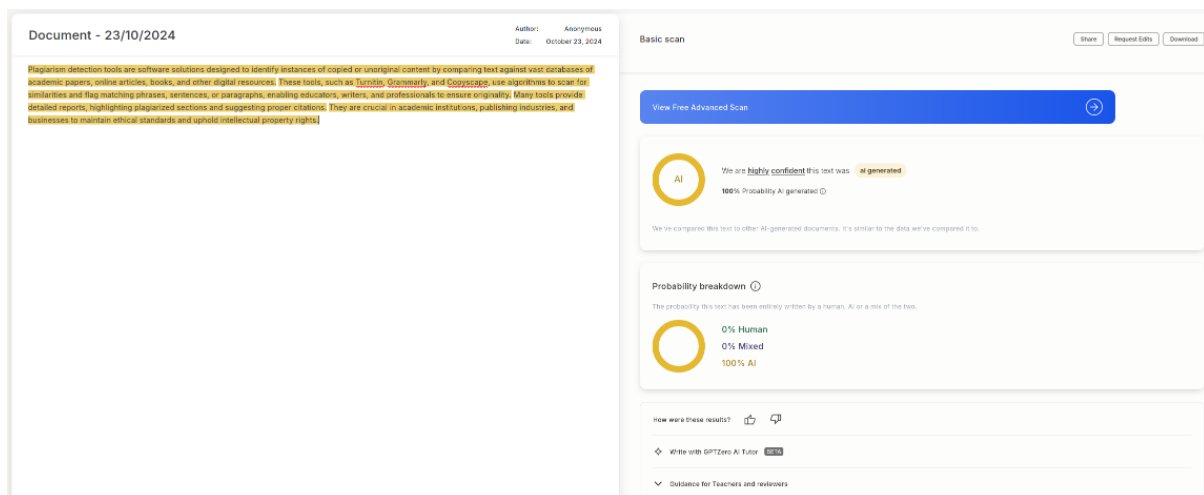
Figure 7: GPTZero evaluation of ChatGPT-generated text, confirming its classification as entirely AI-generated.

Conclusion

The challenge of plagiarism detection in academic and professional settings is more critical than ever, particularly in an era where digital content is easily accessible and often replicated. The tools discussed—Turnitin, DupliChecker, and large language models (LLMs) like Chat-GPT—represent the forefront of technology aimed at ensuring originality in written work.

Turnitin has established itself as a cornerstone in plagiarism detection, utilized by millions of students and educators worldwide. Its extensive database and sophisticated algorithm not only identify similarities in text but also assign a similarity score that helps gauge the originality of submitted work. While Turnitin effectively highlights potential instances of plagiarism through its color-coded system, it also underscores the need for careful interpretation of results. Educators must be equipped to assess the context of flagged sections, as the software may not always distinguish between legitimate citations and plagiarism.

DupliChecker enhances the plagiarism detection landscape by offering a versatile platform that serves various user needs—from students to professional writers. Its AI-driven scanning capabilities and additional features, such as grammar checking and writing improvement suggestions, provide a comprehensive approach to ensuring content integrity. The tool's dual offerings of free and paid plans allow users to select a level of service that meets their specific requirements, making it accessible to a broader audience.

The advent of large language models introduces both opportunities and challenges in plagiarism detection. While tools like ChatGPT can assist in generating and rephrasing content, they can also be misused by students seeking to bypass traditional detection methods. Interestingly, recent studies have indicated that these models can also be effective in identifying plagiarized content, outperforming some conventional tools. This dual capability emphasizes the complexity of the current academic landscape, where technology can both facilitate learning and enable

dishonesty.

In summary, plagiarism detection tools are vital in fostering academic integrity in today's educational landscape. By leveraging these technologies effectively, educators can not only identify potential instances of plagiarism but also cultivate a culture that values originality and critical thinking. Continuous education on the responsible use of these tools, along with a commitment to maintaining high standards of integrity, will be crucial in navigating the complexities of modern academic writing.

References

Biörck, Johann and Eriksson, Sofia (2023). *Detecting Plagiarism with ChatGPT Using Prompt Engineering*.

Bretag, Tracey (Dec. 2014). "Challenges in Addressing Plagiarism in Education". In: *PLOS Medicine* 10.12, pp. 1–4. DOI: 10.1371/journal.pmed.1001574. URL: https://doi.org/10.1371/journal.pmed.1001574.

CopyLeaks (2015). *CopyLeaks - Plagiarism Detection Tool for Web and Education*. Paid. Supports multiple file formats, including rich text and PDFs. Suitable for detecting plagiarism across web content and educational materials. URL: https://www.copyleaks.com/.

Duplichecker (2020). *Duplichecker - Free Online Plagiarism Checker*. Free. Allows users to copy-paste text or upload documents for basic plagiarism checks. Useful for quick and easy detection. URL: https://www.duplichecker.com/.

Grammarly (2016). *Grammarly - Writing Assistant with Plagiarism Detection*. Paid. Provides detailed plagiarism reports that highlight the source, along with writing suggestions for improvement. URL: https://www.grammarly.com/.

iThenticate (1996). *iThenticate - Plagiarism Detection for Academia and Research*. Paid. Supports multiple document formats and is widely used for academic and research plagiarism detection. URL: https://www.ithenticate.com/.

Khalil, Mohammad and Er, Erkan (2023). "Will ChatGPT G et You Caught? Rethinking of Plagiarism Detection". In: *International Conference on Human-Computer Interaction*. Springer, pp. 475–487.

PlagScan (2009). *PlagScan - Plagiarism Checker with Side-by-Side Comparison.* Free. Handles multiple file types and offers side-by-side comparison for a thorough plagiarism review. URL: `https://www.plagscan.com/`.

Pudasaini, Shushanta, Miralles-Pechuán, Luis, Lillis, David, and Salvador, Marisa Llorens (2024). "Survey on Plagiarism Detection in Large Language Models: The Impact of ChatGPT and Gemini on Academic Integrity". In: *arXiv preprint arXiv:2407.13105.*

Turnitin (2000). *Turnitin - Plagiarism Detection Tool.* Paid. Supports both intra-corporal (within a private database) and extra-corporal (across the web) plagiarism detection. URL: `https://www.turnitin.com/`.