

A Comparative Analysis of Quality Metrics in Human vs. AI-Generated Web Content for SEO

Tanja Krunic

The Higher Education Technical School of Professional Studies in Novi Sad
Novi Sad, Serbia City, Country
ORCID: 0009-0007-4352-3433

Abstract— The main focus of this research is the comparison of human-written and AI-generated texts optimized for search engines. For two distinct topics, appropriate keyword lists were created, and based on them, human-authored texts were written and compared to texts generated by the three most popular language model services: ChatGPT, Google Gemini, and Microsoft Copilot. The study tests the originality value of AI-generated text produced by various generators, as well as human-written content. The results indicate that both groups of texts are evaluated as original by two distinct plagiarism checkers. The research analyzes the ability of AI-generated content to bypass detection in its original version and after humanization. The findings indicate that content generated by Google Gemini bypasses AI-content detectors more easily than ChatGPT and Microsoft Copilot, especially after humanization. On the other hand, ChatGPT has shown the lowest probability of bypassing AI-content detectors. This research also analyzes how humanization affects text readability. The results indicate that humanization can further decrease readability, producing text that is less engaging for readers. Ultimately, the study concludes that no single AI-text generator, AI-content detector, or humanization tool is perfect. Human oversight remains very important in obtaining attractive, readable web content.

Keywords— *AI-generated content, SEO, Google Gemini, ChatGPT, Microsoft Copilot, AI detection, Text humanization, Readability*

I. INTRODUCTION

The rapid revolution of Large Language Models (LLMs) and their services, such as ChatGPT, Google Gemini, and Microsoft Copilot, has changed the field of Search Engine Optimization (SEO). These AI tools enable SEO engineers to write content much faster and to be multiple times more productive. However, this presents a new challenge for SEO engineers: besides of producing content, they must ensure that the content is human-like and of high quality. In other words, they are in the middle of a 'cat-and-mouse game' between AI writers and AI detectors.

Following the principle of action and reaction in technology, the development of AI detectors initiated the evolution of humanizers. Content writers are heavily using AI-detectors and text-humanizers to mask their AI-generated texts. Consequently, the majority of them are faced with the same question: how effective are humanizers and detectors at all? Can humanizers paraphrase AI-text to the extent that they easily bypass AI-detectors? Will the text remain readable after

humanization, is it more readable than initially, or is its quality compromised in the humanization process?

In [1], an evaluation of several AI detectors (AI Text Classifier, Content Detector AI, Copyleaks, QuillBot, GPT-2, and GPTZero) using Deepseek-generated content was performed in order to find out their efficiency. The obtained results have shown that only 50% of the tested detectors demonstrated high reliability in identifying AI-generated content. The results from that research also indicate that paraphrasing the text can mask AI-content, i.e., can limit the precision of AI-content detectors.

The efficiency of popular AI-content detectors, including OpenAI, Writer, Copyleaks, GPTZero, and CrossPlag was analyzed in [2]. The AI-content detectors have shown better results when detecting GPT 3.5-generated text content than GPT 4-generated and human-written text content. The results of the study also indicate differences between the efficiency of the AI detectors and conclude that AI-detection tools have to be supplemented with human review.

Another AI-content detector efficiency analysis is presented in [3], where manuscripts from the Annals of Surgical Oncology journal were compared against ChatGPT-generated versions using three different detection tools. The findings indicate that AI detection tools lack absolute reliability. Namely, the majority of human-written texts in this study were estimated at up to 9% AI-generated, and AI-generated articles received AI scores from 12-99.9% (with an average of 43.5%).

The efficiency of traditional antiplagiarism tools vs. modern AI-content detector tools is analyzed in [4]. The study claims traditional antiplagiarism tools to be of low efficiency in detecting AI-content (about 2.7%). On the contrary, modern AI-content tools performed up to 99%.

Recent findings in [5] demonstrate that while pure AI content is detectable with over 90% accuracy, the use of paraphrasing tools like QuillBot significantly reduces the reliability of leading AI-content detector tools such as GPT-Zero. Their study suggests that synonym-based evasion remains a primary challenge for maintaining academic integrity.

This was a motivation for the research herein. Namely, the research presented in this paper evaluates the performance of three leading LLM services in generating SEO-optimized content and tests the effectiveness of popular humanization techniques. By analyzing results across multiple AI content detection platforms and readability metrics, this study aims to identify which AI-

humanization combinations offer the best balance between detection evasion and text quality.

The rest of the paper is organized as follows. Section 2 provides the theoretical background on LLMs, AI-content detection mechanisms, text humanizers, and the role of content originality and readability in SEO. Section 3 describes the research methodology and the specific tools used in the experiment. Section 4 presents the experimental results, followed by a detailed analysis and discussion of the findings. Finally, Section 5 concludes the paper with some directions for future research.

II. THEORETICAL BACKGROUND

To provide a better understanding of the topic, this section gives a brief overview on the theory of AI-content, detection tools, humanization and their role in modern SEO.

A. LLMs as Text Content Sources

LLMs are advanced AI systems trained on large datasets obtained from books and web content. Based on this training, they learn to generate human-like text. They are able to generate vast amounts of content on almost any topic in a very short time, almost instantaneously [6]. According to [7], ChatGPT, Gemini, and Copilot are ranked among the most popular LLM-based services on the market, which was a motivation for using them in this research.

Traditionally, search engines relied on algorithms to match user queries with relevant content. But recently they use LLMs and don't just compare words; they are able to understand and value the meaning, context, and intent of any text. This produces a new challenge for SEO engineers. They have to create content that meets some expectations – it has to address user intent and to cover the relevant topic, rather than just repeat keywords and phrases. Moreover, the LLM-generated text content may be inaccurate, or is highly repetitive and offers no unique value to the reader. So, relying only on LLMs content without human review is not recommended. In other words, the best approach is to combine AI-generated text content with human review and creativity [8].

B. AI-Content Detection Mechanisms

A study conducted in 2024, by [9] claims that more than half of the text content available on the web is AI-generated. Furthermore, more recently research claims that only 25% of web content in 2025, is pure-human generated [10]. The rise of automated content generation has triggered the evolution of AI content detectors — tools intended to analyze text and determine to what extent it is AI-generated. Their algorithms evaluate sentence structure, word usage, and semantic meaning. Furthermore, they compare the text with large datasets of both AI-generated and human-written content to identify predictable AI patterns. They mainly use the same techniques that LLM services use to generate texts, like machine learning (ML) and natural language processing (NLP). ML relies on pattern recognition, which means that it becomes more sophisticated in AI-content detection when it analyses large amounts of texts. On the other hand, NLP analyses the linguistic complexity and semantic depth of the text. The lack of depth is a clear indicator of AI-generated text [11], [12].

C. Text Humanizers and AI-Content Masking Techniques

Text humanizers are tools designed to bypass AI-content detection mechanisms. They mainly rely on adversarial paraphrasing, increasing text perplexity (randomness of word choice) and burstiness (variation in sentence structure) to mimic human writing patterns. These tools manipulate linguistic patterns to mimic human-like irregularities [13], [14]. Recent research in [15] demonstrates that while pure AI-generated content is detectable with high accuracy, humanization tools like QuillBot significantly degrade the reliability of leading detectors. These masking techniques range from synonym replacement to structural reorganization.

D. Content Quality: Readability and Originality in SEO

Text content prepared for the purpose of SEO, is not only to satisfy search engines, but it has to remain readable and attractive to human readers. Modern search engines, particularly Google, have undergone significant algorithmic shifts to prioritize content that demonstrates Expertise, Authoritativeness, and Trustworthiness (E-A-T) [16]. As indicated in that source, this is especially critical for the so called 'Your Money Your Life' (YMYL) topics - pages that cover sensitive topics, such as health and finance, that can significantly influence a person's well-being and security, particularly when presenting inaccurate or unverified information [17]. Automated or unverified content on such pages is often penalized. Originality of text is known to be very important to search engines [18], and it can be easily verified using tools like Copyscape or Duplichecker. On the other hand, if a SEO engineer uses an AI-generated text, and uses humanization tools afterwards, there is no guarantee for success. Namely, the resulting text may bypass AI-content detectors, but may be of decreased readability [19]. So, it is very important to balance low AI-detection with high readability to keep content competitive and trustworthy.

III. THE HUMAN VS. AI-GENERATED WEB CONTENT ANALYSIS

A. Research Methodology

For the purpose of this research, two advertising texts aimed for website publication were both Human and AI-generated using three different AI tools: ChatGPT 4 [20], Microsoft Copilot [21] and Google Gemini [22], were generated. The topic of the first text was memory foam pillow, whereas the second text was advertising a sunscreen product.

Before generating the text, a short list of appropriate keywords for both topics is obtained using three SEO keyword tools Wordstream [23], Mangools [24] and Keyword Tool [25]. For a given topic, each of these three tools provides a large set of keywords. The criteria for generating an appropriate keyword list were relevance for the product, search volume and competition. Below are the keyword lists for the two products:

- Keyword list 1 - Memory foam product: neck pain support memory foam pillow, cloud pillow, side sleepers, cooling memory foam pillow, hypoallergenic pillow" in the case of the memory foam pillow product.
- Keyword list 2 - Sunscreen product: Best sunscreen, sunscreen for dry skin, sunscreen for men, sunscreen

for babies, sunscreen for women, sunscreen during pregnancy, hypoallergenic sunscreen, high SPF, water resistant sunscreen.

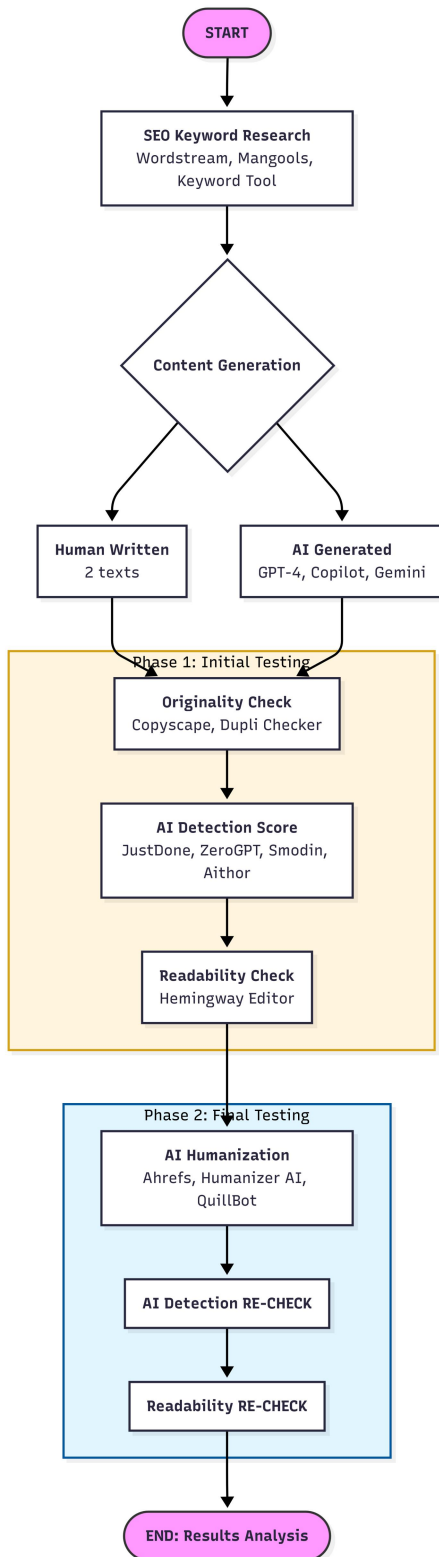


Fig. 1. Research methodology algorithm

Once the keyword lists for both texts were obtained, both texts were firstly human-written, and then AI-generated afterwards. All three aforementioned AI services were provided with the list of keywords, and a simple prompt of the form:

"Create an advertising text in English for [product type goes here] that includes the following keywords [the list of keywords goes here]."

The next step of this analysis was to use tools such as Copyscape [26] and Duplichecker [27] to check the originality of all texts. This is followed by using JustDone [28], ZeroGPT [29], Smodin [30] and Aithor [31] to check the percentage of AI-content of all texts, including both human-written and AI-generated versions of the texts. Then all texts have been checked by the Hemingway editor tool [32] in order to obtain a readability mark. Hemingway rates texts from 0-15. Grade 9 is considered ideal; that is, a text that is easy to read. Therefore, 9 and everything below is classified as Good. The range of 10–13 is considered OK, while 14 and above is classified as Poor. Grades above 15 are mainly designated as Postgraduate meaning that the language is too complicated for most.

AI-generated texts were subsequently humanized using Ahrefs Free AI Text Humanizer [33], Humanizer AI [34], and QuillBot Free AI Humanizer [35], after which readability was reassessed using the same tool. The above described research methodology algorithm is shown on Fig. 1.

B. Research Results

a) *Uniqueness Results:* All texts, including human-written and AI-generated, were considered as original by both Copyscape and Duplichecker. Namely, Copyscape didn't find any match in the analyzed texts from elsewhere on the internet, and Duplichecker assigned all texts as 100% unique. Fig. 2 shows Duplichecker's evaluation of uniqueness for the *memory foam pillow text*, generated with Gemini.

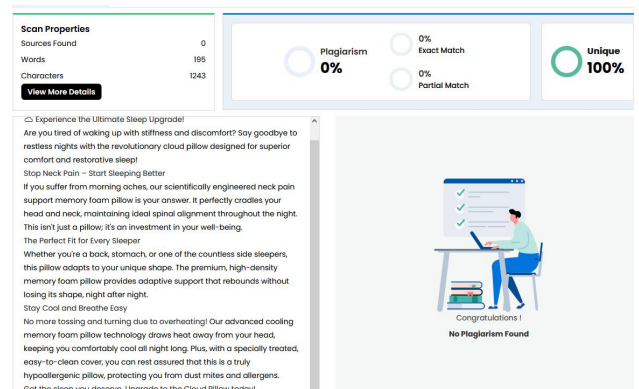


Fig. 2. Duplichecker's evaluation of uniqueness for text generated with Gemini

b) *AI-Generated Content Scores:* Table I shows how different tools estimate the proportion of AI-generated content across the analyzed texts.

TABLE I. AI-GENERATED CONTENT SCORES

Text		AI-Generated Content Scores (%)					
Keyword list	Model/Tool	Humanized by	JustDone	ZeroGPT	Smodin	Aithor	
Keyword list 1 – Memory foam pillow	Human - written	/	100	0	0	0	
	ChatGPT (GPT 4)	/	88	100	100	86	
		Quillbot AI Humanizer	93	100	100	49	
		Ahrefs humanizer	92	100	100	99	
		Humanizer AI	85	100	100	85	
	Google Gemini	/	85	96.41	100	23	
		Quillbot AI Humanizer	82	57.86	37	0	
		Ahrefs humanizer	82	100	32	75	
		Humanizer AI	88	100	40	75	
	Microsoft Copilot	/	92	100	100	99	
		Quillbot AI Humanizer	100	79.38	100	37	
		Ahrefs humanizer	73	100	100	99	
		Humanizer AI	89	100	100	65	
	Keyword list 2 – Sunscreen product	Human - written	/	75	0	0	0
		ChatGPT (GPT 4)	/	87	100	100	56
			Quillbot AI Humanizer	94	100	100	88
Ahrefs humanizer			89	100	90	99	
Humanizer AI			100	100	100	99	
Google Gemini		/	85	97.16	37	0	
		Quillbot AI Humanizer	78	97.32	37	0	
		Ahrefs humanizer	93	97.18	38	0	
		Humanizer AI	72	100	36	0	
Microsoft Copilot		/	87	100	100	56	
		Quillbot AI Humanizer	74	100	54	0	
		Ahrefs humanizer	100	100	100	55	
		Humanizer AI	94	100	100	92	

TABLE II. TEXT READABILITY SCORES

Keyword list	Source	Model/Tool	Readability score / Class (Hemigway editor) Initial assessment	Humanized by	Readability score / Class (Hemigway editor) - Reassessment	
Keyword list 1 - Memory foam pillow	Human - written	/	5 /OK	/	/	
	AI generated	ChatGPT (GPT 4)	14/Poor	Quillbot AI Humanizer	PG ^a / Poor	
				Ahrefs humanizer	14/Poor	
				Humanizer AI	14/Poor	
		Google Gemini	11 /OK	Quillbot AI Humanizer	9 /Good	
				Ahrefs humanizer	9/Good	
				Humanizer AI	10/OK	
	Microsoft Copilot	11/OK	Quillbot AI Humanizer	11/OK		
			Ahrefs humanizer	10/OK		
			Humanizer AI	12/OK		
	Keyword list 2 - Sunscreen product	Human - written	/	8/Good	/	/
		AI generated	ChatGPT (GPT 4)	12/OK	Quillbot AI Humanizer	11/OK
Ahrefs humanizer					11/OK	
Humanizer AI					11/OK	
Google Gemini			11/OK	Quillbot AI Humanizer	10/OK	
				Ahrefs humanizer	11/OK	
				Humanizer AI	PG/ Poor	
Microsoft Copilot		10/OK	Quillbot AI Humanizer	10/OK		
			Ahrefs humanizer	11/OK		
			Humanizer AI	15/Poor		

^a. PG indicates postgraduate-level text complexity.

c) Readability Scores

Table II shows the text readability scores for all content, including human-written texts, the original versions of AI-generated content, and their versions after humanization.

The readability score of the *sunscreen product text* generated with Gemini, and humanized with Humanizer AI, is shown on Fig. 3. As one can see, the postgraduate classification is obtained due to several complicated sentences within the text.

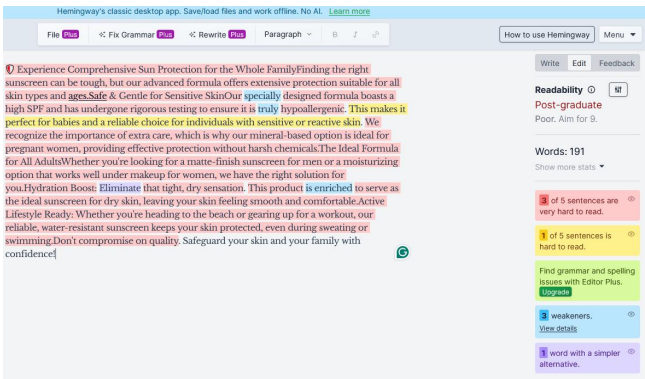


Fig. 3. Readability score of paraphrased AI-text content

C. Discussions

The research followed a strictly defined workflow: keyword selection, content generation, originality checks, AI-content detection, readability assessment, humanization, and readability reassessment of humanized texts.

The obtained results show that both human-written and AI-generated texts pass plagiarism checker tools like Duplichecker and Copyscape with 100% originality score. This confirms that modern LLMs do not simply copy-paste existing content but synthesize unique sentences.

On the other hand, AI-generated texts are still highly detectable as AI content by tools like ZeroGPT and Smodin. Google Gemini demonstrated the highest potential for evading AI detection, frequently achieving the lowest initial AI probability scores among all tested models. This claim can be supported by the examples in Table I, where Gemini achieved an initial AI probability score of 0% by Aithor and 37% by Smodin for the *sunscreen product text*, while other models showed significantly higher detection rates. Similarly, the *memory foam pillow text* generated by Gemini was scored by Aithor as 23% AI-generated, whereas the other models rated it as highly AI-generated. From this data, we can conclude maybe Gemini produces content with linguistic patterns that are less predictable, making it harder for detection tools to identify it as AI-generated. However, as shown in Table I, JustDone flagged both human-written texts as AI-generated content (the *memory foam pillow text* at 100% and the *sunscreen product text* at 75%), whereas all other tools correctly recognized them as 0% AI content. These false-positive results indicate that certain detectors are hypersensitive to SEO-optimized content, since it may have similar characteristics (containing repetitive phrases and structured patterns) to AI-generated content.

As can be seen in Table II, both human-written texts achieved high readability marks according to the Hemingway Editor (Grade 5–8, Class: Good). In contrast, AI-generated texts obtained lower readability marks (10–15), and some were even classified as having postgraduate complexity, being too complex and less likely to attract customers. These marks belong to the classes: OK and Poor. Compared to other AI-content generators, it seems that ChatGPT creates less readable results (14 for the *memory foam pillow text*, and 12 for the *sunscreen product text*).

As shown in Table I, ChatGPT-generated texts are highly resistant to humanization: few AI detection tools recorded a decrease in AI-content percentages, even after the text was processed by Humanizer, Quillbot, or Ahrefs. This suggests that GPT-4’s specific syntactic patterns are not easily altered by standard free humanization tools.

The combination of Google Gemini and Quillbot has shown to be the most successful in masking AI traces in the case of the *memory foam pillow text*, where AI probability dropped to 0% according to Aithor and to 37% according to Smodin. Even though results remained the same for the *sunscreen product text*, this combination demonstrated the highest potential and outperformed the other tested pairs.

As one can see from Table II, text humanization tools have limits when it comes to readability. Namely, in the majority of cases, the readability scores do not drop significantly after the AI-generated texts are humanized by any other of the three humanizer tools used in this research. Moreover, it is noticeable that humanization tools sometimes make the situation even worse. For example, Humanizer AI increased the readability score of the *sunscreen product text* generated by Microsoft Copilot from 10 (OK) to 15 (Poor), and the same text obtained by Google Gemini from 11 (OK) to Postgraduate (Poor). Similarly, Quillbot increased the readability score of the *memory foam pillow text* generated by ChatGPT from 14 (Poor) to Postgraduate (Poor). This happens because these tools often change words randomly with synonyms and distort the sentence structure to trick detectors. As a result, the text becomes less natural to the human eye. On the other hand, in some cases, humanization didn’t have any effect; for example, after humanization of the Gemini-generated *sunscreen product text* with Ahrefs, the initial readability score of 11 remains unchanged. However, in some cases, the readability score shows a positive improvement. For example, the Gemini-generated *memory foam pillow text* shifted from initial score 11 of (OK) to 9 (Good) after humanization with Ahrefs and Quillbot.

IV. CONCLUSIONS

The experiment presented in this paper was conducted using two distinct keyword lists as the primary input. For each list, four texts were analyzed: one human-written text (serving as the control group) and three AI-generated texts produced by different LLMs. Consequently, the dataset consisted of eight texts in total (two human-written and six AI-generated). All texts were subsequently evaluated for originality, readability, and the estimated percentage of AI-generated content they contained. All AI-generated texts were rewritten using three different AI humanizers, and their readability was subsequently re-assessed.

The results demonstrate the fact that while humanization tools can significantly increase the evasion rate against AI detectors, they often compromise readability. This means that the 'cat-and-mouse game' between AI-content generators, humanizers and detectors continues, but the ultimate winner in SEO remains content that provides genuine value to the user. Therefore, AI should be used for drafting, but human editing is essential—not only to evade detectors, but primarily to ensure the clarity of the message.

One direction for future research is to expand the sample size to include a broader range of topics and industries, which would provide more statistically significant data on the effectiveness of humanization tools. Another direction would be to include human readers who can evaluate if the text feels natural and helpful to a real user.

REFERENCES

- [1] H. Alshammari and P. Rao, "Evaluating the Performance of AI Text Detectors, Few-Shot and Chain-of-Thought Prompting Using DeepSeek Generated Text", 2025, <https://doi.org/10.48550/arXiv.2507.17944>
- [2] A. M. Elkhatat, K. Elsaid and S. Almer, "Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text", 2023, *Int J Educ Integr* 19, 17 (2023). <https://doi.org/10.1007/s40979-023-00140-5>
- [3] M. A. Flitcroft., S. A. Sheriff, N. Wolfrath et al. "Performance of Artificial Intelligence Content Detectors Using Human and Artificial Intelligence-Generated Scientific Writing". *Ann Surg Oncol* 31, 6387–6393 (2024). <https://doi.org/10.1245/s10434-024-15549-6>
- [4] C. Maddugoda, "A Comprehensive Review: Detection Techniques for Human-Generated and AI-Generated Texts", Technical Report via ResearchGate, Oct. 2023. [Online]. Available: <https://www.researchgate.net/publication/374825964> (Accessed: Jan. 27, 2026).
- [5] H. Hua and C. J. Yao, "Investigating generative AI models and detection techniques: impacts of tokenization and dataset size on identification of AI-generated text," *Frontiers in Artificial Intelligence*, vol. 7, Nov. 2024. [Online]. Available: <https://doi.org/10.3389/frai.2024.1469197>
- [6] E. Westbrooks, "How LLMs are changing content creation", 2025, <https://www.wrike.com/blog/how-llms-changing-content-creation/>
- [7] "What's the most popular LLM?", 2026, <https://www.thisisdefinition.com/insights/most-popular-llm>
- [8] LLMs are redefining SEO in 2025, 2025, <https://www.tangence.in/blog/large-language-models-in-seo-2025/>
- [9] "More Articles Are Now Created by AI Than Humans", 2024, <https://graphite.io/five-percent/more-articles-are-now-created-by-ai-than-humans>
- [10] R. Law , X. Guan and T. Soulo, "74% of New Webpages Include AI Content (Study of 900k Pages)", 2025, <https://ahrefs.com/blog/what-percentage-of-new-content-is-ai-generated/>
- [11] "How Do AI Detectors Work?", 2025, <https://www.coursera.org/articles/how-do-ai-detectors-work>
- [12] A. Barlow and V. Chien, 2024, "How AI Content Detectors Work To Check For AI", <https://gptzero.me/news/how-ai-detectors-work/>
- [13] M. Fontaine, "How Do AI Humanizers Work?", 2025, <https://gpthuman.ai/how-do-ai-humanizers-work/>
- [14] "How Does AI Humanizer Work", 2026, <https://cleverhumanizer.ai/how-does-ai-humanizer-work>
- [15] R. Marwah, "Humanizing AI-Generated Text: Techniques and Future Directions", 2024, *International Journal of Research in Engineering, Science and Management* Volume 7, Issue 11, November 2024
- [16] L. Ray, "Google's algorithm undergoes enormous shifts to prioritise content focused on expertise, authoritativeness and trustworthiness", 2019, *In the Journal of Digital & Social Media Marketing*, Volume 7, Issue 2. <https://doi.org/10.69554/IHCU2711>.
- [17] "What are YMYL Pages?", 2026, <https://ahrefs.com/seo/glossary/yml-pages>
- [18] M. van de Rakt, "The importance of original content for SEO", 2021, <https://yoast.com/importance-original-content/>
- [19] E. Masrour, B. Emi and M. Spero, "DAMAGE: Detecting Adversarially Modified AI Generated Text", 2025, <https://arxiv.org/pdf/2501.03437>
- [20] ChatGPT, 2025, <https://chatgpt.com/>
- [21] Microsoft Copilot, 2025, <https://copilot.microsoft.com/>
- [22] Google Gemini, 2025, <https://gemini.google.com/app>
- [23] WordStream, 2025, <https://www.wordstream.com/keywords>
- [24] Mangools, 2025, <https://mangools.com/kwfinder/>
- [25] Keyword Tool, 2025, <https://keywordtool.io/>
- [26] Copyscape, The world's most trusted plagiarism checker, 2025, <https://www.copyscape.com/>
- [27] Dupli Checker, 2025, <https://www.duplichecker.com/>
- [28] AI detector by JustDone, 2025, <https://justdone.com/ai-detector>
- [29] ZeroGPT AI detector, 2025, <https://www.zerogpt.com/>
- [30] Smodin, 2025, <https://smodin.io/>
- [31] Aithor, 2025, <https://aithor.com/bs-ba/ai-detektor>
- [32] Hemingway editor, 2025, <https://hemingwayapp.com/>
- [33] Ahrefs free AI text humanizer, 2025, <https://ahrefs.com/writing-tools/ai-humanizer>
- [34] Humanizer AI: Transform AI Text into Natural Human Language, 2025, <https://humanize.org/>
- [35] QuillBot - Free AI Humanizer Tool, 2025, <https://quillbot.com/ai-humanizer>