Journal of Mechatronics, Automation and Identification Technology
Vol. 11, No. 2, pp. 7 – 15, 2026.

A Comparative Analysis of Quality Metrics and Linguistic Shifts across Human, AI-Generated, and Humanized Web Content for SEO

Tanja KRUNIĆ

The Higher Education Technical School of Professional Studies in Novi Sad Department, Academy of Applied Studies, Novi Sad, Serbia
krunic@vtsns.edu.rs

***Abstract*— The main focus of this research is the comparison of human-written, AI-generated, and humanized 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 that generated by ChatGPT and Microsoft Copilot, especially after human-ization. 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. Additionally, a hybrid quantitative and qualitative analysis was performed. The quantitative metrics map the bounda-ries of detection score shifts, while the qualitative evaluation isolates specific linguistic anomalies, identifying repetitive patterns like structural freezing and syntactic inversion. Ultimately, the study concludes that no single AI-text gener-ator, AI-content detector, or humanization tool is perfect. Human oversight remains very important in obtaining at-tractive, readable web content.**

***Keywords*— AI-generated content, SEO, Large Language Models, AI detection, Text humanization, Readability, Marketing text quality, AI copywriting**

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 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, be more readable than initially, or will its quality be 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-generated content, i.e., can limit the precision of AI-generated 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 the primary motivation for the research herein, which represents an extended and significantly revised version of the study originally presented at the 25th International Symposium INFOTEH-JAHORINA 2026 [6]. While the preliminary work eval-uated the performance of three leading LLM services in generating SEO-optimized content and tested the effec-tiveness of popular humanization techniques across de-tection platforms, this expanded paper deepens the methodology. Specifically, the analysis is extended beyond baseline statistical evaluations by introducing a hybrid quantitative and qualitative evaluation framework. This allows the study to not only identify which AI-humanization combinations offer the best balance between detection evasion and text quality, but also to isolate and map repetitive linguistic patterns, such as structural freezing and syntactic inversion.

The rest of the paper is organized as follows. Section 2 provides the theoretical background on LLMs, AI-content detection mechanisms, text humanizers, the role of content originality and readability in SEO, and the impact of LLM stylistic biases on structural text variations. 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 in Section 5. Finally, Section 6 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 of the theory of AI-generated content, detection tools, humanization, and their role in modern SEO, while also examining the stylistic biases and structural anomalies in LLM outputs.

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 [7]. According to [8], 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 used 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 highly repetitive, and offer no unique value to the reader. So, relying only on LLM 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 [9].

B. AI-Content Detection Mechanisms

A study conducted in 2024 [10] reported 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 [11]. 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 text, 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 text. 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 [12], [13].

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 [14], [15]. Recent research in [16] 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) [17]. 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 [18]. Automated or unverified content on such pages is often penalized. Originality of text is known to be very important to search

engines [19], and it can be easily verified using tools like Copyscape or Duplichecker. On the other hand, if an SEO engineer uses 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 [20]. So, it is very important to balance low AI-detection with high readability to keep content competitive and trustworthy.

E. Stylistic Biases and Structural Anomalies in LLM Outputs

The way modern LLMs are trained directly shapes their output, which can sometimes cause stylistic errors [21]. Even though training methods ensure instruction-following, they often force models into a default informational layout with a neutral and predictable syntax [22]. This template can persist even when a task requires a specific style. This creates a contextual failure, where the model falls back on safe, formal structures instead of using persuasive or marketing prose.

Conversely, when a model successfully switches to a highly specialized stylistic mode—such as explicit commercial or promotional syntax—it presents a different challenge for downstream text processing [23]. In commercial text, the marketing intent and specific vocabulary are tightly locked into the sentence structure. When text humanizers try to modify this content, they face structural freezing. Because the meaning is deeply tied to specific marketing expressions, these tools must lock entire word groups to keep the original commercial message. Ultimately, these differences in model formatting and text density directly decide how well AI-generated text responds to humanization.

III. Methodology for Evaluating Web Content from Various Sources

The research methodology used herein is divided into three phases, which are shown in Fig. 1.

For the purpose of this research, advertising texts generated based on two different keyword lists and intended for website publication were analyzed. These included both human-written and AI-generated versions created using three different AI tools: ChatGPT 4 [24], Microsoft Copilot [25] and Google Gemini [26]. 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 [27], Mangools [28] and Keyword Tool [29]. 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.
- 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.

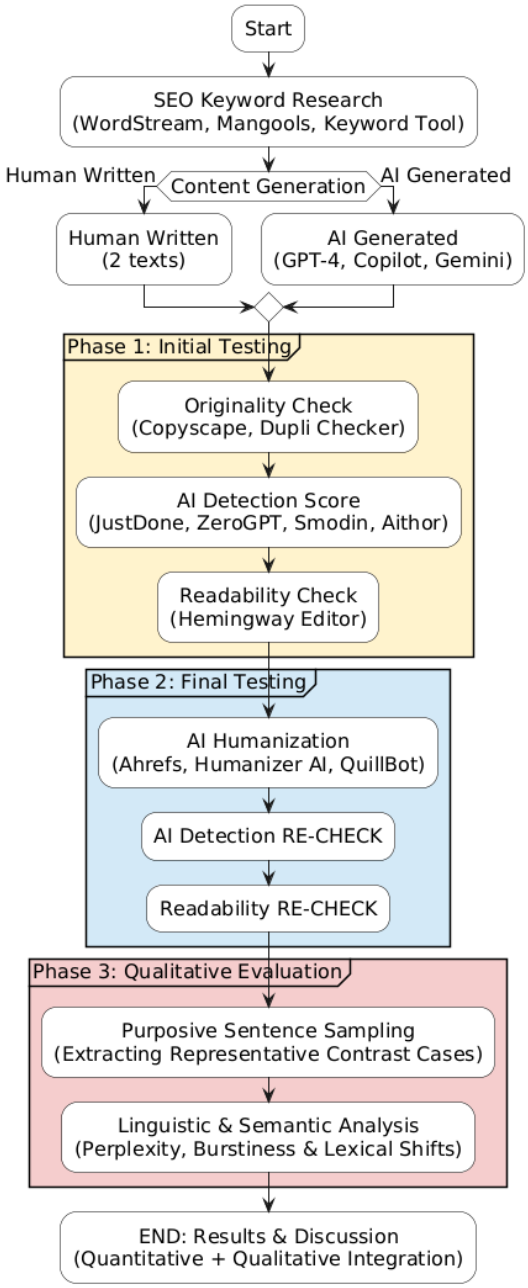

Figure 1 Research methodology algorithm

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 first step of Phase 1 of the analysis was to use tools such as Copyscape [30] and Duplichecker [31] to check the originality of all texts. This was followed by using JustDone [32], ZeroGPT [33], Smodin [34], and Aithor [35] to check the percentage of AI-generated 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 [36] 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 readers.

In Phase 2, the AI-generated texts were subsequently humanized using Ahrefs Free AI Text Humanizer [37], Humanizer AI [38], and QuillBot Free AI Humanizer [39], after which readability was reassessed using the same tool.

In Phase 3, we evaluated the structural rhythm and complexity of the generated content by calculating the coefficient of variation (CV) of sentence lengths and the mean sentence length (μ). This analysis was performed across both the raw AI baselines and their humanized versions to detect structural modifications beyond simple lexical changes.

To provide a deeper linguistic context, this quantitative metric was complemented by a qualitative text analysis based on a structured sample collection framework. Specific sentences from the text were collected to find repeating patterns. In addition, an analysis of the text samples after humanization was performed. The main focus was on how sentences were restructured, how words were reordered, and whether these shifts successfully altered the underlying AI templates.

IV. Experimental Results and Metrics for Human, AI-Generated, and Humanized Texts

A. Uniqueness Results

All texts, including human-written and AI-generated, were considered 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.

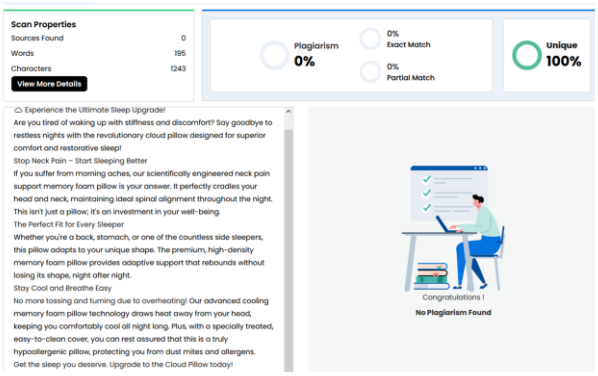

Fig. 2. Duplichecker's evaluation of uniqueness for text generated with Gemini

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*	*Humazied by*	*JustDone*	*ZeroGPT*	*Smodin*	*Aithor*
Keyword list 1 – Memory foam pillow	Human	/	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	/	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

A. *Readability Assessment*

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.

TABLE II. TEXT READABILITY SCORES

Keyword list	Model/Tool	Readability score / Class (Hemigway editor) Initial assessment	Humanized by	Readability score /Class (Hemigway editor) – Reassessment
Keyword list 1 - Memory foam pillow	Human	5 /OK	/	/
	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	8/Good	/	/
	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.

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.

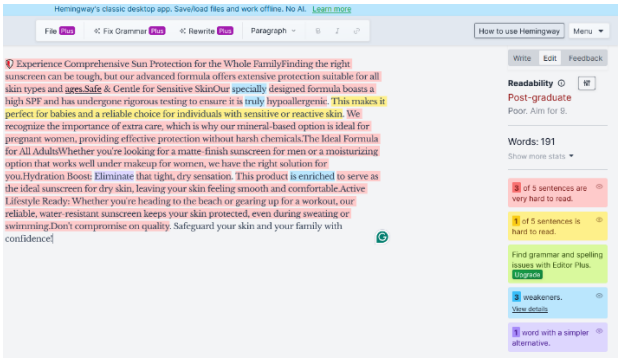

Fig. 3. Readability score of paraphrased AI-text content

B. *Quantitative Text Profiling*

Table III presents the comparative data for sentence length metrics, including the mean (μ), standard deviation (σ), and the Coefficient of Variation (CV), which is expressed as a percentage and calculated using the formula

$$CV = \frac{\sigma}{\mu} \cdot 100\% \,. \quad (1)$$

Note that the sample IDs used in Table III are generated using the pattern

[List ID] [Model ID] [Humanizer ID],

where:

- [List ID] corresponds to the specific keyword list used (K1 for memory foam pillow or K2 for sunscreen).
- [Model ID] indicates the generating source of the text (H for Human, G for ChatGPT, M for Gemini, C for Copilot),

[Humanizer ID] indicates the text humanizer tool applied (Q for Quillbot, H for Humanizer AI, A for Ahrefs), which is omitted for both human-written and raw AI-generated texts.

TABLE III. SENTENCE LENGTH METRICS FOR HUMAN, AI-GENERATED, AND HUMANIZED TEXTS

ID	Total Sentences	Min-Max Range	Mean Length (μ)	Standard Deviation (σ)	Coefficient of Variation (CV)
K1H	8	6-28	12.63	7.52	59.57
K1G	5	15-27	21.4	4.62	21.57
K1GQ	5	17-28	22.8	4.66	20.43
K1GH	6	12-23	17.83	4.12	23.1
K1GA	6	11-25	18.17	4.49	24.72
K1M	10	8-25	16.0	5.21	32.54
K1MQ	10	6-26	16.3	5.66	34.71
K1MH	10	6-23	15.5	5.13	33.07
K1MA	13	5-23	14.31	6.17	43.12
K1C	6	12-27	19.0	5.02	24.42
K1CQ	6	13.30	19.33	5.85	30.28
K1CH	5	12-24	19.4	4.45	22.94
K1CA	7	12-24	18.0	4.93	27.41
K2H	5	8-18	13.2	5.56	27.01
K2G	8	14.25	17.38	3.85	22.17
K2GQ	7	12-22	16.43	3.99	24.31
K2GH	8	14-23	16.13	3.04	18.88
K2GA	8	16-24	18.63	2.92	15.7
K2M	11	6-30	17.18	7.96	46.33

K2MQ	11	6-29	18.09	7.49	41.4
K2MH	10	6-28	17.5	8.42	48.13
K2MA	11	6-27	17.27	7.9	45.74
K2C	8	8-23	13.13	5.28	40.2
K2CQ	8	8-23	14.38	6.21	43.19
K2CH	8	8-19	12.75	4.23	33.21
K2CA	8	9-26	14.63	5.85	40.03

All values were computed using a custom script developed in the Wolfram Cloud environment, demonstrating a robust quantitative approach. Similar structural results can be replicated using any equivalent computational framework. The analysis covers all generated texts for both keyword lists (human, AI-generated, and humanized versions of texts). These texts represent short-form marketing content, containing only 5–11 sentences. This is very common in contemporary digital marketing texts. Namely, SEO content prioritizes concise, highly readable text according to the F-shaped visual scanning pattern typical of online users [40].

However, because perplexity reflects the randomness of word choice, establishing a statistically stable and representative distribution of this metric requires a significantly larger volume of text. Due to the short form of the analyzed samples, the perplexity values measured with [41] showed very little variation, ranging from 72 to 90, where the majority of values were around 80+. Consequently, these language model metrics failed to provide a discriminative baseline for comparison between human-written and AI-generated texts. For this reason, perplexity indicators were excluded from the primary statistical analysis, focusing strictly on structural and stylistic parameters.

C. Qualitative Linguistic Features

The results of the qualitative sentence analysis and the corresponding textual transformations are summarized in Table IV.

TABLE IV. LINGUISTIC FEATURES OF AI-GENERATED AND HUMANIZED TEXTS

ID & Source	**Text Comparison (Raw vs. Humanized)**	**Pattern/Impact**
K1G Marketing	Discover the perfect blend of comfort and support with our neck pain support memory foam pillow, designed to help you wake up refreshed and free of stiffness.	Pattern: K1G has an identical pattern to the Copilot promotional template for List 2 (K2C).
K1GQ Marketing	Our neck pain support memory foam pillow offers the ideal balance of comfort and support, enabling you to wake up feeling rejuvenated and stiff-free.	Impact: The humanized version K1Q applies syntactic inversion to K1G and introduces an unnatural, invented word (stiff-free).
K2C Marketing	Protect your skin and your loved ones with the best sunscreen designed for every lifestyle and need.	Pattern: K2C has an identical pattern to the ChatGPT promotional template for List 1 (K1G).
K2CQ Marketing	Use the best sunscreen for every need and lifestyle to protect your skin and the skin of your loved ones.	Impact: K2CQ is just a syntactic inversion of the raw AI sentence K2C; it shifts the primary CTA from the opening to the sentence end.
K1M Marketing	Say goodbye to restless nights with the revolutionary cloud pillow designed for superior comfort and restorative sleep!	Pattern: Inversion of the opening (*Say/wave goodbye*) into an unnatural closing cliché (*bid adieu*), mirroring the *say/wave goodbye* cliché found across other samples (K1GA, K1C, K1CA, K2M, K2MA).
K1MQ Marketing	With the groundbreaking cloud pillow, which offers exceptional comfort and rejuvenating sleep, bid adieu to sleepless nights!	Impact: Delays the main text focus to the end and replaces a standard phrase with an archaic cliché, neutralizing marketing urgency across samples.
K1M Marketing	Get the sleep you deserve. Upgrade to the Cloud Pillow today!	Pattern: Cosmetic insertion (*quality)* and verbatim copy of the CTA (*Upgrade... today!*) across samples (K1MA, K1C, K1CH, K1CA, K1G,K2C).
K1MA Marketing	Get the quality sleep you deserve. Upgrade to the Cloud Pillow today!	Impact: Structural freezing; the tool adds a single word but leaves the robotic AI pattern intact.
K1G Marketing	Upgrade your sleep and feel the difference every morning with a pillow that redefines comfort.	Pattern: Syntax inversion and systematic synonym rotation (*Upgrade →Invest*, *every → each*) across samples (K1G, K1GQ, K1GH, K1GA).
K1GQ Marketing	Invest in a pillow that redefines comfort to improve your sleep and experience the difference each morning.	Impact: Structural shifting; the tool reorders the clause but remains bound to the identical robotic marketing core.
K2M Marketing	Don't settle for less than the best. Protect your skin and your family with confidence!	Pattern: Minimal lexical swap (*Protect → Safeguard*) within the urgency-driven CTA.
K2M Marketing	Don't settle for anything less than the best. Safeguard your skin and your family with confidence!	Impact: High rigidity; humanizers cannot break the aggressive copywriting anchor, resulting in total structural freezing.

K2G Informational	With the right sunscreen, everyone in the family can enjoy the sun safely and comfortably.	Pattern: Syntactic inversion and clause reordering.
K2GQ Informational	Every member of the family can safely and comfortably enjoy the sun with the appropriate sunscreen.	Impact: High fluidity; the neutral baseline allows a complete structural shift.

V. Discussion

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

A. Analysis of AI Detection Scores, Readability and Humanizer Performance

The obtained results show that both human-written and AI-generated texts pass plagiarism checker tools like Duplichecker and Copyscape with a 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 that 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-generated 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 been 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 positive improvement. For example, the Gemini-generated memory foam pillow text shifted from initial score of 11 (OK) to 9 (Good) after humanization with Ahrefs and Quillbot.

B. Discussion on Sentence Burstiness

Table III shows clear structural differences between human, AI-generated, and humanized texts. In this study, burstiness is measured by the Coefficient of Variation of sentence length, which shows the rhythm and variance of the text. A key finding is the high flexibility of human writing, which depends on the context and inspiration. In the first dataset (K1H), the human author exhibits high structural burstiness (CV=59.57%), mixing short and long sentences. Conversely, in the second dataset (K2H), the human text has a much more uniform distribution (CV=27.01%), adapting to the style of the topic. This difference is also influenced by text length, as K1H contains 8 sentences while K2H contains only 5, making the latter more sensitive to statistical variation.

In contrast, LLMs work under strict rules, though different models behave differently. ChatGPT consistently maintains low burstiness across both lists (CV$\approx$15-24%). This structural monotony reflects an aggressive optimiza-

tion for standard SEO readability guidelines, which prioritize predictive uniformity over stylistic variety.

On the other hand, Gemini and Copilot demonstrate different generative patterns. Gemini exhibits consistently high burstiness across both keyword lists (CV≈32-48%), suggesting that its parameters favor structural variety. This explains the Table I results, where Gemini achieved the highest bypass rates. Copilot shows the highest sensitivity to topic transitions: it is regular like ChatGPT for the texts generated for the first keyword list, but changes to a highly diverse structure (CV>40%) for the second one.

Finally, the data highlights a major limitation of AI content humanizers. While these tools successfully alter word choice to bypass AI text detectors, they systematically fail to reconstruct the structural burstiness of the text. The CVand mean sentence length (μ) of humanized texts stay close to the original AI data. This shows that humanizers mainly change words, while the sentence structure remains largely the same.

C. *Discussion on Qualitative Linguistic and Syntactic Distortions*

As one can see from Table IV, text humanizers do not actually rewrite content with human-like creativity. Instead, they rely on predictable automated rules. Syntactic inversion seems to be the most frequent pattern across the observed humanized sentences. Namely, to trick AI detectors, humanizers often flip the AI-generated sentence structure. The main change in the sentence is the movement of the Call to Action (CTA) from the beginning to the end of the sentence, which is quite unnatural for marketing sentences.

To avoid predictable AI patterns, humanizers also sometimes use artificial and archaic vocabulary. This includes fabricating non-standard, unnatural words like "stiff-free" and replacing modern marketing phrases containing "Say goodbye to restless nights" with outdated phrases like "Bid adieu to sleepless nights", which actively neutralizes the promotional impact.

When dealing with short and aggressive marketing slogans, these tools completely freeze. They cannot change the tight marketing structure, so they only replace a few words with synonyms. As a result, the AI pattern remains completely unchanged. The baseline style of the original text dictates how fluid the humanization will be. Informational texts with a neutral tone allow the tools to achieve much higher structural flexibility. Conversely, in marketing texts, the tools remain tightly bound to rigid promotional templates, making their output highly predictable across different keyword sets.

VI. 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 then re-assessed.

The results demonstrate the fact that while humanization tools can significantly increase the evasion rate against AI detectors, they often compromise readability and pro-motional effectiveness. The findings indicate that these tools struggle with marketing forms; they may destroy the natural layout of promotional texts through syntactic inversions, rely on repetitive templates, and completely freeze when processing short slogans. Furthermore, the study highlights that initial LLM generation errors—such as producing a flat, informational tone instead of persua-sive copy—cannot be fixed by subsequent humanization. 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 pro-vides 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 and targeted tone 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] T. Krunić, „A Comparative Analysis of Quality Metrics in Human vs. AI-Generated Web Content for SEO“, 25th International Symposium INFOTEH-JAHORINA, 133-138, (2026). Available: https://infoteh.etf.ues.rs.ba/zbornik/2026/radovi/376.pdf

[7] E. Westbrooks, How LLMs are changing content creation, 2025, https://www.wrike.com/blog/how-llms-changing-content-creation/

[8] What’s the most popular LLM?, 2026, https://www.thisisdefinition.com/insights/most-popular-llm

[9] LLMs are redefining SEO in 2025, 2025, https://www.tangence.in/blog/large-language-models-in-seo-2025/

[10] More Articles Are Now Created by AI Than Humans, 2024, https://graphite.io/five-percent/more-articles-are-now-created-by-ai-than-humans

[11] 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/

[12] How Do AI Detectors Work?, 2025, https://www.coursera.org/articles/how-do-ai-detectors-work

[13] A. Barlow and V. Chien, 2024, How AI Content Detectors Work To Check For AI, https://gptzero.me/news/how-ai-detectors-work/

[14] M. Fontaine, How Do AI Humanizers Work? , 2025, https://gpthuman.ai/how-do-ai-humanizers-work/

[15] How Does AI Humanizer Work, 2026, https://cleverhumanizer.ai/how-does-ai-humanizer-work

[16] 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

[17] 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.

[18] What are YMYL Pages?, 2026, https://ahrefs.com/seo/glossary/ymyl-pages

[19] M. van de Rakt, "The importance of original content for SEO", 2021, https://yoast.com/importance-original-content/

[20] E. Masrour, B. Emi and M. Spero, "DAMAGE: Detecting Adversarially Modified AI Generated Text", 2025, https://arxiv.org/pdf/2501.03437

[21] OpenAI, "GPT-4 Technical Report", 2023, Available: https://arxiv.org/pdf/2303.08774

[22] F. A. Rodrigues, N. F. Sturm and F. L. Pinheiro, "A linguistic comparison between human- and AI-generated content", iScience, vol 29 (3), 2026, 1-11. Available: https://doi.org/10.1016/j.isci.2026.114976

[23] Gemini Team, Google. (2023). "Gemini: A Family of Highly Capable Multimodal Models." Available: https://arxiv.org/abs/2312.11805

[24] ChatGPT, 2025, https://chatgpt.com/

[25] Microsoft Copilot, 2025, https://copilot.microsoft.com/

[26] Google Gemini, 2025, https://gemini.google.com/app

[27] WordStream, 2025, https://www.wordstream.com/keywords

[28] Mangools, 2025, https://mangools.com/kwfinder/

[29] Keyword Tool, 2025, https://keywordtool.io/

[30] Copyscape, The world's most trusted plagiarism checker, 2025, https://www.copyscape.com/

[31] Dupli Checker, 2025, https://www.duplichecker.com/

[32] AI detector by JustDone, 2025, https://justdone.com/ai-detector

[33] ZeroGPT AI detector, 2025, https://www.zerogpt.com/

[34] Smodin, 2025, https://smodin.io/

[35] Aithor, 2025, https://aithor.com/bs-ba/ai-detektor

[36] Hemingway editor, 2025, https://hemingwayapp.com/

[37] Ahrefs free AI text humanizer, 2025, https://ahrefs.com/writing-tools/ai-humanizer

[38] Humanizer AI: Transform AI Text into Natural Human Language, 2025, https://humanize.org/

[39] QuillBot - Free AI Humanizer Tool, 2025, https://quillbot.com/ai-humanizer

[40] J. Nielsen and K. Pernice, 2010, Eyetracking Web Usability, New Riders. Berkeley, USA, ISBN-10: 0-321-49836-4

[41] Free Perplexity Checker, 2026, https://proofreaderpro.ai/perplexity-checker