

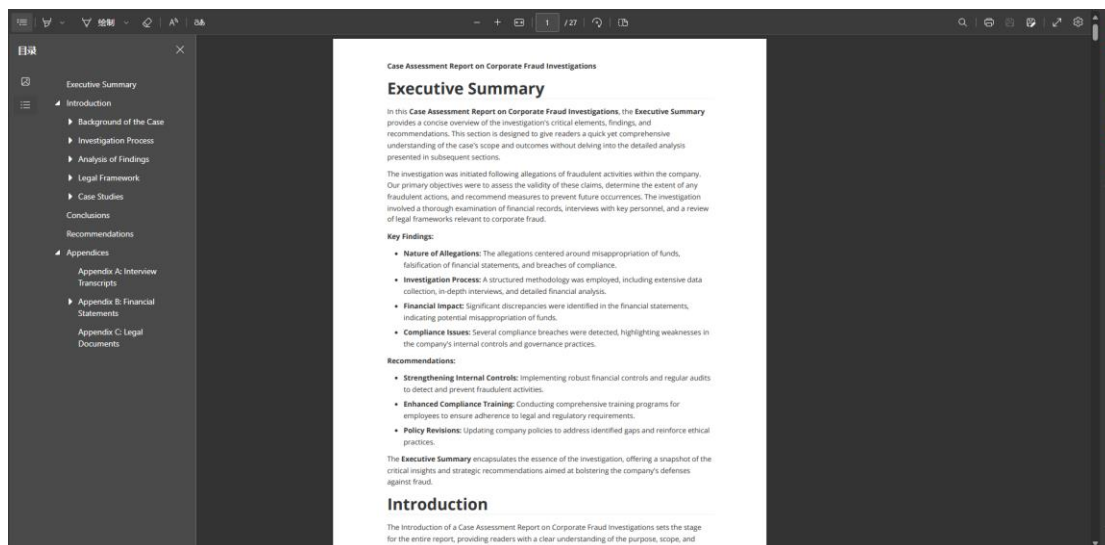
Thank you for participating in our evaluation task.

We have developed an AI algorithm that automatically generates long texts, and we would like your help in evaluating its performance. You will be provided with several articles. Please rate each article based on the given metrics and record your evaluation results in the provided file, [article.xlsx](#).

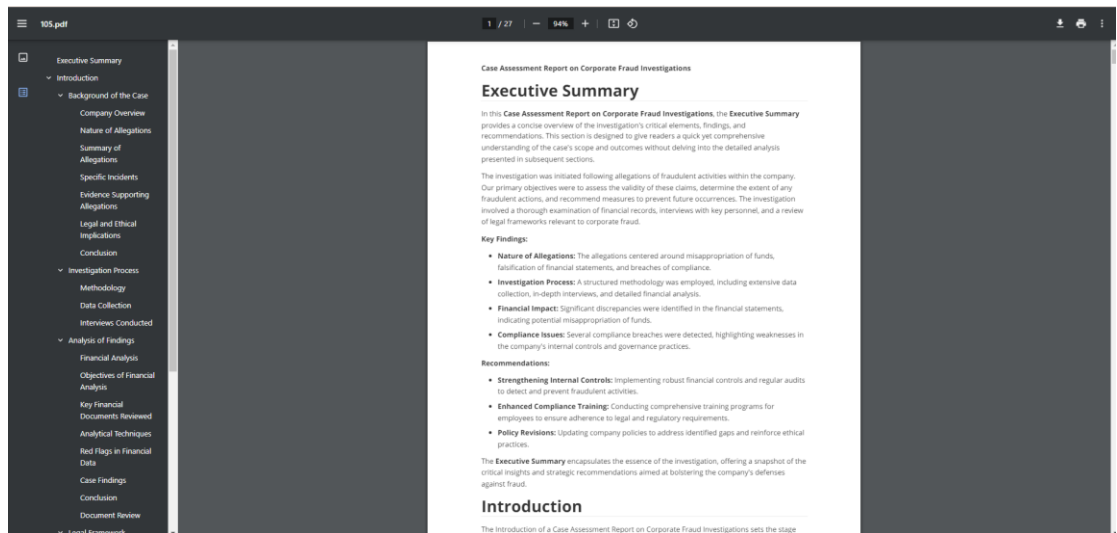
Note:

1. Some of the articles are written by real humans.
2. For comparison purposes, we utilize other AI algorithms besides ours to generate articles, resulting in varying quality levels
3. All PDFs include a table of contents, which you can open in your browser to quickly view the structure of each article.(Highly recommended!)

For edge:



For Chrome:



Demonstration:

	A	C	E	F	G	H	I	J	K	L	M	N
1	new_id	Title	Category	Fluency	Coherence	Consistency	Relevance	Length	Structure	Text_style	Adversarial_success	
2	1	Echoes of the Night	Horror Fiction	2	3	1	1	5	1	2	5	
3	2	The Dark Side of the Moon	Science Fiction	2	3	4	4	5	1	2	4	
4	3	Analysis of the Efficiency of Lean Manufacturing Process	Manufacturing Report	1	5	2	4	2	1	2	4	
5	4	Nihilism and Existentialism: A Comparative Study	Philosophy Paper	1	5	2	1	5	2	4	4	
6	5	Research on long text generation algorithm based on GPT-4	Artificial Intelligence Paper	2	5	4	1	2	2	4	4	
7	6	Modi Wins Third Term: Domestic Challenges and Global Implications	Political Opinion	5	3	4	1	1	2	4	5	
8	7	The Quest for the Golden Relic	Adventure Fiction									
9	8	The Last Colony on Mars	Science Fiction									
10	9	A Deep Dive into 'Inception': A Movie Review	Movie review or Book review									
11	10	Chicago Bridge Construction Project Management Report	Management Report									
12	11	Dating Rumors? Taylor Swift Declares They Are Just Speculation	Entertainment News									
13	12	Behavioral Analysis of Anxiety Disorders: Current Trends and Future Research	Psychology Paper									
14	13	Tourism Industry Recovery: Challenges and Opportunities in Post-Pandemic World	Economic Opinion									
15	14	The Quest for the Golden Relic	Adventure Fiction									
16	15	Battle of Two Cities	Drama									
17	16	The Shadows in the Basement	Horror Fiction									
18	17	The Museum Murder Mystery	Mystery Fiction									
19	18	Love in Paris	Romance Fiction									
20	19	Breakthrough in Quantum Computing: What It Means for the Future	Technology News									
21	20	The Impact of the Industrial Revolution on European Society	History Paper									
22	21	Mastering Python Programming: From Basics to Advanced Topics	Computer Science Textbook									
23	22	Genetic Engineering: Ethical Implications and Future Prospects	a terminology's encyclopedia articles									
24	23	American Immigration: Analyzing the Similarities and Differences Between the 19th and 21st Centuries	Political Opinion									
25	24	The Dragon's Prophecy	Fantasy Fiction									
26	25	Environmental Impact Assessment Report on the Hydroelectric Dam Project	Environment Report									
27	26	Chronicles of the Galactic Empire	Science Fiction									
28	27	Modi Wins Third Term: Domestic Challenges and Global Implications	Political Opinion									
29	28	The Last Emperor	Historical Fiction									
30	29	2024 Election: Pennsylvania and Wisconsin Emerge as Key Battleground States	Political News									
31	30	Japan's New Prime Minister Takes Office	a country's encyclopedia articles									
32	31	KOLO: a improved neural network architecture for image classification	Artificial Intelligence Paper									
33	32	Research on the Anodization and Performance of Nanomaterials	Chemistry Paper									

Mertics:

1. **Fluency:** The smoothness and readability of individual sentences.

Score:

- ☐ 1 (very bad): The sentences are very choppy, awkward, and difficult to read (e.g., "He go store. Later, back home. Tired, he was.").

- ☐ **2 (bad)**: The sentences have noticeable issues with flow and readability (e.g., "He went to the store. Then he come back home. He was tired.").
- ☐ **3 (neutral)**: The sentences are adequately smooth and readable, with minor issues (e.g., "He went to the store and came back home. He felt tired.").
- ☐ **4 (good)**: The sentences are generally smooth and easy to read, with few issues (e.g., "He went to the store, returned home, and felt tired.").
- ☐ **5 (very good)**: The sentences are exceptionally smooth, natural, and highly readable (e.g., "After going to the store, he returned home feeling tired.").

2. **Coherence**: The logical flow and connectivity between sentences.

Score:

- ☐ **1 (very bad)**: The sentences lack logical flow and connectivity, making the text confusing (e.g., "She went to the store. It was raining. He likes pizza.").
- ☐ **2 (bad)**: The sentences have noticeable issues with logical flow and connectivity (e.g., "She went to the store. It started raining. He was reading a book.").
- ☐ **3 (neutral)**: The sentences are somewhat logically connected, with minor issues (e.g., "She went to the store because it was raining. He read a book after that.").
- ☐ **4 (good)**: The sentences generally have a good logical flow and connectivity, with few issues (e.g., "She went to the store as it started raining. When she got back, he was reading a book.").
- ☐ **5 (very good)**: The sentences have an excellent logical flow and are highly connected (e.g., "As it started raining, she decided to go to the store. Upon her return, she found him engrossed in a book.").

3. **Consistency**: The uniformity and alignment of information throughout the text.

Score:

- ☐ **1 (very bad)**: The information is highly inconsistent and poorly aligned throughout the text (e.g., "The report starts with data on apples but ends with unrelated information on cars without any transition.").
- ☐ **2 (bad)**: The information has noticeable inconsistencies and alignment issues (e.g., "The text discusses the benefits of exercise, then suddenly shifts to a different topic without clear connections.").
- ☐ **3 (neutral)**: The information is somewhat consistent, with minor inconsistencies and alignment issues (e.g., "The text mostly sticks to the topic of diet, but occasionally veers off into unrelated anecdotes.").
- ☐ **4 (good)**: The information is generally consistent and well-aligned, with few issues (e.g., "The text follows the theme of healthy living, with clear transitions between diet and exercise.").
- ☐ **5 (very good)**: The information is highly consistent and well-aligned throughout the text (e.g., "The text maintains a clear focus on the benefits of a balanced diet, seamlessly integrating related subtopics.").

4. **Relevance:** The pertinence of the content to the expected topics.

Score:

- ☐ **1 (very bad)**: The content is highly irrelevant to the expected topics (e.g., "An essay supposed to be about climate change discusses a recipe for apple pie.").
- ☐ **2 (bad)**: The content has noticeable irrelevance to the expected topics (e.g., "An article on health benefits of exercise spends a lot of time talking about a vacation trip.").
- ☐ **3 (neutral)**: The content is somewhat relevant to the expected topics, with minor deviations (e.g., "A report on renewable energy includes some unrelated sections on historical events.").
- ☐ **4 (good)**: The content is generally relevant to the expected topics, with few deviations (e.g., "A paper on the impact of technology in education occasionally mentions unrelated business trends.").

- ☐ **5 (very good)**: The content is highly relevant and closely aligned with the expected topics (e.g., "An article on the effects of diet on health remains focused on nutritional information and health outcomes.").

5. **Length**: The extent to which the text meets the expected length requirements for humans (e.g., for essays, reports, articles, etc.).

Score:

- ☐ **1 (very bad)**: The text significantly deviates from the expected length requirements (e.g., an essay expected to be a full discussion is extremely brief or overly long and unfocused).
- ☐ **2 (bad)**: The text has noticeable issues meeting the expected length requirements (e.g., an essay is clearly too short to cover the topic adequately or too long with redundant information).
- ☐ **3 (neutral)**: The text somewhat meets the expected length requirements, with minor deviations (e.g., an essay is somewhat shorter or longer than expected but covers the main points).
- ☐ **4 (good)**: The text generally meets the expected length requirements, with few issues (e.g., an essay is mostly within the expected range and covers the topic well).
- ☐ **5 (very good)**: The text fully meets the expected length requirements (e.g., an essay is well within the expected range and thoroughly covers the topic).

6. **Structure**: The presence of a clear and logical organizational structure.

Score:

- ☐ **1 (very bad)**: The text lacks a clear organizational structure, making it difficult to follow (e.g., "An essay starts discussing the conclusion first, jumps to background information, and then presents the main argument in a disorganized manner.").

- ☐ **2 (bad)**: The text has noticeable issues with its organizational structure, leading to some confusion (e.g., "A report introduces subtopics without any headings or transitions, making it hard to distinguish between different points.").
- ☐ **3 (neutral)**: The text has a somewhat clear organizational structure, with minor issues (e.g., "A research paper follows a basic structure but occasionally places evidence before presenting the main point, causing slight confusion.").
- ☐ **4 (good)**: The text generally has a clear and logical organizational structure, with few issues (e.g., "An article follows a logical flow with an introduction, body, and conclusion, though one or two paragraphs might seem slightly out of place.").
- ☐ **5 (very good)**: The text has an exceptionally clear and logical organizational structure, making it easy to follow (e.g., "A comprehensive essay starts with a clear introduction, follows with well-organized paragraphs that build on each other logically, and concludes with a strong summary and call to action.").

7. **Text Style:** To what extent the writing style matches the expected style for the given type of text.

Score:

- ☐ **1 (very bad)**: The writing style is highly inappropriate for the given type of text (e.g., using slang in a formal academic paper).
- ☐ **2 (bad)**: The writing style has noticeable issues and does not align well with the expected style (e.g., overly casual tone in a business report).
- ☐ **3 (neutral)**: The writing style somewhat matches the expected style, with minor issues (e.g., a few informal phrases in an otherwise formal essay).
- ☐ **4 (good)**: The writing style generally matches the expected style, with few issues (e.g., mostly formal language in an academic paper, with occasional lapses).

- ☐ **5 (very good)**: The writing style perfectly matches the expected style for the given type of text (e.g., consistently formal language in an academic paper or a consistently conversational tone in a blog post).

8. **Adversarial success**: To what extent do you believe that this article was written by a human?

Score:

- ☐ **1 (very likely machine)**: The text is highly likely to have been written by a machine, with clear and obvious indicators (e.g., unnatural language, repetitive patterns).
- ☐ **2 (likely machine)**: The text has noticeable signs of being machine-written (e.g., awkward phrasing, inconsistent style).
- ☐ **3 (uncertain)**: The text somewhat resembles human writing, with some indicators of machine generation (e.g., occasional unnatural sentences).
- ☐ **4 (likely human)**: The text generally appears to be written by a human, with few indicators of machine generation (e.g., mostly natural language, minor lapses).
- ☐ **5 (very likely human)**: The text appears to be written by a human, with no obvious indicators of machine generation (e.g., consistently natural and coherent language).