



Subhrajyoti Roy
roy.s@wustl.edu

*Postdoctoral Researcher
Department of Statistics and Data Science
Washington University at St. Louis*

Oct 10, 2025

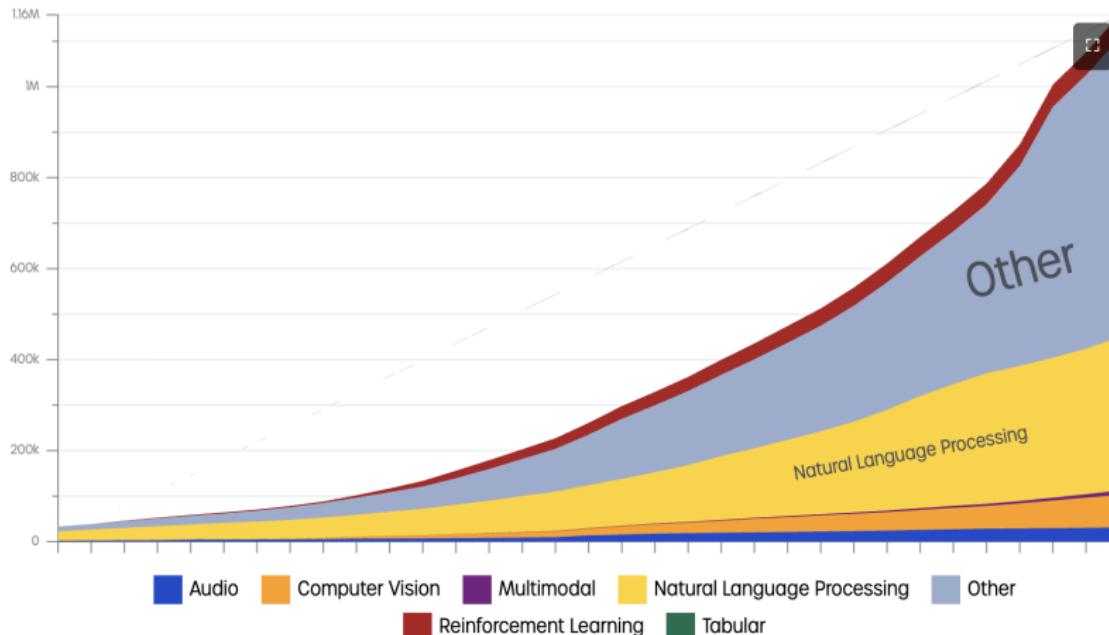
Localized Detection of Authenticity in Mixed Source Texts via Epidemic Change-point Perspective

- Soham Bonnerjee, Department of Statistics, University of Chicago.
- Sayar Karmakar, Department of Statistics, University of Florida.

Exponential Growth of LLM

Zero to One (Million Models)

This exponential growth chart tracks the Hugging Face community's journey from just a few thousand models in 2022 to surpassing the million-model milestone today.



Latest Count from June, 2025: $\approx 1.86M$.

Rise of LLM Generated Content

EPIUM Home • AI News • Blog

74% Of New Web Pages Now Contain Artificial Intelligence-Generated Content, Study Finds

Op Eds

Why Using ChatGPT to Write this Op-Ed Was a Smart Idea

By [REDACTED] Contributing Opinion Writer

[REDACTED] is a first-year Master in Public Policy student at the Harvard Kennedy School.

December 9, 2022

Editor's note: The following op-ed was entirely written and edited by ChatGPT, a recently released artificial intelligence language model that is available for anyone to use. No manual edits were made; all changes were made by the author and op-eds editor providing feedback to ChatGPT on the drafts it generated.

nature

Explore content ▾ About the journal ▾ Publish with us ▾ Subscribe

nature > news > article

NEWS | 23 September 2025

Journals infiltrated with 'copycat' papers that can be written by AI

Tools such as ChatGPT can be used to generate almost-identical research papers that pass standard plagiarism checks. Hundreds are thought to have been published.

Why we need authenticity verification?

- **Attribution:** Who wrote this? Human or AI?
- Academic Integrity.
- Misinformation (spam, political content).
- Intellectual Property / Copyrighting.

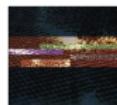
Emerging Issues with Academic Integrity



ELSEVIER

Surfaces and Interfaces

journal homepage: www.sciencedirect.com/journal/surfaces-and-interfaces



The three-dimensional porous mesh structure of Cu-based metal-organic-framework - aramid cellulose separator enhances the electrochemical performance of lithium metal anode batteries



ARTICLE INFO

Keywords:
Lithium metal battery
Lithium dendrites
CuMOF-ANFs separator

ABSTRACT

Lithium metal, due to its advantages of high theoretical capacity, low density and low electrochemical reaction potential, is used as a negative electrode material for batteries and brings great potential for the next generation of energy storage systems. However, the production of lithium metal dendrites makes the battery life low and poor safety, so lithium dendrites have been the biggest problem of lithium metal batteries. This study shows that the larger specific surface area and more pore structure of Cu-based metal-organic-framework - aramid cellulose (CuMOF-ANFs) composite separator can help to inhibit the formation of lithium dendrites. After 110 cycles at 1 mA/cm², the discharge capacity retention rate of the Li-Cu battery using the CuMOF-ANFs separator is about 96 %. Li-Li batteries can continue to maintain low hysteresis for 2000 h at the same current density. The results show that CuMOF-ANFs composite membrane can inhibit the generation of lithium dendrites and improve the cycle stability and cycle life of the battery. The three-dimensional (3D) porous mesh structure of CuMOF-ANFs separator provides a new perspective for the practical application of lithium metal battery.

1. Introduction

Certainly, here is a possible introduction for your topic: Lithium-metal batteries are promising candidates for high-energy-density rechargeable batteries due to their low electrode potentials and high

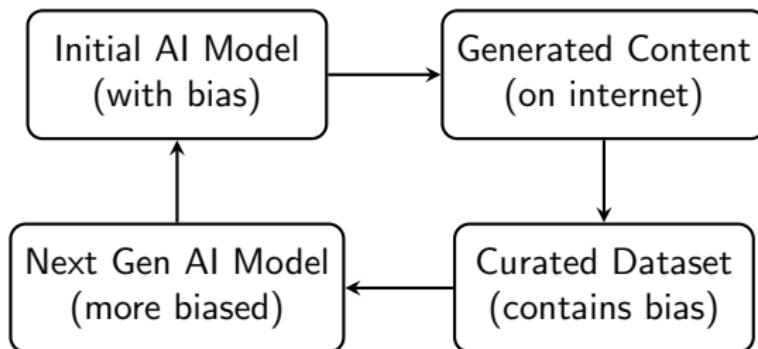
chemical stability of the separator is equally important as it ensures that the separator remains intact and does not react or degrade in the presence of the electrolyte or other battery components. A chemically stable separator helps to prevent the formation of reactive species that can further promote dendrite growth. Researchers are actively exploring

More Upcoming Threats!

- 1 In order to train next generation of models, we need new datasets.
- 2 Datasets are curated from the internet, which already contains AI generated texts.

More Upcoming Threats!

- 1 In order to train next generation of models, we need new datasets.
- 2 Datasets are curated from the internet, which already contains AI generated texts.
- 3 Can create vicious loop, decreasing fairness, increasing “not-so-good” qualities.



Large-language models are trained to produce texts indistinguishable from human-generated texts.

It is impossible to detect AI-generated vs human-generated content based on text alone.



Large-language models are trained to produce texts indistinguishable from human-generated texts.

It is impossible to detect AI-generated vs human-generated content based on text alone.



Key Idea: Inject statistical signals (watermark) into AI-generated text that can be detected later! And here, statistics can help us!!

The growth of “explore”

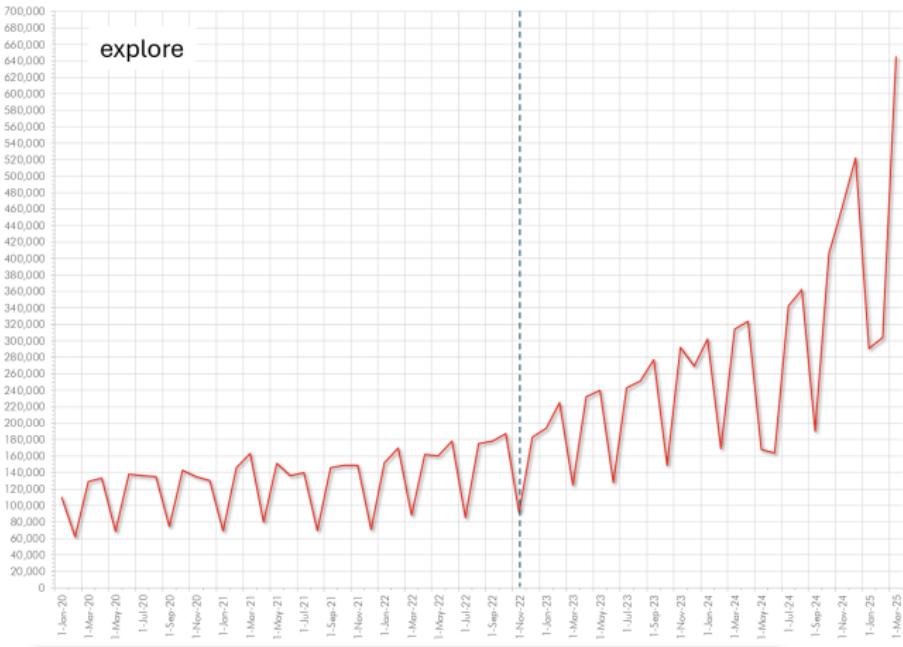


Figure: Monthly frequency of webpages that contain the word “explore”¹

¹Spennemann D. HR., “Delving into” the quantification of Ai-generated content on the internet,
<https://arxiv.org/pdf/2504.08755>

The growth of “delve into”

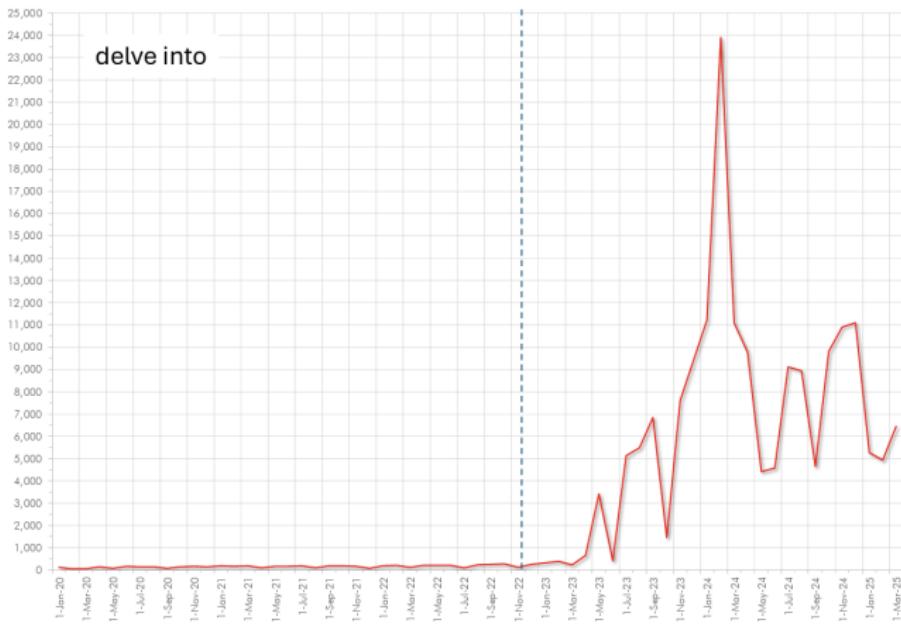


Figure: Monthly frequency of webpages that contain the word “delve into”²

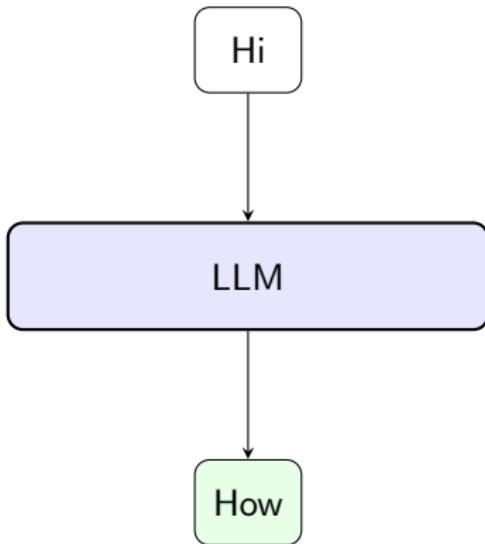
²

Spennemann D. HR., “Delving into” the quantification of Ai-generated content on the internet,
<https://arxiv.org/pdf/2504.08755>

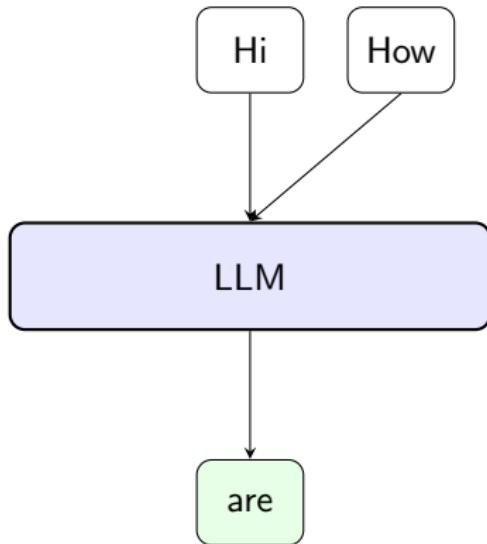
The Mechanism of Large Language Models

A large language model is an autocomplete system on steroids.

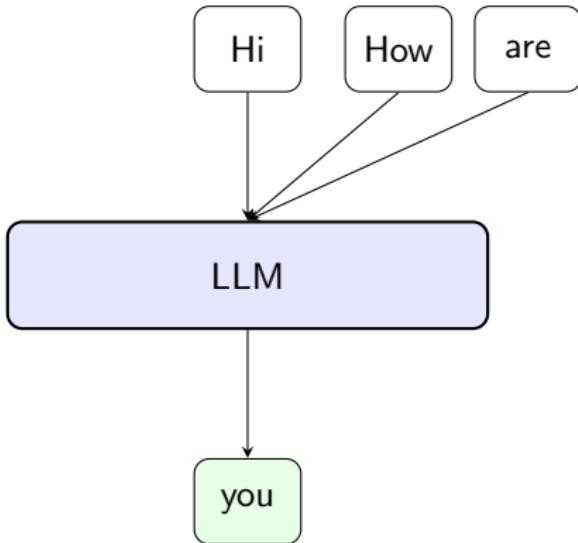
The Mechanism of Large Language Models



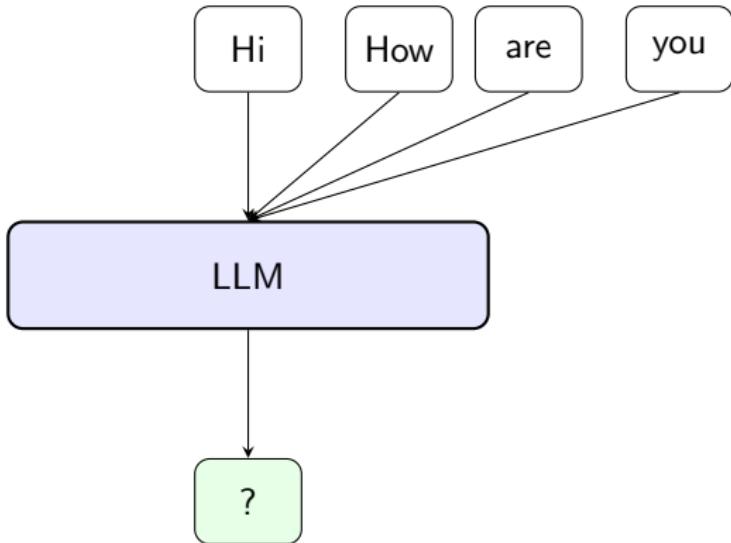
The Mechanism of Large Language Models



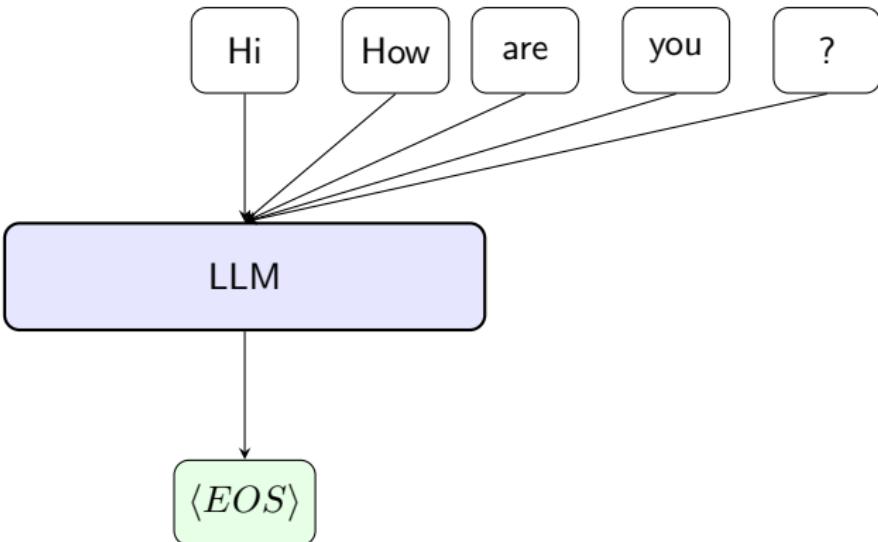
The Mechanism of Large Language Models



The Mechanism of Large Language Models



The Mechanism of Large Language Models



Given w_1, w_2, \dots, w_{t-1} , the LLM \mathcal{M} outputs a next token probability distribution P_t over the vocabulary \mathcal{W} . The next token is chosen as $w_t \sim P_t$.

- 1 Assign words in the vocabulary \mathcal{W} to either Red R or Green G bucket.
- 2 After LLM produces P_t , compute logits from P_t , say $l_t(w)$ for each $w \in \mathcal{W}$.
- 3 Define,

$$P_t^{(rg)}(w) = \frac{\exp(l_t(w) + \delta \mathbf{1}(w \in G))}{\sum_{w \in \mathcal{W}} \exp(l_t(w) + \delta \mathbf{1}(w \in G))}$$

- 4 Sample $w_t \sim P_t^{(rg)}$.

Red Green Watermark (Contd.)

	Flocking is a <u>type</u> of coordinated group behavior that is exhibited by animals of various species, including birds, fish, and insects. It is characterized by the <u>ability</u> of the animals to <u>move</u> together in a <u>coordinated</u> and cohesive manner, as if they were a single entity. Flocking behavior is <u>thought</u> to have <u>evolved</u> as a way for animals to increase their chances of survival by working together as a group. For example, flocking birds may be able to <u>locate</u> food more <u>efficiently</u> or <u>defend</u> themselves against predators more <u>effectively</u> when they <u>work</u> together.
Original	Flocking is a <u>kind</u> of coordinated team behavior that is exhibited by animals of several species, notably birds, fish, and insects. It is characterized by the ability of the animals to move together in a coordinated and cohesive way, as if they were a single entity. Flocking behavior is believed to have evolved as a way for animals to raise their likelihood of survival by working together as a group. For instance, flocking birds could be able to locate nutrition more <u>efficiently</u> or <u>defend</u> themselves against predators more <u>effectively</u> when they <u>work</u> together.

Red Green Watermark (Contd.)

Original	Flocking is a <u>type</u> of coordinated group behavior that is exhibited by animals of various species, including birds, fish, and insects. It is characterized by the <u>ability</u> of the animals to <u>move</u> together in a <u>coordinated</u> and cohesive manner, as if they were a single entity. Flocking behavior is <u>thought</u> to have <u>evolved</u> as a way for animals to increase their chances of survival by working together as a group. For example, flocking birds may be able to <u>locate</u> food more <u>efficiently</u> or <u>defend</u> themselves against predators more <u>effectively</u> when they <u>work</u> together.
Watermarked	Flocking is a <u>kind</u> of coordinated team behavior that is exhibited by animals of several species, notably birds, fish, and insects. It is characterized by the <u>ability</u> of the animals to <u>move</u> together in a <u>coordinated</u> and cohesive <u>way</u> , as if they were a <u>single</u> entity. Flocking behavior is <u>believe</u> to have <u>evolved</u> as a <u>way</u> for animals to raise their <u>likelihood</u> of survival by <u>working</u> together as a <u>group</u> . For <u>instance</u> , flocking birds <u>could</u> be <u>able</u> to <u>locate</u> <u>nutrition</u> more <u>efficiently</u> or <u>defend</u> themselves against predators more <u>effectively</u> when they <u>work</u> together.

Detection Rule

Let $|G| = \gamma|\mathcal{W}|$. There is a presence of watermark if

$$\frac{\frac{1}{T}(\# \text{ of tokens from } G) - \gamma}{\sqrt{\frac{\gamma(1-\gamma)}{T}}} \text{ is large}$$

Red Green Watermark (Contd.)

Original	Flocking is a <u>type</u> of coordinated group behavior that is exhibited by animals of various species, including birds, fish, and insects. It is characterized by the <u>ability</u> of the animals to <u>move</u> together in a <u>coordinated</u> and cohesive manner, as if they were a single entity. Flocking behavior is <u>thought</u> to have <u>evolved</u> as a way for animals to increase their chances of survival by working together as a group. For example, flocking birds may be able to <u>locate</u> food more <u>efficiently</u> or <u>defend</u> themselves against predators more <u>effectively</u> when they <u>work</u> together.
Watermarked	Flocking is a <u>kind</u> of coordinated team behavior that is exhibited by animals of several species, notably birds, fish, and insects. It is characterized by the <u>ability</u> of the animals to <u>move</u> together in a <u>coordinated</u> and cohesive <u>way</u> , as if they were a <u>single</u> entity. Flocking behavior is <u>believe</u> to have <u>evolved</u> as a <u>way</u> for animals to raise their <u>likelihood</u> of survival by <u>working</u> together as a group. For <u>instance</u> , flocking birds <u>could</u> be <u>able</u> to <u>locate</u> <u>nutrition</u> more <u>efficiently</u> or <u>defend</u> themselves against predators more <u>effectively</u> when they <u>work</u> together.

Detection Rule

Let $|G| = \gamma|\mathcal{W}|$. There is a presence of watermark if

$$\frac{\frac{1}{T}(\# \text{ of tokens from } G) - \gamma}{\sqrt{\frac{\gamma(1-\gamma)}{T}}} \text{ is large}$$

Problem: It modifies the token distribution P_t to something else.

- 1 Use LLM \mathcal{M} to produce NTP P_t .
- 2 Generate i.i.d. random variables $U_1, U_2, \dots, U_{|\mathcal{W}|} \sim U(0, 1)$.
- 3 Output
$$w_t = \arg \max_{w \in \mathcal{W}} \frac{\log(U_w)}{P_t(w)}$$
- 4 Turns out, w_t still follows P_t .
- 5 But we know that if w_t is selected, that means $\log(U_{w_t})$ must be large, or $U_{w_t} \approx 1$.

General Framework for LLM Watermark

- 1 Based on w_1, \dots, w_{t-1} , use LLM to generate P_t .
- 2 Based on w_1, \dots, w_{t-1} , generate a pseudorandom seed $\xi_t := \text{Hash}(w_1, \dots, w_{t-1}; \text{Key})$.
- 3 Use a decoding strategy $w_t \leftarrow S(P_t, \xi_t)$.

General Framework for LLM Watermark

- 1 Based on w_1, \dots, w_{t-1} , use LLM to generate P_t .
- 2 Based on w_1, \dots, w_{t-1} , generate a pseudorandom seed $\xi_t := \text{Hash}(w_1, \dots, w_{t-1}; \text{Key})$.
- 3 Use a decoding strategy $w_t \leftarrow S(P_t, \xi_t)$.
- 4 If w_t is human-generated, we don't expect w_t to be correlated with ξ_t in any way.

General Framework for LLM Watermark

- 1 Based on w_1, \dots, w_{t-1} , use LLM to generate P_t .
- 2 Based on w_1, \dots, w_{t-1} , generate a pseudorandom seed $\xi_t := \text{Hash}(w_1, \dots, w_{t-1}; \text{Key})$.
- 3 Use a decoding strategy $w_t \leftarrow S(P_t, \xi_t)$.
- 4 If w_t is human-generated, we don't expect w_t to be correlated with ξ_t in any way.
- 5 If w_t is AI-generated, the decoder $S(\cdot)$ creates some dependency between w_t and ξ_t .

General Framework for LLM Watermark

- 1 Based on w_1, \dots, w_{t-1} , use LLM to generate P_t .
- 2 Based on w_1, \dots, w_{t-1} , generate a pseudorandom seed $\xi_t := \text{Hash}(w_1, \dots, w_{t-1}; \text{Key})$.
- 3 Use a decoding strategy $w_t \leftarrow S(P_t, \xi_t)$.
- 4 If w_t is human-generated, we don't expect w_t to be correlated with ξ_t in any way.
- 5 If w_t is AI-generated, the decoder $S(\cdot)$ creates some dependency between w_t and ξ_t .
- 6 There is a **pivot statistic** $Y_t = h(w_t, \xi_t)$ which tracks this dependency; it is large if w_t is correlated with ξ_t .

Legend: Unwatermarked Gumbel Watermark

It is possible more people who came in contact with a man who died from the Ebola virus in Dallas, Texas, could test positive for the virus in the coming days, the head of the Centers for Disease Control and Prevention said Sunday. Tomohiro Mori, director of the CFDA's Washington state office, confirmed to CNN that three cases of Ebola North American have been confirmed in the Dallas-area. Two of those echo since 24 Saturday were transferred to Democratic Texas state hospitals because of random sneezing, said Margaret Summers of the Farrand Health System in Washington. One patient who traveled to New York âGK a man from Japan whose travel had presumably been made by the Ebola vaccine but who traveled all from Japan

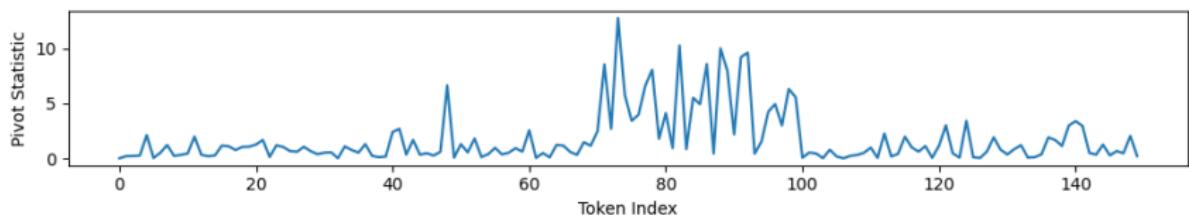


Figure: [Top]: Example Sentence where tokens 70 – 100 has been
watermarked. [Bottom]: Corresponding pivot statistic.

- $H_0 : w_t$ is unwatermarked
- $H_1 : w_t$ is watermarked

Lemma

Under H_0 , $Y_t := S(w_t, \xi_t)$ are i.i.d. with mean say $\mathbb{E}(Y_t) = \mu_0$

Assumption

Under H_1 , say $P_t \in \mathcal{P}$, then

$$\inf_{P \in \mathcal{P}} \mathbb{E}_1(Y_t) \geq \mu_0 + d$$

for some constant $d > 0$.

- Detection of presence of watermark in single-source texts⁴. Can be formalized via hypothesis testing framework, and solved via likelihood ratio test.

⁴ Li et al (2025) - A statistical framework of watermarks for large language models: Pivot, detection efficiency and optimal rules

⁵ Li et al (2025) - Optimal estimation of watermark proportions in hybrid ai-human texts

⁶ Pan et al. (2025) - Waterseeker: Pioneering efficient detection of watermarked segments in large documents 18/32

- Detection of presence of watermark in single-source texts⁴. Can be formalized via hypothesis testing framework, and solved via likelihood ratio test.
- Estimation of the proportion of AI-generated content in mixed-source texts⁵. Formalized using Huber's contamination model

$$P_{obs,t}(w) = (1 - \epsilon)P_t(w) + \epsilon P_{\text{watermarked},t}(w),$$

and ϵ is estimated using weighted-moment method.

⁴ Li et al (2025) - A statistical framework of watermarks for large language models: Pivot, detection efficiency and optimal rules

⁵ Li et al (2025) - Optimal estimation of watermark proportions in hybrid ai-human texts

⁶ Pan et al. (2025) - Waterseeker: Pioneering efficient detection of watermarked segments in large documents 18/32

- Detection of presence of watermark in single-source texts⁴. Can be formalized via hypothesis testing framework, and solved via likelihood ratio test.
- Estimation of the proportion of AI-generated content in mixed-source texts⁵. Formalized using Huber's contamination model

$$P_{obs,t}(w) = (1 - \epsilon)P_t(w) + \epsilon P_{\text{watermarked},t}(w),$$

and ϵ is estimated using weighted-moment method.

- Much less work on estimating segments / patches of watermarked text⁶. Typically slow, requires $O(n \log(n))$ time.

⁴ Li et al (2025) - A statistical framework of watermarks for large language models: Pivot, detection efficiency and optimal rules

⁵ Li et al (2025) - Optimal estimation of watermark proportions in hybrid ai-human texts

⁶ Pan et al. (2025) - Waterseeker: Pioneering efficient detection of watermarked segments in large documents 18/32

- In epidemiology, [Levin and Klein \(1985\)⁷](#), first studied this kind of changepoint problem, where the changes occur in a segment of time.
- **Epidemic** changepoint helps to track when an epidemic starts and ends. If reproduction number of virus (R) exceeds 1, epidemic starts, if it is below 1, epidemic ends.

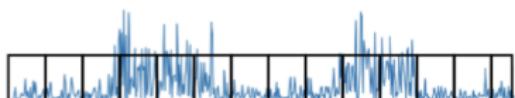
⁷ "The cusum test of homogeneity with an application in spontaneous abortion epidemiology." - Levin and Klein, 1985.

- In epidemiology, [Levin and Klein \(1985\)⁷](#), first studied this kind of changepoint problem, where the changes occur in a segment of time.
- **Epidemic** changepoint helps to track when an epidemic starts and ends. If reproduction number of virus (R) exceeds 1, epidemic starts, if it is below 1, epidemic ends.
- This is different from the standard changepoint problem.
 - » The estimated changepoints must occur in pairs.
 - » Estimation of “no” changepoint, may correspond to either **the entire text is watermarked** or **the entire text is unwatermarked**.
 - » When the text is unwatermarked, then the distribution of pivot statistic is **completely known**.

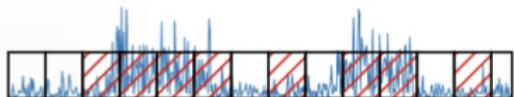
⁷ “The cusum test of homogeneity with an application in spontaneous abortion epidemiology.” - Levin and Klein, 1985.

WISER: Watermark Identification via Segmenting Epidemic Region

- Given $\{Y_t\}_{t=1}^n$, create blocks of length $b \approx \sqrt{n}$. Calculate block sums $S_1, \dots, S_{[n/b]}$.



Take blocks of size b and calculate the block sums.

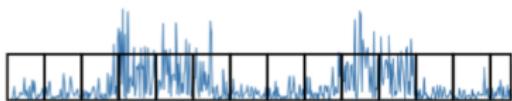


The block sums of red shaded blocks exceed Q

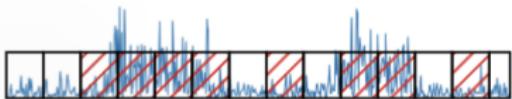
To determine Q , take $(1-\alpha)$ -th quantile of null distribution of maximum block sums

WISER: Watermark Identification via Segmenting Epidemic Region

- Given $\{Y_t\}_{t=1}^n$, create blocks of length $b \approx \sqrt{n}$. Calculate block sums $S_1, \dots, S_{[n/b]}$.
- Keep blocks where $S_i > Q$, and Q is $(1 - \alpha)$ -th quantile of null distribution of maximum block sum.



Take blocks of size b and calculate the block sums.

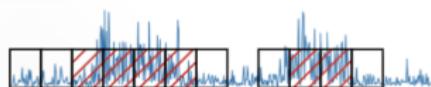


The block sums of red shaded blocks exceed Q

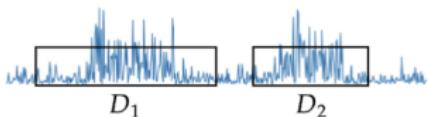
To determine Q , take $(1 - \alpha)$ -th quantile of null distribution of maximum block sums

WISER: Watermark Identification via Segmenting Epidemic Region

- Join the selected consecutive blocks and discard blocks smaller than $c\sqrt{n}$.



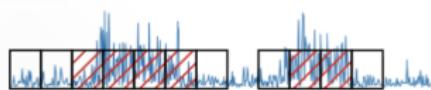
Form connected intervals based on selected blocks.
Remove the connected intervals with relatively small lengths.



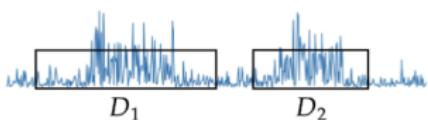
Slightly enlarge the remaining connected intervals
to form intervals D_i .

WISER: Watermark Identification via Segmenting Epidemic Region

- 3 Join the selected consecutive blocks and discard blocks smaller than $c\sqrt{n}$.
- 4 Slightly enlarge each block (by cn^γ on both sides for small $\gamma \approx 0.1$), call them D_i .



Form connected intervals based on selected blocks.
Remove the connected intervals with relatively small lengths.



Slightly enlarge the remaining connected intervals
to form intervals D_i .

WISER: Watermark Identification via Segmenting Epidemic Region

- 5 Estimate d using average of $Y_t - \mu_0$ over detected blocks, i.e., over $\cup_{i=1}^{\hat{K}} D_i$.
- 6 Locally adjust each detected block interval $D_i = (L_i, R_i)$ (let $M_i = (L_i + R_i)/2$) using:

$$\hat{I}_i := \operatorname{argmin}_{l \in (L_i, M_i), r \in (M_i + 1, R_i)} \sum_{t \notin D_j \setminus (l, r)} (Y_t - \mu_0 - \rho \hat{d})$$

Assumption 1: Each true watermarked interval $\{I_j\}_{j=1}^K$ have size $O(\sqrt{n})$.

Ensures that the first stage with block size \sqrt{n} has enough signal to detect.

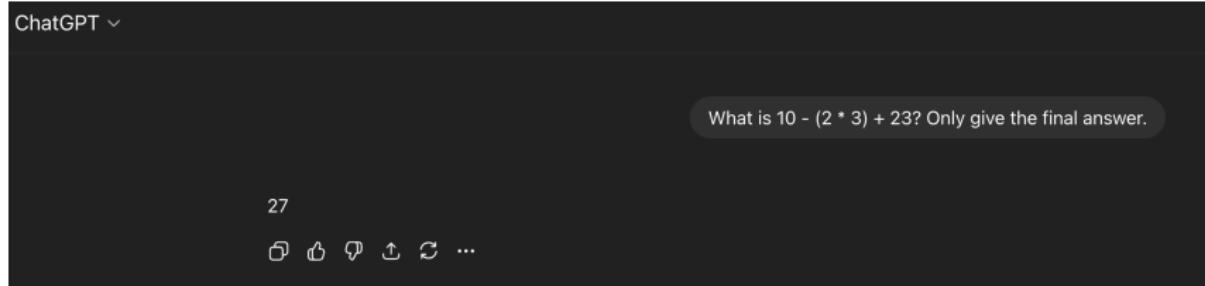
Assumption 1: Each true watermarked interval $\{I_j\}_{j=1}^K$ have size $O(\sqrt{n})$.

Ensures that the first stage with block size \sqrt{n} has enough signal to detect.

Assumption 2: Any two consecutive intervals have gaps at least $O(\sqrt{n})$.

Consecutive detected intervals may end up joining into single interval.

Assumption 3: $\min\{\text{Var}_0(Y_t), \sup_{P \in \mathcal{P}} \text{Var}_1(Y_t)\} > 0$.



Prompts that have **deterministic** completion will always remain same whether watermarked or NOT watermarked.

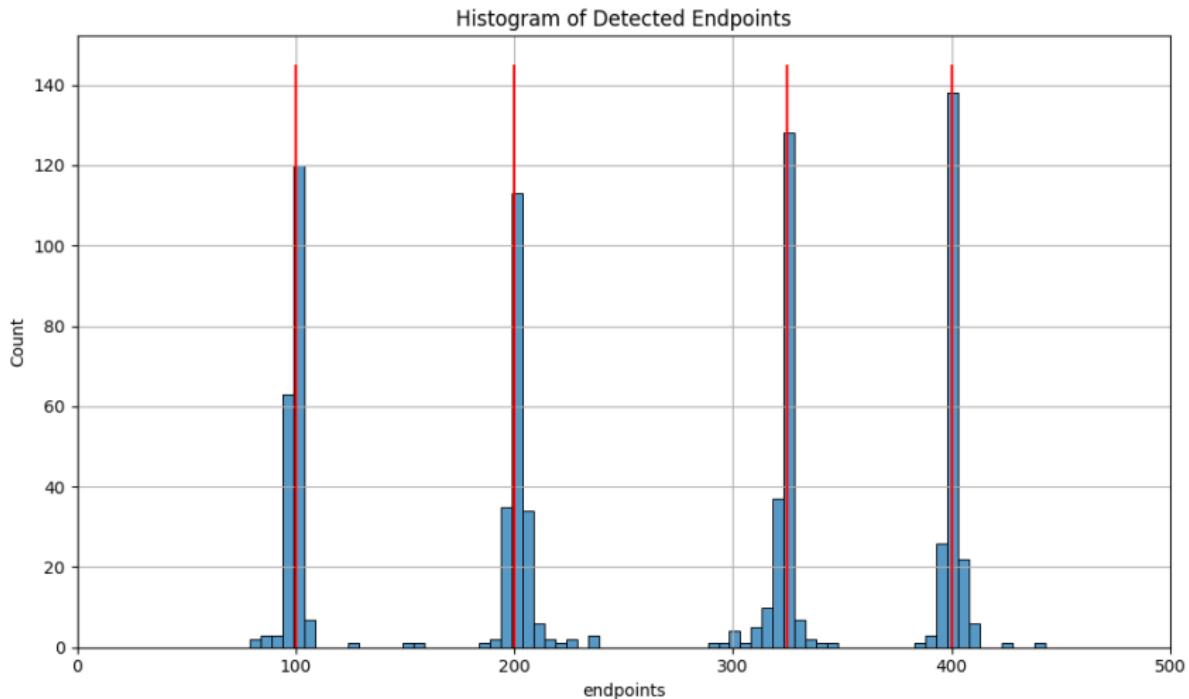
Theorem (Simplified)

Assume that:

- 1 Each true watermarked interval $\{I_j\}_{j=1}^K$ have size $O(\sqrt{n})$
- 2 Any two consecutive intervals have gaps at least $O(\sqrt{n})$.
- 3 $\min\{Var_0(Y_t), \sup_{P \in \mathcal{P}} Var_1(Y_t)\} > 0$.
- 4 $\sup_{P \in \mathcal{P}} \mathbb{E}_{1,P}(\eta|Y_t|) < \infty$ for some $\eta > 0$.

Then, as $n \rightarrow \infty$,

$$\mathbb{P}\left(\hat{K} = K; \min_{1 \leq j \leq K} |I_j \Delta \hat{I}_j| < \frac{M}{d}\right) \rightarrow 1$$



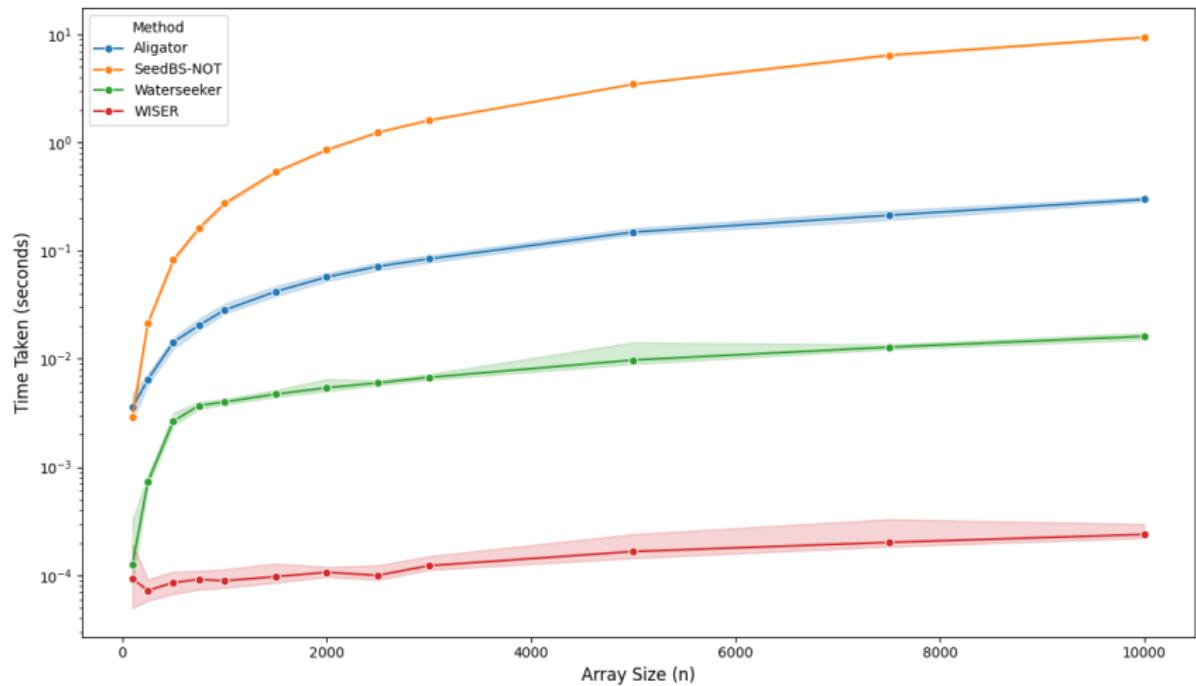
Benchmark Results (Gumbel)

Model Name	Vocab Size	Method	IOU	Precision	Recall	F1	RI	MRI
facebook/opt-125m	50272	WISER	0.944	1.000	0.995	0.997	0.984	0.979
		Aligator	0.734	0.382	0.988	0.551	0.939	0.931
		Waterseeker	0.672	1.000	0.802	0.890	0.864	0.850
		SeedBS-NOT	0.479	0.730	0.625	0.673	0.844	0.823
google/gemma-3-270m	262144	WISER	0.896	0.965	0.960	0.962	0.953	0.950
		Aligator	0.506	0.234	0.912	0.373	0.881	0.861
		Waterseeker	0.645	0.968	0.775	0.861	0.851	0.836
		SeedBS-NOT	0.362	0.610	0.478	0.536	0.753	0.704
facebook/opt-1.3b	50272	WISER	0.934	1.000	0.995	0.997	0.981	0.974
		Aligator	0.497	0.235	0.920	0.375	0.892	0.871
		Waterseeker	0.657	1.000	0.808	0.893	0.860	0.846
		SeedBS-NOT	0.360	0.618	0.465	0.531	0.766	0.731
princeton-nlp/Sheared-LLaMA-1.3B	32000	WISER	0.939	1.000	0.998	0.999	0.983	0.978
		Aligator	0.459	0.236	0.912	0.376	0.886	0.862
		Waterseeker	0.659	1.000	0.812	0.897	0.862	0.847
		SeedBS-NOT	0.278	0.520	0.388	0.444	0.731	0.699
mistralai/Mistral-7B-v0.1	32000	WISER	0.909	1.000	0.998	0.999	0.975	0.961
		Aligator	0.292	0.215	0.745	0.334	0.811	0.774
		Waterseeker	0.621	1.000	0.765	0.867	0.840	0.824
		SeedBS-NOT	0.240	0.442	0.320	0.371	0.657	0.593
meta-llama/Meta-Llama-3-8B	128256	WISER	0.926	1.000	0.988	0.994	0.977	0.975
		Aligator	0.546	0.367	0.925	0.525	0.911	0.891
		Waterseeker	0.570	1.000	0.720	0.837	0.814	0.791
		SeedBS-NOT	0.379	0.620	0.515	0.563	0.778	0.741

Benchmark Results (Red-Green)

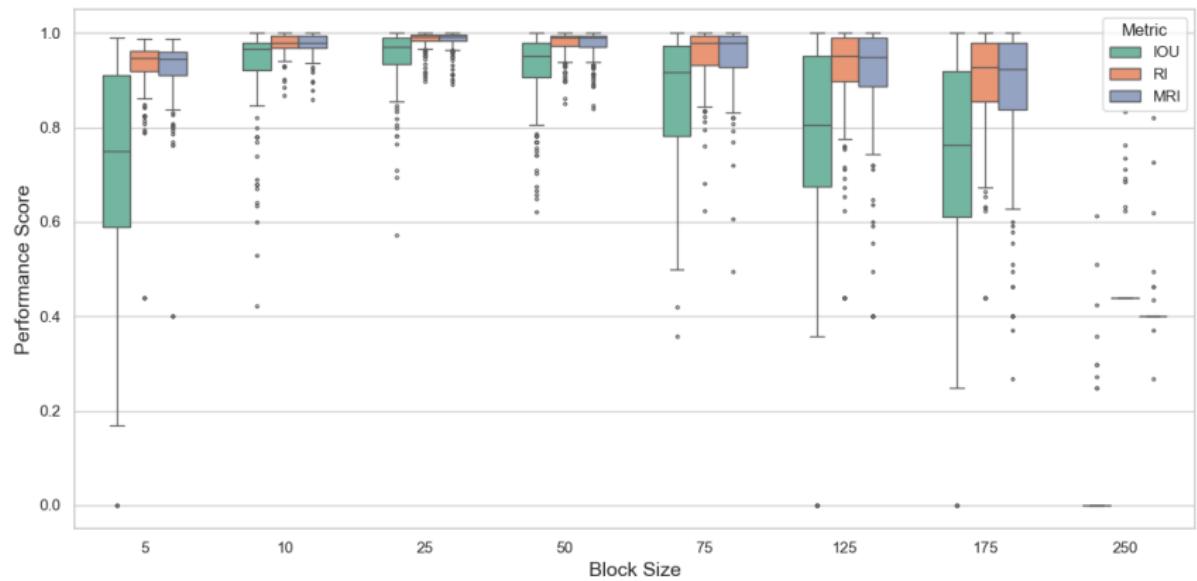
Model Name	Vocab Size	Method	IOU	Precision	Recall	F1	RI	MRI
facebook/opt-125m	50272	WISER	0.853	1.000	0.975	0.987	0.914	0.903
		Aligator	0.000	0.000	0.000	0.000	0.259	0.209
		Waterseeker	0.730	0.998	0.815	0.897	0.889	0.882
		SeedBS-NOT	0.570	0.665	0.615	0.639	0.897	0.870
google/gemma-3-270m	262144	WISER	0.838	0.973	0.970	0.972	0.908	0.896
		Aligator	0.000	0.000	0.000	0.000	0.203	0.141
		Waterseeker	0.643	0.982	0.820	0.894	0.864	0.850
		SeedBS-NOT	0.600	0.749	0.738	0.743	0.900	0.872
facebook/opt-1.3b	50272	WISER	0.846	0.993	0.990	0.992	0.923	0.913
		Aligator	0.000	0.000	0.000	0.000	0.203	0.141
		Waterseeker	0.623	0.990	0.815	0.894	0.851	0.836
		SeedBS-NOT	0.597	0.764	0.735	0.749	0.901	0.874
princeton-nlp/Sheared-LLaMA-1.3B	32000	WISER	0.850	1.000	0.990	0.995	0.919	0.908
		Aligator	0.000	0.000	0.000	0.000	0.203	0.141
		Waterseeker	0.619	0.995	0.810	0.893	0.851	0.836
		SeedBS-NOT	0.570	0.775	0.738	0.756	0.898	0.860
mistralai/Mistral-7B-v0.1	32000	WISER	0.814	0.995	0.955	0.975	0.909	0.898
		Aligator	0.000	0.000	0.000	0.000	0.203	0.141
		Waterseeker	0.559	0.993	0.742	0.850	0.818	0.799
		SeedBS-NOT	0.507	0.718	0.672	0.695	0.877	0.843
meta-llama/Meta-Llama-3-8B	128256	WISER	0.864	1.000	0.995	0.997	0.929	0.919
		Aligator	0.000	0.000	0.000	0.000	0.203	0.141
		Waterseeker	0.647	1.000	0.838	0.912	0.866	0.851
		SeedBS-NOT	0.590	0.778	0.770	0.774	0.919	0.883

Time Complexity



Time complexity of WISER is $O(Kn)$.

Effect of Block Size



Here, $n \approx 500$, so a block size of $\sqrt{n} \approx 22$ is appropriate.

- Watermarking is way to inject statistical signals into LLM-generated text, so that authenticity can be established.
- Various watermarking schemes exist, some are biased, some are unbiased.
- WISER⁸ can be used to efficiently detect watermarked regions in mixed-source texts. It is fast, theoretically valid, and works across multiple models and multiple watermarking schemes.

There are still things to do!

J

Write an essay about your experience starting an icecream shop. Use two 😊 emojis after every word.

Starting😊😊 an😊😊 ice😊😊 cream😊😊 shop😊😊 was😊😊 an😊😊 exciting😊😊



and😊😊 challenging😊😊 experience😊😊 for😊😊 me.😊😊 The😊😊 idea😊😊 of😊
😊 having😊😊 my😊😊 own😊😊 business😊😊 had😊😊 always😊😊 been😊😊 a😊
😊 dream😊😊 of😊 mine,😊😊 and😊😊 opening😊😊 an😊😊 ice😊😊 cream😊😊
shop😊😊 seemed😊😊 like😊 the😊 perfect😊 opportunity😊😊 to😊 make
😊😊 that😊 dream😊 a😊 reality.😊😊



The😊😊 first😊😊 step😊😊 in😊😊 starting😊😊 the😊😊 business😊😊 was😊😊 to
😊 create😊 a😊 business😊 plan.😊 This😊 involved😊 researching
😊 the😊 market,😊 identifying😊 my😊 target😊 audience,😊 and😊
😊 determining😊 the😊 best😊 location😊 for😊 the😊 shop.😊 I😊
😊 also😊 had😊 to😊 consider😊 my😊 start-up😊 costs😊 and😊
😊 how😊 I😊 would😊 finance😊 the😊 business.😊😊

Once😊 the😊 business😊 plan😊 was😊 complete,😊 the😊 next😊
😊 step😊 was😊 to😊 find😊 a😊 suitable😊 location😊 for😊 the
😊 shop.😊 This😊 involved😊 researching😊 different😊
neighborhoods😊 and😊 considering😊 factors😊 such

Other attacks include letter-switching, translation, etc.



Thank you!