iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: IV Month of publication: April 2026

DOI: https://doi.org/10.22214/ijraset.2026.81640

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

Secret Data HidinginLLM Text Generation

Manoj Suklaboina[1], Chaitanya Konda[2], Lokesh Kamepalli[3], Satish Donka[4], Sandeep Pothunuri[5]
[1, 3, 4, 5]*Department of Cyber Security,* [2]*Head of the Data Science & Cyber Security department, ACHARYANAGARJUNA UNIVERSITY GUNTUR – 522508, INDIA*

ABSTRACT: In this work, we explore two different methods for hiding secret information inside text generated by Large LanguageModels (LLMs).Thegoal is to makethehidden content blend naturally into the generated text so that it is difficult for readers to notice. Unlike many existing approaches that rely on predefined cover text for embedding and extraction, our methods follow a coverless approach.This means the text is generated directly from the secret message without needing any pre-written content. In the first method,thesecretmessageisembeddeddirectlyintothegeneratedtext.Beforeembedding,themessageis compresse dusing Huffmancoding toimproveefficiency.Thelanguagemodelthenproduces textbasedon this compressed data. In the second method, each character of the secret message is represented by a word thatstartswiththesameletter,creatingasimpleandeasy-to-understandencodingprocess.Weevaluatethe first method using metrics such as Bits per Word (BpW), Bits per Character (BpC), and Log-Perplexity (PPL).ThesecondmethodisevaluatedusingCharactersperWord(CpW)andPerplexity.Theresultsshow that both methods strike a good balance between hiding capacity and natural-looking text, making them suitable for secure communication.

KEYWORDS: DataHiding,TextSteganography,LLM(LargeLanguagemodels-llama3,mistral),Huffman coding.

I. INTRODUCTION

Withtherapidgrowthofonlinecontent,especially across platforms like social media and messaging systems, new opportunities have emerged for hiding information in different formats such as text,images,andvideos.Amongthesemedia,text has emerged as a particularly significant carrier due to its ubiquity in digital communication, therebydrivingsubstantialinterestintext-oriented steganographic methodologies.

Text steganography entails the unobtrusive integration of confidential information within naturallanguagecontent insuchamannerthatits existence remains imperceptible to unintended recipients.Existingapproachesinthisdomaincan be broadly classified into three principal categories:textmodification-based,textselection- based, and text generation-based techniques. In modification-based schemes, secrecy is achieved through deliberate alterations in textual structure or presentation, including lexical substitution, syntactic transformation, or formatting adjustments; however, such approaches are inherently susceptible to detection through sophisticatedsteganalysismechanisms.Similarly, selection-based and coverless linguistic techniques often suffer from constrained embedding capacity, limiting their practical applicability. In contrast, text generation-based steganography, commonly referred to as generative linguistic steganography, circumvents the need for predefined cover texts by employing Natural Language Processing methodologies to synthesizesteganographiccontentdirectlyfrom thesecretmessage.Thisparadigmenablesflexible andpotentiallyunboundedembeddingcapacity,as the generated text can be extended to accommodate larger payloads, although it may introduce challenges related to maintaining semantic coherence and linguistic fluency in extended outputs. Recognizing these limitations, thepresentstudyaimstoaddressthefundamental challenges associated with existing data hiding techniques, particularly in terms of imperceptibility and payload efficiency, by shifting toward automated, model-driven text generation frameworks. Specifically, this work investigates the integration of Large Language Model-based foundation architectures, such as GPT, in conjunction with Huffman coding, to facilitate efficient and high-quality steganographic text generation.

Unlike conventional word-level embedding strategies,theproposedframeworkoperatesatthe character level, enabling more granular and capacity-efficient concealment of secret bitstreams. The contributions of this research are multifaceted:firstly,itidentifiesandmitigateskey deficienciesintraditionalsteganographicsystems by enhancing both embedding capacity and resistance to detection; secondly, it introduces a hybrid approach that combines deep learning-based text generation with entropy-efficient compression techniques to produce more natural and semantically coherent covertext; and thirdly, it establishes a novel perspective on coverless steganographybyeliminatingdependenceonpre-existing textual carriers and leveragingAI-driven generation mechanisms.

Inthe first proposedapproach,secret information isembeddedatthecharacterlevelbymappingbits to specific letters based on their probabilistic occurrence within the language, thereby improving embedding efficiency while minimizing concealment errors.

Additionally, largelanguagemodelsareemployedtoensurethat the generated text maintains a high degree of linguistic naturalness.
In the second approach, a cipher-based mechanism is utilized wherein each character of thesecretmessagegovernsthegenerationof words through their initial letters, enabling implicit encoding without explicit bit representation.
Collectively,theseapproachessignifyaparadigm shift toward intelligent, adaptive, and high-capacity steganographic systems. The rest of this paper contains: Section 2 presents a comprehensive survey of related work, Section 3 details the proposed methodologies, Section 4 provides a comparative analysis with existing state-of-the-art techniques, Section 5 discusses experimentalevaluationandperformancemetrics, and Section 6 concludes the study with key findings and future research directions.

II. LITERATURE REVIEW

Chang and Clark [4]emphasize the exponential surgeofdatageneratedviasocialmedia,drivenby pervasive internet access and smartphone usage. Thisenvironmentenablesdiscreettransmissionof sensitive information, making such platforms suitableforcovertcommunication.Thesuccessof steganographic systems depends on their resistance to detection, as exposure nullifies their purpose. With the dominance of text across emails, forums, and messaging applications, textual media has emerged as a viable and inconspicuous carrier for hidden communication.
Xiang et al.[5] focused on early rule-based approaches and later improved them using probabilistic models. Their work showed that while traditional methods were limited, machine learning techniques could significantly improve text quality.
Wu et al.[6] investigated both deep learning and Markov chain-based text generation methods for steganography. Their approach utilized statistical language properties to generate fluent text while embedding hidden data. Although deep models improved textual consistency, they introduced higher computational overhead. By refining transition probabilities and limiting state expansions, the proposed Markov-based strategy enhanced efficiency and overall embedding performance.
Grosvald and Orgun [7] highlighted the vulnerabilities of methods that rely on existing cover texts, noting that alterations can expose hiddencontentthroughdetectableinconsistencies. To mitigate this, they proposed a human-assisted approach in which encoded word lists are transformed into natural text by human authors, thereby improving readability and reducing detectability, albeit at the cost of automation.
Yang et al.[8] proposed the RNN-Stega framework,whichusesrecurrentneuralnetworks to learn how language is structured and then generate steganographic text that follows similar statistical patterns. This approach helps produce text that appears locally coherent and natural. However, the model struggles to maintain consistency across longer passages, as it does not fully capture global context, which may make it more vulnerable to advanced detection techniques.
Din et al.[9] discussed statistical machine translation-based systems that rely on corpus-driven templates for text generation. Alongside this, coverless steganography approaches have emerged, avoiding direct modification of carrier text to enhance security. Techniques such as word2vec-based encoding exploit semantic relationships between words to improve embeddingcapacity,thoughchallengesremainin preserving contextual integrity.
Li et al.[10] proposed a knowledge graph-driven steganographic model that transforms structured data into coherent text using transformer architectures. By encoding relationships and entities into natural language, the approach achieves semantically meaningful embedding, though it demands extensive training resources and complex modeling.
Fang et al.[11] developed an LSTM-based steganographic framework that integrates hidden bitstreams into generated text, particularly for socialmediadatasets.Whilethemodeleffectively captures sequential dependencies, it faces challenges in balancing embedding capacity with linguistic fluency.
Farouk and Kabodian [12] presented a method based on AI-generated Arabic poetry, where hidden data is embedded through controlled manipulation ofcharacter-level patterns.Although this approach increases payload capacity, it introduces a trade-off by slightly degrading the naturalness of the generated text.
Ziegler et al.[20] explored transformer-based steganography using GPT models, where vocabulary is partitioned into groups mapped to binary representations. This enables systematic embedding during text generation, producing readable and coherent outputs while maintaining concealment.
Artificial intelligence, encompassing domains such as natural language processing, neural networks,andmachinelearning,hassignificantly advanced text processing capabilities. Its integration into security applications, including steganography, has enabled more adaptive and intelligent data hiding mechanisms.Text steganography has evolved from ancient concealmentmethods,suchashiddeninscriptions and acrostics, to modern digital techniques like whitespace encoding and metadata manipulation. With the advent of large language models, contemporaryapproachesnowembedinformation by subtly altering linguistic patterns within generatedtext.

AsillustratedinFigure1(page4), this evolution reflects a transition from manual techniques to AI-driven methodologies, highlighting the growing role of intelligent systems in secure communication.

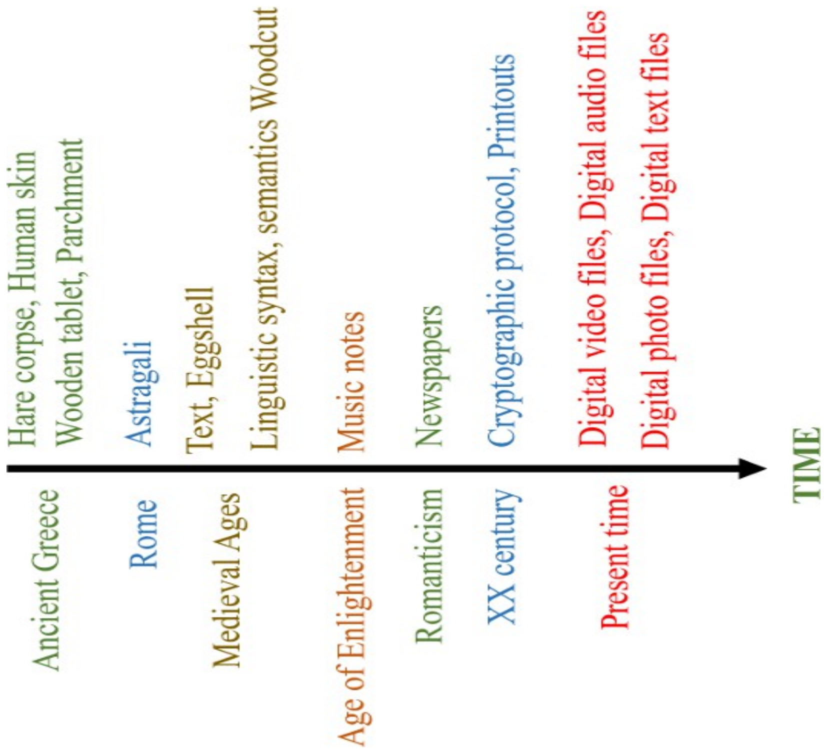


Figure1: EvolutionofDataHidingcarrierovertime.

Seyed Jahanshah Kabudian [23], StegGPT framework employs large language models with Huffman coding to achieve character-level text steganography, significantly enhancing embedding capacity while maintaining natural language quality . Despite these advancements, existing methods—including statistical, RNN-based, and transformer-driven approaches—struggle to balance payload capacity, imperceptibility,andresistancetodetection.Many rely on rigid encoding schemes or single embedding strategies, which introduce detectable patterns or limit flexibility. To address these limitations, this work proposes two coverless LLM-based approaches that improve adaptability andsecurity.Thefirstembedscompresseddataat thecharacterlevelforhigherefficiency,whilethe secondusesacipher-drivenword-initialencoding mechanism,collectively achievingbetter capacity, reduced detectability, and improved linguistic coherence.

III. PROPOSED METHODOLOGY

Text steganography, although one of the earliest forms of information hiding, has often been overshadowed by image and audio-based techniques due to the comparatively limited storage capacity offered by textual data. The restricted space available within textual carriers has consistently posed a significantchallenge for embedding large amounts of secret information. Earlierapproachesprimarilyfocusedoninserting hidden data within sentences, paragraphs, or partial word structures, aiming to minimize storage requirements rather than maximize embedding efficiency. In contrast, the present study adopts a distinct perspective by utilizing individual characters as fundamental units for encoding hidden information.

Inourapproach,individualcharactersareusedas themaincarriersofhiddeninformationinsteadof full words or sentences. The selection of these charactersisguidedbytheirstatisticaloccurrence within a given language, enabling the identification of suitable candidates for informationhiding.Charactersthatappearwith moderate frequency are strategically chosen to maintain a balance between effective data embedding and preservation of textual naturalness.

A key emphasis of this work lies in refining the embedding mechanism while ensuring that the generated text retains its coherence and readability. The overarching goal is to integrate concealedinformationin suchaseamlessmanner that it remains imperceptible to unintended observers. Achieving this objective requires a thoroughunderstandingoflinguisticdistributions andstructuralpatterns,allowingthehiddendatato blendnaturallyintothetextwithoutdisruptingits authenticity, as illustrated in Figure 2.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

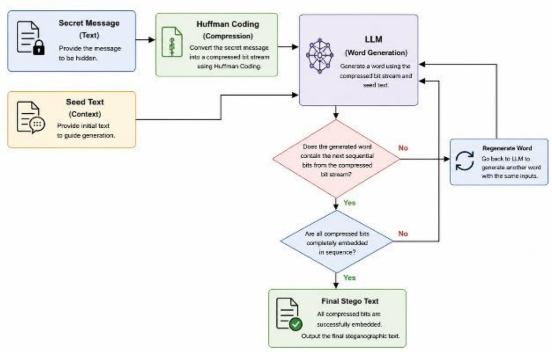


Figure2**:** Blockdiagramoffirstproposedmethodology

1) METHOD1:

a) BitCharacterLevelEmbedding:

Ther are 26 characters in English (A-Z). We collected the frequencies of each character from different paragraphs then we obtained 'E'has the most occurrence followed by 'T', 'A', 'O',….Exceptthefirst4 characterstheremaining aredividedintotwogroupswhichcanholdsecret bits in LLM generated text.

Group0andGroup1consistsofEnglish characters

Group0:{I,S,H,L,C,F,W,P,V,X,J}

Group1:{N,R,D,U,M,Y,G,B,K,Q,Z}

The rest of the English letters are considered as non group letters because those have high frequencies which tends to reduce embedding flexibilityandmaybiased.InFigure3,youwill see the predefined Huffmancodes for compression:

Character	Binary Code
Space	1010
E	1100
T	1011
A	1001
O	0111
I	0110
N	0100
S	0011
R	0010
H	0001
D	11111
L	11110
U	11100
C	11011
M	11010
F	10000
Y	01011
W	01010
G	00001
P	111011
B	111010
V	100010
K	000001
X	000000
Q	1000110
J	10001111
Z	10001110

Figure3:HuffmancodeforEnglishletters.

Note: If we use dynamic Huffman compression then we have to transmit the Huffman codes for the intended recipient. Here, we are using predefined Huffman codes so thatintendedrecipientcanextractthemessageeasilybyjust transmitting Stego text.

Embedding:

- Initially, the secret message is converted into a binary sequence using predefined Huffman coding.
- The model takes two inputs: the seed text (which sets the context) and the compressedbitsequencederivedfromthe secret message.Two character groups, namely Group 0 and Group 1, are defined for encoding. The LLM generates a candidateword,andifallcharacterswithin that word can successfully encode the required bits based on these groups, the word is accepted; otherwise, the LLM regenerates a new word.
- This iterative process continues until all bits from the compressed secret message are completely embedded.
- The final generated text, containing the embedded information, is then ready for transmission.

Extraction:

- The stego text (generated text) is taken as input for the extraction process.
- Eachcharacterisexaminedsequentially:if acharacterbelongs to Group0,abit‘0’is appended to the output; if it belongs to Group 1, a bit ‘1’is appended; otherwise, the character is ignored.
- This procedure is repeated until all characters in the stego text have been processed.
- The resulting sequence of bits is then decoded using Huffman decoding to reconstruct the original characters.
- The recovered text represents the original secret message.

NOTE:Thecharacters that areignored arenot in group 0 or group 1.

Belowistheexampleoffirstproposed methodology:

INPUT:

Secret:agweprsiieyamsrloaryaodcarerltxhbgaqottilursondr g irsimhvphgdiwswhmmrdooiiavemk g hstrgaeixscru

Lengthofsecret:100characters Compressed secret bits:

1001000010101010101100111011001000110110
0110110001011100111010001100101111001111
0010010010111001011111111110111001001011
0000101111010101011000000000111101000001
10011000110011110111011101001101111011100
0010001101110100111110010101000001101001
1000100011011011010000110001011101100010
0001111110110101001010001101010000111010
1101000101111101110111011001101001100010
1100110100000011010000011010000100111011
0010000011001110001100000000011110110010
11100

Total bits: 446 Compressionpercentage:

(800-446)/800=44.25%

OUTPUT:

Stegotextorgenerated text:

real success enhance upon cutting seize tuning chances honed strategies merely us optimized refinement process unfolds edge By fine norms striveRefiningrealizingbutourtruetuningrefine seizing excellence about and way world intricate This evolving grow new achievement flourish growing remain tuning akin changing remain growthitsnovelcommitmentglobaljourneyreach resonate thrive With remain change evolve reaching towards remains have fabric continuous must essential woven grand bloom global reality Striving world also expand only threads novel forward tuned very deeper ongoing tailored fresh enriched interwoven relentless lies remains challenges resilience every rewards thrives pinnacle our trials growth Each has weaving ongoing evolution quest As

BPW:3.878

BPC: 0.641

Log PPL: 4.754

Time:767.31sec

2) METHOD2:

SingleCharacter-BasedGenerationand Retrieval:

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

Inthisapproach,insteadofembeddingbinarybits, the system guides the LLM to generate words based on predefined characters. Each word is generatedsuchthatitsinitiallettercorrespondsto a character from the secret message.This method is particularly suitable for shorter messages, as longer sequences may reduce generation reliability.Seefigure4,whichdepictstheprocess of this approach.

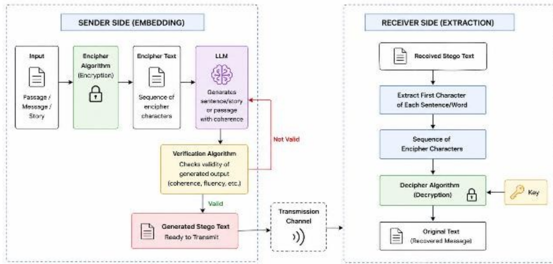


Figure4:processofsinglecharactergenerationand retrieval.

Embedding:

- The original secret message is first encryptedtoensureconfidentiality.Inthis implementation, a Caesar cipher is employed; however, more advanced cryptographic techniques can be used for stronger security.
- The resulting cipher text is processed character by character in sequence.
- For each character, the LLM generates a word that begins with the corresponding character.
- This process continues iteratively until all characters in the cipher text have been encoded into generated words.

Extraction:

- Thegeneratedstegotextistakenasinput.
- The first character of each word in the stego text is extracted sequentially to reconstruct the cipher text.
- The reconstructed cipher text is then decrypted using the corresponding decryption method to recover the original secret message.

Belowistheexampleforsecondapproach:
INPUT:
Secret: tragedies Cipher:usbhfejftOUTPUT:
Generatedtext:

Sunlightisuplifting,serene,beaming,harmonious,fulfilling,enlightening,jubilant,fascinating,transcendent.
CPW:1.0
PPL:5.5

IV. PERFORMANCE METRICS AND RESULTS

1) METHOD1:

1-BitCharacterLevel Embedding:
Here,BPW(BitperWord),BPC(BitperCharacter)arecalculatedtochecktheembeddingefficiency.Those are calculated as
BpW = Number of hidden bits / Number of words in Stego text BpC=Numberofhiddenbits/NumberofcharactersinStegotext
Forevaluatingimperceptibilityweareusinglogppl(logarithmicperplexibility)aswearehidingsecret using group 0 or group 1 which are always referred by bits either 0 or 1.

$$\log PPL(X) = -\frac{1}{N}\sum_{i=1}^{N} \log p\theta(xi|x<i)$$

whereX=x1,x2,...,xNis thegeneratedsequenceinstegotext,andNisthesequencelength,logPθ(xi|x<i) isthelog-probabilityofthetokenxiconditionedonpreceding tokensx<iaccordingtoourmodel.youwill be able to see the sample secrets, number of bits after Huffman compression and the stego text generated by LLM(llama3 or Mistral) and their metrics are shown in the table 1.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

Table1:sampletestcasesfor method1

No.	Secret	No. of bits	Stego	Results
1	zcbdokbjgb	59	depthsdarkestraysbrokenrealmsbeautymystery glow splendor realms delicate darkness hidden night.	BPW:3.471 BPC:0.573 LogPPL:2.89 Time:57.75s
2	uwxcwqfcww	53	robustvaluesadoptaccessisnoreverysystemsentails audited this learning system evaluations.	BPW:3.312 BPC:0.541 LogPPL:2.833 Time:63.07s
3	yspxrohfic	47	computers language goal rules follow simple natural like replicates make learning	BPW:3.133 BPC:0.461 LogPPL:2.773 Time:51.8s
4	acahionhnf	42	MSsymbolnatureiconicworldNestledMoreover spectacle against waterfalls	BPW:3.0 BPC:0.56 LogPPL:2.708 Time:61.37s
5	awztqcgmhr	51	blackpanthersDespiteorangepanthergenetic contribute tigers solid mutation While black s	BPW:3.4 BPC:0.573 LogPPL:2.773 Time:200.55s
6	dnlxgxcckz	55	grandrealm line beautyaesthetic walls passage still their upon by reverence living silent quest	BPW:2.619 BPC:0.495 LogPPL:3.091 Time:70.93sec
7	zjpercdrdnkmpdrihrfgfczupfsqkr	156	differentbalanceHinduenergyharmonywithin standingnumerousprofoundbeginShivawisdom rhythm reminder underpins intricate cosmic temporal lives life much tapestry through vibrant Lord battling mirrors This like but lesson rich existence just	BPW:4.0 BPC:0.653 LogPPL:3.689 Time:247.02sec
8	oddetwlaibdgbsopeatttmnubscmgsacppct oc waw	235	global environment nations measures scale growth climate country management monetary by limited maintain from arena provide commitment global beyond merely efficient regulate different economy regulatorybodiesemployencouragelongencouraged sustainvibrantfostersencouragesdevelopmentuniqu e engine while serves markets thereby prudent beacon bordersmanagedeconomyextendsmarketmaking beaconsvariousdiverseonlyas	BPW:4.052 BPC:0.603 LogPPL:4.078 Time:383.78sec

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

9	cnuemkahwgchyqorhtimesxnhgqhhsxanu sntjerhnsrdd o euoab ah wtd eh tahrtwkipe	358	gentle known changed finer their now gave this like daily glade would within over crisp find being thatchedsoulmemoriessymphonyancientpassingsun commitment forever person richly roof itself light storied help vanish fabric threads existence high hum trees roofs whispers sentinel through wisdom across absorbroofsfromorangeresiliencepassedthemselves edge secrets remark memory breeze stones awakened roof rain behind releasing drifted ebbed could world witnesswarmthriversenseabsorbednaturegentle eyes damp villagers rhythm quilt roofs crickets temperedsentinelstwilightquietaboutabove tranquil	BPW:3.545 BPC:0.615 LogPPL:4.625 Time:529.42sec
10	agweprsiieyamsrloaryaodcarerltxhbgaqot tilursondr g irsimhvphgdiwswhmmrdooiiavemk g hstrgaeixscru	446	realsuccessenhanceuponcuttingseizetuningchances honed strategies merely us optimized refinement process unfolds edge By fine norms strive Refining realizing but our true tuning refine seizing excellence aboutandwayworldintricateThisevolvinggrownew achievement flourish growing remain tuning akin changingremain growthitsnovelcommitmentglobal journey reach resonate thrive With remain change evolve reaching towards remains have fabric continuous must essential woven grand bloom global reality Striving world also expand only threads novel forward tuned very deeper ongoing tailored fresh enrichedinterwovenrelentlessliesremainschallenge s resilience every rewards thrives pinnacle our trials growthEachhasweavingongoingevolutionquestAs	BPW:3.878 BPC:0.641 LogPPL:4.754 Time:767.31s

Weconsideredsecrettextlengthsof10,20,30,40,50,60,70,80,90,100andcalculatedthemeansof BPC,BPW,Time taken and log PPL.Their values are 0.52,3.03,330 seconds and 3.5 for LLAMA3 model.

2) METHOD2:
SingleCharacter-BasedGenerationandRetrieval:
Here,CPWisusedtoknowtheembeddingefficiencyandppltoknowhow perceptiblethegeneratedtext is. Those are calculated as

$$\mathrm{PPL}(X)=\exp\{-\frac{1}{N}\sum_{i=1}^{N}\log p\theta(xi|x<i)\}$$

Thetable2showsthe5sampletestcasesofsecret length10withthestego/generatedtextandtheir CPW and ppl values.

Table2:sampletestcasesformethod2

No.	Secret	Stego	Results
1	zcbdokbjgb	Zephyr, cloudless, bright,delightful, outstanding,kinetic, breezy, joyous, gracious, beautifully.	CpW:1.0 PPL:6.91
2	uwxcwqfcww	understandingwisdom,xenialcompassion,wonderquietness, fidelity courage, wise wisdom.	CpW:1.0 PPL:7.87
3	yspxrohfic	yokedstrengthpoweringxenialityreverenceostensiblyharnessing fantastic inspiring courage.	CpW:1.0 PPL:6.45
4	acahionhnf	abundant, coveted, astonishing,heavenly, inscrutable, overflowing,nurturing, harvests, nurturing, fantastic.	CpW:1.0 PPL:7.11
5	awztqcgmhr	Alert,Warning;Zephyr,Tremor,Quiver-Consequence; Guarding Momentum; Heighten, Rigor."	CpW:1.0 PPL:8.01

We considered secret text of length 10 only because if the text size increases then model is unable to get stego correctly and its performance degrades. Themeans ofCPW, Time taken and PPL are 1.0, 52 seconds and 7.7 for LLAMA3 model.

COMPARISION:

For 1-Bit Character level Embedding the average resultsareshownbelowforLLAMA3,MISTRAL andcurrentexistedmodelGPT-2shownintable3.

Table3:comparisionbetweenproposedandexistedmodels

		BPW (high)	BPC (high)	log PPL(low)
proposed	Llama3	3.03	0.52	3.5
proposed	Mistral	3.08	0.52	3.6
existed	GPT2	4.27	0.78	26.2

For Single character generation and retrieval the results is shown over 1-Bit Character level generation and Retrieval. As there are different metrics we are evaluating its efficiency based on ppland averagetimetakenforresponseshownin thetable 4.

Table 4: comparsion of method 1 and method 2 for shorter secret messages with ppl and time taken for LLM response.

	LLAMA3		MISTRAL	
	ppl	Time in sec	ppl	Time in sec
METHOD1	3.2 103	103	3.3	82
METHOD2	7.7	52	7.2	60

V. CONCLUSION

The results indicate that the proposed methods improve how naturally the hidden text appears, while still maintaining a reasonable level of data embedding.Inthe1-bitcharacter-levelembedding method, models such as LLaMA3 and Mistral exhibit lower Bits per Word (BpW) and Bits per Character(BpC)comparedtoGPT-2,indicatinga reduction in embedding capacity; however, this trade-off results in a substantial decrease in perplexity, thereby producing more natural and less detectable stegotext.Furthermore, in the single-character generation and retrieval approach, the results indicate that although this method does not support traditional embedding metrics such as BpW and BpC, it provides competitive perplexity scores and reduced response times, particularly for shorter messages.

LLaMA3 and Mistral demonstrate consistent performance, achieving a balance between generation quality and computational efficiency.Overall,thefindingshighlightacritical trade-off between embedding capacity and imperceptibility. While higher-capacity methods tendtointroducedetectableartifacts,theproposed approaches prioritize naturalness and security, making them more suitable for covert communication scenarios. These results validate the effectiveness of leveraging modern LLMs for text steganography and suggest that adaptive strategiescanfurtherenhancebothefficiencyand robustness in future work.

REFERENCES

[1] R.Thabit,N.I.Udzir,S.M.Yasin,A.Asmawi,N. A. Roslan, and R. Din, "A comparative analysis of Arabictextsteganography,"Appl.Sci.,vol.11,no. 15,p.6851,Jul.2021,doi:10.3390/app11156851.

[2] S.Jusoh,A.Mustapha,A.Ismail,andR.Din,''A review of Arabic text steganography: Past and present,'' Indonesian J. Electr. Eng. Comput. Sci., vol.17,no.2,p.1040,Feb.2020,doi:10.11591/ijeecs.v17.i2.pp10401046.

[3] A. M. Alshamsi, S. M. Albaloushi, M. Y. Alkhoori, H. A. Almheiri, and N. Ababneh, ''A reviewofArabictextsteganography,''inProc.4th Int. Conf. Data Storage Data Eng., Feb. 2021, pp. 6–11, doi: 10.1145/3456146.3456148.

[4] C.-Y. Chang and S. Clark, ''Practical linguistic steganography using contextual synonym substitution and a novel vertex coding method,'' Comput. Linguistics, vol. 40, no. 2, pp. 403–448, Jun. 2014, doi: 10.1162/COLI_a_00176.

[5] L. Xiang, S. Yang, Y. Liu, Q. Li, and C. Zhu, ''Novel linguistic steganography based on character-level text generation,'' Mathematics, vol. 8, no. 9, p. 1558, Sep. 2020, doi: 10.3390/math8091558.

[6] N.Wu,P.Shang,J.Fan,Z.Yang,W.Ma,andZ. Liu, ''Research on coverless text steganography basedonsinglebitrules,''J.Phys.,Conf.Ser.,vol. 1237,no.2,Jun.2019,Art.no.022077,doi:10.1088/17426596/1237/2/022077.

[7] M. Grosvald and C. O. Orgun, ''Free from the cover text: A human generated natural language approach to text-based steganography,'' J. Inf. HidingMultimediaSignalProcess.,vol.2,no.2,pp.1–9,2011.

[8] Z.-L.Yang,X.-Q.Guo,Z.-M.Chen,Y.-F.Huang,and Y.-J. Zhang, ''RNN stega: Linguistic steganographybasedonrecurrentneuralnetworks, ''IEEETrans.Inf.ForensicsSecurity,vol.14,no. 5,pp. 1280–1295, May 2019, doi: 10.1109/TIFS.2018.2871746.

[9] R. Din, S. Utama, and A. Mustapha, ''Evaluation review on effectiveness and security performancesoftextsteganographytechnique,'' IndonesianJ.Electr.Eng.Comput.Sci.,vol.11,no. 2,p.747,Aug.2018,doi:10.11591/ijeecs.v11.i2.pp747-754.

[10] Y.Li,J. Zhang,Z.Yang, and R.Zhang, ''Topic-aware neural linguistic steganography based on knowledge graphs,'' ACM/IMS Trans. Data Sci., vol. 2, no. 2, pp. 10:1–10:13, Apr. 2021, doi: 10.1145/3418598.

[11] T. Fang, M. Jaggi, and K. Argyraki, ''GeneratingsteganographictextwithLSTMs,''in Proc.ACL,StudentRes.Workshop,2017,pp.100–106. [Online]. Available: https://aclanthology.org/P17-3017/

[12] O.F.A.AdeebandS.J.Kabudian,''Arabictext steganographybasedondeeplearningmethods,'' IEEEAccess,vol.10,pp.94403–94416,2022,doi: 10.1109/ACCESS.2022.3201019.

[13] S. T. Klein and D. Shapira, ''Huffman coding with non-sorted frequencies,'' in Proc. Data Compress. Conf. (DCC), Snowbird, UT, USA, Mar. 2008, p. 526, doi: 10.1109/DCC.2008.73.

[14] D. Tamir, ''Delta-huffman coding of unbounded integers,'' in Proc. Data Compress. Conf.,Snowbird,UT,USA,Mar.2018,p.428,doi: 10.1109/DCC.2018.00081.

[15] R. Arshad, A. Saleem, and D. Khan, ''PerformancecomparisonofHuffmancodingand doubleHuffmancoding,''inProc.6thInt.Conf. Innov.Comput.Technol.(INTECH),Aug.2016,pp. 361–364,doi:10.1109/INTECH.2016.7845058.

[16] A. Chan, Y.-S. Ong, B. Pung, A. Zhang, and J. Fu, ''CoCon: A self-supervised approach for controlled text generation,'' 2020, arXiv:2006.03535.

[17] J. Li, Z. Li, L. Mou, X. Jiang, M. R. Lyu, and I. King, ''Unsupervised text generation by learning from search,'' in Proc. 34th Conf. Neural Inf. Process. Syst. (NeurIPS), 2020.

[18] M. A. Niculescu, S. Ruseti, and M. Dascalu, ''RoGPT2:RomanianGPT2fortextgeneration,''in Proc.IEEE33rdInt.Conf.ToolsArtif.Intell.(ICTAI), Nov. 2021, pp. 1154–1161, doi: 10.1109/ICTAI52525.2021.00183.

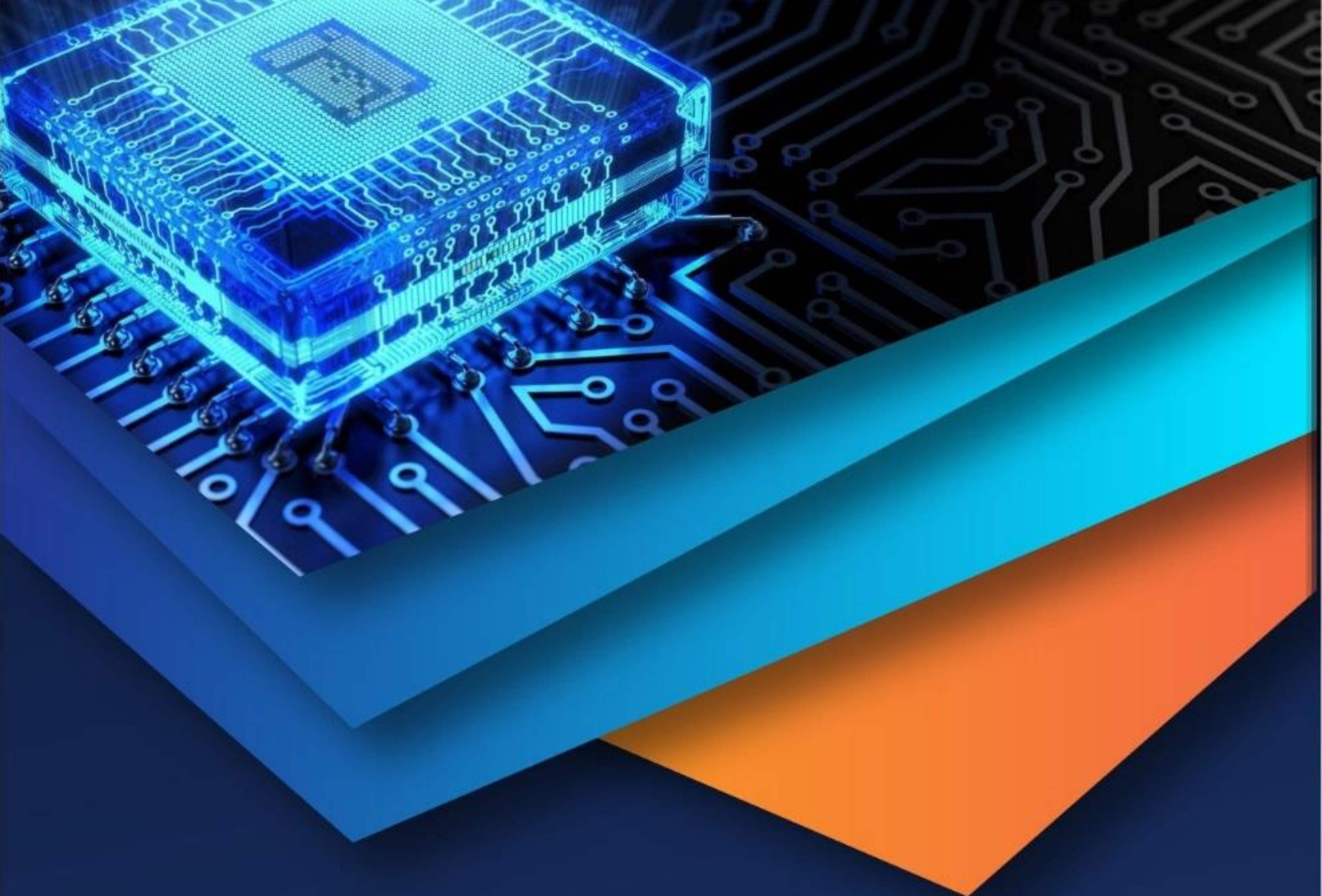
[19] H. H. Moraldo, ''An approach for text steganographybasedonMarkovchains,''inProc. Workshop de Seguridad Informatica (WsegI), 2012.

[20] Z. Ziegler, Y. Deng, and A. Rush, ''Neural linguistic steganography,'' in Proc. Conf. Empirical Methods Natural Lang. Process. 9th Int. Joint Conf. Natural Lang. Process. (EMNLP-IJCNLP), Hong Kong, 2019, pp. 1210–1215.

[21] J.Yang,Z.Yang,S.Zhang,H.Tu,andY.Huang, ''SeSy: Lin guistic steganalysis framework integratingsemanticandsyntacticfeatures,''IEEE SignalProcess.Lett.,vol.29,pp.31–35,2022,doi: 10.1109/LSP.2021.3122901.

[22] F.Daneshfarand S.J.Kabudian,''Speech emotion recognition using multi layer sparse auto-encoder extreme learning machine and spectral/spectro temporal features with new weighting method fordataimbalance,''in Proc. 11th Int. Conf. Comput. Eng. Knowl. (ICCKE), Mashhad, Iran, Oct. 2021, pp. 419–423, doi: 10.1109/ICCKE54056.2021.9721524.

[23] Omer Farooq ahmedadeeb and seyedjahanshahkabudian, "StegGPT: A Novel foundationmodelbasedcharacterlevellinguistic steganographymethodutilizinglargelanguage models"inProc.IEEEACCESS,vol.13,pp.89915-89925, doi : 10.1109/ACCESS.2025.3568339.



10.22214/IJRASET

INDEX COPERNICUS 45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY