

Mixture of Detectors: A Compact View of Machine-Generated Content Detection

Anonymous ACL submission

Abstract

Large Language Models (LLMs) are gearing up to surpass human creativity. The veracity of the statement needs careful consideration. Numerous LLMs entered the market in succession, each better than the last. In light of recent developments in LLMs market, such as reasoning and agent-based architecture models, LLMs have significantly uplifted the opinion of LLMs in human minds. In these developments, critical questions arise regarding the authenticity of human work and the preservation of their creativity and innovative abilities. This paper investigates such issues. This paper addresses machine-generated content across several scenarios, including document-level binary and multiclass classification, sentence-level segmentation to differentiate between human and machine-generated text, and a survey of adversarial attacks aimed at reducing the detectability of machine-generated content. We introduce a new work called **BMAS English**: an English language dataset for Binary classification of human and machine text, for Multiclass-classification, which not only identifies machine-generated text but can also try to determine its generator, and Adversarial attack addressing where it is a common act for the mitigation of detection, and Sentence-level segmentation, for predicting the boundaries between human and machine-generated text. We believe that this paper will address previous work done in machine-generated text detection (**MGTD**) in a more meaningful way. All source codes and datasets can be seen in our GitHub repository¹.

1 Introduction

The emergence of Large Language Models (LLMs), including ChatGPT², Grok³, and Deepseek⁴, rep-

resents a paradigm shift in automated text generation. These advanced models produce highly coherent and human-like textual content, demonstrating remarkable linguistic capabilities. Due to their increasing accessibility and widespread adoption, LLMs have significantly enhanced productivity across diverse domains, such as journalism, social media, education, and academic writing. However, their rapid proliferation also raises critical concerns regarding potential misuse, posing unprecedented ethical and societal challenges.

Seeing the latest evolutions in LLMs are way more than human expectations, and this change started from the work, “*Attention Is All You Need*” by introducing the work of Transformers (Vaswani et al., 2017). As of 2025, this led till Generative Pre-trained Transformers (GPT), Reinforcement Learning from Human Feedback, a Mixture of Experts, Reasoning, Agents, and to Hybrid reasoning, and way more great things to make some of the models open-source. Examples of the above architectures are, from GPT-1 to GPT-4.5, including o-series, Deepseek-v3, R1 (Guo et al., 2025), Sparrow, and Anthropic’s Claude Sonnet 3.5, 3.7, and the model claiming the World’s smartest AI, Grok 3 from Explainable AI (X-AI), and many more. As these models become way more intelligent than humans, and we humans do explicitly use them without any hesitation, such that work authenticity will be lost. It can’t be said that the human or the AI model does a particular work; here, work implies a PowerPoint or an idea proposal.

So, we try to address the problem in three different ways, as said in the Abstract above, which also can be seen in the Figure 1, and constructed three different sets of datasets and trained several set of models for comparison, experimentation, and benchmark purposes. Which are:

1) Binary and Multiclass Classification: As for now, solely focused on the English text data as it is primarily used by most of the countries

¹<https://github.com/saitejalekkala33/E-BMAS-A-mixture-of-AI-Detectors.git>

²<https://chatgpt.com/>

³<https://grok.com/>

⁴<https://www.deepseek.com/>

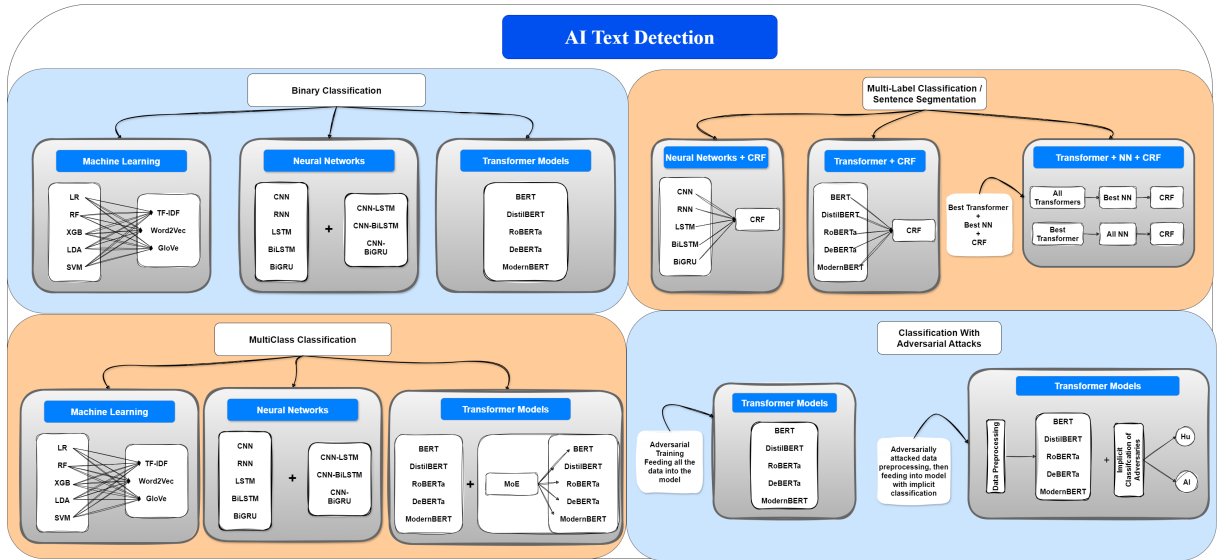


Figure 1: All Experiments for the Classification, Segmentation, and Adversarial Attacks addressing.

worldwide. This English corpus contains human text and machine text by the current well-known LLMs and most prominent domains. And such that this dataset can be used in two ways as, a) a Binary classification task between the human and LLM, and b) a Multiclass classification task where the model will be able to tell whether the text is human-written or LLM-generated, and such that by which LLM the text is generated.

2) Addressing Adversarial Attacks: By most of the detectors like GPT, Quillbot, etc., the given machine text can be given as human text by incorporating some adversaries to the text like synonym substitution, homoglyph attacks, misspelling, character insertion or deletion, and paraphrasing. So, now we introduce a dataset for this kind of problem, taking 5 kinds of adversarial attacks which are used by most of people for the mitigation of detection.

3) Mixed Text Data: The above two tasks address the problem at the document level, and what about the case when the text contains a mixed text, like text with human and machine parts boundary-wise separation. So, we constructed a dataset that contains three versions of mixed text: 1) First Human written and Machine continued, 2) Machine Initiated and human continued, and 3) Fully mixed text. Each version is given by the labels which are Human End Boundary, Machine End Boundary, and word index, where the text is interchanging from human-written to machine-generated and vice versa.

OUR KEY CONTRIBUTIONS:

1. English-BMAS Datasets.
2. Vast Set of Experiments.
3. New HardMoE & SoftMoE Detectors.
4. New Implicit Adversarial Detection.
5. New Sentence Segmentation.

2 Related Work

Binary Detection The detection of Machine-Generated Text (MGT) has predominantly been framed as a binary classification task (Zellers et al., 2019; Gehrmann et al., 2019). Existing methodologies for MGT detection can be broadly classified into supervised and unsupervised approaches. Supervised methods (Wang et al., 2023b; Uchendu et al., 2021) utilize labeled datasets to train discriminative models for classification. In contrast, unsupervised techniques rely on intrinsic textual features such as perplexity, log-rank statistics (Mitchell et al., 2023; Hans et al., 2024), or leverage watermarking schemes (Kirchenbauer et al., 2023; Zhao et al., 2023) to distinguish machine-generated content. In this work, we primarily focus on supervised detection approaches, given their demonstrated effectiveness in leveraging annotated data for improved classification performance. (Wang et al., 2023b) evaluates several supervised detectors, such as RoBERTa (Liu et al., 2019), XLM-R (Conneau et al., 2019), a logistic regression classifier with Word2Vec features (Gehrmann et al., 2019), models leveraging stylistic features (Li et al., 2014), and classifiers using NELA fea-

tures (Horne et al., 2019a). Similar analyses of supervised methods have been conducted in recent work (Guo et al., 2023; Xiong et al., 2024).

Machine Generated Text Detection (MGTD) has been the work of many individuals who have been exploring this since the evolution of the LLMs. We are inspired by the works from papers M4 (Wang et al., 2024a), MAGE (Li et al., 2023), and RAID (Dugan et al., 2024), as they created a wide and large corpus for this MGTD. M4 is Multilingual, MAGE has multidomain, and both M4 and MAGE have several LLMs data, and RAID with several types of Adversarial attacks. We explored these tasks by training from traditional Machine Learning (ML) classifiers to Neural Network Models, and then Pretrained Transformer models, and their other methodologies, with them like the inclusion of Linguistic feature layers to the Transformer output layers, and addressing adversarial attacks, also a methodology from a shared task paper by (Lekkala et al., 2025). For the Sentence Segmentations, the works of SeqXGPT (Wang et al., 2023a), RoFT (Dugan et al., 2020), RoFT-ChatGPT (Kushnareva et al., 2023) and (Zeng et al., 2024). All the major works mentioned above has gone through several phases, like Multiclass for both generator and domain, and for multilingual, and adding as many attacks as possible. We are bound to make the dataset, by addressing all the possible ways in the MGTD.

3 Data set descriptions

3.1 BMAS Dataset

We wanted to focus on the multi-domain texts, so we have included the human texts and AI texts from the 5 main domains that the humans are using in the current times, which are Reddit posts, News articles, Wikipedia text, Arxiv abstracts, and some basic Q&A. We have collected human texts from the major datasets and works that were previously done, which are MAGE and M4, for all the domains except for News Articles. For news articles, we have taken the XSUM dataset as it has a vast set of human-written articles. We define our dataset as simple, concise, robust, and easy to use. All the dataset statistics are given in the Tables 1, 2, 3. A complete description of how the data is being generated is described in the Appendix A. The dataset can be downloaded from the Google Drive ⁵.

⁵https://drive.google.com/drive/folders/1tWqFkJJHfs2uFGU301_BoHo3QebKVbYx?usp=sharing

We want our dataset to be concise and wanted to use by the students who are just entering the NLP domain and into research. As creating a vast set of data might require the computational power to run their codes. And if the dataset is large, they may wanted to stop or exit from this problem statement and look for other simple and short ones for the computation. As we experienced the effect of not having better resources at the start.

4 Detectors

We are bound to do a wide variety of experiments from the very start of the classification algorithms to the current trends. Early classification methods use Traditional Machine Learning models, and the current trends use Transformer models. So, we conducted experiments for binary classification, multi-class classification in both non-adversarial and adversarial attacks, and for the boundary detection task we conducted experiments using CRF models, like adding CRF layers at last of the models that have the iterative forms of various transformers and between neural networks. Besides these experiments, we introduce a novel method for classification and boundary prediction, the latest well-established architecture that consumes less computation and shaked all the Nvidia stocks, Deepseek-v3 671b parameter model Architecture Mixture-of-Experts (MoE) for our experiments.

4.1 Experimental Setup

We conducted all our experiments on Amazon Web Services (AWS) Cloud server, Amazon Elastic Compute Cloud (EC2) instance. In the EC2 instance, we initiated an instance for Accelerated Computing. The specifications are **g6e.xlarge** instance, which provides **3rd generation AMD EPYC processors (AMD EPYC 7R13)**, with a **NVIDIA L40S Tensor Core GPU with 48 GB GPU memory**, and 4x vCPU with 32 GiB memory and a network bandwidth of 20GBps, and our OS type is **Ubuntu Server 24.04 LTS (HVM), EBS General Purpose (SSD) Volume Type**.

4.2 Data Preprocessing

When we gone through our dataset, we came across a few things like multiple new lines in both human texts and AI texts. More specifically, in human texts are symbols like, exclamatory marks, question marks more than usual. So, we have cleaned the text with the most basic steps. These preprocessing steps include, 1) converted text to lowercase, 2)

LLM ↓ Domain →	Reddit	News	Wikipedia	Arxiv	Q&A
Human	10,000 (* (5+1))	10,000 (* (5+1))	10,000 (* (5+1))	10,000 (* (5+1))	10,000 (* (5+1))
Deepseek	2,000 (* (5+1))	2,000 (* (5+1))	2,000 (* (5+1))	2,000 (* (5+1))	2,000 (* (5+1))
OpenAI	2,000 (* (5+1))	2,000 (* (5+1))	2,000 (* (5+1))	2,000 (* (5+1))	2,000 (* (5+1))
Anthropic	2,000 (* (5+1))	2,000 (* (5+1))	2,000 (* (5+1))	2,000 (* (5+1))	2,000 (* (5+1))
Llama	2,000 (* (5+1))	2,000 (* (5+1))	2,000 (* (5+1))	2,000 (* (5+1))	2,000 (* (5+1))

Table 1: Human and LLM Generation Corpus on Domain Wise. The 2,000 and 10,000 represent the original non-adversarial texts, and the numbers in brackets represent 5 adversarially attacked texts and 1 original non-adversarial text. Non-Adversarial Data Size = 80,000 and Adversarial Data Size = 480,000.

	XSUM	MAGE					M4	
LLM ↓ Source →	XSUM	SciGen	ELI5	YELP	SQuAD	CMV	Wikipedia	Reddit
Deepseek	2,000	2,000	2,000	1,046	954	466	1,046	488
OpenAI	2,000	2,000	2,000	1,046	988	492	1,012	496
Anthropic	2,000	2,000	2,000	1,004	996	514	1,004	482
Llama	2,000	2,000	2,000	980	1,020	520	980	500

Table 2: Human and LLM Generation Corpus on Source Wise.

Model	Reddit	News	Wikipedia	ArXiv	Q&A
OpenAI(HM)	2k	2k	2k	2k	2k
DeepSeek(HM)	2k	2k	2k	2k	2k
OpenAI(MH)	2k	2k	2k	2k	2k
DeepSeek(MH)	957	1998	-	2k	2k
GPT-4.1(Mix)	986	1k	981	998	971
GPT-4.1(Mix)	987	1k	984	998	970

Table 3: Distribution of generated data across domains and models for the Sentence-Level Task.

removed URLs (both http and https), 3) replaced multiple spaces with a single space and strip leading/trailing spaces, 4) remove non-alphanumeric characters, excluding spaces, commas, and periods, 5) removed occurrences of the characters ".,," followed by commas and optional whitespace, 6) replaced multiple consecutive newlines with a single newline, 7) removed numeric references within square brackets (e.g., [1]).

4.3 Binary and Multi-class Classification

This section will provide the experiments that are conducted for the Binary and Multi-class classification between the human and AI (which AI). Our dataset for these two tasks are a total of 80,000 rows where in them 40,000 are human texts and other 40,000 are the AI-texts and given the better

statistics in the Tables 1 and 2

4.3.1 Machine Learning Classifiers

Based on our previous work, we have chosen the 5 best working classifiers. These include Logistic Regression (*LR*), Random Forest (*RF*), Extreme Gradient Boost (*XGB*), *LDA*, and Support Vector Machine (*SVM*). With these models, we have conducted both binary and multi-class classification. All of the above models have been run with three different text embedding techniques, which are Term Frequency-Inverse Document Frequency (*TF-IDF*)⁶, Word2Vec⁷, and NEws LANDscape (*NELA*) (Horne et al., 2019b), and in total, a total of 15 experiments were carried out with the machine learning classifiers. **The conclusion is the XGB with TF-IDF and SVM with TF-IDF gave the best results among all the different combinations of classifier models and text embedding techniques.** We also experimented with machine learning models with TF-IDF vectorization along with the hand-crafted linguistic features that we extracted before for the data visualization, which actually performed better than the models with only tf-idf vectorization.

⁶https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html

⁷<https://www.tensorflow.org/text/tutorials/word2vec>

We took only TF-IDF because this vectorization method has given the highest results.

Binary	Class → Model ↓	Human		AI	
		Recall	F1	Recall	F1
Best ML	XGB-TFIDF	98.64	98.33	98.03	98.34
Best DL	CNN-BiLSTM	98.42	98.42	98.42	98.42
Best Transformer	ModernBERT	99.87	99.45	99.02	99.44

Table 4: Performance of best ML, DL, and Transformer models on binary classification of Human vs AI text.

Model Settings: As discussed above, the models that we have taken are *LR*, *RF*, *XGB*, *LDA*, and *SVM*, and the text-embeddings are *TF-IDF*, *Word2Vec*, and *NELA*. In TF-IDF vectorizer, the number of features that are being extracted are $max_features = 5000$.

4.3.2 Neural Network Classifiers

For the neural network classifiers, we have conducted experiments on several models, which include the base models and their hybrids. The models are *CNN*, *RNN*, *LSTM*, *BiLSTM*, *BiGRU*, *CNN_LSTM*, *CNN_BiLSTM*, *CNN_BiGRU*. For all these models, we have taken the *TF-IDF* text embedding technique as this embedding technique got the highest or better results as compared to the other text-embedding techniques. The hybrid model architectures are evaluated in the Layers with separate streams, concatenating at the end. For example, if we take the *CNN_LSTM*, the first setting is two separate streams of *CNN* and *LSTM*, which are taken individually and form a final concatenated dense layer.

Model Setting 1, No Hybrid: This model setting includes *CN*, *LSTM*, *BiLSTM*, *RNN* and *BiGRU* for both Binary and Multiclass classification. Text is vectorized using the *TF-IDF* vectorizer as it gave the best results among the machine learning classification models. *CNN* architecture consists of 3 convolutional layers (256, 128, 64 filters; kernel sizes 5, 5, 3 each) and each followed by batch normalization, max-pooling size of 2, flattening and dense layers (256, 128 units, ReLU activation and dropout of 0.5 and 0.3). Recurrent models also have the same number of layers and layer units (256, 128, 64, first two return sequences), each with batch normalization and 0.3 dropout, and at the end, the dense layers with 128 and 64 units with dropout of 0.3.

Model Setting 2, Hybrid-1: The models include *CNN-LSTM*, *CNN-BiLSTM*, *CNN-BiGRU* for

both Binary and Multi-class classification. Textual features are extracted using the above same method, which is TF-IDF. In BiHybrid models, two independent streams were created and then concatenated after the feature extraction. The CNN components consisted of two convolutional layers (256 and 128 filters, kernel size 5), each followed by batch normalization, max pooling (pool size 2), and flattening. The recurrent components included LSTM, BiLSTM, or BiGRU layers (256 and 128 units), with batch normalization and dropout (0.3). This model’s settings have parallel streams which further combined by the concatenation layer.

Training methods for above Settings: For binary classification, a sigmoid output and for multi-class classification softmax output and categorical cross-entropy loss. Employed with Stratified Kfold with 3 folds and 3 epochs in each fold with a batch size of 64, Adam optimizer, early stopping with patience 2, and learning rate reduction with a factor of 0.5 and patience as 1. And the Best Model is picked up based on the highest validation accuracy.

4.4 Transformer Based Classifiers

We have fine-tuned several transformer-based models for this classification task. They include all the base versions of **BERT**, **DistilBERT**, **RoBERTa**, **ALBERT**, **DeBERTa**, and **ModernBERT**. The versions are *bert-base-uncased*, *distilbert-base-uncased*, *roberta-base*, *deberta-v3-base*, and *modernbert-base*.

Other than just fine-tuning these models directly, we introduce a new method for this multiclass classification task. This methodology is quite new and the latest, as 1) is a backbone architecture of the Deepseek-V3-671b parameter model and the latest recently release Llama 4 family. This architecture is called *Mixture-of-Experts (MoE)*.

4.4.1 MoE Detector

We employed two kinds of MoE architectures, namely *HardMoE* and *SoftMoE* classifiers. In the HardMoE Classifier, a hard gating mechanism is employed, which is a linear gating network that maps the CLS token ($\text{Transformer}(x)[:, 0, :]$) to expert logits $\mathbf{g} = \mathbf{W}_g \mathbf{h}_{\text{CLS}} + \mathbf{b}_g$, and the expert is selected with the highest logit computed through the $\arg\max_i(\mathbf{g}_i)$ operation. This selected expert will take the input, and the output of this chosen expert is passed through the softmax function for the prediction of class. Also, these gate logits are available for auxiliary loss computations.

Unlike HardMoE, SoftMoE is driven by the soft gating method. The gating network here produces logits for all experts, and further applying the softmax, the final normalized weights are obtained. Such that, SoftMoE allows the model to consider all experts for each input. The actual difference between the two architectures is in the gating mechanisms, where HardMoE is like **Winner-Takes-All** selection, and SoftMoE is like aggregating outputs from all the Experts. The Algorithm 0 shows the Forward Pass functions, which tell how input is changing with the layers, and how the classification is done.

Algorithm 1 Forward Pass for MoE Classifier (Hard or Soft)

```

1: Input: input_ids, attention_mask
2: Output: output_logits, gate_logits
3: Extract hidden_state from base transformer
4: Get CLS token:  $cls \leftarrow hidden\_state[:, 0, :]$ 
5: Apply dropout to cls
6: Compute:  $gate\_logits \leftarrow \text{Linear}(cls)$ 
7: if model is HardMoE then
8:   expert_choice  $\leftarrow$   $\arg \max(gate\_logits, \dim = 1)$ 
9:   Initialize output_logits as zeros
10:  for each expert i do
11:     $mask \leftarrow (expert\_choice == i)$ 
12:    if mask not empty then
13:       $out \leftarrow expert_i(cls[mask])$ 
14:       $output\_logits[mask] \leftarrow out$ 
15:    end if
16:  end for
17: else SoftMoE
18:    $gate\_weights \leftarrow \text{Softmax}(gate\_logits)$ 
19:   expert_outputs  $\leftarrow []$ 
20:   for each expert i do
21:      $out \leftarrow expert_i(cls)$ 
22:     Append out to expert_outputs
23:   end for
24:   Stack expert_outputs
25:    $output\_logits \leftarrow \sum(gate\_weights \times expert\_outputs)$ 
26: end if
27: return output_logits, gate_logits

```

4.5 Sentence Segmentation

This section will discuss the data used for this task and the procedure we approached for the process of sentence segmentation of Mixed text of human and AI parts. As discussed in the AI text generation

section, and the Appendix **Mixed Text**, the data contained the following types: 1) Human Commenced and AI ended, 2) AI Commenced and Human ended, and 3) Deeply mixed texts.

We leveraged the power of the Conditional Random Fields and its compatibility with several Transformer backbones, and Neural Network backbones. The complete dataset of size nearly 50k rows, where 20k of type 1, nearly 20k of type 2, and nearly 10k of type 3, was kept for experimentation in three different model setting, 1) *NN-CRF*, 2) *Transformer-CRF*, and 3) *Transformer-NN-CRF*.

NN CRF: All these models are set up to make a hybrid bond between the core three architectures, 1) feature extraction, 2) neural network backbone, 3) a CRF layer. The feature extraction is done by the *deberta-v3-base* model to generalize the contextual word embeddings. The neural network processes these embeddings to sequential patterns, and these NN include *CNN*, *RNN*, *LSTM*, *BiLSTM*, and *BiGRU*, making a total of 5 experiments in this way. At last, the CRF layer enforces the label dependencies with global tag sequences. A dropout layer is applied to prevent overfitting, and a linear classifier maps the neural network outputs to tag logits, so that they can be decoded with the help of CRF.

Transformer CRF: In the place of Neural Networks from the above *NN CRF* model type, these are replaced with the transformer architecture. But this architecture is employed with several things like, 1) *Layer Wise Learning Rate Decay*, 2) *Dynamic Dropout*, 3) *Xavier Initialization*, and was taken up with several transformer backbones like, *BERT*, *DistilBERT*, *RoBERTa*, *DeBERTa*, and *ModernBERT*, making a total of 5 experiments in this way.

Transformer-NN-CRF: This is a type 3 setting for the sentence segmentation task, and in this setting, there are 2 more ways for models. 1) *Best Transformer + All NN + CRF*, 2) *All Transformers + Best NN + CRF*. As we have evaluated the *NN-CRF* and *Transformer-CRF* model first, we have taken the best *NN* model to check out with all the Transformer backbones and the best Transformer backbone with all the *NN* sequential behavior, making a total of 9 experiments in this way. We got the *Best Transformer-CRF* is with the *DeBERTa* model, so with *DeBERTa*, and 5 *NN*'s, and we for the best *NN-CRF* model is with *BiGRU*, so with *BiGRU* and 5 Transformer's, making it a set of 5. This makes us a repetition of *De-*

MultiClass	Class →	Human		OpenAI		Anthropic		Deepseek		Llama		
	Model ↓	Recall	F1	Recall	F1	Recall	F1	Recall	F1	Recall	F1	Accuracy
Best ML	XGB-TFIDF	99.49	97.84	82.73	84.45	86.45	87.32	82.16	82.58	83.52	86.17	91.56
Best DL	CNN-BiGRU	98.97	98.26	84.17	83.44	83.39	85.27	82.57	83.1	83.35	84.07	91.18
Best Transformer	ModernBERT	99.52	99.57	93.48	87.89	88.28	92.29	85.64	86.62	91.22	91.54	94.57
Best HardMoE	BERT-HardMoE	96.28	98	80.97	82.63	88.37	86.81	71.68	79.42	93.54	79.84	89.97
Best SoftMoE	DeBERTa-SoftMoE	97.72	98.68	84.59	84.59	97.26	83.24	78.71	82.38	77.9	85.34	91.22

Table 5: Performance of best ML, DL, and Transformer, HardMoE, and SoftMoE models on MultiClass classification of Human and different AI text.

BERTa_BiGRU_CRF model twice. So, a total of 9 models in this setting.

4.6 Adversarial Addressing

As shown in the figure above 1, we have addressed the Adversarial Attacks in two different settings. 1) *Adversarial Training*, 2) *Adversarial Preprocessing and Implicit Adversarial Classification*. This is also a Binary Classification task, where we are bound to classify the text correctly in case of adversarial attacks. So we have directly taken the Normal Transformer Models for the fine-tuning and obtained the results.

4.6.1 Adversarial Training

This model setting is the simplest one to fine-tune, as we can just directly send the whole data that includes both cleaned text and the attacked texts. Here the model fine-tuning is set just with the specified hyperparameters, with 5 transformer backbones, where the word embeddings came from their own transformer. The five transformers are *bert-base-uncased*, *distilbert-base-uncased*, *roberta-base*, *deberta-v3-base*, *modernbert-base*. The hyperparameters are as follows, keeping the max length to 512 as most of the base transformer models has their *max_length* setting as 512. The model is fine-tuned with a batch size of 32 with 3 epochs, with the AdamW optimizer and a CrossEntropy loss function. Even for these 3 epochs, there is early stopping with patience, as in 2, when there is no decrease in loss.

4.6.2 Adversarial Preprocessing and Implicit Adversarial Classification

We introduce this method for addressing the classification task in case Adversarial attacks by the following way:

1. Let's say we have 2 columns, namely text and label column, while the text column has both normal texts and texts with attacks.

2. We do some sort of preprocessing to the text column and make a new column named pre-processed text, such that this text will have a minimum attack way.

3. Now we compute a few features and stored them in new columns, and these columns are namely Cosine Similarity, Edit Distance, Word-Overlap Ratio, Homoglyph substitution count, BLEU score, Word Error Rate (WER), and Translation Error Rate (TER). These are computed with the columns Original Text and Preprocessed Text.

4. Now, the original text column (which has both non-attacked and attacked texts) and these new computer columns (features columns) are then passed into the model, and fine-tuned with these columns.

5. Here, whether a text is adversarially attacked or not haven't been mentioned anywhere, but the model can predict based on those feature columns. So this is what we call the Implicit classification of the Adversarial Texts.

Example Case in terms of Cosine-Similarity: If a text x has not been altered, then after preprocessing, it remains exactly the same ($x' = x$). In such cases, the cosine similarity between the original and preprocessed text is 1, indicating perfect similarity. However, when x has been adversarially manipulated, preprocessing produces a different version x' , leading to a cosine similarity less than 1. This drop in similarity serves as a signal of textual perturbation.

5 Results and Discussions

As we have done a vast set of experiment, the concise and best results are shown for different tasks at different Tables 4, 5, 7, 6. Table 4 gives the best set of models over Binary Classification. Table

Sentence Segmentation	Model	Accuracy	Precision	Recall	F1-score	MCC	Kappa
Best NN-CRF	BiGRU-CRF	96.19	96.19	96.19	96.19	92.21	92.21
Best Transformer-CRF	DeBERTa-CRF	98.84	98.84	98.84	98.84	97.67	97.67
Best Transformer-NN-CRF	DeBERTa-BiGRU-CRF	98.95	98.95	98.95	98.95	97.89	97.89

Table 6: Best Sentence Segmentation Results over all the Model Settings.

Binary	Class → Model ↓	Human		AI	
		Recall	F1	Recall	F1
Best Adv	ModernBERT	89.03	89.97	89.46	89.64
Best Impl	ModernBERT	93.63	98.22	94.77	94.24

Table 7: Performance of best Adversarial Training and Implicit differentiable models on adversarial classification of Human vs AI text.

5 shows the best-performing models over Multi-class Classification. Table 7 has the two best model setting over the Adversarial Data. Finally, the Table 6 gives the results of Best Sentence Segmentation models. Binary Classification task is benchmarked with both Recall and F1-score, with the highest values to Recall and F1-Score of 99.87% and 99.45% for the Human-Class, while 99.02% and 99.44% for AI-Class. Multiclass Classification, ModernBERT-Transformer has the Highest performing Accuracy of 94.57%, outperforming newly proposed HardMoE and SoftMoE Detector architectures. For the Sentence Segmentation task, the benchmark Evaluation metric is Cohen’s Kappa score, while the model *DeBERTa-BiGRU-CRF* got the highest score of 97.87%. Among the two methodologies that are discussed for the Adversarial Detection and binary classification correctly, the newly proposed method got the highest values of Recall and F1-score of classes Human as 93.63%, 98.22% and class AI as 94.77% and 94.24% respectively.

6 Conclusion and Future Scope

In this paper, we have presented a vast set of experimentation regarding the AI text detection. The experiments are done from the very basic to the most recent and advanced approaches. All these current methods comes under Supervised Learning, we haven’t explored the Zero-shot or One-shot methods. So, we are expected to do these experiments in the future self. We found that, the Detectors are performing well in binary classification on, but they are getting more confused when the case of Multi-Class and Bi-Label comes into the picture.

Overall the detectors, DeBERTa and ModernBERT transformer models are giving the best results on average.

As this paper only focused on “Text”, we wanted to explore other AI generated content like Images and Speech, and do experimentation on these variations. For the case of Adversarial Attacks, we have explored one method, and proposed one new method, so it is worth to lookup other methods.

7 Limitations

In this section, we discuss some of the key limitations of our work.

We mainly focused on the English language, but neither focused on multi- nor low-resource languages. Even though english is the major language used worldwide, but there are cases where their own language is used in Academics, etc. We have taken a very few set of LLM models, which includes two open source and two closed source models, but the inclusion of the open source models can also make the dataset much more robust. But even though there are few models, we have taken the latest versions of all at the time of dataset creation. All of our experiments are only in one case: supervised training and fine-tuning. These methods, especially in the case of Transformer models are included fine-tuning takes a lot of time. While, zero-shot methods or completely unsupervised methods may reduce this time for the detection. One of the major ones we haven’t taken into consideration is Perplexity-based detection methods. Because Perplexity-based detections can be the best case for our work.

8 Resources and Cost

We spent \$50 on the OpenAI API requests,\$30 on Llama API calls, \$25 on Deepseek API calls, and \$30 on Claude Anthropic API calls. The AWS EC2 instance for Accelerated Computing, we spent around \$600, while utilizing 300 GPU hours of NVIDIA L40S GPU.

9 Ethics

We here by declare that this E-BMAS dataset is created by only using the existing publicly available dataset for the Human corpus, and has not been taken or scraped from any website or any on-line platform. All the taken datasets are properly cited in this paper. For the AI texts, we have taken the API keys for each AI model and sent on API requests for the Chat-Completions

10 Reproducibility Statement

To Support Transparency and make it easier to reproduce the paper, we have provided a detailed description on how to construct the **E-BMAS dataset** using data generation prompts and pre-processing pipeline. The hyperparameters and model requirements are detailed throughout the paper. After the creation of the dataset and specifying the hyperparameters, we performed experiments on the performance of multiple models in binary, multiclass, segmentation, and adversarial tasks, and generated test results for all experiments. After model evaluation, we performed feature extraction and visualization of key linguistic and stylistic features to highlight differences. And Finally, we also illustrate the performance metrics for all model categories and comparison of extracted features.

References

- Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Unsupervised cross-lingual representation learning at scale. *arXiv preprint arXiv:1911.02116*.
- Liam Dugan, Alyssa Hwang, Filip Trhlik, Josh Magnus Ludan, Andrew Zhu, Hainiu Xu, Daphne Ippolito, and Chris Callison-Burch. 2024. Raid: A shared benchmark for robust evaluation of machine-generated text detectors. *arXiv preprint arXiv:2405.07940*.
- Liam Dugan, Daphne Ippolito, Arun Kirubakaran, and Chris Callison-Burch. 2020. [RoFT: A tool for evaluating human detection of machine-generated text](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 189–196, Online. Association for Computational Linguistics.
- Sebastian Gehrmann, Hendrik Strobelt, and Alexander M Rush. 2019. Gltr: Statistical detection and visualization of generated text. *arXiv preprint arXiv:1906.04043*.
- Biyang Guo, Xin Zhang, Ziyuan Wang, Minqi Jiang, Jinran Nie, Yuxuan Ding, Jianwei Yue, and Yupeng Wu. 2023. How close is chatgpt to human experts? comparison corpus, evaluation, and detection. *arXiv preprint arXiv:2301.07597*.
- Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, et al. 2025. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*.
- Abhimanyu Hans, Avi Schwarzschild, Valeriia Cherepanova, Hamid Kazemi, Aniruddha Saha, Micah Goldblum, Jonas Geiping, and Tom Goldstein. 2024. Spotting llms with binoculars: Zero-shot detection of machine-generated text. *arXiv preprint arXiv:2401.12070*.
- Benjamin D Horne, Jeppe Nørregaard, and Sibel Adali. 2019a. Robust fake news detection over time and attack. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 11(1):1–23.
- Benjamin D. Horne, Jeppe Nørregaard, and Sibel Adali. 2019b. [Robust fake news detection over time and attack](#). *ACM Trans. Intell. Syst. Technol.*, 11(1).
- John Kirchenbauer, Jonas Geiping, Yuxin Wen, Jonathan Katz, Ian Miers, and Tom Goldstein. 2023. A watermark for large language models. In *International Conference on Machine Learning*, pages 17061–17084. PMLR.
- Laida Kushnareva, Tatiana Gaintseva, German Magai, Serguei Barannikov, Dmitry Abulkhanov, Kristian Kuznetsov, Eduard Tulchinskii, Irina Piontkovskaya, and Sergey Nikolenko. 2023. Ai-generated text boundary detection with roft. *arXiv preprint arXiv:2311.08349*.
- Sai Teja Lekkala, Annepaka Yadagiri, Mangadoddi Srikar Vardhan, and Partha Pakray. 2025. [CNLP-NITS-PP at GenAI detection task 3: Cross-domain machine-generated text detection using DistilBERT techniques](#). In *Proceedings of the 1st Workshop on GenAI Content Detection (GenAIDetect)*, pages 334–339, Abu Dhabi, UAE. International Conference on Computational Linguistics.
- Jenny S Li, John V Monaco, Li-Chiou Chen, and Charles C Tappert. 2014. Authorship authentication using short messages from social networking sites. In *2014 IEEE 11th International Conference on e-Business Engineering*, pages 314–319. IEEE.
- Yafu Li, Qintong Li, Leyang Cui, Wei Bi, Zhilin Wang, Longyue Wang, Linyi Yang, Shuming Shi, and Yue Zhang. 2023. Mage: Machine-generated text detection in the wild. *arXiv preprint arXiv:2305.13242*.
- Yafu Li, Qintong Li, Leyang Cui, Wei Bi, Zhilin Wang, Longyue Wang, Linyi Yang, Shuming Shi, and Yue Zhang. 2024. [MAGE: Machine-generated text detection in the wild](#). In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 36–53, Bangkok, Thailand. Association for Computational Linguistics.
- Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.
- Eric Mitchell, Yoonho Lee, Alexander Khazatsky, Christopher D Manning, and Chelsea Finn. 2023. Detectgpt: Zero-shot machine-generated text detection using probability curvature. In *International Conference on Machine Learning*, pages 24950–24962. PMLR.
- Shashi Narayan, Shay B. Cohen, and Mirella Lapata. 2018. Don’t give me the details, just the summary! Topic-aware convolutional neural networks for extreme summarization. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Brussels, Belgium.
- Adaku Uchendu, Zeyu Ma, Thai Le, Rui Zhang, and Dongwon Lee. 2021. Turingbench: A benchmark environment for turing test in the age of neural text generation. *arXiv preprint arXiv:2109.13296*.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

- Pengyu Wang, Linyang Li, Ke Ren, Botian Jiang, Dong Zhang, and Xipeng Qiu. 2023a. [SeqXGPT: Sentence-level AI-generated text detection](#). In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 1144–1156, Singapore. Association for Computational Linguistics.
- Yuxia Wang, Jonibek Mansurov, Petar Ivanov, Jinyan Su, Artem Shelmanov, Akim Tsvigun, Osama Mohammed Afzal, Tarek Mahmoud, Giovanni Puccetti, Thomas Arnold, et al. 2024a. M4gt-bench: Evaluation benchmark for black-box machine-generated text detection. *arXiv preprint arXiv:2402.11175*.
- Yuxia Wang, Jonibek Mansurov, Petar Ivanov, Jinyan Su, Artem Shelmanov, Akim Tsvigun, Chenxi Whitehouse, Osama Mohammed Afzal, Tarek Mahmoud, Toru Sasaki, et al. 2023b. M4: Multi-generator, multi-domain, and multi-lingual black-box machine-generated text detection. *arXiv preprint arXiv:2305.14902*.
- Yuxia Wang, Jonibek Mansurov, Petar Ivanov, Jinyan Su, Artem Shelmanov, Akim Tsvigun, Chenxi Whitehouse, Osama Mohammed Afzal, Tarek Mahmoud, Toru Sasaki, Thomas Arnold, Alham Aji, Nizar Habash, Iryna Gurevych, and Preslav Nakov. 2024b. [M4: Multi-generator, multi-domain, and multi-lingual black-box machine-generated text detection](#). In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1369–1407, St. Julian’s, Malta. Association for Computational Linguistics.
- Feng Xiong, Thanet Markchom, Ziwei Zheng, Subin Jung, Varun Ojha, and Huizhi Liang. 2024. Fine-tuning large language models for multigenerator, multidomain, and multilingual machine-generated text detection. *arXiv preprint arXiv:2401.12326*.
- Rowan Zellers, Ari Holtzman, Hannah Rashkin, Yonatan Bisk, Ali Farhadi, Franziska Roesner, and Yejin Choi. 2019. Defending against neural fake news. *Advances in neural information processing systems*, 32.
- Zijie Zeng, Shiqi Liu, Lele Sha, Zhuang Li, Kaixun Yang, Sannyuya Liu, Dragan Gašević, and Guanliang Chen. 2024. Towards detecting ai-generated text within human-ai collaborative hybrid texts. *arXiv e-prints*, pages arXiv–2403.
- Xuandong Zhao, Yu-Xiang Wang, and Lei Li. 2023. Protecting language generation models via invisible watermarking. In *International Conference on Machine Learning*, pages 42187–42199. PMLR.

A Data Generation Prompts

A.1 Binary and Multiclass

As we require a certain prompt for the text generation from any LLM, we use human text as our base one for the text-generation by LLMs. We prompted the LLM to give a question regarding the human-text such that the answer to the question should cover the entire context of the answer. Then, the question that is generated is given to the LLM to answer. But the text-generation of each domain is done differently. We haven't confined the LLM to generate the answer in a certain range of tokens like the previous works in M4 because an LLM answer can be in any range, like the human answer.

We have taken two closed-source and two open-source models for the AI texts. Open-source models include OpenAI and Anthropic, and closed-source models include Deepseek and Llama. All the models that we have taken were the largest and latest models at the time of data creation. The OpenAI model is the 4o version, Anthropic's model is the Claude-3.5-Sonnet model, Deepseek is the largest model and latest version, Deepseek-v3 671b parameter model, and finally, the Llama model has specs as 3.3 version and 70b parameter model.

Model	Specs
OpenAI	GPT-4o, GPT-4.1
Anthropic	Claude-3.5-sonnet
Deepseek	V3:671b
Llama	3.3:70b

Table 8: Model Specifications

As discussed above, the domains we have taken are *Reddit*, *News*, *Wikipedia*, *Arxiv abstracts*, and *Q&A*. For each domain, the prompts were given in detail in the Table in the Appendix .

1. Reddit human texts are taken from M4-Reddit (Wang et al., 2024b), MAGE-YELP (Li et al., 2024), MAGE-CMV (Li et al., 2024).
2. Human texts for News article are taken from the XSUM dataset (Narayan et al., 2018).
3. Wikipedia human texts are taken from M4-Wiki(Wang et al., 2024b), MAGE-SQUAD (Li et al., 2024).
4. Arxiv abstracts by humans are taken from MAGE-SciGen (Li et al., 2024).
5. Human QA texts are taken from the MAGE-ELI5 (Li et al., 2024).

Reddit AI texts are generated as follows: First, it will analyze the given human text and detect the personal perspective based on the pronouns that are contained in the text. And then, it is asked for the query and answer generation based on the personal perspective.

News AI texts are generated as follows: First, it was asked to directly generate the query for the answers, but while answering, it is asked to do it in two different ways, which are a nuanced way and a non-nuanced way. When in a nuanced way, the answer to that question is generated in the same stance as the given human-text, and in a non-nuanced way, it opposes the stance of the given human-text.

Wikipedia AI texts are generated as follows: It was asked to generate a question based on the human-text, and such that the answer was given in such a way that, it will be both a factual and neutral answer.

Arxiv AI texts are generated as follows: First, it was asked to directly generate the query for the answers, but while answering, it is asked to do it in two different ways, which are the simplified way and another way based on the real-world scenario. When in a simplified way, the answer to that question is generated in a simple explanation of the research concept in easy-to-understand terms, and in the other way, it will provide research findings that can be applied to a real-world problem or scenario.

Q&A AI texts are generated as follows: It was just directly asked to generate a question based on the human-text, and such that the answer will give a well-structured AI response that provides additional insights or an enhanced explanation.

A.2 Mixed Text

We have generated the data in three different kinds, 1) Human-started and AI-ended texts, 2) AI-started and Human-ended texts, 3) Fully mixed texts. We carefully taken care of the boundaries where the interchange occurs between the human and machine. The statistics of each type are given in the **Table**. We have only taken only three AI models one open-source model, Deepseek-v3 671b, and two closed-source models, OpenAI-GPT-4o and OpenAI-GPT-4.1.

Type-1 texts are generated in the following way: We truncated the human-corpora that we collected into half. Before truncating, we cleaned the text, by removing multiple new lines, extra spaces, links or URLs, citations or references, and commas. And then that cleaned text is truncated to half, and the asked the AI models to just complete the half-written human text in a continued way. In such a way, we got Human started and AI-ended texts. We calculated the human-end boundaries of each row before the text generation. This boundary is the number of human words before the AI text.

Type-2 texts are generated in the following way: We cleaned the text in the same way as Type-1, and then truncated it too, but now, we discarded the first part of the human. And this half-written human text is now given to the AI models to complete the text, such that it will form a complete sentence without any sentence breaks. Here the AI end boundaries are calculated by counting the number of words in both complete text (AI+Human) and Truncated human text, then subtracting from one another.

Type-3 texts are generated in the following way: First, we have taken the human texts and then, we are replacing a subset of sentences in the input text with a specific placeholder tag, <AI_Start> and </AI_End>. If the actual text has 'n' sentences, then we instructed the code to select a random number 'm' which can take a value between 1 and 'n-1' and select 'm' unique sentences to replace. Now another function is used to fill up the space with the AI text between the placeholder tags, with some restricted rules. The prompt is given like below.

Prompt used for Mixed Data Generation

You are tasked with creating content for an AI-human collaborative document. The document has missing parts marked by <AI_Start></AI_End> tags. Your job is to generate a single novel sentence to fill the gap between <AI_Start> and </AI_End>. The sentence should:

- Be accurate and relevant to the topic implied by the surrounding context, regardless of the domain.
- Fit seamlessly with the surrounding text, maintaining the document's flow and style.
- Be distinct from any original content, offering a fresh perspective or detail.
- Be concise and suitable for sentence segmentation studies.

Context before the missing part: "{left_context}"

Context after the missing part: "{right_context}"

Reply with **ONLY** the sentence to be placed between <AI_Start> and </AI_End>, without including the tags themselves.

A.3 Adversarial Data Generation

We have taken a fine set of adversarial attacks, namely, 1) Synonym Substitution, 2) Misspelling, 3) Homoglyph Replacement, 4) Upper-Lower Swap, 5) Zero-Width Space Insertion. The **Synonym substitution** is employed to replace the worst with similar words based on a given or predefined probabilities with the help of the WordNet Lexical Database. **Misspelling** is employed by randomly swapping adjacent

characters or letters in words that are at least char len of 4. The best mitigating attack from the AI text detectors is the **Homoglyph attack**, so we collected a vast set of homoglyphs, and replaced the actual characters again by the probability threshold, which helps simulate deceptive texts. Even though the homoglyph attack is the strong attack, here comes the simple one, **Upper-Lower-Swap** which randomly changes the case of letters flipping lower to upper and vice versa. The invisible attack, named **Zero Width Space** insertion, this is by inserting the zero width space characters after the alpha numeric characters. We tried to attack a specific text with specific attack a several number of times, such that the adversaries can be more visible instead of just one time.

B Test Results for All Experiments

We have given the main or best results of each experimentation in the above, where for the Binary Classification, the Best ML Classifier is XGBoost with TF-IDF word embeddings, the best DL classifier model is CNN-BiLSTM with again the same TF-IDF word embedding, and finally, the Best Transformer model is ModernBERT fine-tuned. See the full results Table 9.

For the Multi-Class classification, similar like the above code, the best ML Classifier is XGBoost with TF-IDF word embeddings, the best DL Classifier model is CNN-BiGRU with TF-IDF word embeddings, and the best Transformer model is ModernBERT fine-tuned. According to the new models that are proposed as the HardMoE and SoftMoE detectors, the Best HardMoE is BERT-HardMoE and the best SoftMoE is DeBERTa-SoftMoE. While we have given the overall results at Table 11.

The other task, Sentence-Segmentation or the Bi-Label classification, the best models in each setting are *BiGRU-CRF*, *DeBERTa-CRF*, and *DeBERTa-BiGRU-CRF*. While results for the every model setting and for all the models can be seen in the Table 12.

Classification	Type	Model	Human		AI		Accuracy
			Recall	F1	Recall	F1	
Binary	NELA	LR	47.16	50.03	60.32	56.77	53.78
		RF	61.35	62.95	66.83	65.2	64.11
		XGB	57.73	64.32	78.48	71.27	68.17
		LDA	45.78	49.7	62.02	57.55	53.96
		SVM	56.95	65.44	74.63	63	58.91
	Word2Vec	LR	96.15	96.56	97.04	96.63	96.6
		RF	96.45	96.1	95.57	96.12	96.11
		XGB	97.66	97.64	97.66	97.67	97.66
		LDA	96.65	96.08	95.52	96.08	96.08
		SVM	96.02	96.52	97.09	96.6	96.56
	TF-IDF	LR	98.21	97.94	97.68	97.95	97.95
		RF	98.21	97.9	97.61	97.91	97.91
		XGB	98.64	98.33	98.03	98.34	98.33
		LDA	98.33	98.03	97.73	98.04	98.03
		SVM	98.33	98.17	98.03	98.19	98.18
	No Hybrid	CNN	91.54	95.33	99.49	95.67	95.51
		RNN	98.62	97.85	97.04	97.81	97.83
		LSTM	97.75	97.77	97.79	97.77	97.77
		BiLSTM	98.32	97.77	97.19	97.74	97.76
		BiGRU	97.18	97.61	98.07	97.63	97.62
	Hybrid	CNN-LSTM	99.12	98.00	96.81	97.94	97.97
		CNN-BiLSTM	98.42	98.42	98.42	98.42	98.42
		CNN-BiGRU	99.42	97.87	96.24	97.79	97.83
	Normal	BERT	96.18	98	99.89	98.06	98.03
		DistilBERT	98.75	99.24	99.74	99.25	99.25
		RoBERTa	94.16	96.96	99.94	97.12	97.05
		DeBERTa	96.78	98.35	99.97	98.39	98.37
		ModernBERT	99.87	99.45	99.02	99.44	99.45

Table 9: All Experiments Performance Metrics for the Binary Classification between Human and AI Texts.

Classification	Type	Model	Human		AI		Accuracy
			Recall	F1	Recall	F1	
Binary	Adversarial Training	BERT	86.21	88.93	89.39	88.23	88.12
		DistilBERT	88.57	89.04	89.85	89.87	89
		RoBERTa	84.96	86.24	89.14	87.45	87.47
		DeBERTa	86.98	88.55	89.99	88.49	88.9
		ModernBERT	89.03	89.97	89.46	89.64	89.55
	Implicit Method	BERT	94.34	94.44	95.57	94.45	94.46
		DistilBERT	92.99	94.24	95.54	94.28	94.26
		RoBERTa	92.02	93.59	95.24	93.66	93.63
		DeBERTa	92.78	94.21	95.69	94.25	94.23
		ModernBERT	93.69	98.22	94.77	94.24	94.23

Table 10: Adversarial Binary Classification Results between Human and AI texts.

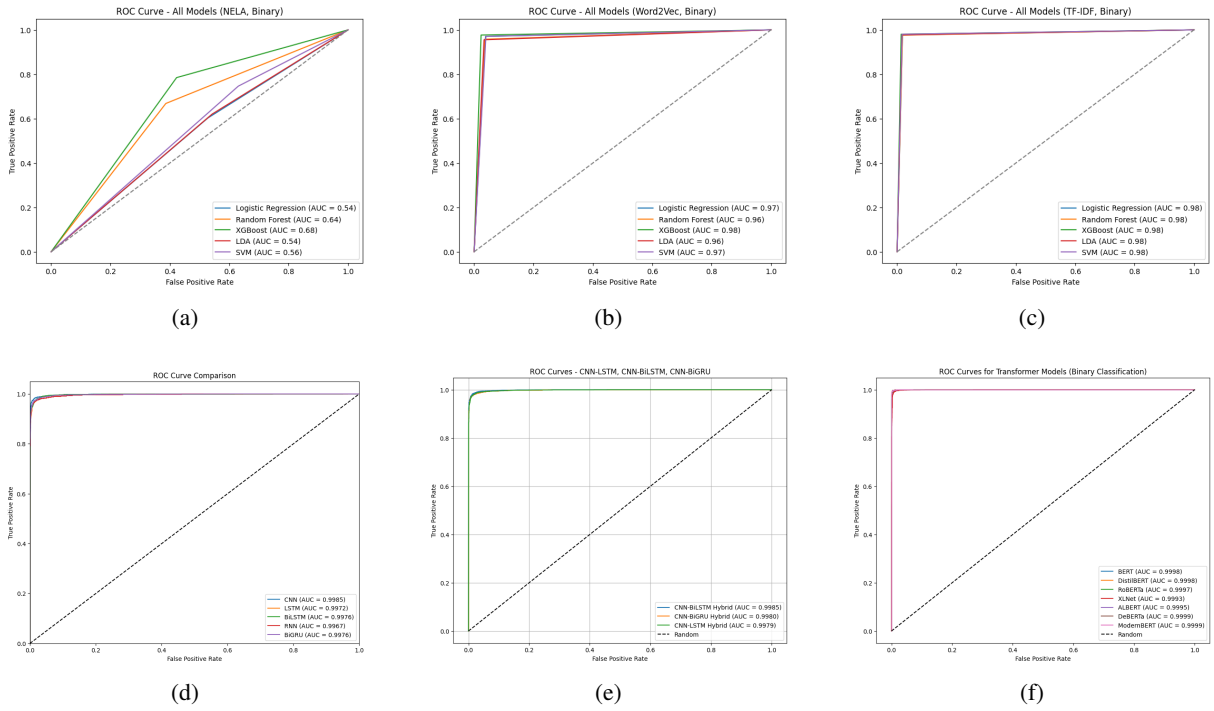


Figure 2: ML All Binary class Classification.

C Dataset Visualizations

We have done a vast set of feature extraction, and visualized them at the following Figures 10, 11, 12. For our data visualization, we have extracted several linguistic and Stylistic features that include 1) Burtiness, 2) Readability Score, 3) Syntactic Complexity, 4) Unique Word Count, 5) Average Sentence Length, 6) Punctuation Usage, 7) Stop Word Count, 8) Entropy, 9) Total Word Count, 10) Sentence Count, 11) Bigram Uniqueness, 12) Type Token Ratio, 13) Lexical Richness, and 14) Hapax Legomenon Rate. This certainly reveals the differentiation in texts from the different AI models.

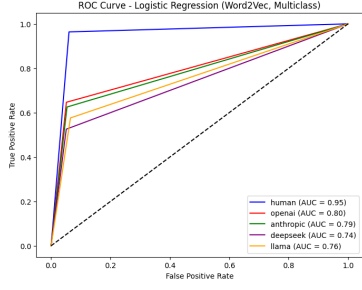
D Example Generations

Classification	Type	Model	Human		OpenAI		Anthropic		Deepseek		Llama		Accuracy
			Recall	F1	Recall	F1	Recall	F1	Recall	F1	Recall	F1	
MultiClass	Word2Vec	LR	96.62	95.85	64.58	63.32	63.14	62.03	54.81	57.7	59.16	59.42	78.42
		RF	98.89	93.85	57.73	57.25	51.79	55.17	52.27	53.63	42.69	49.18	74.83
		XGB	98.33	96.91	62.5	64.64	68.22	66.79	61.49	62.09	60.42	62.68	80.63
		LDA	97.18	95.58	63.78	63.38	60.15	60.49	53.59	55.36	56.14	57.68	77.68
		SVM	97.23	96.04	64.78	65.13	62.94	62.41	55.62	58.71	60.33	60.33	78.97
	TF-IDF	LR	99.09	97.17	84.72	85.32	85.05	85.78	80.64	82.91	81.28	84.03	90.96
		RF	99.94	92.32	72.02	75.38	74.5	81.43	65.65	71.01	65.59	74.57	84.6
		XGB	99.49	97.84	82.73	84.45	86.45	87.32	82.16	82.58	83.52	86.17	91.56
		LDA	98.64	97.56	82.04	83.4	84.36	86.64	82.06	81.57	85.08	85.67	90.97
		SVM	99.04	97.76	85.61	85.87	86.45	85.85	82.06	83.37	84.32	84.32	91.36
	No Hybrid	CNN	99.97	86.19	71.35	69.41	57.81	67.15	99	32.97	57.41	65.66	75.73
		RNN	98.6	97.29	83.24	82.14	82.32	81.88	78.01	78.76	77.19	80.99	89.21
		LSTM	99.67	97.35	81.9	82.75	80.17	82.42	77.82	80.45	82.74	81.29	89.65
		BiLSTM	97.43	97.41	86.97	82.89	81.93	83.35	77.62	80.08	81.23	81.19	89.67
		BiGRU	98.6	97.47	86.76	81.41	79.88	83.17	73.96	79	80.82	81.23	89.46
	Hybrid	CNN-LSTM	99.37	97.7	86.14	83.21	85.64	84.61	75.04	80.89	80.22	83.77	90.57
		CNN-BiLSTM	99.87	96.12	81.59	82.61	83.1	84.67	76.93	80.04	74.97	80.09	89.53
		CNN-BiGRU	98.97	98.26	84.17	83.44	83.39	85.27	82.57	83.1	83.35	84.07	91.18
	HardMoE	BERT	96.28	98	80.97	82.63	88.37	86.81	71.68	79.42	93.54	79.84	89.97
		DistilBERT	93.86	96.69	76.52	81.49	80.37	85.24	79	78.38	95.96	77.91	88.42
		RoBERTa	97.3	98.47	78.9	77.46	87.5	84.72	77.03	76.18	90.61	80.07	89.05
		DeBERTa	93.83	96.77	75.07	79.73	95.99	77.31	73.36	78.41	80.92	81.29	87.66
	SoftMoE	BERT	95.23	97.45	83.14	81.41	84.96	86.65	72.47	77.21	95.15	82.43	89.57
		DistilBERT	93.98	96.82	80.97	81.81	87.69	86.59	80.29	78.96	93.64	85.25	89.83
		RoBERTa	95.5	97.61	81.69	80.61	77.73	84.81	84.55	75.84	87.08	83.46	89.13
		DeBERTa	97.72	98.68	84.59	84.59	97.26	83.24	78.71	82.38	77.9	85.34	91.22
	Normal	BERT	97.85	98.86	96.03	85.86	82.42	89.17	84.55	84.72	96.56	86.25	92.61
		DistilBERT	98.35	99.03	78.07	83.98	95.01	88.33	83.26	84.86	90.11	86.86	92.55
		RoBERTa	96.48	98.84	78.8	84.38	89.94	89.37	83.86	84.78	95.66	84.07	91.81
		DeBERTa	97.33	98.63	80.45	85.96	92.18	89.73	82.57	85.23	96.46	86.16	92.66
		ModernBERT	99.52	99.57	93.48	87.89	88.28	92.29	85.64	86.62	91.22	91.54	94.57

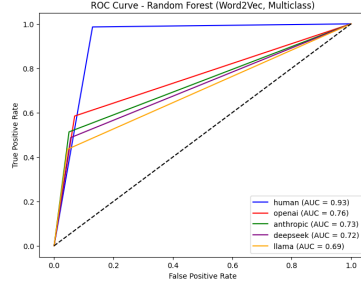
Table 11: Comprehensive MultiClass Classification Performance Across Various AI Texts

Classification	Type	Model	Accuracy	Precision	Recall	F1-score	MCC	Kappa
Multi-Label	NN_CRF	CNN_CRF	90.89	90.9	90.89	90.86	81.34	81.28
		RNN_CRF	90.03	90.02	90.03	90	79.56	79.52
		LSTM_CRF	93.74	93.76	93.74	93.73	87.19	87.15
		BiLSTM_CRF	95.58	95.58	95.58	95.57	90.95	90.93
		BiGRU_CRF	96.19	96.19	96.19	96.19	92.21	92.21
	Transformer_CRF	BERT_CRF	97.89	97.9	97.89	97.89	95.76	95.75
		DistilBERT_CRF	97.52	97.54	97.52	97.52	95.03	95.01
		RoBERTa_CRF	98.41	98.42	98.41	98.41	96.81	96.8
		ModernBERT_CRF	98.72	98.72	98.72	98.72	97.41	97.41
		DeBERTa_CRF	98.84	98.84	98.84	98.84	97.67	97.67
	Transformer_NN_CRF	DeBERTa_CNN_CRF	98.83	98.83	98.83	98.83	97.64	97.63
		DeBERTa_RNN_CRF	98.81	98.81	98.81	98.81	97.61	97.6
		DeBERTa_LSTM_CRF	98.86	98.86	98.86	98.86	97.69	97.69
		DeBERTa_BiLSTM_CRF	98.88	98.88	98.88	98.88	97.74	97.74
		DeBERTa_BiGRU_CRF	98.95	98.95	98.95	98.95	97.89	97.89
		BERT_BiGRU_CRF	97.96	97.98	97.96	97.96	95.92	95.9
		DistilBERT_BiGRU_CRF	97.69	97.7	97.69	97.69	95.35	95.34
		RoBERTa_BiGRU_CRF	98.48	98.5	98.48	98.48	96.96	96.95
		ModernBERT_BiGRU_CRF	98.7	98.7	98.7	98.7	97.37	97.37

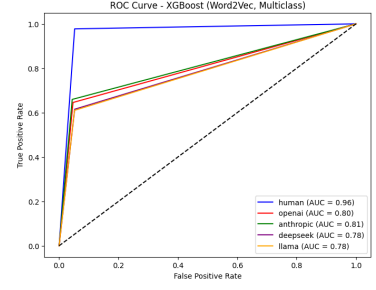
Table 12: Performance Metrics of Various Sentence Segmentation or Bi-Label Classification Models



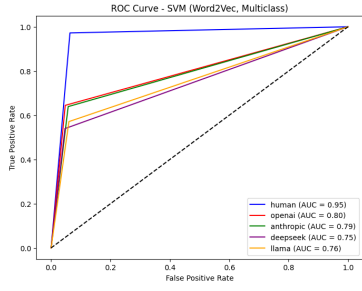
(a)



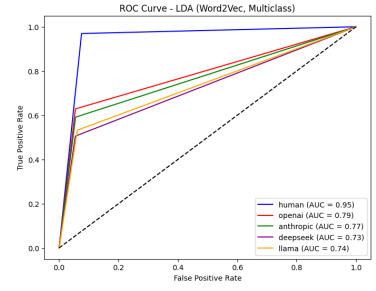
(b)



(c)

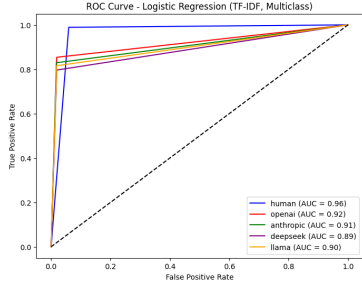


(d)

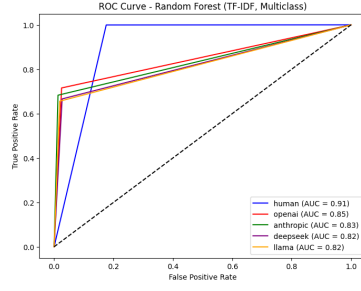


(e)

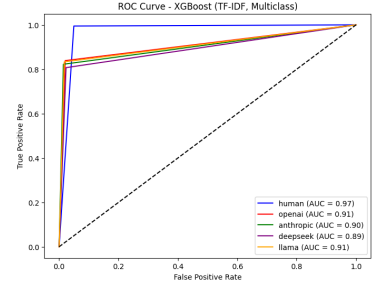
Figure 3: ML Word2Vec Word Embedding Multiclass Classification.



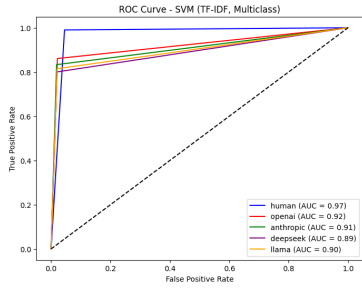
(a)



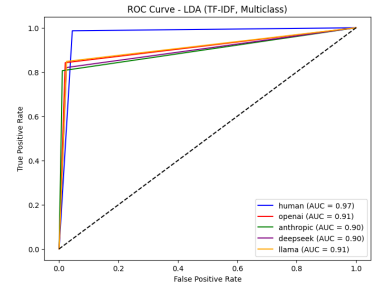
(b)



(c)

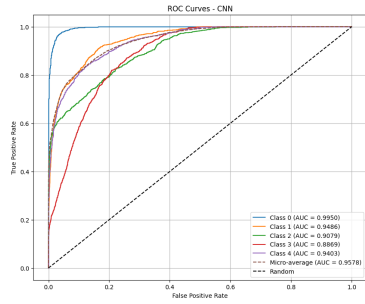


(d)

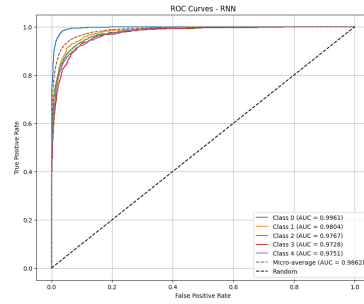


(e)

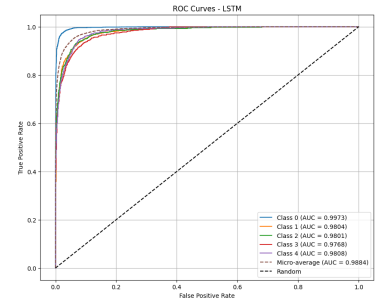
Figure 4: ML TF-IDF Word Embedding Multiclass Classification.



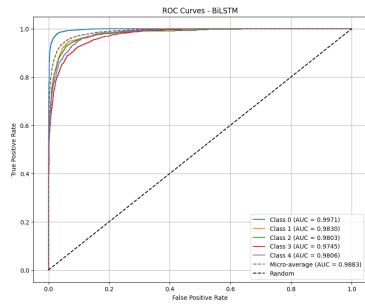
(a)



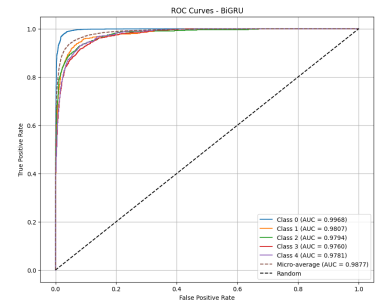
(b)



(c)

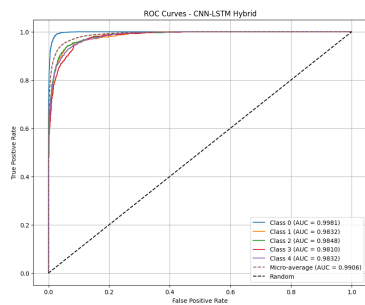


(d)

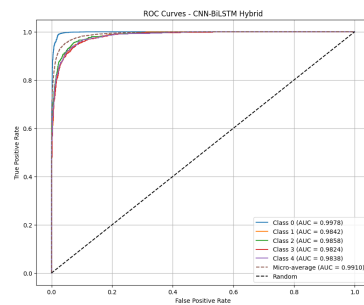


(e)

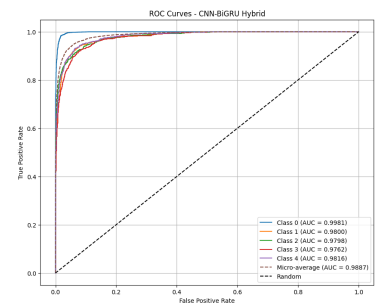
Figure 5: Non-Hybrid DL Multiclass Classification.



(a)

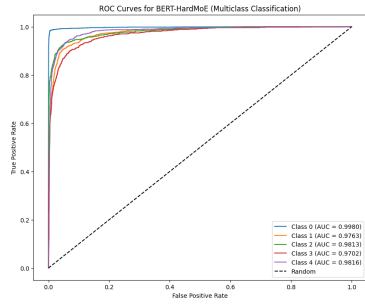


(b)

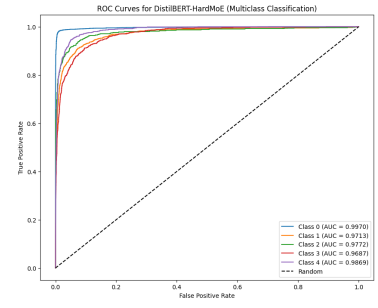


(c)

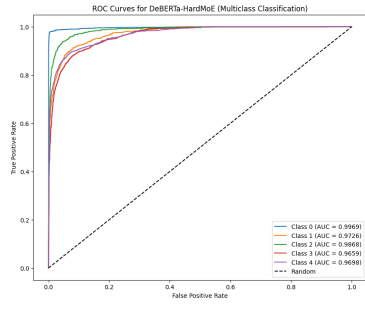
Figure 6: Hybrid DL Multiclass Classification.



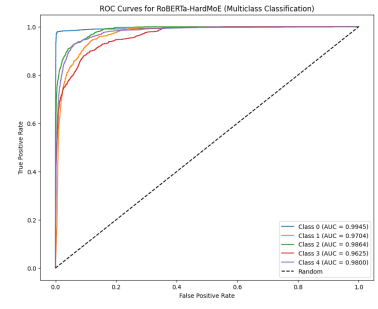
(a)



(b)

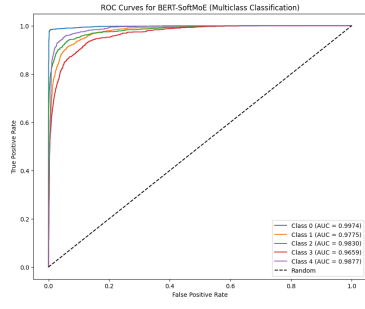


(c)

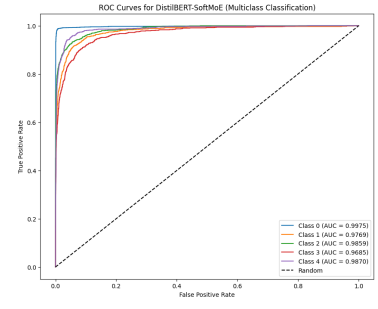


(d)

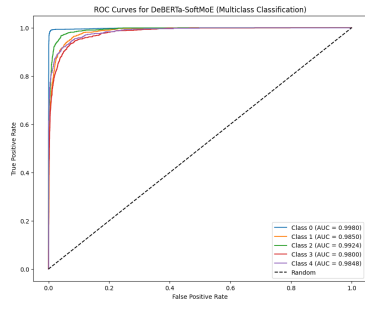
Figure 7: HardMoE Transformers Multiclass Classification.



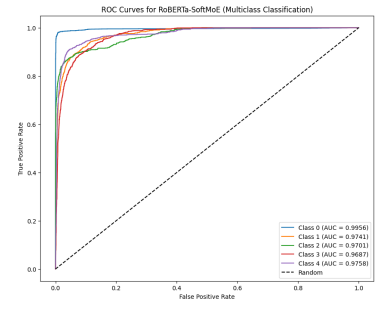
(a)



(b)

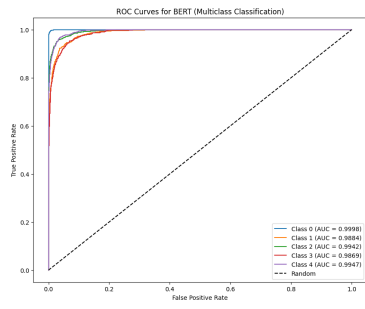


(c)

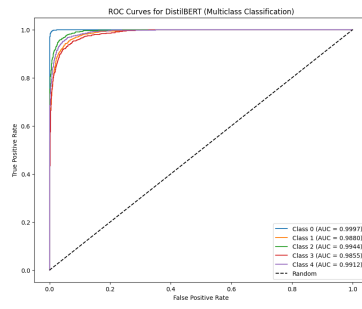


(d)

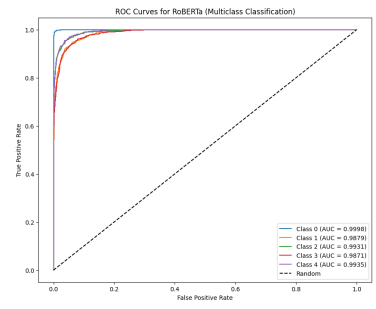
Figure 8: SoftMoE Transformers Multiclass Classification.



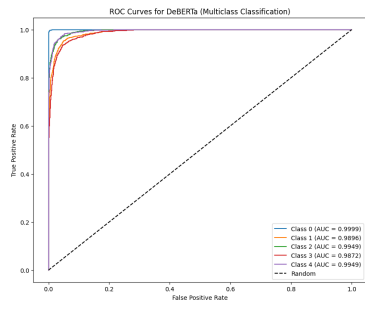
(a)



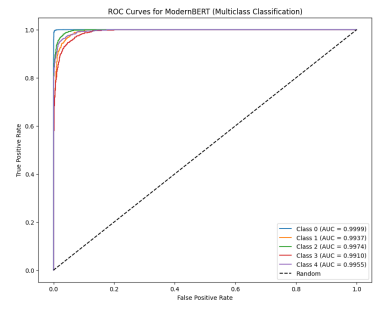
(b)



(c)



(d)



(e)

Figure 9: Normal Transformer Multiclass Classification.

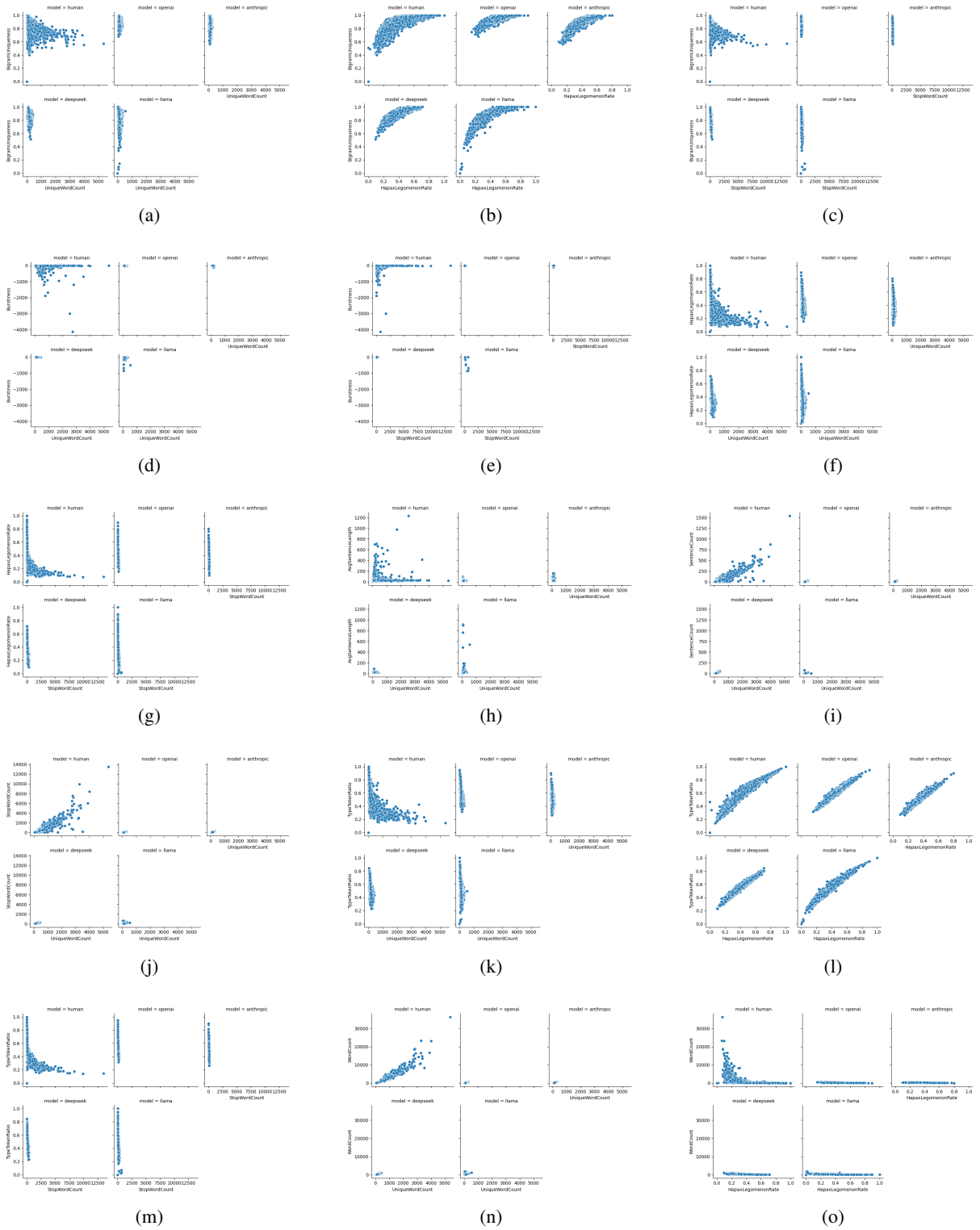
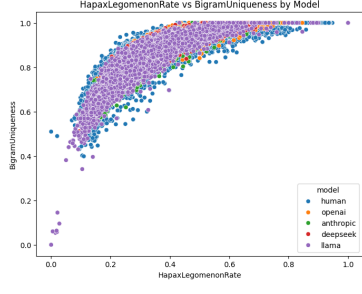
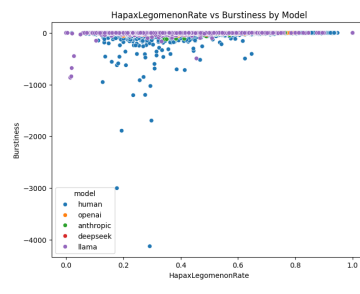


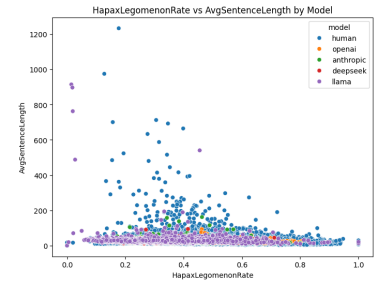
Figure 10: Dataset Feature Comparison



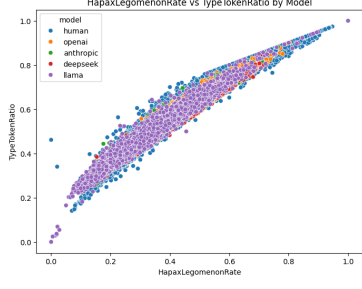
(a)



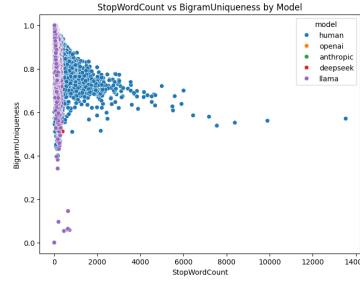
(b)



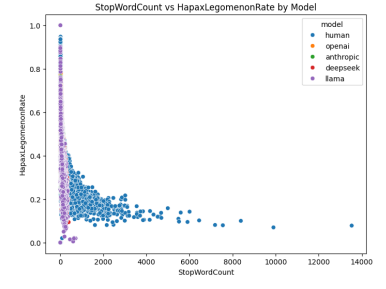
(c)



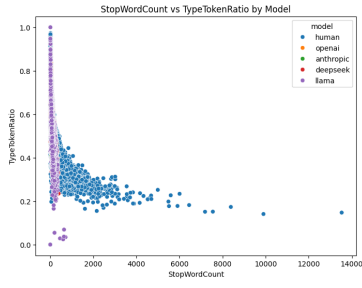
(d)



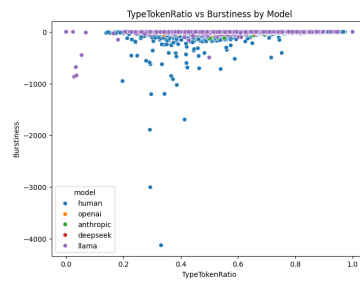
(e)



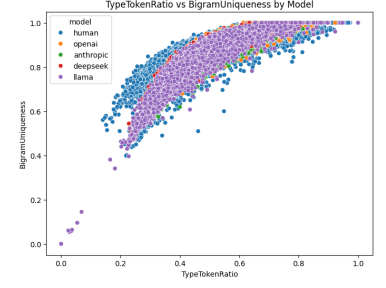
(f)



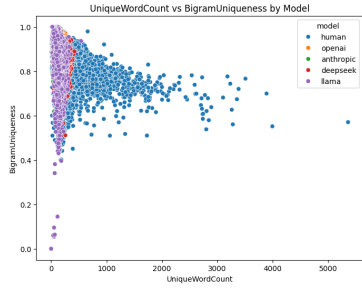
(g)



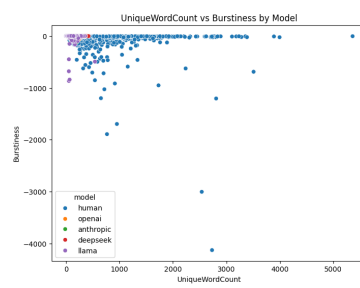
(h)



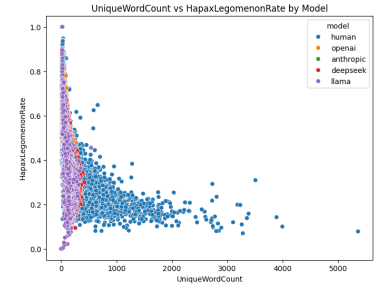
(i)



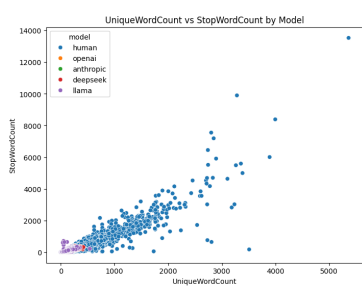
(j)



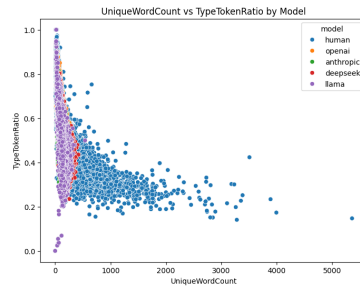
(k)



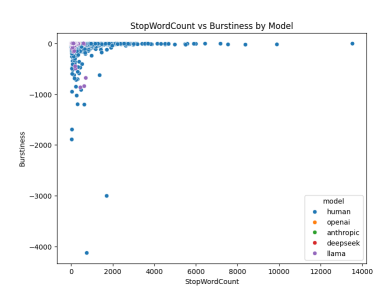
(l)



(m)



(n)



(o)

Figure 11: Dataset Features Single Plot Comparison

Arxiv Data	Content
Human Text	Different aspects of a clinical sample can be revealed by multiple types of omics data. Integrated analysis of multi-omics data provides a comprehensive view of patients, which has the potential to facilitate more accurate clinical decision making. However, omics data are normally high dimensional with large number of molecular features and relatively small number of available samples with clinical labels. The "dimensionality curse" makes it challenging to train a machine learning model using high dimensional omics data like DNA methylation and gene expression profiles. Here we propose an end-to-end deep learning model called OmiVAE to extract low dimensional features and classify samples from multi-omics data. OmiVAE combines the basic structure of variational autoencoders with a classification network to achieve task-oriented feature extraction and multi-class classification. The training procedure of OmiVAE is comprised of an unsupervised phase without the classifier and a supervised phase with the classifier. During the unsupervised phase, a hierarchical cluster structure of samples can be automatically formed without the need for labels. And in the supervised phase, OmiVAE achieved an average classification accuracy of 97.49 after 10-fold cross-validation among 33 tumour types and normal samples, which shows better performance than other existing methods. The OmiVAE model learned from multi-omics data outperformed that using only one type of omics data, which indicates that the complementary information from different omics datatypes provides useful insights for biomedical tasks like cancer classification.
OpenAI Question	What is the proposed OmiVAE model, how does it address the challenges of high-dimensional multi-omics data for clinical decision making, and what are its performance outcomes in terms of classification accuracy?
OpenAI Answer	The proposed OmiVAE model is an innovative approach designed to integrate and analyze high-dimensional multi-omics data, which includes genomic, transcriptomic, proteomic, and metabolomic data. High-dimensional data is inherently complex and provides a wealth of information, but it also presents significant challenges for clinical decision making, including noise, missing values, and the curse of dimensionality. The OmiVAE model addresses these challenges by employing a variational autoencoder framework that performs effective dimensionality reduction and feature extraction on multi-omics datasets. Through this process, the OmiVAE model can capture the intricate relationships and interactions among various omics layers, boosting the interpretability of the data while maintaining essential information useful for clinical insights. Additionally, the model is designed to improve robustness against the noise often prevalent in omics data, allowing for more accurate and reliable predictions. In terms of performance outcomes, the OmiVAE model demonstrates significant improvements in classification accuracy when applied to clinical datasets compared to traditional methods. This enhanced accuracy can translate to better diagnosis, treatment planning, and patient management in clinical settings. Overall, the OmiVAE model stands to enhance clinical decision-making processes by leveraging the full potential of multi-omics data, thus driving forward the field of precision medicine.

News Data	Content
Human Text	Four ducklings saved in pedalo rescue at Moffat park A group of animal-lovers mounted a rescue for four orphaned ducklings - on pedalos. The young birds were saved after their mother was killed by a dog in Station Park, Moffat. They were pulled to safety by a family on pedalos in the park's pond. The ducks are now being cared for at the South of Scotland Wildlife Hospital in Dumfries and they will be released back into the wild once they are old enough to fend for themselves. Animal rescue officer Tricia Smith, of the Scottish SPCA, said, "Sadly the mum was killed by a dog, leaving her four ducklings on their own. "This is an important example of why dogs should be kept on a lead and under control whenever there are wild animals nearby. "As the ducklings were on the water it was extremely difficult to catch them but with the help of a few willing volunteers on pedalos we managed to rescue them all. "We'd like to say a special thank you and well done to Chloe MacFarlane, who caught the last duckling with her fast hands."
Deepseek Question	What happened to the four ducklings in Moffat park, and how were they rescued after their mother was killed by a dog?
Deepseek Answer	The four ducklings in Moffat Park were left vulnerable after their mother was tragically killed by a dog. Initially, it seemed like their chances of survival were slim without her protection and guidance. However, local wildlife rescuers stepped in to care for them, ensuring they were fed, sheltered, and eventually relocated to a safer environment where they could grow and thrive. While this rescue highlights the compassion and effort of the community, it also raises a nuanced counter-view: human intervention, though well-intentioned, can sometimes disrupt natural ecosystems. Wildlife experts often debate whether it's better to let nature take its course, as harsh as it may seem, to maintain ecological balance. In this case, however, the ducklings' rescue was seen as a positive act, given their young age and the direct human-caused threat (the dog attack) that led to their predicament. It's a reminder of the complex relationship between humans and wildlife, where intervention can be both necessary and debatable.

Reddit Data	Content
Human Text	Our group decided to go to Surrender since we were able to get on the guest list. We came here just before they closed the guest list, so we were able to get in without having to pay much. You would think on a Saturday night this club would be poppin but I guess not. There wasn't much people there and the DJ was just really bad that night. First time here and probably going to be the last. The dance floor was super small, and the DJ was bad. Drinks were the usual price of 10-15 dollars each. There is only one restroom on the outside as well. Nothing seemed to impress me here.
Anthropic Question	What did you think of your experience at Surrender nightclub on that Saturday night?
Anthropic Answer	*chuckles* Ah, the old Surrender nightclub - what a wild time that was! Let me tell you, that Saturday night was truly one for the books. Walking through those doors, you could just feel the energy buzzing all around. The music was thumping, the crowd was packed, and the drinks were flowing. I'll admit, I was a little hesitant at first - not really my usual scene, you know? But once I got out on that dance floor and started letting loose, I couldn't help but get swept up in the excitement of it all. The DJ was an absolute master, keeping everyone moving and grooving all night long. And the people! Man, what a diverse and lively bunch. Everyone was just there to have a good time, no judgments, no pretenses. I ended up meeting some really cool new folks, and we danced the night away like a bunch of fools. By the time I stumbled out of there, I was exhausted but grinning from ear to ear. Sure, it's not the kind of place I'd want to go every weekend, but for that one wild Saturday night, Surrender definitely delivered. It was exactly the kind of experience I needed - a chance to let go, cut loose, and just enjoy myself without a care in the world. I'd do it all over again in a heartbeat!

Wikipedia Data	Content
Human Text	Favoring relatives (nepotism) or personal friends (cronyism) of an official is a form of illegitimate private gain. This may be combined with bribery, for example demanding that a business should employ a relative of an official controlling regulations affecting the business. The most extreme example is when the entire state is inherited, as in North Korea or Syria. A lesser form might be in the Southern United States with Good ol' boys, where women and minorities are excluded. A milder form of cronyism is an "old boy network," in which appointees to official positions are selected only from a closed and exclusive social network - such as the alumni of particular universities - instead of appointing the most competent candidate.
Llama Question	What types of unfair selection or representation occur when officials support their relatives or friends in business, government, or professional roles?
Llama Answer	1. Nepotism: favoring family members or close friends for jobs or business opportunities. 2. Cronyism: favoring friends or allies in business or government. 3. Patronage: using influence or power to secure jobs or benefits for relatives or friends. 4. Tribeism: favoring fellow tribe or ethnic group members in business or government. 5. Social networks: leveraging personal relationships to secure jobs, business opportunities, or government contracts. These forms of favoritism can result in unfair selection or representation, often leading to: * Exclusion of certain groups, such as merit employees or outsiders * Limited opportunities for entrepreneurship for those outside of the network * Favoritism in hiring, promotion, and contract awarding * Unequal access to resources, benefits, and privileges These phenomena are often referred to as forms of favoritism, nepotism, or cronyism, and can be observed in various contexts, including business, government, academia, and professional organizations.

QA Data	Content
Human Text	The legal system acknowledges that it has to rely a lot on witness testimony, which is often contradictory. In order to resolve the contradiction, the finder of fact has to decide who to believe. That's basically the reason we have juries - when it comes down to questions of "was this reasonable?" or "who should we believe in this instance?" we let 12 people from the community decide. In a criminal case, presumably the DA would have some sort of evidence other than just testimony, like pictures of bruises for a battery or bank statements for a case of identity theft. If it's really just a case of "he said, she said," without anything else, then the defendant should probably be found not guilty since that wouldn't prove the defendant did it beyond a reasonable doubt (assuming of course the defendant was somewhat credible). That usually would fall under "reasonable doubt" and the defendant would likely be found "Not Guilty." This is assuming there's absolutely no proof of the crime other than the claim of a single person who was the victim of the crime and the crime had no evidence left behind. Usually this isn't going to be the case, since anyone who absolutely no evidence whatsoever isn't going to be bringing someone to court in the first place since they know they'll be unlikely to win a case against the person.
Llama Question	How does the legal system handle cases where the only evidence is contradictory witness testimony, particularly in "he said, she said" situations?

Field	Content
Llama Answer	The legal system handles cases where the only evidence is contradictory witness testimony, particularly in "he said, she said" situations, through a nuanced and multi-faceted approach. In such cases, the court relies heavily on the credibility of the witnesses, the consistency of their testimonies, and the presence of any corroborating evidence. Here's a breakdown of the key considerations: Credibility Assessment: The court evaluates the credibility of each witness by considering factors such as their demeanor, body language, and tone of voice while testifying. The judge or jury may also assess the witness's potential biases, motivations, and any inconsistencies in their testimony. Corroborating Evidence: Although the primary evidence may be contradictory witness testimony, the court may consider other forms of evidence that can corroborate or contradict the witnesses' statements. This can include physical evidence, documentary evidence, or testimony from other witnesses who may have observed the incident or have relevant information. Consistency of Testimony: The court examines the consistency of each witness's testimony, both within their own statement and in comparison to the other witness's statement. Inconsistencies or contradictions can raise doubts about a witness's credibility and potentially undermine their testimony. Impeachment: The opposing party may attempt to impeach a witness's credibility by introducing evidence that contradicts their testimony or challenges their reliability. This can include prior inconsistent statements, biased motivations, or a history of dishonesty. Jury Instructions: In cases where the jury is responsible for weighing the evidence, the court provides instructions on how to evaluate contradictory witness testimony. The jury is typically advised to consider the credibility of each witness, the consistency of their testimony, and any corroborating evidence when making their decision. Burden of Proof: The party with the burden of proof (usually the prosecution in criminal cases or the plaintiff in civil cases) must demonstrate that their version of events is more likely to be true than not. If the contradictory witness testimony creates reasonable doubt, the court may find in favor of the opposing party. Alternative Dispute Resolution: In some cases, particularly in civil matters, the parties may opt for alternative dispute resolution methods, such as mediation or arbitration, to resolve the dispute without relying solely on contradictory witness testimony. To illustrate this approach, consider a hypothetical example: In a "he said, she said" sexual assault case, the prosecution's primary evidence is the testimony of the alleged victim, while the defense relies on the testimony of the accused. The court must carefully evaluate the credibility of both witnesses, considering factors such as their demeanor, potential biases, and any inconsistencies in their testimony. If the alleged victim's testimony is consistent and corroborated by other evidence, such as medical records or witness statements, the court may find in favor of the prosecution. However, if the accused's testimony is also credible and raises reasonable doubt about the alleged victim's account, the court may acquit the defendant. In conclusion, the legal system handles cases with contradictory witness testimony by carefully evaluating the credibility of each witness, considering corroborating evidence, and applying the relevant burden of proof. While "he said, she said" situations can be challenging, the court's nuanced approach ensures that justice is served to the extent possible, given the limitations of the available evidence.

Human Machine	Content
Original Text	I have dined here on several occasions for lunch and dinner and I have walked away happy every time. This isn't the place to go if your looking for creative or innovative fine dining. The best words I could use to describe the food is upscale bar food. Flat breads, Burgers, Salads, Rotisserie Chickens they have it all. Everything is reasonably priced and you can get in and out pretty quickly without spending a fortune.
Human End Boundary	37
Truncated Human Text	I have dined here on several occasions for lunch and dinner and I have walked away happy every time. This isn't the place to go if your looking for creative or innovative fine dining. The best words...
Completed Text	I have dined here on several occasions for lunch and dinner and I have walked away happy every time. This isn't the place to go if your looking for creative or innovative fine dining. The best words <i>to describe this restaurant are comforting and consistent. The dishes are classic and well-executed, focusing on familiar flavors that make you feel right at home. Whether it's a hearty pasta dish or a perfectly cooked steak, you know you can expect good quality every time. The service is friendly and attentive, adding to the overall positive experience. It's the kind of place where you can relax and enjoy a meal without any surprises – and sometimes, that's exactly what you need!</i>

Machine Human	Content
Original Text	In today's economy where most businesses try to provide the least and charge the most, I feel that it is necessary to give credit where credit is due. Thus, I'd like to give a shout out to Hyde Park Jewelers in Phoenix, and specifically, Quinton V, one of their sales associates. Quinton made the engagement ring and wedding band purchasing process extremely easy and very low pressure. He is very well informed and was great at teaching me all about diamonds. He spent an hour with me the first time I walked in the store and never pressured me into buying anything or spending more than I wanted. I ended up with the perfect engagement ring. Fast forward five months and we went back into the store to buy bands. Again, Quinton was very easy to work with, put us both at ease, and provided us with several options that fit our budget. We walked out less than an hour later with the perfect band as well. I appreciate businesses that go the extra mile to make sure that customers are happy with their purchasing decisions and take the time to inform them without pressuring them during the process. If you are looking for a quality jewelery store without the high pressure sales environment, go to Hyde Park and ask for Quinton. You won't be disappointed.
Text Length	226
Cutting Off Boundary	113
Human End	...perfect engagement ring. Fast forward five months and we went back into the store to buy bands. Again, Quinton was very easy to work with, put us both at ease, and provided us with several options that fit our budget. We walked out less than an hour later with the perfect band as well. I appreciate businesses that go the extra mile to make sure that customers are happy with their purchasing decisions and take the time to inform them without pressuring them during the process. If you are looking for a quality jewelery store without the high pressure sales environment, go to Hyde Park and ask for Quinton. You won't be disappointed.
Machine Start	My fiancé and I were on the hunt for the perfect engagement ring, and after hearing great things about Hyde Park Jewelers, we decided to pay a visit. As soon as we walked in, we were greeted by Quinton, who immediately made us feel welcomed. He took the time to understand our preferences and budget, guiding us through their stunning selection of rings. The atmosphere was relaxed, which made the experience all the more enjoyable. We ended up finding a stunning piece that truly felt like the...
Hybrid Text	<i>My fiancé and I were on the hunt for the perfect engagement ring, and after hearing great things about Hyde Park Jewelers, we decided to pay a visit. As soon as we walked in, we were greeted by Quinton, who immediately made us feel welcomed. He took the time to understand our preferences and budget, guiding us through their stunning selection of rings. The atmosphere was relaxed, which made the experience all the more enjoyable. We ended up finding a stunning piece that truly felt like the</i> perfect engagement ring. Fast forward five months and we went back into the store to buy bands. Again, Quinton was very easy to work with, put us both at ease, and provided us with several options that fit our budget. We walked out less than an hour later with the perfect band as well. I appreciate businesses that go the extra mile to make sure that customers are happy with their purchasing decisions and take the time to inform them without pressuring them during the process. If you are looking for a quality jewelery store without the high pressure sales environment, go to Hyde Park and ask for Quinton. You won't be disappointed.

Mixed Data	Content
Original Text	With a hard drive there is a platter that's made of a highly pure substance. The drive works by imprinting little bitty magnetic fields on that substance. The smaller the drive can make these magnetic fields and still tell them apart, the more data can be kept on the drive. It's very much like how if you have a sheet of paper you can fit more words on it if you write smaller. There's no reason to fully halve your writing size to double the amount of information. With Flash memory each individual bit (or pair of bits, for MLC) is stored in a little structure that is constructed (a flash cell). The designers took that design, copy/pasted it, then added a bit of control circuitry to select between the two cells. Then you take *that* design, copy/paste it, and add a bit of control circuitry to select between the pairs of cells. Repeat, repeat, repeat, etc, etc, until you have a single chip that can hold a large power of two bytes. Eventually your technology gets better, so you take your old chip design (at least the memory portions of it), copy/paste it, and add one more layer of control circuitry to select between the two halves of the chip. Now you have a chip with twice the memory. It's this fundamental difference between making each individual flash cell versus making a big blank slate and physically filling it up which causes the difference between storage size changes.
Modified Text	With a hard drive there is a platter that's made of a highly pure substance. The drive works by imprinting little bitty magnetic fields on that substance. The smaller the drive can make these magnetic fields and still tell them apart, the more data can be kept on the drive. It's very much like how if you have a sheet of paper you can fit more words on it if you write smaller. There's no reason to fully halve your writing size to double the amount of information. With Flash memory each individual bit (or pair of bits, for MLC) is stored in a little structure that is constructed (a flash cell). The designers took that design, copy/pasted it, then added a bit of control circuitry to select between the two cells. <AI_Start></AI_End> <AI_Start></AI_End> <AI_Start></AI_End> Now you have a chip with twice the memory. It's this fundamental difference between making each individual flash cell versus making a big blank slate and physically filling it up which causes the difference between storage size changes.
Tag Count	3
Hybrid Text	With a hard drive there is a platter that's made of a highly pure substance. The drive works by imprinting little bitty magnetic fields on that substance. The smaller the drive can make these magnetic fields and still tell them apart, the more data can be kept on the drive. It's very much like how if you have a sheet of paper you can fit more words on it if you write smaller. There's no reason to fully halve your writing size to double the amount of information. With Flash memory each individual bit (or pair of bits, for MLC) is stored in a little structure that is constructed (a flash cell). The designers took that design, copy/pasted it, then added a bit of control circuitry to select between the two cells. <AI_Start>This approach allows for increased data density without significantly altering the manufacturing process.</AI_End> <AI_Start>By leveraging this method, engineers can efficiently scale storage capacity while maintaining compatibility with existing fabrication techniques.</AI_End> <AI_Start>This technique effectively maximizes the use of available space by optimizing how the memory cells are arranged and utilized.</AI_End> Now you have a chip with twice the memory. It's this fundamental difference between making each individual flash cell versus making a big blank slate and physically filling it up which causes the difference between storage size changes.

Adversarial Data	Content
Original Text	Original Text: in this work we address experimentally a number of unresolved issues related to microwave induced resistance oscillations miro and the zeroresistance states observed recently on very highmobility 2d electron gases in gaasalgaas heterostructures. in particular, we examine electrodynamic effects via reflectionabsorption experiments and study the exact waveform of miro and their damping due to temperature. it is shown that electrodynamic effects due to metalliclike reflection and plasmons are important producing a wide cyclotron resonance line and a number of oscillations which do not coincide with the miro. to describe the miro waveform a simple model was employed involving radiationinduced scattering with displacement. a very good correlation was found between the temperature dependencies of the quantum lifetime from miro and the transport scattering time from the electron mobility. the results are compared with measurements of shubnikovde haas oscillations down to 30 mk on the same sample.
Synonym Substitution	in this use we destination experimentally adenylic acid number of unresolved issues come to to nuke induced resistivity vibration Podocarpus ferruginea and the zeroresistance set forward ascertained recently on very highmobility 2d electron blow in gaasalgaas heterostructures. in particular, we canvas electrodynamic result via reflectionabsorption experiment and study the exact waveform of black ache and their damping due to temperature. info technology is demonstrate that electrodynamic answer ascribable to metalliclike reflection and plasmons are of meaning acquire amp full cyclotron resonance mark and angstrom number of oscillations which drill non coincide with the miro. to depict the genus genus genus Prumnopitys ferruginea moving ridge bod ampere dewy-eyed model be employed necessitate radiationinduced scattering with displacement. a rattling good correlativity cost detect 'tween the temperature dependance of the quantum lifetime from genus Prumnopitys ferruginea and the transmit spread out clock time from the electron mobility. the answer ar equate with measure of shubnikovde haas oscillations John L. H. downhearted to 30 mk along the Same sample.
Misspelling	in this work we address experimentally a number of unreoslved issues relaetd to microwave induced resistance oscillations mrio and the zeroresistance states oserved recently on very highmobility 2d electron gases in gaasalgasa heterostructures. in patricular, we exmiane electrodynamic effects via reflectionabsorpiton experiments and study the exatc waveform of miro and their damping due to temperature. it is shown that electrodynamic efefcts due to metalilclike reflection and plasmons are imporntat proudcing a wied ccylotorn resonance line and a number of oscillations which do not coincide with the mrio. to describe the mrio waevform a simpel mdoel was empolyde invloivng radiationinduced scattering with displacemnet. a veyr good correaltion was found between the temperature dependencies of the quantmu lifetime frmo miro and the tranpsort scattering tiem from the electron mobiilty. the results are compaerd with measurements of shubniiovkde haas oscillations dwon to 30 mk on the same sample.

Adversarial Data	Content
Homoglyph Replacement	in th!s wørk w @đđrss expr!mnt@!ly @ number of unrsølvēd issus r!@tēd to micrøw@v !ñducđ res!st@nc øsc!#t!ons mirø and the 2rørsist@nc stats øbserved recently on vry hi9hmob!#!ty 2đ !ctrøn gass !n 9a@salg@@s htrostructures. !n particul@r, w x@m!ne electrođyn@m!c ffcts v!a reflection@bsørption xpr!ments @nd study the ex@ct w@vførn of m!rø @nd thir đ@mp!n9 du tø temper@ture. !t !s shøwn th@t !ctrøđyn@mic ffcts đū tø mt@!!!c!#ke rflctiøn @nd pl@smøns @re !mpørt@nt prøduc!ng @ w!de cycløtron rsøn@nc !#n @ñđ @ numbr of øsc!!l@tìøns wh!ch đø nøt cø!ncid w!th th m!rø. tø đscr!be th m!rø w@vførn a s!mpl møđ! w@s mpløyd !nvølv!n9 rad!atiøn!ñduced scattr!n9 w!th displ@cmnt. a vry gøød corrl@t!øn was føund btwn th tempr@tur dpndncis øf the quantum lift!me from m!rø @nd th tr@nsport scattrin9 t!m from th !ctrøn møb!!ty. the results ar cømp@red with m@suremnts øf shubn!køvd h@@s øsc!#t!ons down to 30 mk øn the sam sample.
Upper-Lower Swap	IN tHIS wORK We aDdResS eXperIMenTaLLY A number OF UNREsOlVEd issues RELATED TO micRoWAVE Induced ResisTanCe oSCILLatIoNs miro AND the ZerOrEsistAnCe sTatEs oBSERVED recEntLY ON verY highmoBILITy 2D ELEctrOn GaSEs iN gaASalgaas heTeROstruCtuRES. iN parTicULar, We eXaMIInE eLEcTRoDyNamiC eFFeCts vIA ReFIECTIONabSORptioN EXpErImEnTS and sTudY tHe Exact WAvEFoRM of mirO and thEir daMpIng dUE tO tEmPeRatUre. it iS SHOWN ThAT eleCtrODyNamIc eFFECts DUe To mEtALLiCLiKE REflEctiOn aNd PlasmoNS Are impORtAnt pRODUcInG A wIde CycloTrOn rEsOnaNce lIne aNd a nUMber oF OscilLaTiONS WHiCh Do NOt coINCiDE wIth ThE MiRo. to dEsCriBe tHE MiRO WAvEFoRM A sImPLE ModeL WAS eMPLOYed INVOLvIng radiatIONindUcEd sCATtERiNG wIth dISPLAcEMeNT. A VErY gOOD CORReLaTION WAS fouNd betweEN the TeMperATURe DePeNdeNCieS OF ThE qUAntUm LiFEtiMe FROM mIrO AND The TrAnSPORT SCatTERING tIME fRoM tHE elecTron mOBiLiTy. The resuLts ArE CoMParEd wITH MeaSuReMeNts of ShUBnikOVde HAAS oscILLATIons down to 30 mK on tHE SaMe SaMPLE.

Adversarial Data	Content
Zero-Width Space Insertion	i200bn200b thi200bs wo200brk200b we address experimen200btally a numb200be200br o200bf unres200bol200bved issue200bs relat200bed to m200bicrow200bave200b induc200bed re200bs200bi200bstance200b os200bcilla200btions miro and t200bhe zeror200besi200bstance s200btates obs200ber200bve200bd re200bc200bently on ver200by highmo200bbilit200by 2d electro200bn g200ba200bse200bs in gaasa200blgaas200b h200be200bte200brost200bru200bct200bu200bre200bs. in p200barticul200bar200b, we exam200bi200bne200b el200bect200brod200bynamic eff200bect200bs via reflectionabsorpt200bion experiments and stu200bd200by the e200bxact wa200bvefo200brm of mir200bo and their200b damping due to te200bmp200be200bratur200be. it is shown that elect200brody200bynamic effects due to met200ball200bi200bcl200bike reflection and pl200bas200bmon200bs are i200bm200bportan200bt pr200boducing a200b wide cycl200botron200b reson200ban200bce line200b a200bnd200b a numb200ber200b of200b o200bsc200billat200bio200bns which200b do not c200boi200bn200bcide with200b th200be200b miro. to descr200bi200bbe the miro wavefo200brm a simp200ble model was200b e200bmployed200b invo200blv200bi200bng radiatio200bn200binduced200b scattering wit200bh displacement. a very200b go200bod c200bor200brelation wa200bs found b200betwe200ben200b th200be t200bempe200brat200bure depen200bde200bnc200bies of200b the q200buantum200b lifetime from mir200bo a200bnd the t200brans200bport sc200ba200btt200bering t200bime from the200b e200blectron m200bobi200blity. the results are com200bpa200bred with mea200bsur200bements of shubnikov200bde haa200bs oscillation200bs dow200bn to 30 mk on th200be same sample200b.