

EDMAL - ENHANCED DETECTION OF AI GENERATED TEXT USING MACHINE LEARNING

Ang Jun Ray

Introduction

Recent developments in Large Language Models (LLMs) like ChatGPT are problematic for several reasons, which include malicious use in unsafe applications like generating fake news [1]. It could create tensions between values, potential harm [2] and much more. As such, it is crucial that people are aware what content is AI-generated, to safeguard themselves from the dangers AI-generated content poses. Therefore, we aimed to enhance existing detection of AI generated text by focusing on and enhancing an approach titled: “DNA-GPT: Divergent N-Gram Analysis For Training-Free Detection Of GPT-Generated Text” [3]. Building upon the empirical observation that “given appropriate preceding text, LLMs tend to output highly similar text across multiple runs of generations” [3], we devised and tested 4 different methods to improve DNA-GPT’s current approach.

Methodology

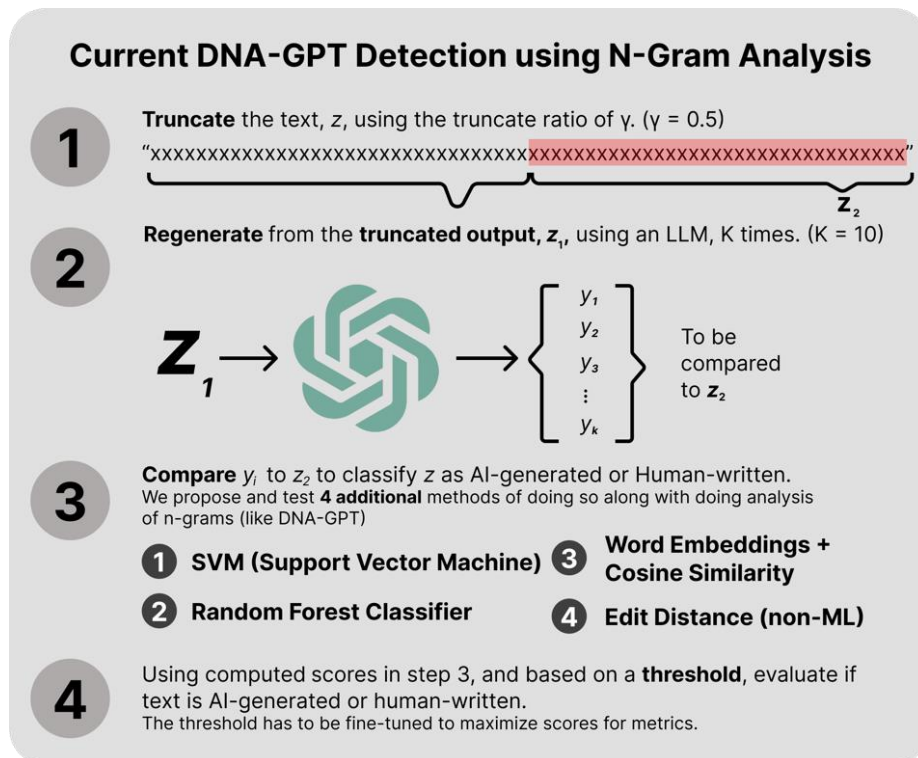


Figure 1: An infographic showcasing DNA-GPT’s current approach to detection, and our plans to enhance it. (Adapted from Figure 1 in DNA-GPT’s paper [3].)

As seen in Figure 1, DNA-GPT’s current approach involved a 4-step detection process. We aim to improve this by using 4 different methods to approach the third step, where y is compared to z_2 .

Task Definition

One Model, Baseline and Detection Scenario. Due to time constraints, we were only able to evaluate our approach’s performance on 1 model during regeneration¹, OpenAI’s GPT-3.5-Turbo [6]. This model demonstrates decent performance on multiple benchmarks, while remaining efficient and cost-effective [7]. Furthermore, we established our baseline as DNA-GPT’s original detection approach [7] and compared our results with theirs as a sanity check. Finally, we focused our efforts on Black-Box Detection. In the real-world scenario, it is more likely that these Detection Algorithms will be used by external sources who have no access to the closed-source LLMs used to generate text. As such, we felt that it is much more crucial that we focus on Black-Box Detection.

Two Metrics. To align with DNA-GPT [3] and previous studies focused on AI text-generation detection [8], we focused on Area Under The ROC Curve (AUROC) score for evaluating detection algorithms’ effectiveness and TPR (True Positive Rate) scores at a fixed 1% FPR (False Positive Rate) [9]. The TPR at 1% FPR metric could further ensure the reliability of detection algorithms for real-life deployment since it is crucial to maintain a high TPR while minimizing the FPR. [3]

Three Datasets². Firstly, we pruned and filtered the Reddit long-form question-answer dataset from the ELI5 community’s [4] texts to have a minimum of 500 words. After generating the AI section of the dataset with the questions provided by the ELI5 dataset, we evaluated the methods’ performance on ELI5, which DNA-GPT used. We evaluated on this data set to ensure we can achieve similar results to those published in DNA-GPT [3].

¹ Embeddings-Ada2 by OpenAI was used when testing our 3rd method, cosine similarity. Since it was not used during regeneration, it was not included above.

² More information about the dataset is available in Appendix A.

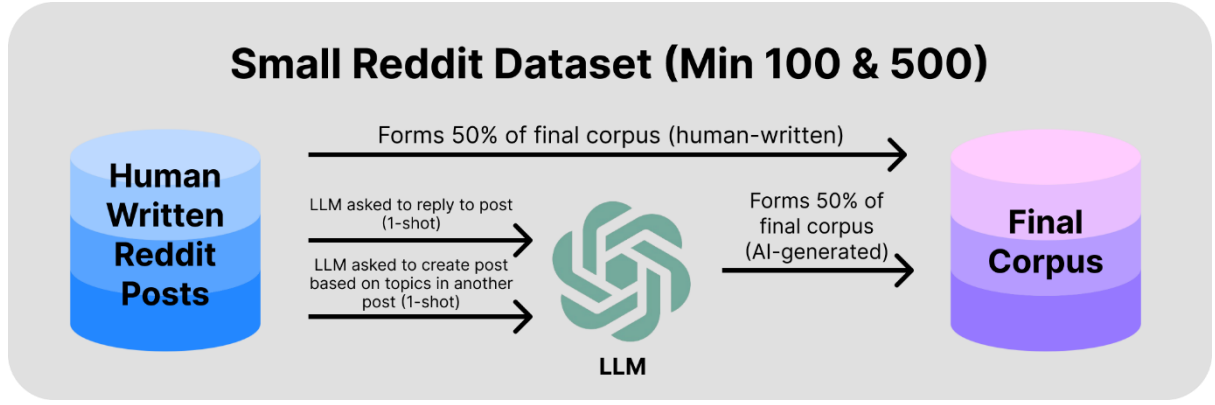


Figure 2: An infographic showcasing how the 2 Reddit Small Datasets used, Reddit-Minimum-500 and Reddit-Minimum-100 were formulated.

Next, as seen in Figure 2, we collected 2 other datasets based on the Reddit Small Corpus. [5] This process involved combining human-written Reddit posts with AI-generated ones, the latter being produced through the generation of 1-shot replies and new posts. Using this method, we generated 2 datasets differing in minimum word count. This was done to test if the performance could be significantly affected by the amount of context provided to the LLM during regeneration, in step 2.

Four Methods. We targeted the third step in DNA-GPT’s original detection approach. Other than n-gram analysis, we propose and experiment with 3 other machine-learning methods and 1 ad hoc method for the same task. The methods are grouped into 2 groups as follows:

Group 1: Extensions of N-Gram Analysis

In DNA-GPT, n-gram scores are calculated for all $1 \leq N \leq 25$. The final score is defined by the following function, where x is defined as a vector of the n-gram scores for all n-grams:

$$\text{score}_z = f(x) = \sum_{i=1}^{25} x_i \cdot i \log(i)$$

The goal of function $f(x)$ is ultimately to calculate a score such that all score_z can, as accurately as possibly, be classified into 2 distinct groups, AI-generated and human-written. We aim to use 2 machine learning-based methods, SVM (support vector machines) and the random forest classifier to replace $f(x)$. Since these require machine learning, we had to split the data with a train-test split ratio of 80%. Furthermore, we tuned the hyperparameters³ of the models to optimize its performance.

³ More information about the hyperparameter tuning results can be seen in Appendix B

1. SVM (support vector machines)

An SVM finds the optimal hyperplane that best separates data points of the 2 classes in a high-dimensional space, while maximize the margin between classes, enhancing its generalization ability. In this case, the SVM is trained by taking feature vector x as input and returning a label of 1 (AI-generated) or 0 (Human-written) as $score_z$.

2. Random forest classifier

A random forest classifier builds multiple decision trees during training and outputs a class based on the mode of the classes. It combines the strengths of multiple decision trees while mitigating overfitting. In this case, we also trained it by taking feature vector x as input and returning a label of 1 (AI-generated) or 0 (Human-written) as $score_z$.

Group 2: Alternative Approaches

These methods aim to provide $score_z$ without the use of N-Gram Analysis.

1. Word Embeddings with Cosine Similarity

Word embeddings aim to represent text with a vector such that its semantic meaning is captured. Embeddings were calculated for all y_i and z using OpenAI's Embeddings-Ada2 API [6]. To accurately measure the similarity of y_i and z , we harnessed cosine similarity to measure the angle between their embeddings:

$$Sim(S_a, S_b) = \cos(\theta) = \frac{\vec{s}_a \cdot \vec{s}_b}{\|S_a\| \cdot \|S_b\|} [10]$$

After obtaining their cosine similarity score, we took the mean of the cosine similarity scores as $score_z$.

2. Edit Distance (Ad hoc)

The edit distance represents the minimum number of transformations required to convert the string from S_a to S_b . [10] We tested the levenshtein edit distance, which is defined mathematically as:

$$lev_{a,b}(i,j) = \begin{cases} \max(i,j) & \text{if } \min(i,j) = 0 \\ \min \begin{cases} lev_{a,b}(i-1,j) + 1 \\ lev_{a,b}(i,j-1) + 1 \\ lev_{a,b}(i-1,j-1) + 1_{a_i \neq b_j} \end{cases} & \text{otherwise} \end{cases}$$

While this does not use machine learning, we tested it nonetheless as it is like that of the N-Gram Analysis that DNA-GPT used. Both methods analyse the lexical structure of the text literally, and thus would be an interesting comparison.

Results and Discussion

Here, we analyse the results⁴ by comparing each method's effectiveness on all datasets:

Datasets	ELI5 (Min 500)		Reddit Small (Min 500)		Reddit Small (Min 100)	
Method	AUROC	TPR	AUROC	TPR	AUROC	TPR
DNA-GPT (original)	96.85	63.50	-	-	-	-
DNA-GPT	98.07	59.08	88.32	8.06	71.50	1.36
Random Forest	97.20	61.04	87.60	12.16	71.83	2.89
SVM	97.91	56.78	84.09	8.04	71.62	2.64
Cosine Similarity	90.75	38.41	58.07	1.11	57.58	1.13
Edit Distance	95.45	34.19	77.29	3.96	60.58	0.78

Table 1: Scores for AUROC and TPR metrics for all 3 datasets and 5 (including DNA-GPT's original score) methods, with the best-scores (with a margin of 1%) bolded.

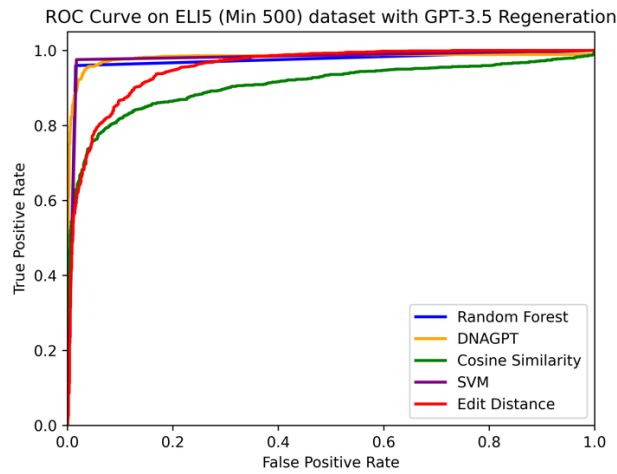
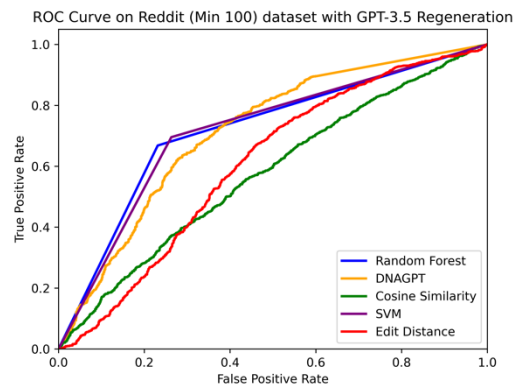
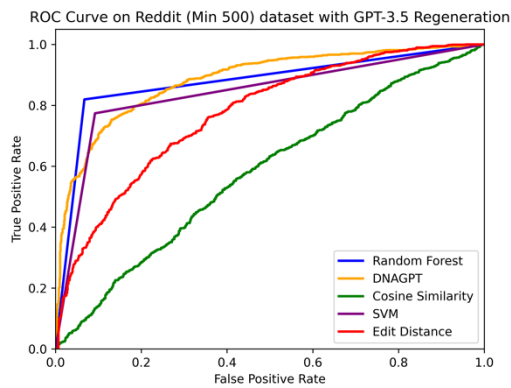


Figure 3: ROC Curve of all 4 methods and DNA-GPT's method for the **ELI5** dataset with a minimum of **500** characters.



⁴ For qualitative samples of the results, view Appendix C. For graphs of how each method scored each sample with different datasets, view Appendix D.

Figure 4: ROC Curve of all 4 methods and DNA-GPT’s method for the **Reddit Small** dataset with a minimum of **500** characters.

Figure 5: ROC Curve of all 4 methods and DNA-GPT’s method for the **Reddit Small** dataset with a minimum of **100** characters.

DNA-GPT: We were able to closely match the results provided in DNA-GPT’s paper for ELI5 [3] with a mean difference of about 2.82 for both TPR and AUROC. We think this is due to the differences in the pruning and generation of the ELI5 dataset. While our methods aimed to improve DNA-GPT’s performance, its original method proved to be extremely competitive with the rest of our suggested methods. We think this is because DNA-GPT’s original function, $f(x)$, to calculate $Score_z$, was rather accurate when using a logarithmic function to predict the importance of each N-Gram when calculating $Score_z$.

Random forest classifier and SVM: As seen in Table 1, across all 3 datasets, the random forest classifier method consistently performs the best among the other 5 methods. The SVM (support vector machines) performs slightly worse compared to the random forest classifier, which likely occurred as the random forest classifier relies on multiple models and can better correlate the scores of different N-grams. When comparing these 2 methods to DNA-GPT, random forest classifier does perform on par or better than DNA-GPT in multiple areas, while SVM’s performance is inferior. However, as these methods require training, and we used a rather humble dataset size⁵, we are confident that if we used a significantly larger dataset, both models’ performance would be able to far exceed that of DNA-GPT.

Cosine similarity with Word Embeddings: Next, regarding the cosine similarity method, it performed significantly worse than SVM, random forest classifier and DNA-GPT. This is likely because the semantic meaning of y_i is very likely to match z_2 , with z_1 as context, irrelevant of whether z is human-written or AI-generated. However, lexically, y_i is likely to differ from z_2 , which is not considered by this method.

Edit distance and DNA-GPT: Comparing the edit distance method to DNA-GPT’s N-Gram Analysis, we can conclude that edit distance was much less effective. Unfortunately, 2 words of similar lexical nature, like “Stationary” and “Stationery”, are drastically different in

⁵ Dataset size information is available in Appendix A.

semantic meaning. Thus, edit distance treats them as extremely similar words, while they are semantically completely different. Since in N-Gram Analysis, they are treated as two completely different words, it avoids the issues edit distance faces. This explains the drastic difference in performance between the two methods.

Limitations and Future Work

When comparing Figure 3 and 4, while both datasets had the same restrictions in length, performances for all methods are significantly better for ELI5. As the Reddit dataset contains random posts, while the ELI5 dataset contains essays attempting to explain concepts thoroughly, the ELI5 dataset is much more consistent in nature. This highlights a significant flaw in Step 2, in that it requires text to be highly consistent. When comparing Figure 4 and 5, we can infer that performances for all methods were significantly impacted by the length of the text. As seen in Table 1, regarding our best performing method, the random forest classifier, its performance decreased by 15.77 for the AUROC metric, while decreasing by a significant 9.27 for the TPR metric. Thus, Step 2 requires a large amount of context for AI detection to be accurate. These 2 major flaws identified by our results could be looked into as future work.

Conclusion

Through testing of the various methods, we can conclude that it is beneficial to use machine learning to better interpret the N-Gram Scores, like the random forest classifier and SVM did. Furthermore, we were able to generally outperform DNA-GPT with the random forest classifier. Through the addition of significantly more training data and tuning of both methods, we are confident that their results could be significantly improved. In addition, the mediocre performance of cosine similarity and edit distance shows that the lexical nature of y_i and z_2 is significantly more important than that of the semantic nature in Step 3. Finally, our results also revealed significant limitation in Step 2, which could be explored in future work.

References

- [1] Węcel, K., Sawiński, M., Stróżyna, M., Lewoniewski, W., Księżniak, E., Stolarski, P., & Abramowicz, W. (2023). Artificial intelligence—friend or foe in fake news campaigns. *The Poznań University of Economics Review*, 9(2).
- [2] Samuel Gehman, Suchin Gururangan, Maarten Sap, Yejin Choi, and Noah A. Smith. 2020. RealToxicityPrompts: Evaluating neural toxic degeneration in language models. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 3356–3369, Online. Association for Computational Linguistics
- [3] Yang, X. (2023, May 27). DNA-GPT: Divergent N-Gram Analysis for Training-Free Detection of GPT-Generated Text. *arXiv.org*.
- [4] Angela Fan, Yacine Jernite, Ethan Perez, David Grangier, Jason Weston, and Michael Auli. ELI5: Long form question answering. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 3558–3567, Florence, Italy, July 2019. Association for Computational Linguistics. doi: 10.18653/v1/P19-1346. URL
- [5] The ConvoKit Developers. (n.d.). Reddit Corpus (small) [Dataset]. In *ConvoKit*.
- [6] OpenAI. (n.d.). *OpenAI API* [Software].
- [7] Emmanuel, C. (2023, August 3). GPT-3.5 and GPT-4 comparison: - Chude Emmanuel - Medium. *Medium*.
- [8] Eric Mitchell, Yoonho Lee, Alexander Khazatsky, Christopher D Manning, and Chelsea Finn. Detectgpt: Zero-shot machine-generated text detection using probability curvature. *arXiv preprint arXiv:2301.11305*, 2023.
- [9] Kalpesh Krishna, Yixiao Song, Marzena Karpinska, John Wieting, and Mohit Iyyer. Paraphrasing evades detectors of ai-generated text, but retrieval is an effective defense. *arXiv preprint arXiv:2303.13408*, 2023

- [10] Wang, J., & Dong, Y. (2020). Measurement of text similarity: a survey. *Information*, 11(9), 421.
- [11] scikit-learn: machine learning in Python — scikit-learn 1.3.2 documentation. (n.d.).

Appendix A – Dataset Information

Dataset	Source of Human Section	Source of AI Section	Min Word Count	Size
ELI5 (Min 500)	ELI5 Huggingface	GPT-3.5 generated responses to questions in the ELI5 dataset	500	2986
Reddit Small (Min 500)	Reddit Small Corpus	1-shot replies to posts in Reddit Small Corpus and	500	2156
Reddit Small (Min 100)	Reddit Small Corpus	1-shot new posts to topics from posts in Reddit Small Corpus	100	2500

For methods which involved machine learning, a test-train split of 80% was used.

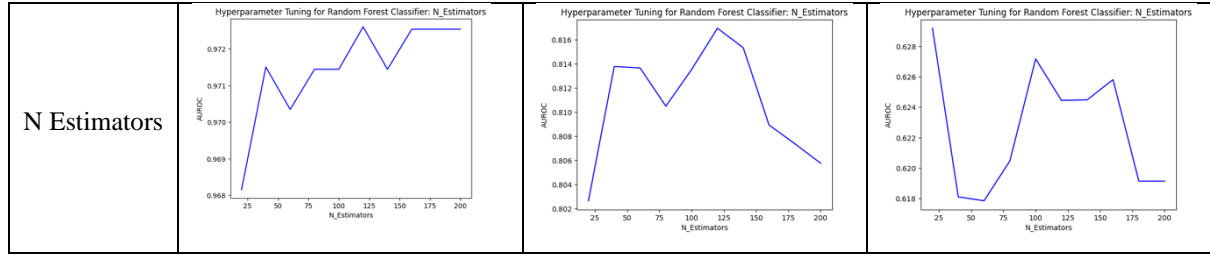
Appendix B – Hyperparameter Tuning

This applies only for the random forest classifier and SVM methods, which require machine learning.

To create and train both models, we harnessed SKLearn’s [11] random forest classifier and SVM classes. These have several built-in hyperparameters that could be fine-tuned to optimise performance on each dataset. More info about these hyperparameters can be found in SKLearn’s documentation. Unfortunately, due to time constraints, we were only able to tune a few selected important hyperparameters.

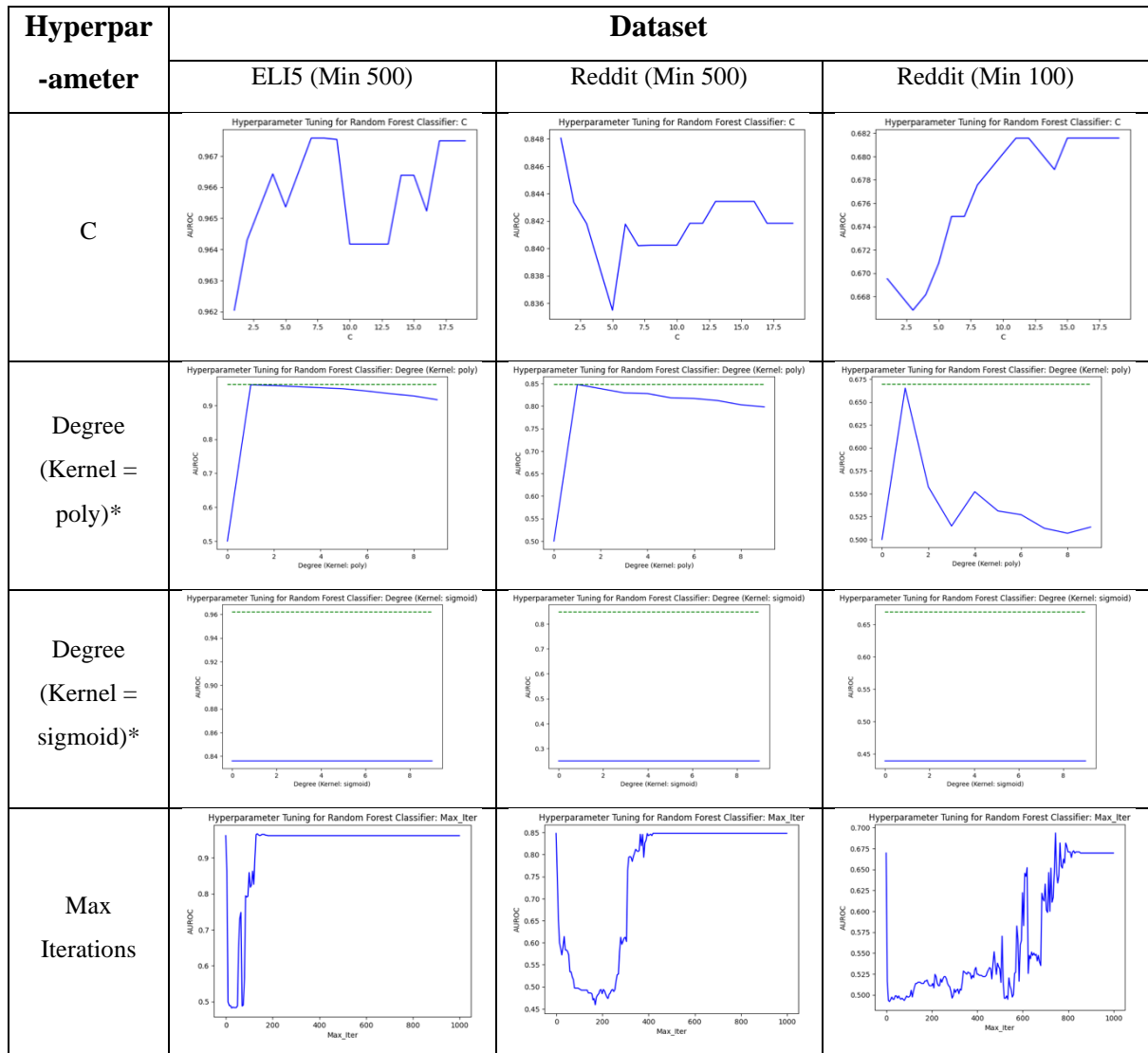
In the tables below, we provide graphs of how the AUROC metric performed at different values for each hyperparameter. For the random forest classifier method:

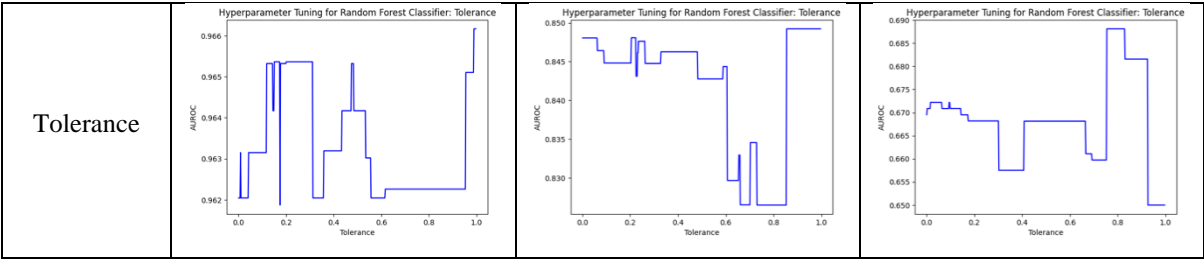
Hyperparameter	Dataset		
	ELI5 (Min 500)	Reddit (Min 500)	Reddit (Min 100)
Max Depth			
Max Leaf Nodes			
Min Samples Leaf			
Min Samples Split			



For the SVM method:

*For the hyperparameter degree, we tuned it based on whether the kernel is set to sigmoid or poly. The green-dashed line refers to the AUROC score when the kernel is set to the default rbf kernel.





After generating the models’ performance on AUROC for the hyperparameters, we set the optimal value for each hyperparameter, to optimize the models’ performance.

Appendix C – Qualitative Samples

Here, we present multiple samples across all 3 datasets:

**Note: The sample numbers refer to that of the AI or Human section of the dataset, and not the entire dataset. In the case of methods using machine learning, the sample number applies to the test section of the dataset. Furthermore, final classifications for non-classifier-based methods are not provided, as it depends on FPR we are targeting.*

Details	Original Text (highlighted text represents z ₂)	Regenerated Text (1 sample from K = 10, $\gamma = 0.5$)
Dataset: ELI5 (Min 500) Sample Number*: 187 Type of Original Text: AI-generated Score_Z: 1 (predicted as human) Prediction Method Used: Random Forest Classifier	<p>no, genetic analysis has not refuted or overturned any significant widely held historical theories. however, it has certainly provided new insights and additional evidence that have led to the revision and refinement of some historical theories.</p> <p>genetic analysis, also known as genetic testing or dna analysis, involves studying an individual's genetic material to understand their ancestry, genetic traits, and potential health risks. it has become a powerful tool in various fields, including anthropology, archaeology, and history. by analyzing ancient dna from bones, teeth, and other remains, researchers can gain valuable information about human migration patterns, population movements, and relationships between different groups of people.</p> <p>one area where genetic analysis has made significant contributions is in the study of human origins and migration. prior to the advent of dna analysis, historians and archaeologists relied on linguistic, archaeological, and written evidence to reconstruct the history of human populations. however, genetic analysis has provided a new layer of evidence that has enriched our understanding of human migration and population movements.</p> <p>for example, prior to genetic analysis, the prevailing theory suggested that modern humans originated in africa and migrated to other parts of the world in a single wave around 60,000 years ago. however, genetic studies conducted over the past few decades have revealed a more complex picture. these studies have found evidence of multiple waves of migration out of africa, with some groups interbreeding with archaic humans like neanderthals and denisovans.</p> <p>another significant contribution of genetic analysis to historical understanding is the study of human admixture. admixture refers to the mixing of genetic material between different populations. through genetic analysis, researchers have been able to identify instances of admixture between different groups of people, which has challenged previous assumptions about human populations being isolated and homogeneous.</p> <p>for instance, genetic analysis has shown that modern humans of non-african descent carry a small percentage of neanderthal dna, indicating interbreeding between early humans and neanderthals when they encountered each other in eurasia. this finding not only challenges the notion of human "purity" but also provides evidence for interactions and interbreeding between different hominin species.</p> <p>moreover, genetic analysis has shed light on historical events and population movements that were previously poorly understood or lacking in historical records. for example, the genetic analysis of ancient skeletons discovered in mass gravesites has helped identify victims of past epidemics, such as the black death in medieval europe. by comparing the dna of these individuals to modern populations, researchers can trace the genetic legacy of these devastating pandemics.</p> <p>in conclusion, while genetic analysis has not refuted or overturned any significant widely held historical theories, it has certainly provided new insights and evidence that have refined and expanded our understanding of human history. by analyzing ancient dna, geneticists have contributed to the study of human origins, migration patterns, admixture, and historical events. this interdisciplinary approach, combining genetics with traditional historical methods, has allowed researchers to paint a more nuanced and accurate picture of our shared past.</p>	<p>historical diseases and their impact on human populations. by analyzing ancient dna, scientists have been able to identify the presence of certain genetic variants associated with diseases such as tuberculosis, leprosy, and malaria in ancient human remains. this has provided insights into the spread and evolution of these diseases throughout history.</p> <p>genetic analysis has also shed light on the origins and migrations of specific populations. for instance, studies have revealed the genetic legacy of the mongol empire and its impact on the populations of central asia and europe. similarly, genetic analysis has helped trace the ancestry of indigenous peoples in the americas, providing evidence for their migration patterns and connections to other populations.</p> <p>in addition to these contributions, genetic analysis has also been used to investigate historical individuals and their relationships. for example, the dna analysis of the remains of king richard iii of england confirmed his identity and provided insights into his genetic traits and predispositions.</p> <p>it is important to note that while genetic analysis has provided valuable insights into historical questions, it is just one piece of the puzzle. it should be used in conjunction with other forms of evidence, such as archaeological and written records, to create a comprehensive understanding of the past.</p> <p>in conclusion, genetic analysis has not refuted or overturned widely held historical theories, but it has certainly enhanced our understanding of human history. by analyzing ancient dna, researchers have gained new insights into human migration patterns, the spread of diseases, the origins of specific populations, and the relationships between historical individuals. however, it is crucial to integrate genetic analysis with other forms of evidence to create a more complete and accurate picture of the past.</p>
Dataset: Reddit Small (Min 500) Sample Number*: 546 Type of Original Text: human-written Score_Z: 1 (predicted as human) Prediction Method Used: SVM	<p>okay, fine.. if we want to make exceptions and move the goal posts.. i'm not making exceptions. i guess technically its moving the goal posts, but i don't see how setting the goal to be "literally the most popular thing" is reasonable. who said god of war was a failure? you are literally the only one in this thread that used the words "failure" to describe (or not describe) gow. ea said cancelling the star wars game was an "easy decision", and that: "as we kept reviewing the game, it continued to look toward a much more linear game [which] people don't like as much today as they did five years ago or ten years ago." translating here, they said that single player games aren't economically viable. again, a single player game is the best selling game of the year so far. i'm just saying they're absolutely still economically viable. the problem is with ea and their dogshit games. not with the market moving away from split player titles. they just want to put forth the minimum possible effort. buy a big license, slap together some shitty half - assed game, and then sell it and corresponding microtransactions because people will buy something based off the license.> has ea actually released anything new this year yet? i brought in <cur> 5 billion in revenue in 2017. i don't consider not offering anything at all until september to somehow be a "win" based off of that. actually, madden 19 released last month and is breaking records within the franchise. it broke a " digital sales record " for the franchise because digital sales rose by 7 %. they declined to release official sales numbers, but vgchartz has them estimated at about 1. 1m copies so far between both the ps4 or xb1 versions. god of war did 3. 1 (?) m in one month, and its on one platform was not 2. looks like madden 18 sold about 2. 8m copies on ps4 total, and the xb1 presumably did less. its safe tracks say that god of war outsold last years madden, and probably this year's too. point is not that god of war is.. bad but a failure what you're implying we're saying... the point is multiplayer games make way more money than single player games and have larger install bases. this is what ea said. this is not untrue. no, the point is that ea acted as if it was somehow financially unviable to make and release single player games, which is completely untrue. they're a dogshit company that is repeatedly doubling down on predatory business practices. they are releasing fewer video</p>	<p>new games a successful strategy. they seem to rely heavily on established franchises and lack innovation. it's no wonder they face criticism for their lackluster releases. the success of god of war proves that there is still a demand for single-player games. people appreciate a well-crafted narrative and immersive gameplay experience. ea's decision to cancel the star wars game shows a disconnect with the market. their claim that linear games are no longer popular is unfounded, as evidenced by the overwhelming positive reception of titles like red dead redemption 2 and the last of us part ii. the problem lies with ea's approach, not the viability of single-player games. they prioritize profit over quality and innovation. instead of investing in new and exciting projects, they opt for quick cash grabs with minimal effort. this strategy may yield short-term gains, but it damages their reputation in the long run. gamers are becoming more discerning and expect more from the industry. ea needs to step up their game and deliver experiences that truly resonate with players. only then can they regain the trust and respect they have lost.</p>

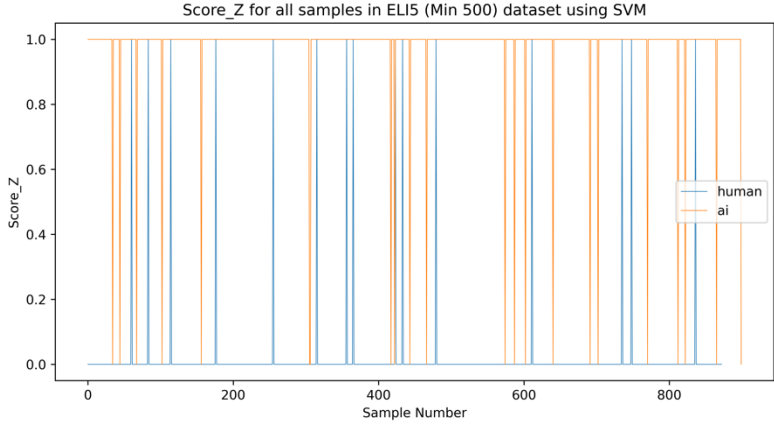
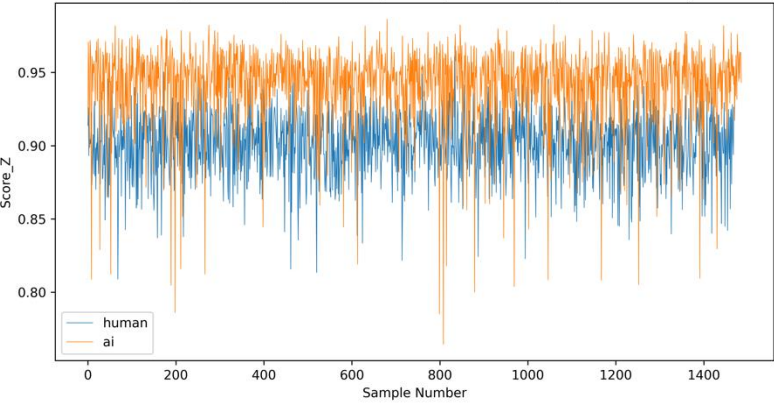
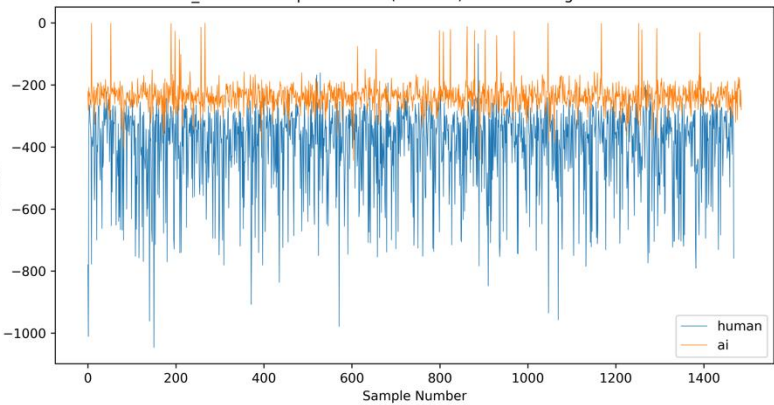
	games each year, despite laying off employees and making more money year after year after year. they're looking for a scapegoat for their own failures. don't give them one.	
Dataset: Reddit Small (Min 500) Sample Number*: 126 Type of Original Text: AI-generated Score_Z: 0.912 Prediction Method Used: Cosine Similarity	<p>alright, listen up, folks! we're gonna dive right into the case of rod reiss, and i gotta say, i've seen a lot of hate towards the guy. but let's not forget, rod also has a duty to fulfill in this story. yeah, his thoughts may not align with the survey corps, but he's got his own ideals, just like eren, for example.</p> <p>the power passed down in the reiss family, man, that's something sacred to rod. he sees it as the incarnation of god on earth, something that should be protected and worshipped at all costs. when grisha attacked the reiss family, he not only trampled all over rod's deepest beliefs, but he also slaughtered everything around him. he basically tore apart rod's entire reason for living.</p> <p>now, i gotta say, i condemn rod's treatment of historia and her mother. that was messed up, using her for his own interests. definitely not a cool move, my dude. but let's be real here, is rod the only one in this messed up universe who coerces people? hell no! we've seen levi and hange forcing people to do their bidding, and the fandom doesn't seem to have a problem with that.</p> <p>we can't forget that principles and values differ from person to person. the survey corps doesn't hold the monopoly on absolute truth. rod, kenny, flegel, and sannhes all have their own principles and values that diverge from the survey corps. and hey, rod's attachment to his beliefs, to god, and the sacred duty he feels, that's some deep stuff. he just wants to bring this god back to earth, man.</p> <p>for rod, the justice of this god, embodied in the reiss family, is what's good. and let's not forget that he wanted to save humanity from the titans back in the day. he used to share the same ideals as the survey corps. but somewhere along the line, he must have thought, "hey, if the truth is in the hands of this all-powerful person and they're not doing anything, then it must be the right thing to do." he sees humans as weak and ignorant, while this "god" has all the knowledge and strength.</p> <p>look, it's important for us to try and understand where these characters are coming from, even if their beliefs don't align with ours. justice, my friends, is not some fixed thing. it can be interpreted and exercised in different ways depending on the person, their goals, their beliefs, and the era they're in.</p> <p>so, before we go bashing rod reiss, let's take a step back and try to see things from his perspective. we don't have to agree with him, but understanding where he's coming from might just give us a new perspective on this whole crazy story.</p>	<p>own perspectives and motivations. and even though their actions may seem questionable, they are driven by their own convictions.</p> <p>rod's ultimate goal was to reclaim the founding titan's power and use it for the sake of humanity. he believed that by doing so, he could bring about true peace and prosperity. now, whether or not his plan would have actually worked is up for debate, but we can't deny that he had good intentions, at least in his own twisted way.</p> <p>in the end, rod's downfall came from his desperation and inability to let go of his own desires. he became consumed by the power and lost sight of what truly mattered. it's a tragic story, really, and a cautionary tale about the dangers of clinging to power and sacrificing everything else in the process.</p> <p>so, as we analyze rod reiss and his role in the grand scheme of things, let's remember that there are no purely good or purely evil characters in this world. everyone has their own reasons, their own struggles, and their own flaws. and it's through understanding and empathy that we can truly appreciate the complexity of their stories.</p> <p>so, let's not be quick to judge rod reiss. instead, let's try to understand his perspective and the circumstances that shaped him. after all, in this cruel and unforgiving world, it's not always easy to make the right choices.</p>
Dataset: Reddit Small (Min 100) Sample Number*: 601 Type of Original Text: human-written Score_Z: 601 Prediction Method Used: Edit Distance	<p>maintenance isn't too different from any other japanese bike. i bought my vstrom with ~ 22000 miles on it, have put on nearly another 22000, but have not yet needed new chain, sprockets, or rocker pads. some of that is down to proper chain maintenance, some to the fact it isn't a crotch rocket with a super powerful engine and i don't ride like a jackass (usually :). tires are just as with any other motorcycle - some last longer than others, and you're very doing well to get 10000 + miles out of a rear tire. the dl1000 engine is very reliable. i had my valve clearance checked at about 37000 miles, and it was spot - on. hadn't changed a bit since the last valve clearance check (around 15000 miles). the major pain in the ass has been electrical issues. by that i mean, charging system problems. my 2007 dl1000 was (probably) on its original stator but i bought the bike at about 22000 miles. by ~ 25000 miles, the stator was burned out. rotor magnets had also shifted out of position. short story short, i ended up replacing / repairing / upgrading the entire charging network, from stator to battery. i also installed a charging voltage meter, so i can tell if there's a problem before i get stranded (this is just basic equipment for any bike imo). haven't had any charging problems for about a year / the last 12000 miles. i'm probably making it sound worse as it is. the charging system was a larger issue for the first few generations of dl1000 and dl650. some vstrom owners never had a problem. all this has been dealt with, through design revisions and recalls, in newer generations of vstrom (roughly 2010 - present). the guys with recent - model dl1000 / dl650's don't seem to have charging issues. if it's been too few years old, you should be fine (probably install a voltage meter anyway!). i've had to replace both of my power hoses (inlet and outlet). at over 10 years old, they had lost their elasticity and were leaking slightly. under <cur> 20 each. easy to replace, not a big deal. as for replacement parts : in general they're affordable, about on par with any other japanese bike. suzuki does not seem to go out of their way to gouge you. you can often find used parts on ebay, if you're on a budget. some of the more - commonly - broken controls : <cur> 40 - <cur> 70 each for things like the clutch & brake levers, shift lever, rear brake pedal, anything you're likely to break off or damage in a drop / low - speed crash. all of which i have done at least once, except for the brake lever. you can also do a cheap fix with a lever saver. not as pretty as a pristine factory clutch lever, but much cheaper. get crash bars (aka "engine guard"), whether you plan to ride unpaved roads or not. these prevent a lot of damage when you drop the bike (and you almost inevitably will!). if you do plan to go on dirt, get a skid plate as well. those two things will pay for themselves, eventually. seat height: it's rather tall, in the neighborhood of 31", depending on year and aftermarket seats. so that should be fine for you. luggage : my vstrom came with givi / kappa monokey side and top cases. i like the givi monokey system, because replacement parts are fairly cheap and easy to replace. for example, i replaced all 3 locks (so they could all use the same key) for around <cur> 30. i've mangled both of my side cases in offroad shenanigans, yet repaired them for about <cur> 20 each time. windshield: california scientific makes an extra tall windshield for vstroms. being a taller dude, you may want this. i'm 6'2 and i just got their "xl" vstrom windshield. it's sooooo nice on the highway. i can sit up straight, see over the top of the windshield, and yet not have a face full of buffeting. lots more info is at the stromtrooper forums (<url>). there are specific sub - forums for different model years, so you can do more research when you find a used vstrom on craigslist etc. hope this helps!</p>	<p>pressure, it was a relatively easy fix. i replaced them with high-quality silicone hoses, which should last much longer. other than that, i haven't had any major issues with my v-strom.</p> <p>in terms of performance, the dl1000 is a fantastic bike. it has enough power to handle highway speeds with ease and can even handle some off-road adventures if you're inclined. the suspension is adjustable and provides a smooth and comfortable ride, even on longer trips. the seating position is upright and ergonomic, making it great for both commuting and touring.</p> <p>one of the standout features of the v-strom is its fuel efficiency. the bike has a large fuel tank capacity and can easily achieve over 200 miles on a single tank, depending on your riding style. this makes it perfect for long-distance rides without the constant need for refueling.</p> <p>the v-strom also offers a range of accessories and aftermarket parts to customize your riding experience. from luggage options to heated grips, you can easily tailor the bike to your specific needs. additionally, the v-strom has a strong and supportive online community, which is always helpful when it comes to finding advice and recommendations.</p> <p>in terms of overall reliability, the v-strom has a solid reputation. with proper maintenance and regular servicing, it can easily reach high mileage without any major issues. the engine is known for its durability, and many riders have reported reaching well over 100,000 miles without any major problems.</p> <p>in conclusion, the suzuki v-strom dl1000 is a reliable and versatile motorcycle that offers a great balance of performance and comfort. while there have been some electrical issues in earlier models, they have been addressed in newer generations. with proper maintenance, this bike can provide you with many years of enjoyable riding experiences. so, if you're considering a v-strom, don't hesitate, as it's a bike that can take you on countless adventures.</p>

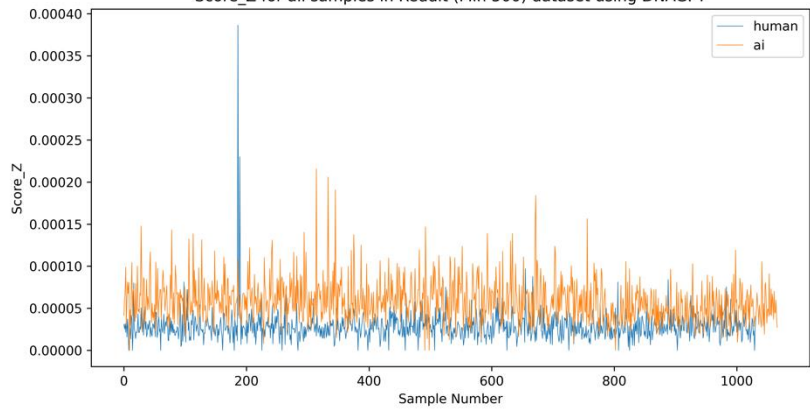
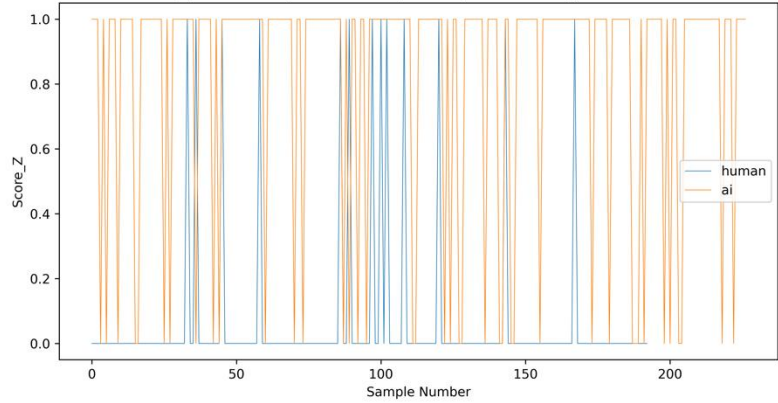
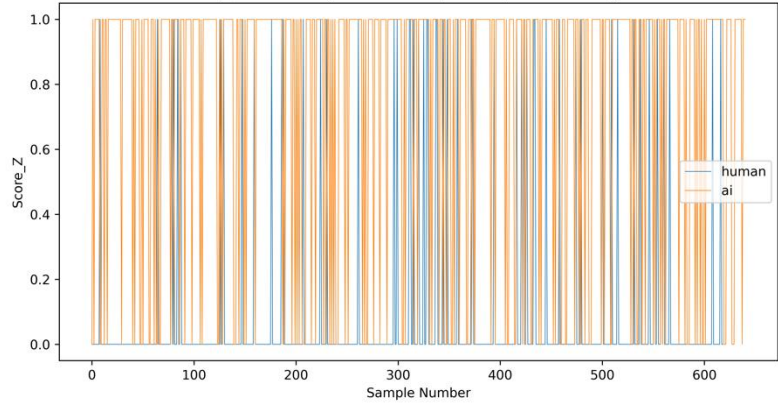
Appendix D – Graphs of Score_Z against Sample Number

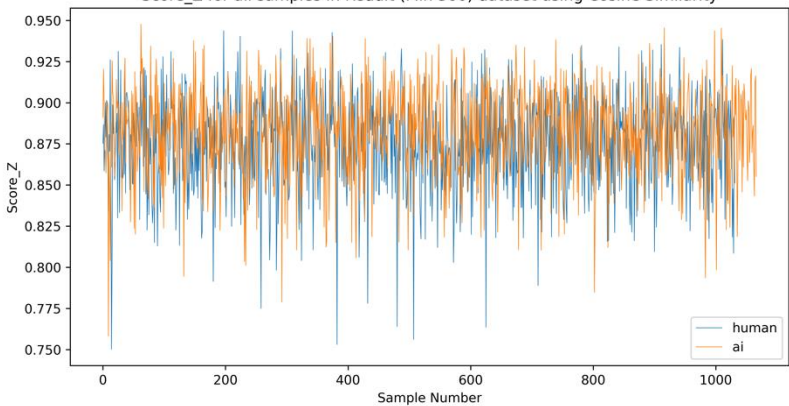
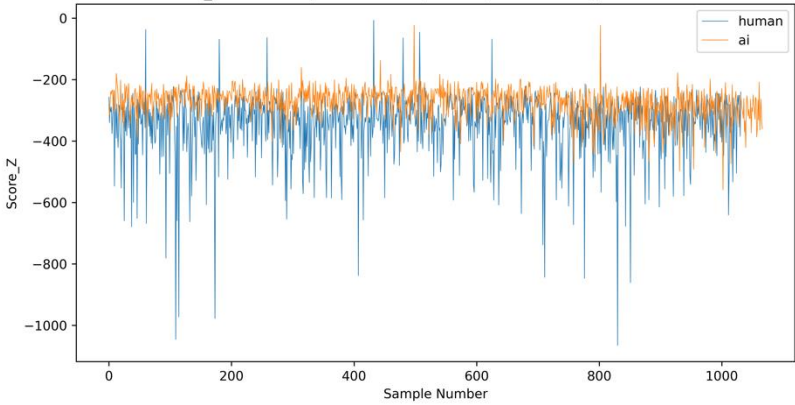
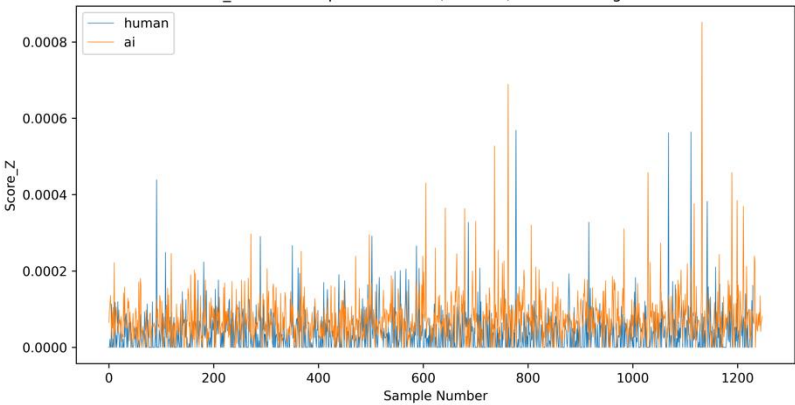
We present graphs of how each method scored samples from each dataset, to provide a more visual representation of the methods' performance.

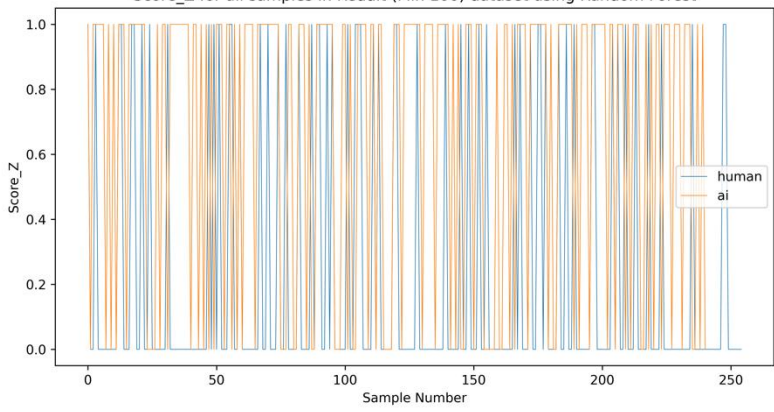
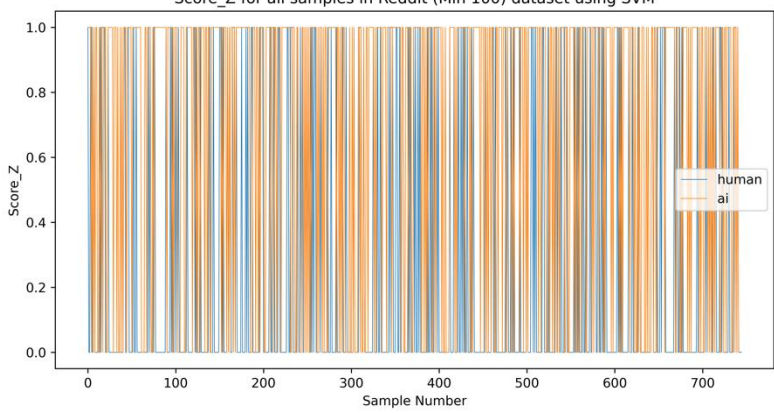
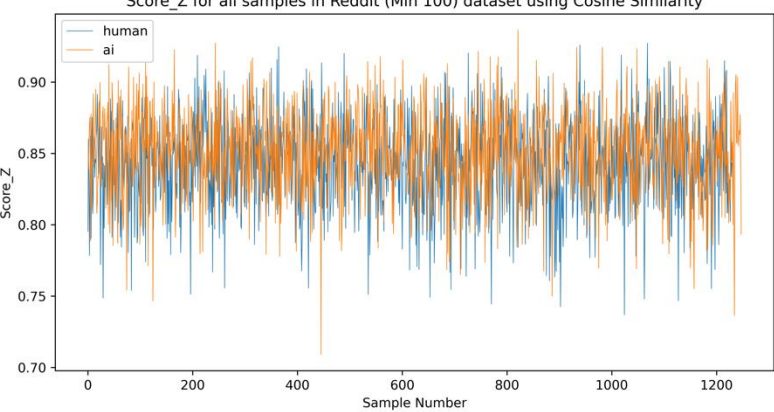
**Note: The sample numbers refer to that of the AI or Human section of the dataset, and not the entire dataset. In the case of methods using machine learning, the sample number applies to the test section of the dataset.*

Methods	Datasets
	ELI5 (Min 500)
DNA-GPT	<p>Score_Z for all samples in ELI5 (Min 500) dataset using DNAGPT</p>
Random Forest	<p>Score_Z for all samples in ELI5 (Min 500) dataset using Random Forest</p>

SVM	
Cosine Similarity	
Edit Distance	

Methods	Datasets
	Reddit Small (Min 500)
DNA-GPT	<p>Score_Z for all samples in Reddit (Min 500) dataset using DNAGPT</p> 
Random Forest	<p>Score_Z for all samples in Reddit (Min 500) dataset using Random Forest</p> 
SVM	<p>Score_Z for all samples in Reddit (Min 500) dataset using SVM</p> 

Cosine Similarity	<div>Score_Z for all samples in Reddit (Min 500) dataset using Cosine Similarity</div> 
Edit Distance	<div>Score_Z for all samples in Reddit (Min 500) dataset using Edit Distance</div> 
Methods	Datasets
	Reddit (Min 100)
DNA-GPT	<div>Score_Z for all samples in Reddit (Min 100) dataset using DNAGPT</div> 

Random Forest	<p>Score_Z for all samples in Reddit (Min 100) dataset using Random Forest</p> 
SVM	<p>Score_Z for all samples in Reddit (Min 100) dataset using SVM</p> 
Cosine Similarity	<p>Score_Z for all samples in Reddit (Min 100) dataset using Cosine Similarity</p> 

Edit Distance

