Evidence of Policy Violation by Reviewer wNfJ

Paper: #6307

1. Summary

This report presents evidence that the review provided by Reviewer wNfJ (Score: 2) was **fully generated using an LLM**, violating ICLR's review policy. Our analysis includes a control group: the same detection tools correctly identified all other reviewers (WqcR, rGrB, CKa4, F3q3) as "Fully Human," eliminating the possibility that the paper's technical jargon triggered a false positive. The summary of detection results is shown in Table 1.

Table 1: Summary of LLM detection result.

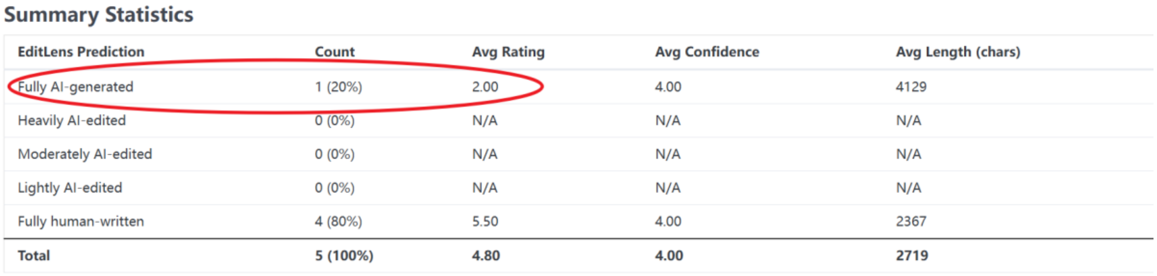
Reviewer	Rating	Pangram Labs	GPTZero
wNfJ	2	Fully AI-generated	100% AI generated
WqcR	4	Fully human-written	100% Human
rGrB	6		
CKa4	6		
F3q3	6		

2. Evidence

2a. Pangram Labs

Pangram Labs (https://www.pangram.com/) is a leading enterprise-grade detection solution trusted by global organizations and academic institutions. Notably, Pangram provides a specialized detection module for ICLR reviews (https://iclr.pangram.com), specifically calibrated to handle the technical density and formal tone of ICLR submissions.

As shown in Figure 1, the tool successfully filtered our submission's reviews. Only **Reviewer wNfJ** (Score: 2) was flagged as "**Fully AI-generated**" (Red). The table reveals a perfect correlation between the AI flag and the low quality: the AI-generated review corresponds to the outlier score of 2.00, whereas the verified human reviews maintain a high average rating of 5.50.



Summary Statistics

EditLens Prediction	Count	Avg Rating	Avg Confidence	Avg Length (chars)
Fully AI-generated	1 (20%)	2.00	4.00	4129
Heavily AI-edited	0 (0%)	N/A	N/A	N/A
Moderately AI-edited	0 (0%)	N/A	N/A	N/A
Lightly AI-edited	0 (0%)	N/A	N/A	N/A
Fully human-written	4 (80%)	5.50	4.00	2367
Total	**5 (100%)**	**4.80**	**4.00**	**2719**

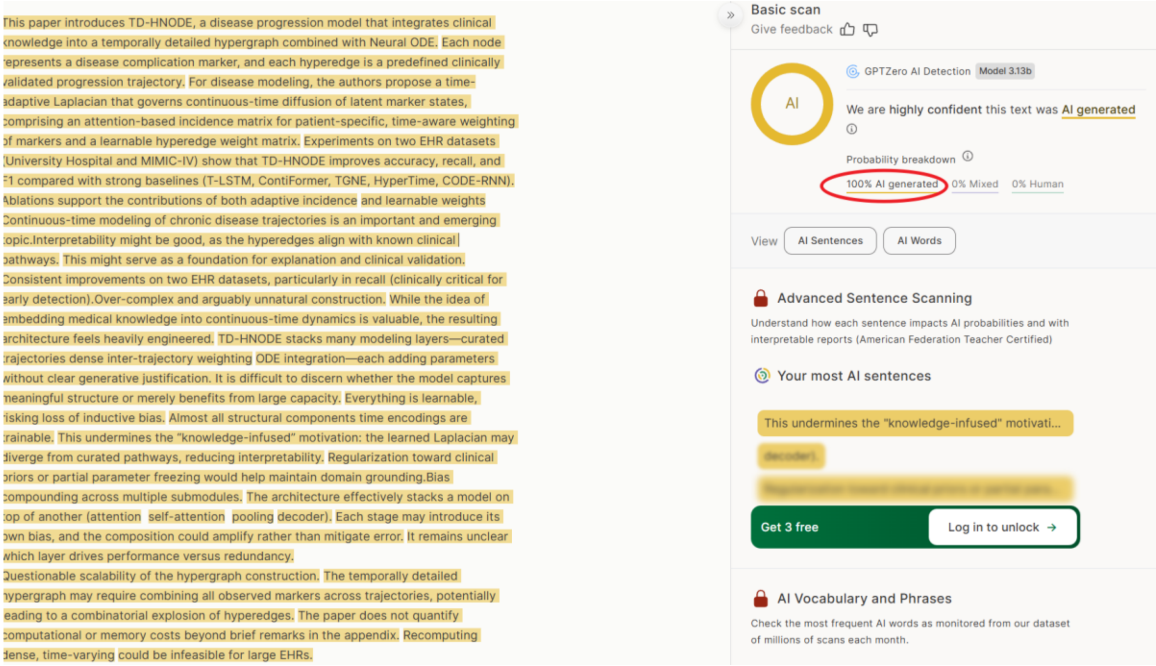
Figure 1: Screen shot from Pangram Labs detection result for paper # 6307.

Official report link: https://iclr.pangram.com/reviews?submission_number=6307

2b. GPTZero

GPTZero (https://gptzero.me/) is widely regarded as the "gold standard" for AI detection, used by educators and institutions worldwide. It utilizes perplexity and burstiness metrics to distinguish between human and machine writing patterns.

Figure 2 identifies **Reviewer wNfJ as "100% AI-generated."** In contrast, Figures 3 and 4 confirm that reviews from other reviewers were correctly identified as "100% Human."



This paper introduces TD-HNODE, a disease progression model that integrates clinical knowledge into a temporally detailed hypergraph combined with Neural ODE. Each node represents a disease complication marker, and each hyperedge is a predefined clinically validated progression trajectory. For disease modeling, the authors propose a time-adaptive Laplacian that governs continuous-time diffusion of latent marker states, comprising an attention-based incidence matrix for patient-specific, time-aware weighting of markers and a learnable hyperedge weight matrix. Experiments on two EHR datasets (University Hospital and MIMIC-IV) show that TD-HNODE improves accuracy, recall, and F1 compared with strong baselines (T-LSTM, ContiFormer, TGNE, HyperTime, CODE-RNN). Ablations support the contributions of both adaptive incidence and learnable weights Continuous-time modeling of chronic disease trajectories is an important and emerging topic.Interpretability might be good, as the hyperedges align with known clinical pathways. This might serve as a foundation for explanation and clinical validation. Consistent improvements on two EHR datasets, particularly in recall (clinically critical for early detection).Over-complex and arguably unnatural construction. While the idea of embedding medical knowledge into continuous-time dynamics is valuable, the resulting architecture feels heavily engineered. TD-HNODE stacks many modeling layers—curated trajectories dense inter-trajectory weighting ODE integration—each adding parameters without clear generative justification. It is difficult to discern whether the model captures meaningful structure or merely benefits from large capacity. Everything is learnable, risking loss of inductive bias. Almost all structural components time encodings are trainable. This undermines the "knowledge-infused" motivation: the learned Laplacian may diverge from curated pathways, reducing interpretability. Regularization toward clinical priors or partial parameter freezing would help maintain domain grounding.Bias compounding across multiple submodules. The architecture effectively stacks a model on top of another (attention self-attention pooling decoder). Each stage may introduce its own bias, and the composition could amplify rather than mitigate error. It remains unclear which layer drives performance versus redundancy.
Questionable scalability of the hypergraph construction. The temporally detailed hypergraph may require combining all observed markers across trajectories, potentially leading to a combinatorial explosion of hyperedges. The paper does not quantify computational or memory costs beyond brief remarks in the appendix. Recomputing dense, time-varying could be infeasible for large EHRs.

Basic scan

Give feedback

GPTZero AI Detection Model 3.13b

We are **highly confident** this text was AI generated

Probability breakdown

100% AI generated 0% Mixed 0% Human

View AI Sentences AI Words

Advanced Sentence Scanning

Understand how each sentence impacts AI probabilities and with interpretable reports (American Federation Teacher Certified)

Your most AI sentences

This undermines the "knowledge-infused" motivati...

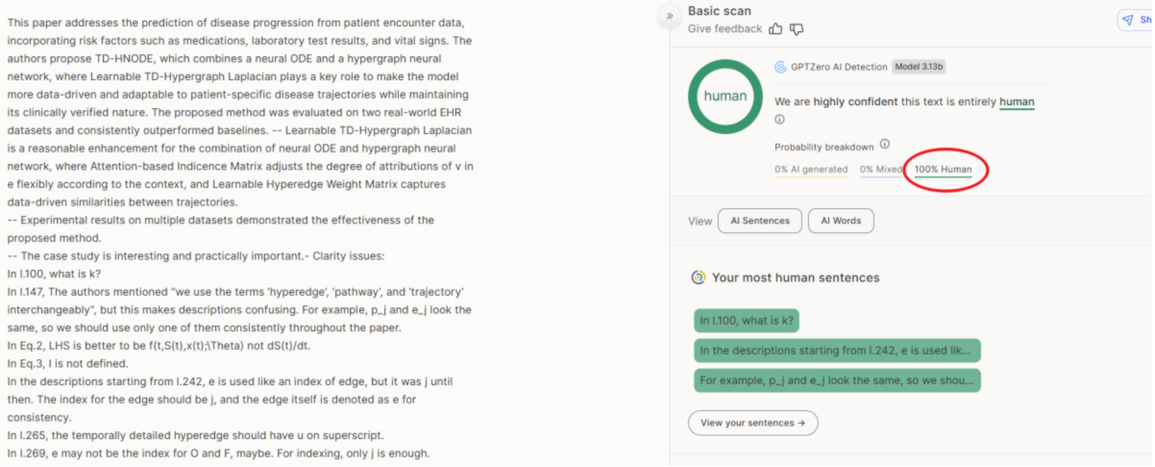
decoder).

Regularization toward clinical priors or partial para...

AI Vocabulary and Phrases

Check the most frequent AI words as monitored from our dataset of millions of scans each month.

Figure 2: GPTZero detection result (100% AI generated) for **Reviewer wNfJ.**



This paper addresses the prediction of disease progression from patient encounter data, incorporating risk factors such as medications, laboratory test results, and vital signs. The authors propose TD-HNODE, which combines a neural ODE and a hypergraph neural network, where Learnable TD-Hypergraph Laplacian plays a key role to make the model more data-driven and adaptable to patient-specific disease trajectories while maintaining its clinically verified nature. The proposed method was evaluated on two real-world EHR datasets and consistently outperformed baselines. -- Learnable TD-Hypergraph Laplacian is a reasonable enhancement for the combination of neural ODE and hypergraph neural network, where Attention-based Indicence Matrix adjusts the degree of attributions of v in e flexibly according to the context, and Learnable Hyperedge Weight Matrix captures data-driven similarities between trajectories.
-- Experimental results on multiple datasets demonstrated the effectiveness of the proposed method.
-- The case study is interesting and practically important.- Clarity issues:
In l.100, what is k?
In l.147, The authors mentioned "we use the terms 'hyperedge', 'pathway', and 'trajectory' interchangeably", but this makes descriptions confusing. For example, p_j and e_j look the same, so we should use only one of them consistently throughout the paper.
In Eq.2, LHS is better to be f(t,S(t),x(t);\Theta) not dS(t)/dt.
In Eq.3, I is not defined.
In the descriptions starting from l.242, e is used like an index of edge, but it was j until then. The index for the edge should be j, and the edge itself is denoted as e for consistency.
In l.265, the temporally detailed hyperedge should have u on superscript.
In l.269, e may not be the index for O and F, maybe. For indexing, only j is enough.

Basic scan

Give feedback

GPTZero AI Detection Model 3.13b

We are **highly confident** this text is entirely human

Probability breakdown

0% AI generated 0% Mixed 100% Human

View AI Sentences AI Words

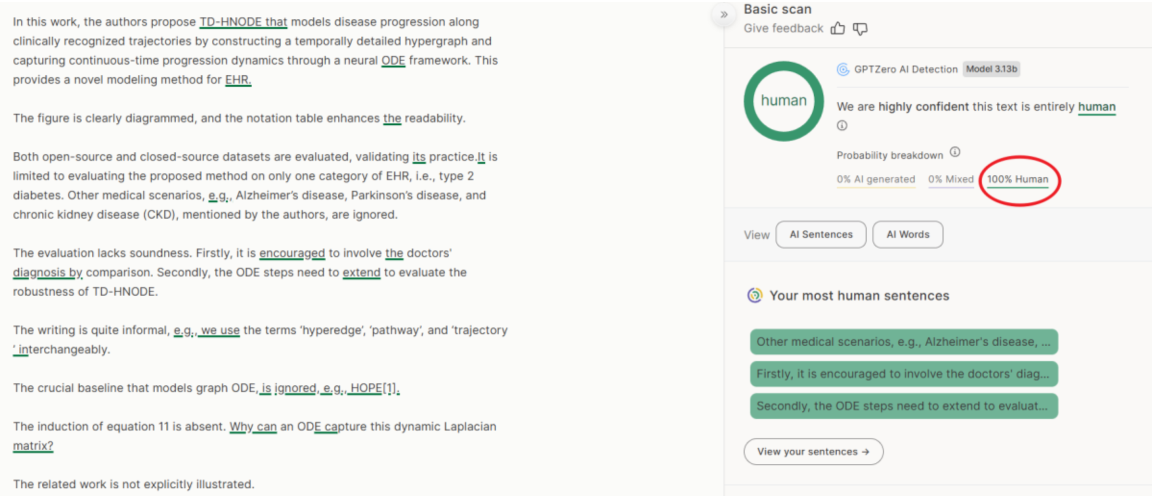
Your most human sentences

In l.100, what is k?

In the descriptions starting from l.242, e is used lik...

For example, p_j and e_j look the same, so we shou...

Figure 3: Control check: GPTZero detection result (100% Human) for Reviewer rGrB.



In this work, the authors propose TD-HNODE that models disease progression along clinically recognized trajectories by constructing a temporally detailed hypergraph and capturing continuous-time progression dynamics through a neural ODE framework. This provides a novel modeling method for EHR.

The figure is clearly diagrammed, and the notation table enhances the readability.

Both open-source and closed-source datasets are evaluated, validating its practice.It is limited to evaluating the proposed method on only one category of EHR, i.e., type 2 diabetes. Other medical scenarios, e.g., Alzheimer's disease, Parkinson's disease, and chronic kidney disease (CKD), mentioned by the authors, are ignored.

The evaluation lacks soundness. Firstly, it is encouraged to involve the doctors' diagnosis by comparison. Secondly, the ODE steps need to extend to evaluate the robustness of TD-HNODE.

The writing is quite informal, e.g., we use the terms 'hyperedge', 'pathway', and 'trajectory ' interchangeably.

The crucial baseline that models graph ODE, is ignored, e.g., HOPE[1].

The induction of equation 11 is absent. Why can an ODE capture this dynamic Laplacian matrix?

The related work is not explicitly illustrated.

We are highly confident this text is entirely human

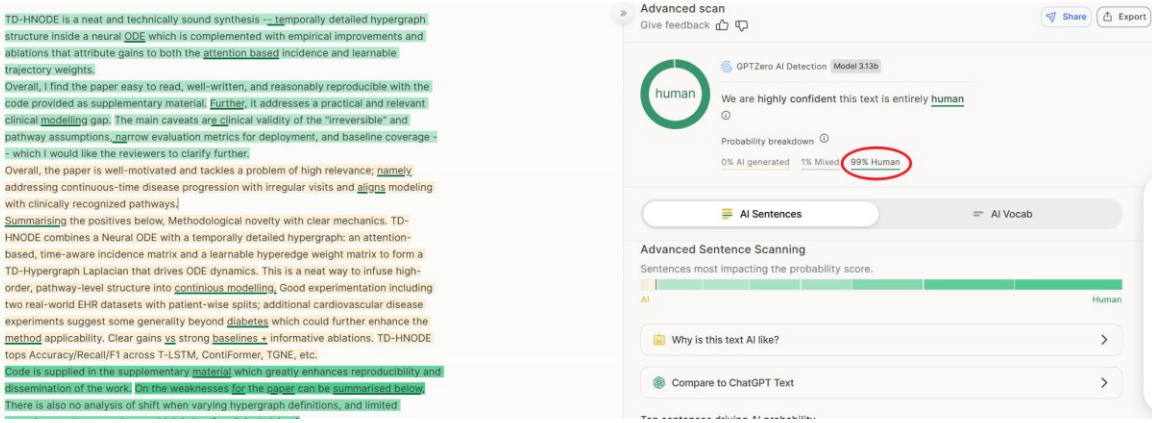
Probability breakdown

0% AI generated 0% Mixed 100% Human

Your most human sentences

- Other medical scenarios, e.g., Alzheimer's disease, ...
- Firstly, it is encouraged to involve the doctors' diag...
- Secondly, the ODE steps need to extend to evaluat...

Figure 4: Control check: GPTZero detection result (100% Human) for Reviewer WqcR.



TD-HNODE is a neat and technically sound synthesis -- temporally detailed hypergraph structure inside a neural ODE which is complemented with empirical improvements and ablations that attribute gains to both the attention based incidence and learnable trajectory weights.

Overall, I find the paper easy to read, well-written, and reasonably reproducible with the code provided as supplementary material. Further, it addresses a practical and relevant clinical modelling gap. The main caveats are clinical validity of the "irreversible" and pathway assumptions, narrow evaluation metrics for deployment, and baseline coverage -- which I would like the reviewers to clarify further.

Overall, the paper is well-motivated and tackles a problem of high relevance; namely addressing continuous-time disease progression with irregular visits and aligns modeling with clinically recognized pathways.

Summarising the positives below, Methodological novelty with clear mechanics. TD-HNODE combines a Neural ODE with a temporally detailed hypergraph: an attention-based, time-aware incidence matrix and a learnable hyperedge weight matrix to form a TD-Hypergraph Laplacian that drives ODE dynamics. This is a neat way to infuse high-order, pathway-level structure into continious modelling. Good experimentation including two real-world EHR datasets with patient-wise splits; additional cardiovascular disease experiments suggest some generality beyond diabetes which could further enhance the method applicability. Clear gains vs strong baselines + informative ablations. TD-HNODE tops Accuracy/Recall/F1 across T-LSTM, ContiFormer, TGNE, etc.

Code is supplied in the supplementary material which greatly enhances reproducibility and dissemination of the work. On the weaknesses for the paper can be summarised below.

There is also no analysis of shift when varying hypergraph definitions, and limited

We are highly confident this text is entirely human

Probability breakdown

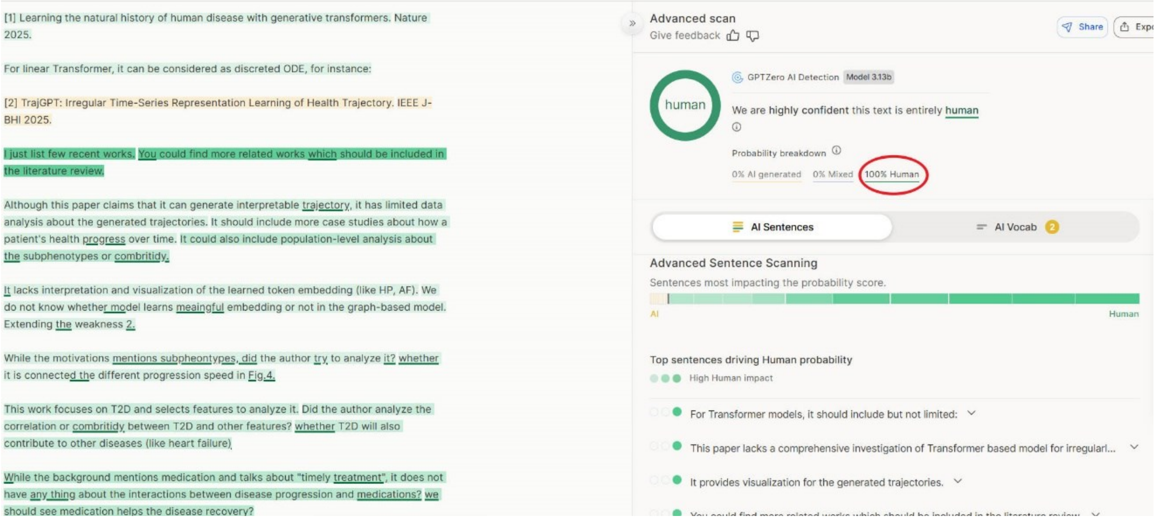
0% AI generated 1% Mixed 99% Human

Advanced Sentence Scanning

Sentences most impacting the probability score.

AI
Human
Why is this text AI like?
Compare to ChatGPT Text
Top sentences driving AI probability

Figure 5: Control check: GPTZero detection result (99% Human) for Reviewer CKa4.



[1] Learning the natural history of human disease with generative transformers. Nature 2025.

For linear Transformer, it can be considered as discreted ODE, for instance:

[2] TrajGPT: Irregular Time-Series Representation Learning of Health Trajectory. IEEE J-BHI 2025.

I just list few recent works. You could find more related works which should be included in the literature review.

Although this paper claims that it can generate interpretable trajectory, it has limited data analysis about the generated trajectories. It should include more case studies about how a patient's health progress over time. It could also include population-level analysis about the subphenotypes or combritidy.

It lacks interpretation and visualization of the learned token embedding (like HP, AF). We do not know whether model learns meaingful embedding or not in the graph-based model. Extending the weakness 2.

While the motivations mentions subpheontypes, did the author try to analyze it? whether it is connected the different progression speed in Fig.4.

This work focuses on T2D and selects features to analyze it. Did the author analyze the correlation or combritidy between T2D and other features? whether T2D will also contribute to other diseases (like heart failure)

While the background mentions medication and talks about "timely treatment", it does not have any thing about the interactions between disease progression and medications? we should see medication helps the disease recovery?

We are highly confident this text is entirely human

Probability breakdown

0% AI generated 0% Mixed 100% Human

Advanced Sentence Scanning

Sentences most impacting the probability score.

Top sentences driving Human probability

- For Transformer models, it should include but not limited:
- This paper lacks a comprehensive investigation of Transformer based model for irregularl...
- It provides visualization for the generated trajectories.
- You could find more related works which should be included in the literature review.

Figure 6: Control check: GPTZero detection result (100% Human) for Reviewer F3q3.