# *Enhanced language models for predicting and understanding HIV Care disengagement: A case study in Tanzania*

1313


Waverly Wei[1*], **Junzhe Shao[2*],** Rita Qiuran Lyu[2], Rebecca Hemono[2], Xinwei Ma[3], Joseph Giorgio[4], Zeyu Zheng[5], Emmanuel Katabaro[6], Matilda Mlowe[6], Amon Sabasaba[6], Caroline Lister[7], Siraji Shabani[8], Sandra I. McCoy[2], Jingshen Wang[2]

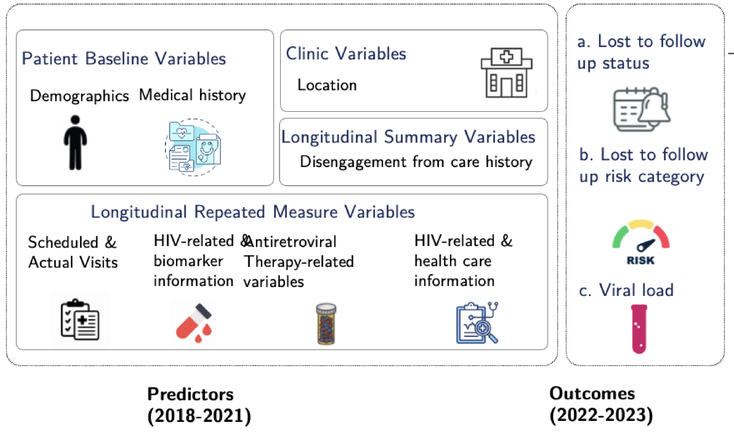
[1] Marshall School of Business, University of Southern California, California, USA,[2] School of Public Health, University of California, Berkeley, California, USA, [3] Department of Economics, University of California, San Diego, California, US,[4] Helen Wills Neuroscience Institute, University of California, Berkeley, California, USA, [5] College of Engineering, University of California, Berkeley, California, USA,[6] Health for a Prosperous Nation, Dar es Salaam, Tanzania, [7] Dodoma Referral Hospital, Dodoma, Tanzania,[8] Ministry of Health, Dodoma, Tanzania


## BACKGROUND

Despite significant progress in HIV treatment, maintaining patient engagement in care remains a critical challenge. While UNAIDS aims to achieve the “95-95-95” targets by 2030, many people living with HIV (PLHIV), particularly in sub-Saharan Africa, tend to discontinue care within two years. Current one-size-fits-all interventions often lack efficient targeting, and resources could be better allocated to those at the highest risk. The increasing availability of electronic medical records (EMRs) in sub-Saharan Africa, combined with advancements in AI and ML techniques, offers promising solutions for predicting which PLHIV may require additional support.

## METHODS

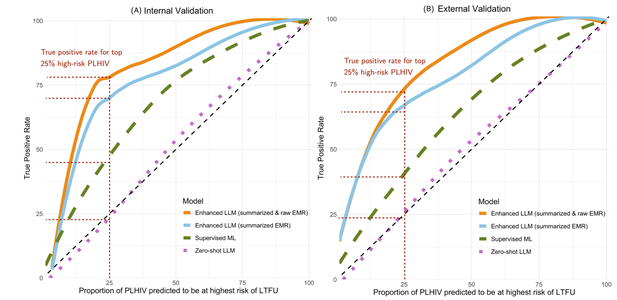
We developed a novel AI model by enhancing a **pre-trained LLM (LLaMA 3.1)** using routinely collected EMRs from Tanzania‘s National HIV Care and Treatment Program to identify people at risk of disengaging from HIV care or developing adverse outcomes. Outcomes included risk of ART non-adherence, non-suppressed viral load, and loss to follow-up. Additionally, a team of HIV physicians in Tanzania assessed the LLM’s predictions, and LLM provided justifications for a subset of patient records to evaluate their clinical relevance and reasoning.



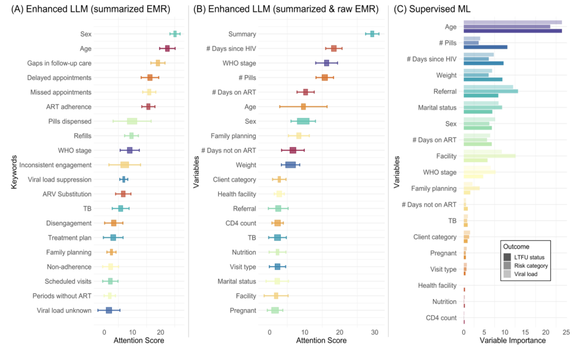

Despite increased ART access, disengagement from care remains a significant issue, particularly in sub-Saharan Africa. Traditional machine learning (ML) models have shown moderate success in predicting care disengagement, which would enable early intervention. We develop an **enhanced large language model (LLM)** fine-tuned with **electronic medical records (EMRs)** to predict people at **risk of disengaging from HIV care** in Tanzania more accurately and to provide interpretative insights into modifiable risk factors.

## RESULTS

- The enhanced LLMs achieved higher AUC than the supervised ML model and zero-shot LLMs in internal and external validation.
- When focusing on the 25% of PLHIV predicted as most likely to LTFU, the model correctly identified 78% of PLHIV genuinely at risk in internal validation and 73% (7,105 of 9,733) in external validation.
- Attention score analysis indicated that the enhanced LLM focused on keywords such as gaps in follow-up care and ART adherence.
- The human expert evaluation showed alignment between clinician assessments and the LLM's predictions in 65% of cases, with experts finding the model’s justifications reasonable and clinically relevant in 92.3% of aligned cases.




| **Internal** validation in the Kagera Region | | | **External** validation in the Geita Region | | |
|---|---|---|---|---|---|
| Risk category | Viral suppression | LTFU | Risk category | Viral suppression | LTFU |
| Enhanced LLM with **summarized** CTC3 medical records | | | | | |
| 0.73 | 0.65 | 0.76 | 0.66 | 0.62 | 0.68 |
| Enhanced LLM with **summarized and raw** CTC3 medical records | | | | | |
| **0.74** | **0.68** | **0.77** | **0.67** | **0.64** | **0.71** |
| Supervised machine learning (gradient boosting) | | | | | |
| 0.72 | 0.68 | 0.71 | 0.66 | 0.64 | 0.65 |




**Step 1.** Serialization of tabular data

**Prompt Instruction**

As a question answering agent, your task is to predict if people living with HIV (PLHIV) have a detectable viral load, is at risk of ART non-adherence, or will be lost to follow-up for more than 28 days in year 2022-2023. The prediction result can be formalized into an output sentence in the following format: This PLHIV will have **[suppressed/detectable/unknown]** viral load, is **[possibly/unlikely]** to be lost to follow-up for over 28 days, and is at **[high/moderate/low/no/unknown]** risk of ART non-adherence based on their days not on ART in year 2022-2023. Give me your reasonings for outcome prediction.

**Prompt Input**

Raw EMR

Patient Medical Records: The PLHIV is a 48 year-old Male.The patient is alive.Scheduled appointments:[xxx].Number of ARV pills prescribed on each visit date:[xxx]. Number of days not on ART are:[xxx].The patient ART time since diagnosis of HIV: 2219,2251,2279,2370,2461,2548,2640,2734.The patient has been in ART for:[xxx].ARV status: [xxx]. CD4 status: [xxx].Pregnant status: [xxx].TB TPT status:[xxx].WHO stage:[xxx].Viral load:[xxx].Weight: [xxx].Health facility type: [xxx].Visit type: [xxx]. Client category:[xxx].Marital status: [xxx].Family planning status: [xxx].Nutrition status: [xxx].

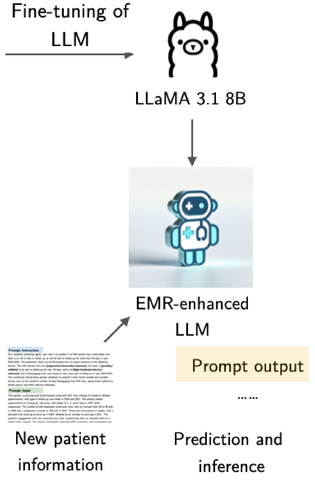
Summarized EMR

The patient, a [xxx]-year-old [male/female] living with HIV, has a history of missed or delayed appointments, with gaps in follow-up care noted in 2020 and 2021. The patient missed appointments on [xxx],[xxx], and [xxx], with delays of 2, 2, and 4 days in ARV refills, respectively. The number of pills dispensed varied over time, with an increase from 30 to 90 pills in 2020 and a subsequent increase to 180 pills in 2021. There were fluctuations in weight, with a decrease from [xxx] kg to [xxx] kg in 20 followed by an increase to [xxx] kg in 2021. The patient's engagement with care improved over tim transitioning from an unstable client to a stable client category. The patient consistently continued ART treatment, with suppressed viral load noted in some visits.

**Prompt Output**

This PLHIV will have [suppressed] viral load, is [unlikely] to be lost to follow-up for over 28 days, and is at [no] risk of ART non-adherence based on their days not on ART in the year 2022-2023.

**Step 2.** Fine-tuning of LLM




## CONCLUSIONS

- The results underscore the potential of LLMs in providing actionable public health insights that can guide targeted interventions to enhance patient retention.
- This study represents the first application of fine-tuned LLMs in predicting disengagement from HIV care using large-scale EMR data.

## ADDITIONAL KEY INFORMATION


The study was supported by a grant from the US National Institutes of Health (NIH): NIH NIMH 1R01MH125746.