

Subject: Here is the way GPT thinks we should do it

Date: Thursday, July 10, 2025 at 1:22:36 PM Central Daylight Time

From: Martin Paulus

To: Dale Peasley, Mikey Ferguson

To generate a **patient query entry for an LLM using FHIR data**, particularly for the purpose of:

1. **Computing an encoder vector** (i.e., a numerical patient embedding for downstream modeling tasks)
2. **Predicting risk of Treatment-Resistant Depression (TRD)**

we need to:

- Convert **FHIR-based structured EHR data** into **serialized input** (usually Markdown or plain text with structure)
- Maintain **temporal order, coding semantics, and clinical context**
- Make the serialization **LLM-friendly**, meaning it's concise, flattened, and ready for language modeling or embedding

Step-by-Step Framework

1. Define Scope of Relevant FHIR Resources

To capture information relevant to TRD prediction, include:

FHIR Resource	Key Fields
Patient	Age, sex, race/ethnicity, socioeconomic status
Condition	Major depressive disorder (onset, course), comorbid anxiety, PTSD, etc.
MedicationRequest	Antidepressant class, dose, timing, number of trials, augmentation meds
Observation	PHQ-9 scores, sleep/appetite/energy ratings, weight/BMI
Encounter	Visit frequency (outpatient/inpatient/ED), gaps in care
Procedure	ECT, TMS, psychotherapy documented
DiagnosticReport	Imaging (e.g., MRI), if relevant
AllergyIntolerance	Treatment-limiting allergies (e.g., SSRI)

2. Convert FHIR to Text Prompt (Markdown or Plain Text)

Flatten into a **chronological, structured text block**. Here's an example:

LLM Patient Input Example for TRD Risk Prediction

Patient Demographics

- Age: 42
- Sex: Female
- Race: Black
- Insurance: Medicaid

Diagnoses (Condition Resource)

- Major depressive disorder (ICD-10: F33.1) — diagnosed 2021-03-01
- Generalized anxiety disorder (ICD-10: F41.1) — diagnosed 2019-11-15
- Obesity (ICD-10: E66.9) — diagnosed 2022-05-10

Antidepressant Trials (MedicationRequest)

1. Sertraline 50–200mg daily — started 2021-03-10, stopped 2021-06-15 (ineffective)
2. Bupropion XL 150mg → 300mg daily — 2021-06-20 to 2021-12-10 (partial response)
3. Venlafaxine XR 75mg → 225mg — 2022-01-05 to 2022-07-01 (poor tolerability)
4. Mirtazapine 15mg qhs — 2022-08-15 to present (ongoing)

Augmentation & Other Psychiatric Medications

- Aripiprazole 2mg — added 2022-10-01 (partial benefit)
- Hydroxyzine PRN
- No lithium, ECT, or TMS

PHQ-9 Scores (Observation)

- 2021-03-01: 18
- 2021-06-01: 15
- 2021-10-01: 17
- 2022-03-01: 19
- 2022-09-01: 16
- 2023-03-01: 18

Clinical Events (Encounter + Procedure)

- 14 outpatient psych visits over 2 years
- No psychiatric hospitalizations
- No psychotherapy procedures documented

Comorbidities

- Type 2 Diabetes
- Smoking history: Yes
- BMI: 33.8 (2023-06-01)

3. Use Case 1: Encoder Vector

To generate a **patient embedding**, you can pass the entire block into an **LLM embedding model** (e.g., text-embedding-3-large, Gemma, or MedGemma 27B) using:

```
embedding = LLM_embedder(text_block)
```

This vector can then be:

- Clustered (e.g., for phenotyping)
- Used as input to a classifier
- Passed into a downstream TRD prediction model

4. Use Case 2: Risk Prediction Query Prompt

You can either use a **prompting format** for zero-shot/few-shot inference:

Instruction:

You are a clinical assistant. Given the patient data below, estimate the risk of treatment-resistant depression.

Patient Information:

<Insert structured patient summary from above>

Answer:

Or build a **supervised training set** (serialized patient → TRD label) for fine-tuning a classifier.

Notes on TRD Operationalization

- Often defined as **failure of ≥ 2 antidepressants** with adequate dose and duration
 - More advanced models may incorporate:
 - Polypharmacy
 - Comorbid anxiety/PTSD
 - High PHQ-9 persistence
 - Lack of remission markers (e.g., continued suicidal ideation, functional impairment)
-

Summary

Component	Purpose
FHIR data → structured text	Enables LLM-friendly serialization
Encoder vector	Use embeddings for downstream ML/representation tasks
Prompt-based diagnosis	Estimate TRD risk from serialized text
Optional fine-tuning	Train classifier using embeddings + TRD outcome labels

The Art of War: In the midst of chaos, there is also opportunity.
Martin

Scientific Director and President, Laureate Institute for Brain Research
6655 S Yale Ave Tulsa, OK 74136-3326 P 918 502 5120 F 918 502 5135
email: mpaulus@laureateinstitute.org X:@mpwpaulus
web: <http://www.laureateinstitute.org>

Professor of Neuroscience, Oxley College of Health and Natural Sciences
University of Tulsa, email: mpp4692@utulsa.edu

Adjunct Professor, Department of Psychiatry, University of California, San Diego
email: mpaulus@health.ucsd.edu