

Supplementary Materials:

Title:

Toward Interpretable Glucose Forecasting for Type 2 Diabetes: A Comparative Study among Traditional, Deep, and Large Language Models.

Dataset sample:

Example of patient data matrix (patientID:2000_0_20201230, first3samples):

CGM_t1	CGM_t2	CGM_t3	CGM_t4	CGM_t5	CGM_t6	CGM_t7	CGM_t8	CGM_t9	CGM_t10
138.6	120.6	106.2	97.2	88.2	75.6	64.8	59.4	59.4	68.4
104.4	104.4	97.2	102.6	115.2	117.0	104.4	102.6	122.4	131.4
133.2	133.2	142.2	142.2	138.6	129.6	115.2	102.6	97.2	91.8
CGM_t11	CGM_t12	CGM_t13	CGM_t14	CGM_t15	CGM_t16	CGM_t17	CGM_t18	CGM_t19	CGM_t20
72.0	77.4	81.0	88.2	93.6	86.4	84.6	88.2	90.0	93.6
99.0	102.6	104.4	109.8	109.8	138.6	120.6	106.2	97.2	88.2
75.6	73.8	70.2	68.4	72.0	77.4	81.0	88.2	93.6	86.4
Future30	Future60	Future90	BMI	Age	Weight	Smoking	Duration of diabetes	Gender	
104.4	102.6	117.0	23.6	57.0	67.4	0.0	25.0	2	
133.2	142.2	129.6	23.6	57.0	67.4	0.0	25.0	2	
55.8	59.4	77.4	23.6	57.0	67.4	0.0	25.0	2	

```
=== Sample data for patient 2000_0_20201230 ===
  CGM_t1  CGM_t2  CGM_t3  CGM_t4  CGM_t5  CGM_t6  CGM_t7  CGM_t8  CGM_t9  \
0   138.6   120.6   106.2   97.2    88.2    75.6    64.8    59.4    59.4
1   104.4   104.4    97.2   102.6   115.2   117.0   104.4   102.6   122.4
2   133.2   133.2   142.2   142.2   138.6   129.6   115.2   102.6    97.2

  CGM_t10  ...  Future30  Future60  Future90  Patient Number  BMI (kg/m2)  \
0     68.4  ...    104.4    102.6    117.0    2.000020e+12    23.6
1    131.4  ...    133.2    142.2    129.6    2.000020e+12    23.6
2     91.8  ...     55.8     59.4     77.4    2.000020e+12    23.6

  Age (years)  Weight (kg)  Smoking History (pack year)  \
0         57.0         67.4                        0.0
1         57.0         67.4                        0.0
2         57.0         67.4                        0.0

  Duration of diabetes (years)  Gender (Female=1, Male=2)
0                        25.0                        2.0
1                        25.0                        2.0
2                        25.0                        2.0

[3 rows x 30 columns]
=====
```

XGBoost_result:

2

RandomForest_result:

mlflow2.22.0

ExperimentsModelsPrompts

GitHubDocs

Test_RandomForest3 >

Test_RandomForest3

Register model

OverviewModel metricsSystem metricsTracesArtifacts

Parameters (9)

Q Search parameters

Parameter	Value
batch_size	32
epochs	100
model_type	randomforest
scaler_type	standard
transform_gender	True
use_shuffle	True
use_time2vec	False
window_size	16
y_shape	(5549, 3)

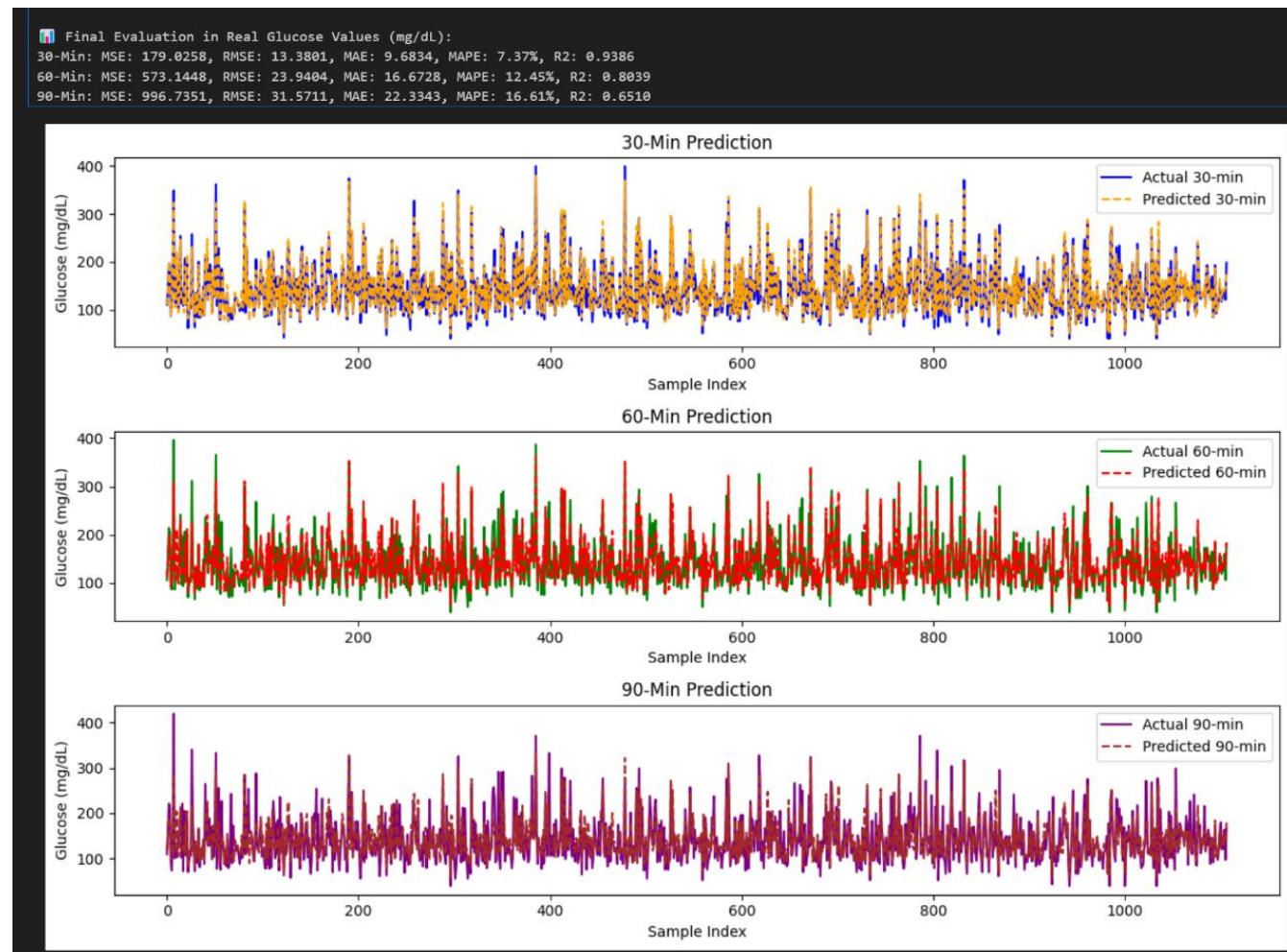
Metrics (21)

Q Search metrics

Metric	Value
30_mae	10.56694638668357
30_mape	0.07774678703541597
30_mse	218.7281433928949
30_r2	0.9165915145044214
30_rmse	14.789460551111894
60_mae	18.457482708933718
60_mape	0.13802896967160988
60_mse	664.0667269884725
60_r2	0.7535956226726029
60_rmse	25.769492175603162
90_mae	23.629798270893367
90_mape	0.1801596528386804
90_mse	1004.6717208025938
90_r2	0.6225816078420336
90_rmse	31.696556923467156
mean_mae	17.551409122170217
mean_mape	0.13197846984856876
mean_mse	629.1555303946537
mean_r2	0.7642562483396859
mean_rmse	24.08516988339407

b) Deep learning:

LSTM_result:



Time2vec_result:

mlflow2.2.2.0ExperimentsModelsPrompts

Test_Time2Vec3 >

Test_Time2Vec3

Register model

OverviewModel metricsSystem metricsTracesArtifacts

Parameters (10)

Q Search parameters

Parameter	Value
batch_size	32
epochs	100
model_type	time2vec
scaler_type	standard
time2vec_kernel	2
transform_gender	True
use_shuffle	True
use_time2vec	True
window_dze	20
y_shape	(4427, 3)

Metrics (21)

Q Search metrics

Metric	Value
30_mae	10.660277057017092
30_mape	0.07785109330221063
30_mse	232.24281399256404
30_r2	0.9203026892720833
30_rmse	15.23951488704821
60_mae	17.638116619106665
60_mape	0.1288028414154335
60_mse	651.6001572283456
60_r2	0.777056854358108
60_rmse	25.52645994313245
90_mae	23.153432499748583
90_mape	0.16970964442310654
90_mse	1081.1714752955281
90_r2	0.6214883580969277
90_rmse	32.88117204868963
mean_mae	17.15060872529078
mean_mape	0.12545452638025023
mean_mse	655.0048155054792
mean_r2	0.7729493005757062
mean_rmse	24.54904895962343

GRU_result:

mlflow2.2.2.0ExperimentsModelsPrompts

Test_gru_result >

Test_gru_result

Register model

OverviewModel metricsSystem metricsTracesArtifacts

Parameters (9)

Q Search parameters

Parameter	Value
batch_size	32
epochs	100
model_type	gru
scaler_type	standard
transform_gender	True
use_shuffle	True
use_time2vec	False
window_size	20
y_shape	(4427, 3)

Metrics (21)

Q Search metrics

Metric	Value
30_mae	10.289735588342518
30_mape	0.07836589397244952
30_mse	215.2786235732027
30_r2	0.926124184162955
30_rmse	14.672376207458788
60_mae	17.436510963060883
60_mape	0.13080930590319415
60_mse	611.0697822679886
60_r2	0.7909242071932314
60_rmse	24.719825692508202
90_mae	22.840263435618773
90_mape	0.17080455366301323
90_mse	1021.152507664539
90_r2	0.6425006382970904
90_rmse	31.955476958802212
mean_mae	16.8555032900739
mean_mape	0.12665991784621897
mean_mse	615.8336378352434
mean_r2	0.7865163432177589
mean_rmse	23.782559619589733

transformer_result:

Test_Transformer2

Overview Model metrics System metrics Traces Artifacts

Parameters (9)

Parameter	Value
batch_size	32
epochs	100
model_type	transformer
scaler_type	standard
transform_gender	True
use_shuffle	True
use_time2vec	False
window_size	20
y_shape	(4427, 3)

Metrics (21)

Metric	Value
30_mae	10.61779492592747
30_mape	0.07523849507467675
30_rmse	239.58847079452815
30_r2	0.9177819262715733
30_rmse	15.4786456382504
60_mae	16.912644379571848
60_mape	0.12194518928921728
60_rmse	602.0920670466462
60_r2	0.7939959069924409
60_rmse	24.53756440738661
90_mae	22.544683029877786
90_mape	0.1645278292802762
90_rmse	1015.6639745086379
90_r2	0.6444221408005779
90_rmse	31.869483436488864
mean_mae	16.695040778459035
mean_mape	0.12057050454805675
mean_rmse	619.1148374499375
mean_r2	0.785399991354864
mean_rmse	23.961897827375292

Ensemble_LSTM_Trans_result:

Test_Ensemble_LSTM_TRANS2

Overview Model metrics System metrics Traces Artifacts

Parameters (9)

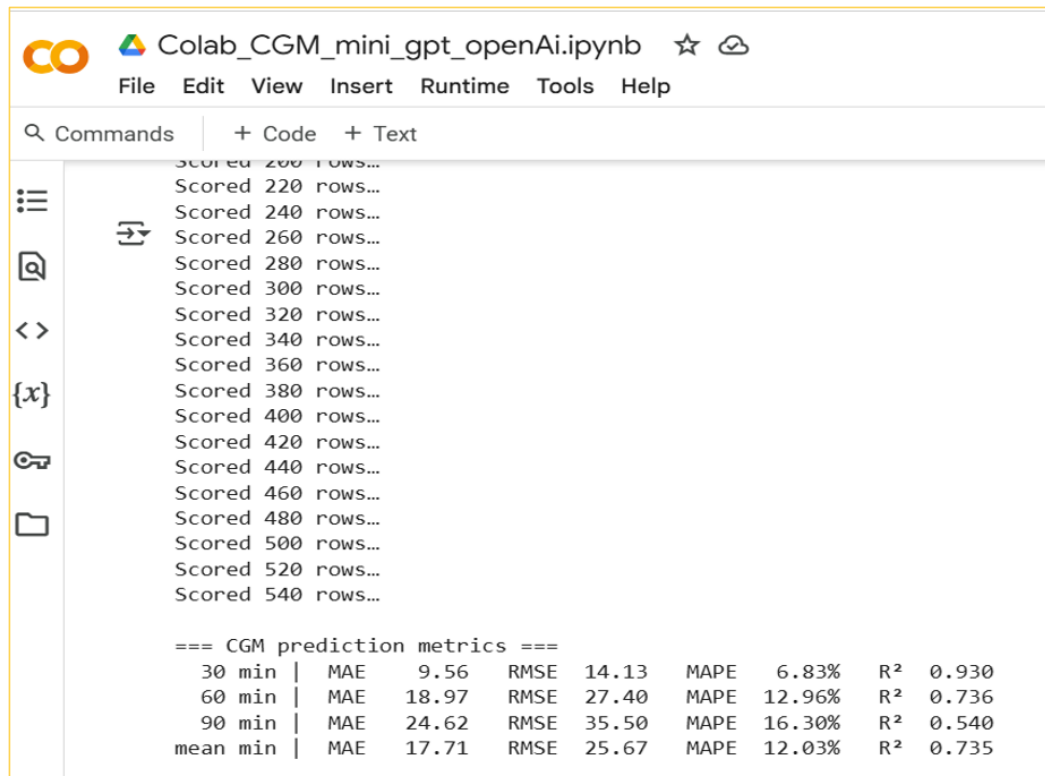
Parameter	Value
batch_size	32
epochs	100
model_type	ensemble
scaler_type	standard
transform_gender	True
use_shuffle	True
use_time2vec	False
window_size	20
y_shape	(4427, 3)

Metrics (21)

Metric	Value
30_mae	10.608981933180234
30_mape	0.0814110757593797
30_rmse	209.52493438265762
30_r2	0.9280986416170633
30_rmse	14.474976144459017
60_mae	17.730087368490466
60_mape	0.13444508676253056
60_rmse	612.566732197366
60_r2	0.7904120300207406
60_rmse	24.75008549878901
90_mae	23.493425249687267
90_mape	0.17696707146357968
90_rmse	1050.5769737679007
90_r2	0.6321993093854481
90_rmse	32.412605167864875
mean_mae	17.27749818378599
mean_mape	0.1309410779951633
mean_rmse	624.2228801159748
mean_r2	0.7835699936744174
mean_rmse	23.87922270370963

c) LLMs:

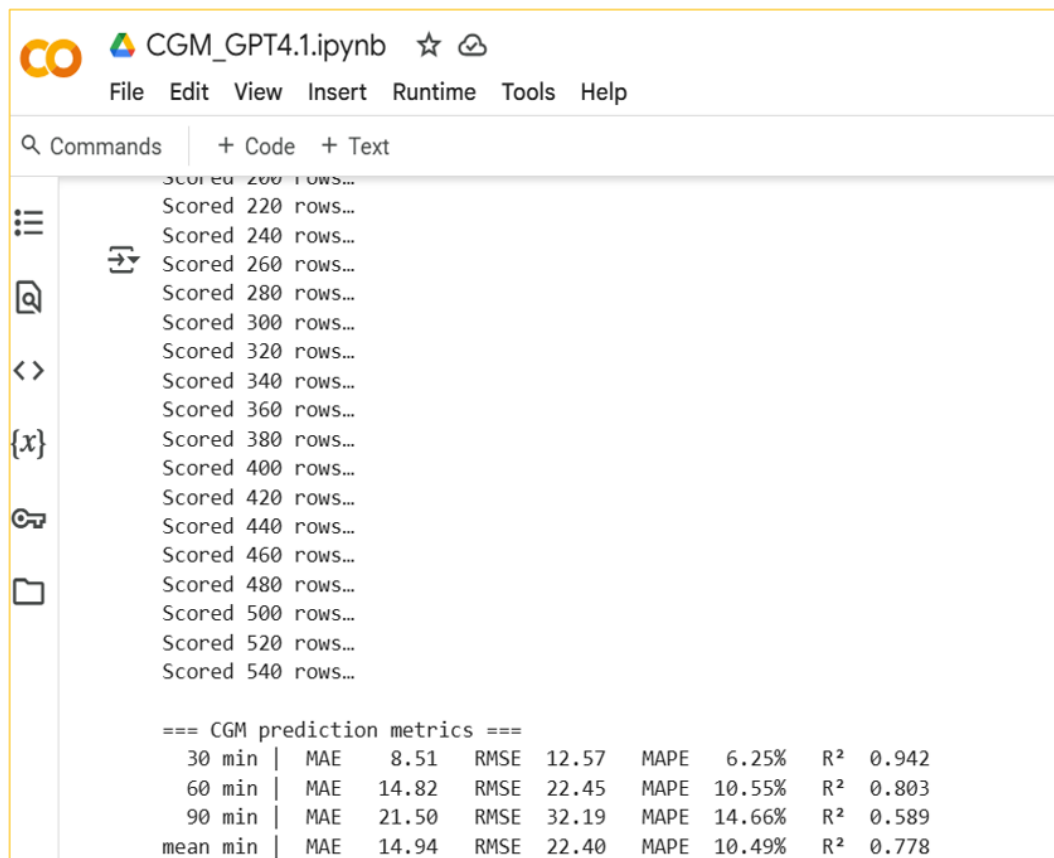
Mini_GPT_result:



```
Scored 200 rows...
Scored 220 rows...
Scored 240 rows...
Scored 260 rows...
Scored 280 rows...
Scored 300 rows...
Scored 320 rows...
Scored 340 rows...
Scored 360 rows...
Scored 380 rows...
Scored 400 rows...
Scored 420 rows...
Scored 440 rows...
Scored 460 rows...
Scored 480 rows...
Scored 500 rows...
Scored 520 rows...
Scored 540 rows...

=== CGM prediction metrics ===
30 min | MAE    9.56  RMSE  14.13  MAPE   6.83%  R²   0.930
60 min | MAE   18.97  RMSE  27.40  MAPE  12.96%  R²   0.736
90 min | MAE   24.62  RMSE  35.50  MAPE  16.30%  R²   0.540
mean min | MAE   17.71  RMSE  25.67  MAPE  12.03%  R²   0.735
```

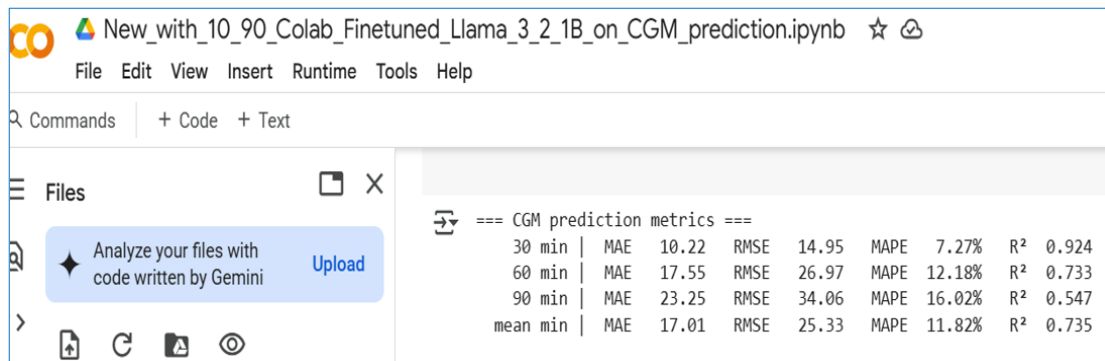
GPT_4.1_result:



```
Scored 200 rows...
Scored 220 rows...
Scored 240 rows...
Scored 260 rows...
Scored 280 rows...
Scored 300 rows...
Scored 320 rows...
Scored 340 rows...
Scored 360 rows...
Scored 380 rows...
Scored 400 rows...
Scored 420 rows...
Scored 440 rows...
Scored 460 rows...
Scored 480 rows...
Scored 500 rows...
Scored 520 rows...
Scored 540 rows...

=== CGM prediction metrics ===
30 min | MAE    8.51  RMSE  12.57  MAPE   6.25%  R²   0.942
60 min | MAE   14.82  RMSE  22.45  MAPE  10.55%  R²   0.803
90 min | MAE   21.50  RMSE  32.19  MAPE  14.66%  R²   0.589
mean min | MAE   14.94  RMSE  22.40  MAPE  10.49%  R²   0.778
```

LLAMA-1B_result:



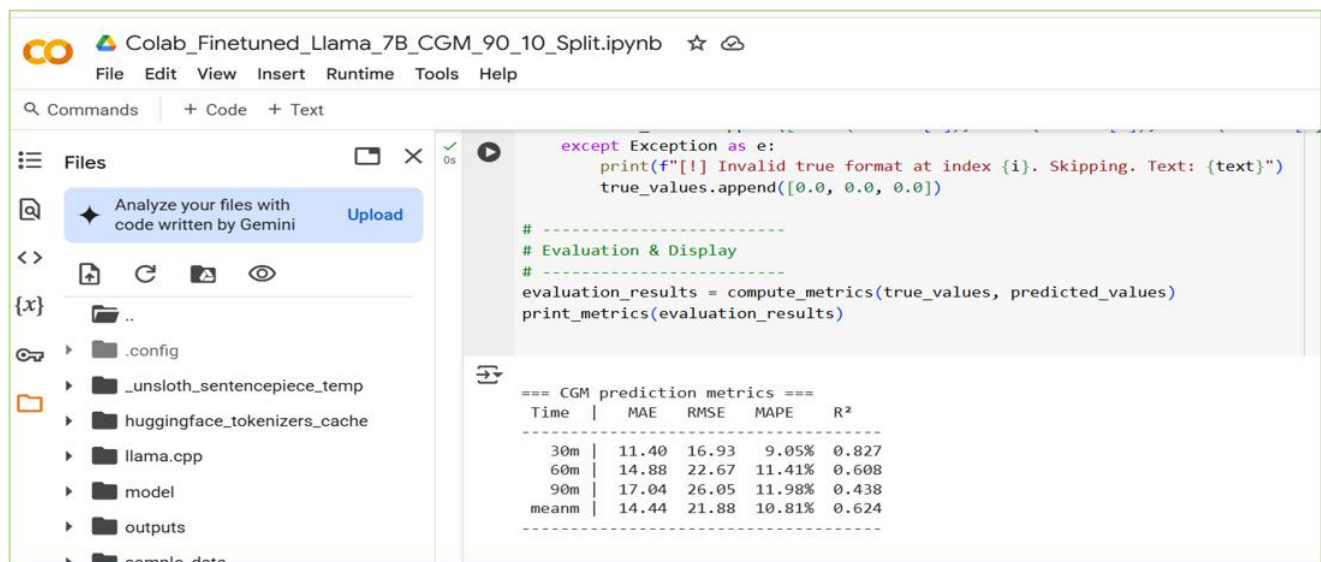
Colab notebook titled "New_with_10_90_Colab_Finetuned_Llama_3_2_1B_on_CGM_prediction.ipynb". The interface shows a menu bar (File, Edit, View, Insert, Runtime, Tools, Help) and a toolbar with "Commands", "+ Code", and "+ Text". A sidebar on the left displays "Files" and a prompt to "Analyze your files with code written by Gemini" with an "Upload" button. The main code cell contains the following output:

```

=== CGM prediction metrics ===
30 min | MAE  10.22  RMSE  14.95  MAPE  7.27%  R²  0.924
60 min | MAE  17.55  RMSE  26.97  MAPE  12.18%  R²  0.733
90 min | MAE  23.25  RMSE  34.06  MAPE  16.02%  R²  0.547
mean min | MAE  17.01  RMSE  25.33  MAPE  11.82%  R²  0.735

```

LLAMA-7B_result:



Colab notebook titled "Colab_Finetuned_Llama_7B_CGM_90_10_Split.ipynb". The interface shows a menu bar (File, Edit, View, Insert, Runtime, Tools, Help) and a toolbar with "Commands", "+ Code", and "+ Text". A sidebar on the left displays "Files" and a prompt to "Analyze your files with code written by Gemini" with an "Upload" button. The main code cell contains the following code and output:

```

except Exception as e:
    print(f"[!] Invalid true format at index {i}. Skipping. Text: {text}")
    true_values.append([0.0, 0.0, 0.0])

# -----
# Evaluation & Display
# -----
evaluation_results = compute_metrics(true_values, predicted_values)
print_metrics(evaluation_results)

```

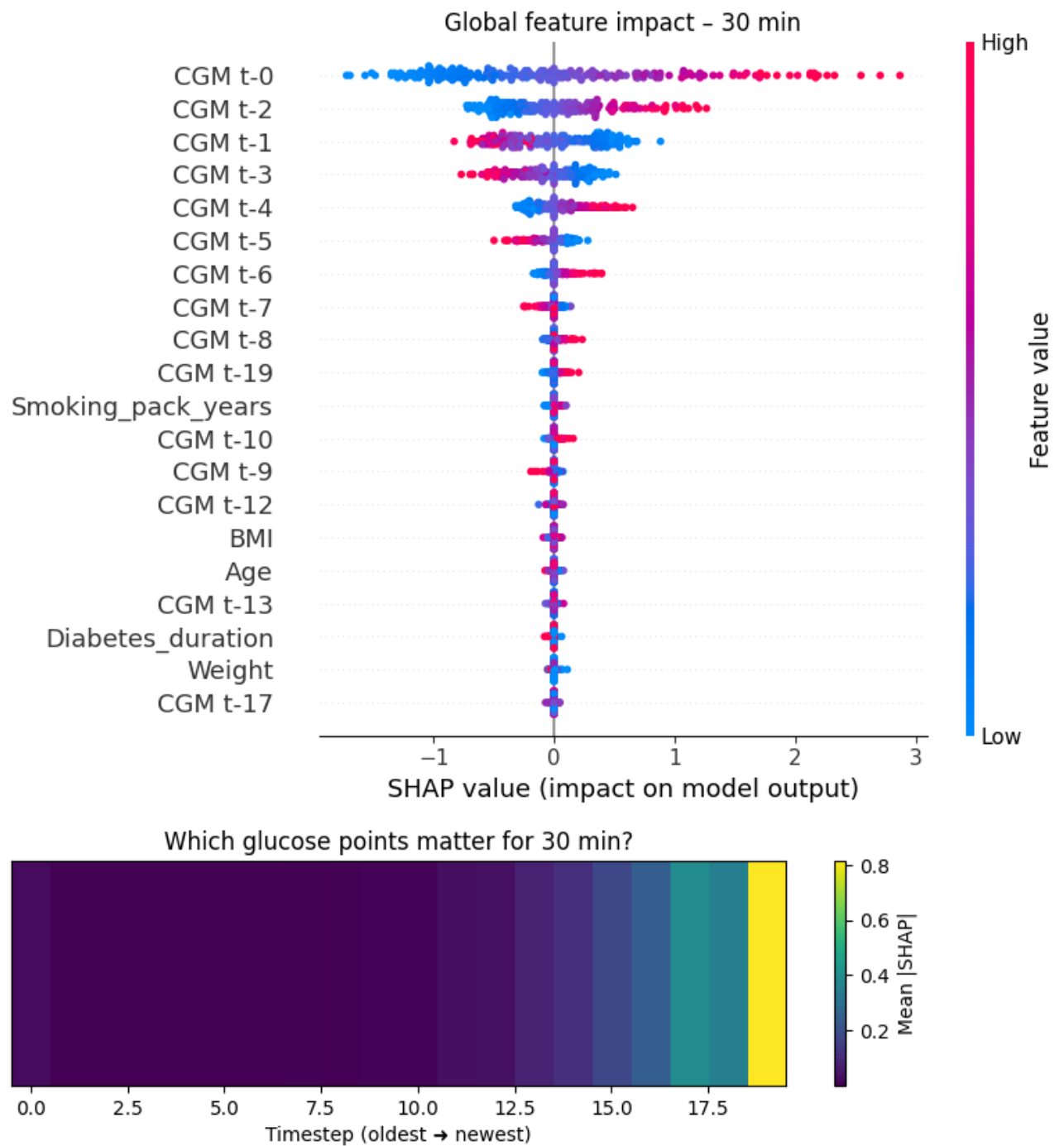
```

=== CGM prediction metrics ===
Time | MAE  RMSE  MAPE  R²
-----
30m | 11.40 16.93  9.05%  0.827
60m | 14.88 22.67 11.41%  0.608
90m | 17.04 26.05 11.98%  0.438
meanm | 14.44 21.88 10.81%  0.624
-----

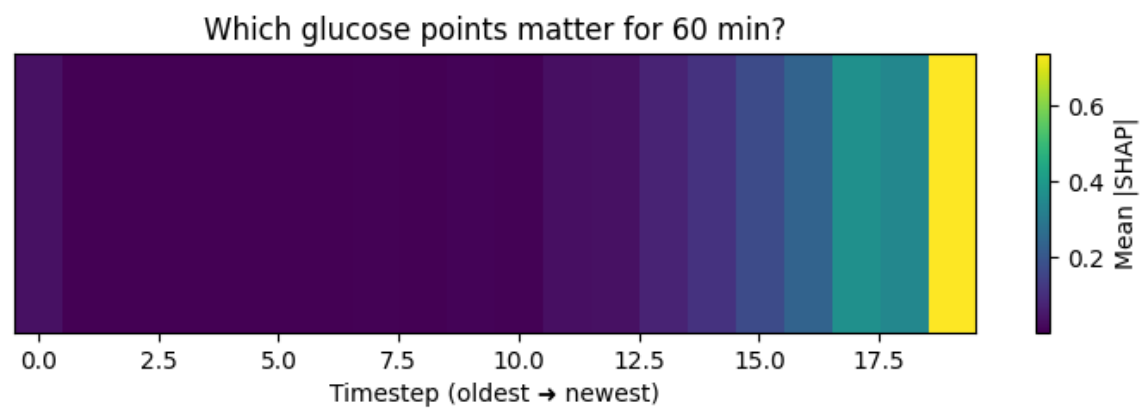
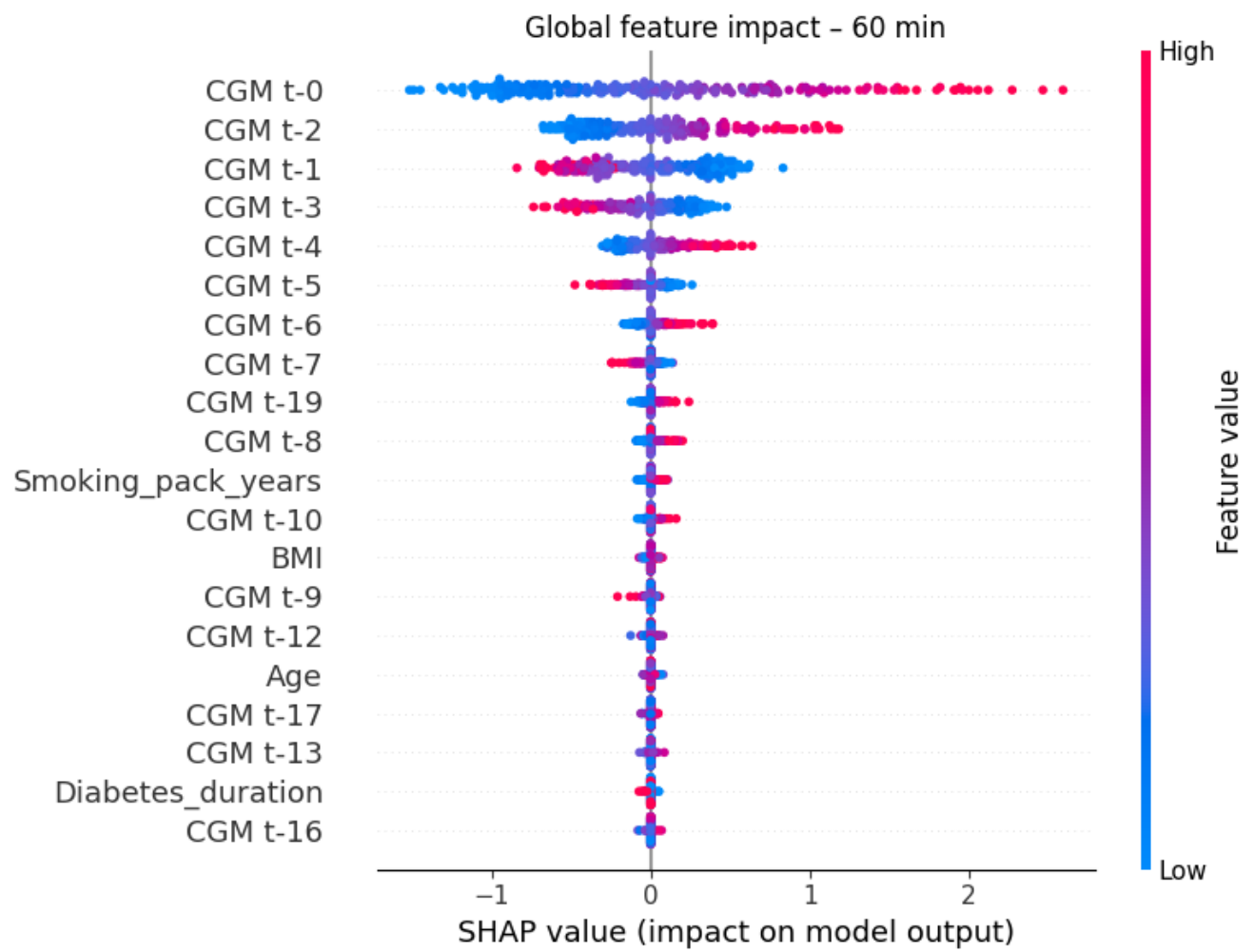
```


2. XAI result:

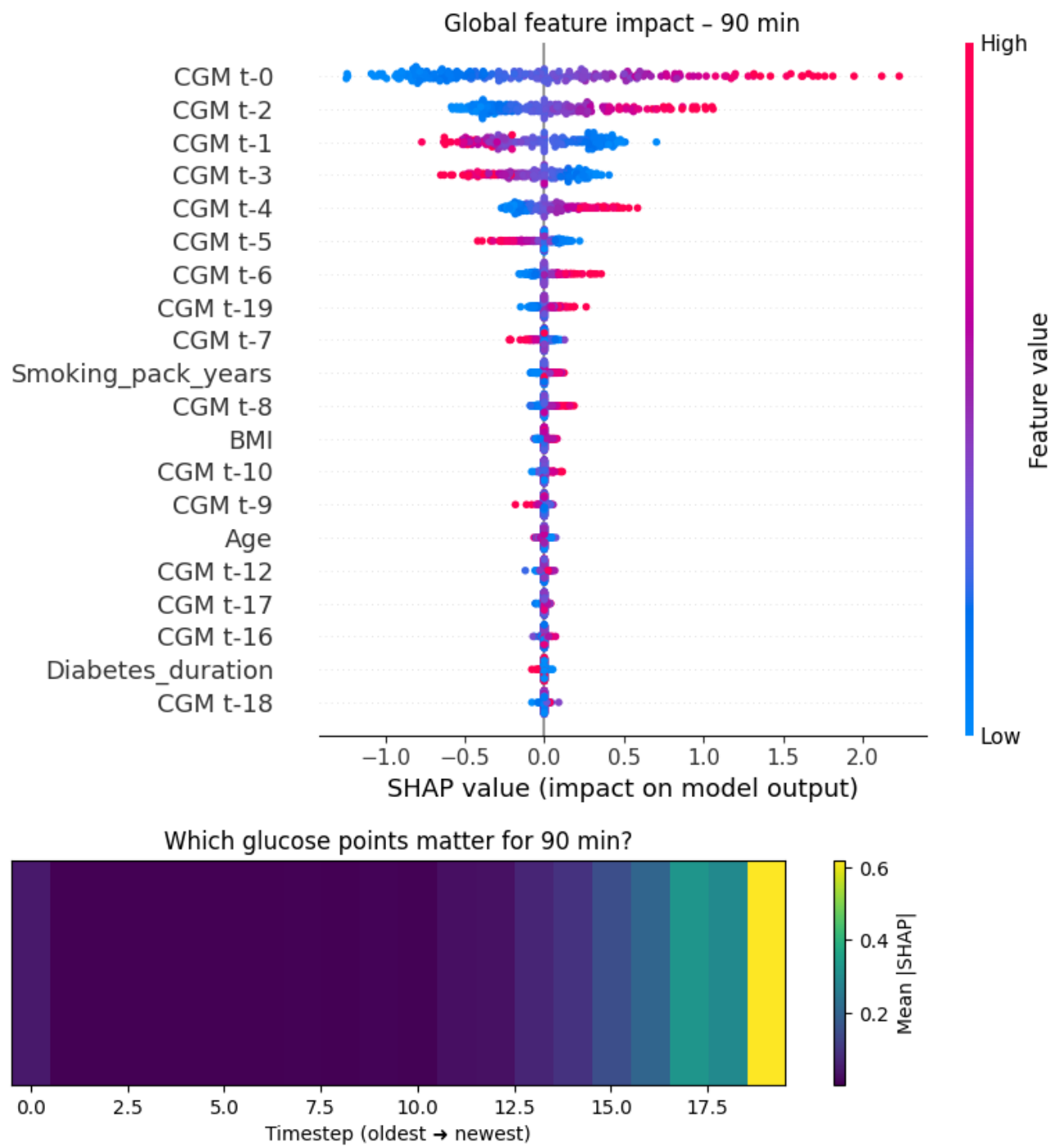
SHAP_30min:



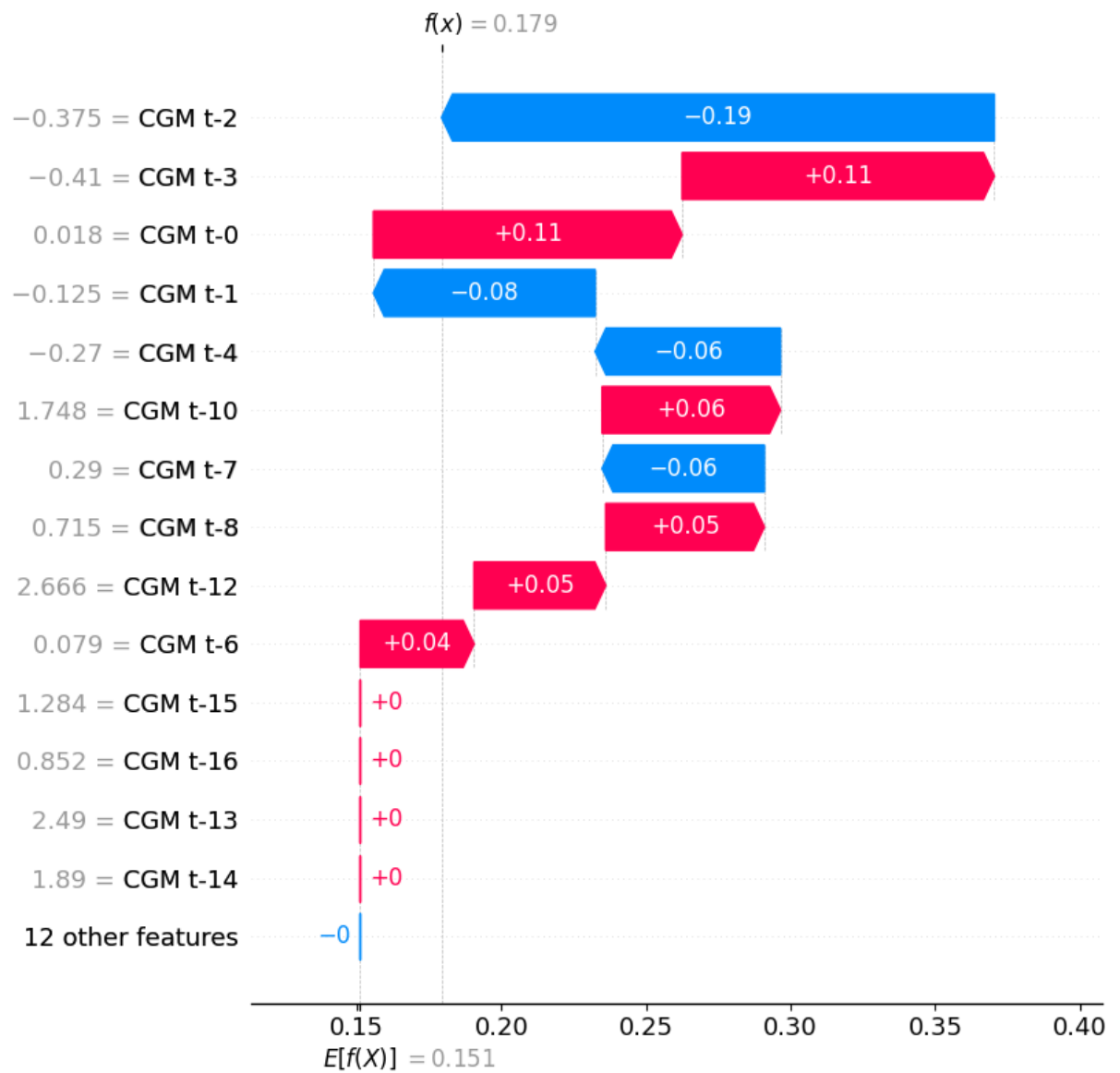
SHAP_60min:



SHAP_90min:



SHAP_Waterfall:



3. GPT Interpretability results:

prompt 1:

rawan / Default project

Playground Dashboard Docs API reference

CGM v1

Compare Optimize Evaluate Update

Model gpt-4.1
text.format: text temp: 1.00 tokens: 2048 top_p: 1.00 store: true

Variables Create...

Tools Create...

System message
Describe desired model behavior (tone, tool usage, response style)

Add messages to prompt

User

You are a medical AI that predicts blood glucose. Given the patient characteristics and CGM history of 20 values recorded each 15 min, predict blood glucose values for the next 30, 60, and 90 minutes. Return predictions in mg/dL. The patient features are: (BMI (kg/m2): 26.4; Age (years): 52.0; Weight (kg): 81.0; Smoking History (pack year): 0.0; Duration of diabetes (years): 20.0; Gender (Female=1, Male=2): 2.0). The CGM recoding of past values each 15 min of the patient 20: 113.4, 115.2, 120.6, 122.4, 118.8, 109.8, 102.6, 93.6, 86.4, 81.0, 75.6, 72.0, 73.8, 77.4, 77.4, 77.4, 88.2, 102.6, 104.4, 97.2

Assistant

Let's analyze the provided data and give you a blood glucose prediction for the next 30, 60, and 90 minutes.

Patient Characteristics

- BMI: 26.4 (overweight range)
- Age: 52 years
- Weight: 81.0 kg
- Smoking: 0.0 pack-years (non-smoker)
- Duration of Diabetes: 20 years (long-term)
- Gender: Male

CGM Glucose Values (last 5 hours)

20 sequential readings (most recent last, interval: 15 min):

Collapse

Chat with your prompt...

Auto-clear

rawan / Default project

Playground Dashboard Docs API reference

CGM v1

Compare Optimize Evaluate Update

Model gpt-4.1
text.format: text temp: 1.00 tokens: 2048 top_p: 1.00 store: true

Variables Create...

Tools Create...

System message
Describe desired model behavior (tone, tool usage, response style)

Add messages to prompt

CGM Glucose Values (last 5 hours)

20 sequential readings (most recent last, interval: 15 min):

113.4, 115.2, 120.6, 122.4, 118.8, 109.8, 102.6, 93.6, 86.4, 81.0, 75.6, 72.0, 73.8, 77.4, 77.4, 77.4, 88.2, 102.6, 104.4, 97.2

Recency & Trend:

- First 5 readings: slightly high-normal, then steady drop.
- Mid-sequence: steady, low-normal range (72-77).
- Last 4 readings: rising trend again.

Recent readings (last 4 values):
88.2, 102.6, 104.4, 97.2 (most recent)

Extrapolation/Prediction Approach

You have a clear nadir (-72), followed by a steady rise back up to ~104, then slight drop. No sudden jumps. The increase is modest and could slow down. With a long diabetic history and no obesity/smoking factors, the patient may have some impaired insulin response but no acute insult.

Blood glucose typically does not oscillate wildly unless there is carbohydrate intake/insulin given, which is not specified.

Collapse

Chat with your prompt...

Auto-clear

rawan / Default project

PlaygroundDashboardDocsAPI reference

CGM v1

CompareOptimizeEvaluateUpdate

Modelgpt-4.1
text_format: text temp: 1.00 tokens: 2048 top_p: 1.00 store: true

VariablesCreate...

ToolsCreate...

System message
Describe desired model behavior (tone, tool usage, response style)

Add messages to prompt

Calculation

Let's calculate the recent rate of change as a simple linear extrapolation:

- Change from 88.2 → 102.6 ($\Delta=14.4$ in 15 min)
- 102.6 → 104.4 ($\Delta=1.8$ in 15 min)
- 104.4 → 97.2 ($\Delta= -7.2$ in 15 min)

Average change over last 3 intervals:

- $(14.4 + 1.8 + (-7.2)) / 3 = 3.0$ mg/dL per 15 min

Extrapolation (assuming trend continues):

- 30 min ahead (2 intervals): $97.2 + 2 \times 3.0 = 103.2$ mg/dL
- 60 min ahead (4 intervals): $97.2 + 4 \times 3.0 = 109.2$ mg/dL
- 90 min ahead (6 intervals): $97.2 + 6 \times 3.0 = 115.2$ mg/dL

Consideration for Smoothing

Since glucose trends usually revert to the mean and the uptick might be due to a recent meal (not specified), the actual increases may be slightly less. To avoid overshooting, let's smooth by a factor of 0.75:

- $3.0 \times 0.75 = 2.25$ mg/dL per 15 min

Collapse

Chat with your prompt...

Auto-clear

rawan / Default project

PlaygroundDashboardDocsAPI reference

CGM v1

CompareOptimizeEvaluateUpdate

Modelgpt-4.1
text_format: text temp: 1.00 tokens: 2048 top_p: 1.00 store: true

VariablesCreate...

ToolsCreate...

System message
Describe desired model behavior (tone, tool usage, response style)

Add messages to prompt

Since glucose trends usually revert to the mean and the uptick might be due to a recent meal (not specified), the actual increases may be slightly less. To avoid overshooting, let's smooth by a factor of 0.75:

- $3.0 \times 0.75 = 2.25$ mg/dL per 15 min

Adjusted Predictions

- 30 min: $97.2 + 2 \times 2.25 = 101.7$ mg/dL
- 60 min: $97.2 + 4 \times 2.25 = 106.2$ mg/dL
- 90 min: $97.2 + 6 \times 2.25 = 110.7$ mg/dL

Predictions

Next 30 minutes: ~101.7 mg/dL

Next 60 minutes: ~106.2 mg/dL

Next 90 minutes: ~110.7 mg/dL

These predictions assume no carbohydrate intake or insulin administration occurs and are based on the recent trend and patient background.

Collapse

Chat with your prompt...

Auto-clear

prompt 1_explanation:

rawan / Default project

Playground Dashboard Docs API reference

CGM v1

Model gpt-4.1
text.format: text temp: 1.00 tokens: 2048 top.p: 1.00 store: true

Variables Create...

Tools Create...

System message
Describe desired model behavior (tone, tool usage, response style)

Add messages to prompt

Compare Optimize Evaluate Update

User
Can you explain how you obtained these results, given that the true glucose values at 30, 60, and 90 minutes are 95.4, 99.0, and 100.8 respectively? Additionally, what had the greatest influence on your predictions? the glucose readings or the static patient features? and why?

Assistant
Thank you for your follow-up and for sharing the true future glucose values.

1. Explanation of How the Results Were Obtained

- Recent History Focus:
Given the absence of information about food intake, insulin, or other acute interventions, I primarily leveraged the most recent sequence of CGM (glucose) values—this provides the strongest evidence for near-future glucose levels.
- Trend Analysis:
I calculated the rate of change over the last three 15-minute intervals:
 - 88.2 → 102.6: +14.4 mg/dL
 - 102.6 → 104.4: +1.8 mg/dL
 - 104.4 → 97.2: -7.2 mg/dL The simple average of these is +3.0 mg/dL per 15 minutes, which I then applied forward in time for 2, 4, and 6 intervals (30, 60, 90 mins).
- Smoothing Adjustment:
Realizing this approach can over-predict future values (especially if the increases are just a short-lived bump), I introduced a smoothing factor (here: 0.75), resulting in an "adjusted" rate (+2.25 mg/dL/15 min).

Collapse ^

Chat with your prompt...

Auto-clear

rawan / Default project

Playground Dashboard Docs API reference

CGM v1

Model gpt-4.1
text.format: text temp: 1.00 tokens: 2048 top.p: 1.00 store: true

Variables Create...

Tools Create...

System message
Describe desired model behavior (tone, tool usage, response style)

Add messages to prompt

Compare Optimize Evaluate Update

- Smoothing Adjustment:
Realizing this approach can over-predict future values (especially if the increases are just a short-lived bump), I introduced a smoothing factor (here: 0.75), resulting in an "adjusted" rate (+2.25 mg/dL/15 min).
- Extrapolation:
Starting from the most recent glucose (97.2 mg/dL), I projected future values by adding this adjusted trend.

2. Comparison to True Values

- Predicted (mine):
 - 30 min: ~101.7 mg/dL
 - 60 min: ~106.2 mg/dL
 - 90 min: ~110.7 mg/dL
- True Values:
 - 30 min: 95.4
 - 60 min: 99.0
 - 90 min: 100.8

Observation:
The model's prediction for the 90-minute value (~110.7 mg/dL) is notably higher than the true value (100.8 mg/dL), indicating a slight over-prediction.

Collapse ^

Chat with your prompt...

Auto-clear

rawan / Default project

PlaygroundDashboardDocsAPI reference

CGM v1

CompareOptimizeEvaluateUpdate

Modelgpt-4.1
text-format: text temp: 1.00 tokens: 2048 top_p: 1.00 store: true

VariablesCreate...

ToolsCreate...

System message
Describe desired model behavior (tone, tool usage, response style)

Add messages to prompt

Observation:

My predictions overshot the actual values, mainly because the recent sharp rise was not sustained—it was either an outlier or a short-term fluctuation, and the broader pattern was a stabilization at a lower value.

3. Greatest Influence: Glucose Readings vs. Patient Features

The greatest influence on my prediction comes from the CGM (recent glucose readings), much more so than the patient features.

Why?

- CGM Data Represents Immediate Physiology:
The glucose values are the most direct, high-resolution indicators of what's happening metabolically in real time. They show current trends (rising, falling, stable). Near-term predictions (within 90 minutes) are almost entirely driven by the *momentum* and *pattern* of the most recent numbers.
- Patient Features are Baseline Modifiers:
Age, BMI, duration, gender, and smoking status modify overall risk (likelihood of hypo/hyperglycemia, response to interventions, baseline insulin resistance), but for **short-term prediction** based only on the last few hours, they are much less influential.
- Mathematical Reasoning:
If a model is built, such as a regression or neural net, the coefficients for lagged glucose values (recent CGM data) are always far larger for short-term forecasting than those for static features, except under

Chat with your prompt...Auto-clear

rawan / Default project

PlaygroundDashboardDocsAPI reference

CGM v1

CompareOptimizeEvaluateUpdate

Modelgpt-4.1
text-format: text temp: 1.00 tokens: 2048 top_p: 1.00 store: true

VariablesCreate...

ToolsCreate...

System message
Describe desired model behavior (tone, tool usage, response style)

Add messages to prompt

CGM data is always far larger for short-term forecasting than those for static features, except under very specific scenarios (e.g., acute illness, major weight loss/gain, or after decades-long changes).

- Your Example's Error Source:
I overestimated the future because I placed too much weight on a brief, steep increase (recent CGM values), while the longer-term average would suggest more stability near the low-normal range.

Summary Table

Greatest Influence	
Short-term BG prediction (30-90 min)	Recent CGM values
Long-term risk (years)	Patient features + CGM trends

In practice: For near-term blood glucose predictions with CGM data available, **recent glucose readings always dominate the prediction**, while patient characteristics are more important for longer-term risk stratification, insulin dosing algorithms, and general diabetes management plans.

If you want the prediction to be more accurate with respect to only stable periods (ignoring outliers in recent data), you might use a moving average of the last 4-5 readings, or apply more weight to established baseline levels rather than short-term excursions.

Chat with your prompt...Auto-clear

prompt 2:

rawan / Default project

Playground Dashboard Docs API reference

CGM v1

Model gpt-4.1
text..format: text temp: 1.00 tokens: 2048 top_p: 1.00 store: true

Variables Create...

Tools Create...

System message
Describe desired model behavior (tone, tool usage, response style)

Add messages to prompt

User

You are a medical AI that predicts blood glucose. Given the patient characteristics and CGM history of 20 values recorded each 15 min, predict blood glucose values for the next 30, 60, and 90 minutes. Return predictions in mg/dL. The patient features are: (BMI (kg/m2): 23.6; Age (years): 57.0; Weight (kg): 67.4; Smoking History (pack year): 0.0; Duration of diabetes (years): 25.0; Gender (Female=1, Male=2): 2.0). The CGM recoding of past values each 15 min of the patient 20: 153.0, 147.6, 145.8, 142.2, 142.2, 144.0, 149.4, 147.6, 133.2, 117.0, 104.4, 97.2, 93.6, 90.0, 93.6, 104.4, 108.0, 106.2, 104.4, 108.0

Assistant

Let's analyze and predict the next blood glucose values.

Patient Characteristics

- BMI: 23.6 (Normal)
- Age: 57 years
- Weight: 67.4 kg
- Smoking History: 0
- Duration of diabetes: 25 years (long-standing)
- Gender: Male

Chat with your prompt...

rawan / Default project

Playground Dashboard Docs API reference

CGM v1

Model gpt-4.1
text..format: text temp: 1.00 tokens: 2048 top_p: 1.00 store: true

Variables Create...

Tools Create...

System message
Describe desired model behavior (tone, tool usage, response style)

Add messages to prompt

CGM History (Last 5 hours, 15 min intervals; newest last)

153.0, 147.6, 145.8, 142.2, 142.2, 144.0, 149.4, 147.6, 133.2, 117.0, 104.4, 97.2, 93.6, 90.0, 93.6, 104.4, 108.0, 106.2, 104.4, 108.0

Observe the recent 5 values:

- 93.6, 104.4, 108.0, 106.2, 104.4, 108.0

Notice:

- Early: Downward trend, then stabilization.
- Later: Mild upward drift, no major spike.

Calculate Recent Trends

Last 4 intervals:

- 93.6 → 104.4 = +10.8
- 104.4 → 108.0 = +3.6
- 108.0 → 106.2 = -1.8
- 106.2 → 104.4 = -1.8

Chat with your prompt...

rawan / Default project

PlaygroundDashboardDocsAPI reference

CGM v1

CompareOptimizeEvaluateUpdate

Modelgpt-4.1
text.format: text temp: 1.00 tokens: 2048 top.p: 1.00 store: true

VariablesCreate...

ToolsCreate...

System message
Describe desired model behavior (tone, tool usage, response style)

Add messages to prompt

106.2 → 104.4 = -1.8

104.4 → 108.0 = +3.6

Let's average the change over last 4 intervals:

$(+10.8 + 3.6 + (-1.8) + (-1.8) + 3.6) / 5 = 2.88 \text{ mg/dL per 15 min}$

Current value (most recent): 108.0 mg/dL

Prediction

Project forward using trend per 15 min (cautiously, as the rate of rise is modest):

30 min (2 × 2.88)

$= 108.0 + 2 \times 2.88 = 113.8 \text{ mg/dL}$

60 min (4 × 2.88)

$= 108.0 + 4 \times 2.88 = 119.5 \text{ mg/dL}$

90 min (6 × 2.88)

Collapse

Chat with your prompt...

Auto-clear

rawan / Default project

PlaygroundDashboardDocsAPI reference

CGM v1

CompareOptimizeEvaluateUpdate

Modelgpt-4.1
text.format: text temp: 1.00 tokens: 2048 top.p: 1.00 store: true

VariablesCreate...

ToolsCreate...

System message
Describe desired model behavior (tone, tool usage, response style)

Add messages to prompt

$= 108.0 + 6 \times 2.88 = 125.3 \text{ mg/dL}$

Predicted Glucose (mg/dL)

30 minutes: 113.8

60 minutes: 119.5

90 minutes: 125.3

Reasoning

The recent values show a gentle rise after a significant drop.

No evidence of sharp spikes or abrupt change.

The prediction is almost entirely driven by the immediate CGM values and their short-term trend, as longer-term patient characteristics mainly affect baseline, not short-term direction.

If you want a more conservative forecast (expecting the rise to slow), you could use a smaller slope, but

Collapse

Chat with your prompt...

Auto-clear

rawan / Default project

PlaygroundDashboardDocsAPI reference

CGM v1

CompareOptimizeEvaluateUpdate

Modelgpt-4.1
text.format: text temp: 1.00 tokens: 2048 top.p: 1.00 store: true

VariablesCreate...

ToolsCreate...

System message
Describe desired model behavior (tone, tool usage, response style)

Add messages to prompt

Predicted Glucose (mg/dL)

30 minutes: 113.8

60 minutes: 119.5

90 minutes: 125.3

Reasoning

The recent values show a gentle rise after a significant drop.

No evidence of sharp spikes or abrupt change.

The prediction is **almost entirely driven by the immediate CGM values and their short-term trend**, as longer-term patient characteristics mainly affect baseline, not short-term direction.

If you want a more conservative forecast (expecting the rise to slow), you could use a smaller slope, but based on the current pattern, the provided numbers are the most likely based on historical data.

Collapse

Chat with your prompt...

Auto-clear

19

prompt 2_explanation:

rawan / Default project

Playground Dashboard Docs API reference

CGM v1

Model gpt-4.1
text.format: text temp: 1.00 tokens: 2048 top_p: 1.00 store: true

Variables Create...

Tools Create...

System message
Describe desired model behavior (tone, tool usage, response style)

Add messages to prompt

User

Can you explain how you obtained these results, given that the true glucose values at 30, 60, and 90 minutes are 120.6, 129.6 and 135.0 respectively? Additionally, what had the greatest influence on your predictions? the glucose readings or the static patient features? and why?

Assistant

How were the results obtained?

Step-by-step process:

1. Review the CGM trend:
I examined the most recent values from your CGM series, particularly focusing on the last 5-6 points. This section went from a low (around 90) and rose toward 108 mg/dL at the present, showing a modest but consistent upward trend.

2. Calculate average rate of change:
To project the next values, I computed the average increase per 15 minutes over these recent readings:

- 93.6 → 104.4: +10.8
- 104.4 → 108.0: +3.6
- 108.0 → 106.2: -1.8
- 106.2 → 104.4: -1.8
- 104.4 → 108.0: +3.6

Averaged over these five intervals: +2.88 mg/dL per 15 minutes

3. Linear extrapolation:
Collapse

Chat with your prompt...

Auto-clear

rawan / Default project

Playground Dashboard Docs API reference

CGM v1

Model gpt-4.1
text.format: text temp: 1.00 tokens: 2048 top_p: 1.00 store: true

Variables Create...

Tools Create...

System message
Describe desired model behavior (tone, tool usage, response style)

Add messages to prompt

3. Linear extrapolation:
Collapse

4. Predictions:

- 30 min: $108.0 + 2 \times 2.88 = 113.8$
- 60 min: $108.0 + 4 \times 2.88 = 119.5$
- 90 min: $108.0 + 6 \times 2.88 = 125.3$

Comparison with True Values

Your provided true future values:

- 30 min: 120.6
- 60 min: 129.6
- 90 min: 135.0

My predictions are too low.
This happened because the underlying trend in the last few points under-represented the true, more significant rise that followed. The most recent CGM readings suggested only a gentle increase, so a simple linear projection under-shot the stronger upcoming rise.

Collapse

Chat with your prompt...

Auto-clear

rawan / Default project

Playground Dashboard Docs API reference

CGM v1

Compare Optimize Evaluate Update

Modelgpt-4.1text.format: text temp: 1.00 tokens: 2048 top_p: 1.00 store: true

VariablesCreate...

ToolsCreate...

System messageDescribe desired model behavior (tone, tool usage, response style)

Add messages to prompt

What influences the predictions most: Recent CGM vs. Static Patient Features?

The CGM data (recent glucose readings) had by far the greatest influence—much more than static patient features.

Why?

- Short-term forecasting (30–90 minutes):
 - Short-term trends in glucose are best predicted by the *immediate trajectory* revealed in recent CGM values. The “momentum” of glucose—whether it’s rising, falling, or stable—reflected in the recent points is the best predictor unless interrupted by a meal or insulin (not specified here).
- Patient characteristics:
 - Features like BMI, age, diabetes duration, etc. influence baseline risks (how high/low glucose ranges are likely to be over months/years), how aggressive therapy might need to be, and risk of variability.
 - For predicting the next few points, this “static” information doesn’t move the trend—it only shapes the context or may slightly modulate the rate if many similar patients are used in a model. For example, someone with very long diabetes might have slower correction, but this adjustment is small compared to the effect of recent glucose direction.
- Mathematical reality:
 - If you train forecasting models, coefficients for recent glucose values (lags) are much larger than those for static patient traits for “next several values” prediction.

Collapse

Chat with your prompt...Auto-clear

rawan / Default project

Playground Dashboard Docs API reference

CGM v1

Compare Optimize Evaluate Update

Modelgpt-4.1text.format: text temp: 1.00 tokens: 2048 top_p: 1.00 store: true

VariablesCreate...

ToolsCreate...

System messageDescribe desired model behavior (tone, tool usage, response style)

Add messages to prompt

those for static patient traits for “next several values” prediction.

Why did my model underpredict future glucose?

- Recent CGM values under-represented the upcoming rise.
- If there was a meal, stress, or less insulin recently, the model couldn’t know: With only past glucose readings, the model can’t see future inflection points if they aren’t reflected yet in the past data.
- Forecasts for 90 min are less certain than for 30 min—unforeseen inflections or ongoing (perhaps delayed) glucose absorption from a meal can make future values higher.

In summary:

- The series of recent CGM glucose readings had the greatest influence on the prediction, because blood glucose dynamics change hour-to-hour based primarily on recent direction, not baseline patient features.
- Patient features are more important for defining general risk and setting long-term management, not for near-term point estimates if recent CGM is available.

Collapse

Chat with your prompt...Auto-clear

Prompt 3:

rawan / Default project

Playground Dashboard Docs API reference

CGM v1

Compare Optimize Evaluate Update

Modelgpt-4.1text.format: text temp: 1.00 tokens: 2048 top_p: 1.00 store: true

VariablesCreate...

ToolsCreate...

System messageDescribe desired model behavior (tone, tool usage, response style)

Add messages to prompt

User

You are a medical AI that predicts blood glucose. Given the patient characteristics and CGM history of 20 values recorded each 15 min, predict blood glucose values for the next 30, 60, and 90 minutes. Return predictions in mg/dL. The patient features are: (BMI (kg/m2): 34.1; Age (years): 42.0; Weight (kg): 84.1; Smoking History (pack year): 0.0; Duration of diabetes (years): 7.0; Gender (Female=1, Male=2): 1.0). The CGM recoding of past values each 15 min of the patient 20: 127.8, 126.0, 120.6, 117.0, 129.6, 131.4, 126.0, 124.2, 122.4, 118.8, 115.2, 118.8, 113.4, 120.6, 145.8, 174.6, 190.8, 196.2, 190.8

Assistant

Let's analyze your data and make glucose predictions for the next 30, 60, and 90 minutes.

Patient Characteristics

- BMI: 34.1 (Obese)
- Age: 42 years
- Weight: 84.1 kg
- Smoking History: 0.0 pack-years
- Duration of diabetes: 7.0 years (moderate duration)
- Gender: Female

Collapse

Chat with your prompt...

Auto-clear

rawan / Default project

Playground Dashboard Docs API reference

CGM v1

Compare Optimize Evaluate Update

Modelgpt-4.1text.format: text temp: 1.00 tokens: 2048 top_p: 1.00 store: true

VariablesCreate...

ToolsCreate...

System messageDescribe desired model behavior (tone, tool usage, response style)

Add messages to prompt

Recent CGM History (past 5 hours, every 15 min, newest last)

127.8, 126.0, 120.6, 117.0, 129.6, 131.4, 126.0, 124.2, 122.4, 118.8, 115.2, 118.8, 113.4, 120.6, 145.8, 174.6, 190.8, 196.2, 190.8

Recent Trend (last 6 values, newest last):

- 120.6, 145.8, 174.6, 190.8, 196.2, 190.8
- There is a clear sharp rise over the past 90 minutes, briefly peaking and plateauing just under 200 mg/dL.

Calculate Recent Rate of Change

15-min intervals:

- 120.6 → 145.8: +25.2
- 145.8 → 174.6: +28.8
- 174.6 → 190.8: +16.2
- 190.8 → 196.2: +5.4
- 196.2 → 190.8: -5.4

Average recent rise (last four intervals):

Collapse

Chat with your prompt...

Auto-clear

22

rawan / Default project

PlaygroundDashboardDocsAPI reference

CGM v1

CompareOptimizeEvaluateUpdate

Modelgpt-4.1
text.format: text temp: 1.00 tokens: 2048 top_p: 1.00 store: true

VariablesCreate...

ToolsCreate...

System message
Describe desired model behavior (tone, tool usage, response style)

Add messages to prompt

Average recent rise (last four intervals):
 $(+28.8 + 16.2 + 5.4 + (-5.4)) / 4 = 11.25$ mg/dL per 15 minutes

However, the last value decreased, indicating the peak may be passing, so it's wise to assign less weight to the earlier, sharper rises.

Let's also consider a smoothed average over the last three intervals:
 $(16.2 + 5.4 + (-5.4)) / 3 = 5.4$ mg/dL per 15 min

The last two intervals: +5.4, -5.4—showing a possible start of plateau or slight drop.

Prediction Approach

Given that the peak may be reached or slightly passed, a sustained plateau or slight decline is more likely than continued sharp rise.

Let's average the last two changes:
 $(+5.4 + (-5.4)) / 2 = 0$ mg/dL per 15 min.

To accommodate a possible gentle decline or short plateau, let's project stability or a very gentle drop, as this is common following a postprandial glucose spike.

Current value: 190.8 mg/dL
Collapse

Chat with your prompt...
Auto-clear

rawan / Default project

PlaygroundDashboardDocsAPI reference

CGM v1

CompareOptimizeEvaluateUpdate

Modelgpt-4.1
text.format: text temp: 1.00 tokens: 2048 top_p: 1.00 store: true

VariablesCreate...

ToolsCreate...

System message
Describe desired model behavior (tone, tool usage, response style)

Add messages to prompt

Predicted Values

30 min (2 steps):
* $190.8 + (2 \times 0) = 190.8$ mg/dL

60 min (4 steps):
* $190.8 + (4 \times 0) = 190.8$ mg/dL

90 min (6 steps):
* $190.8 + (6 \times 0) = 190.8$ mg/dL

But, as a realistic adjustment (glucose usually drops a little after a peak), let's model a gentle correction:

* Expected decline: about 2 mg/dL per 15 min (conservative estimate after peak)

Adjusted predictions:

* 30 min: $190.8 - (2 \times 2) = 186.8$ mg/dL

* 60 min: $190.8 - (4 \times 2) = 182.8$ mg/dL

* 90 min: $190.8 - (6 \times 2) = 178.8$ mg/dL

Predicted Glucose (mg/dL)
Collapse

Chat with your prompt...
Auto-clear

rawan / Default project

PlaygroundDashboardDocsAPI reference

CGM v1

Modelgpt-4.1
text.format: text temp: 1.00 tokens: 2048 top_p: 1.00 store: true

VariablesCreate...

ToolsCreate...

System message
Describe desired model behavior (tone, tool usage, response style)

Add messages to prompt

CompareOptimizeEvaluateUpdate

Predicted Glucose (mg/dL)

- 30 minutes: 186.8
- 60 minutes: 182.8
- 90 minutes: 178.8

Reasoning:

- There was a sharp rise to a peak, now possibly correcting downward or plateauing.
- Small declines are typical post-peak unless food or insulin alters the trend.

Greatest influence on these predictions:
Recent CGM trends. The patient's static features (BMI, age, gender, duration, etc.) help define baseline risk and the magnitude of changes over time, but for short-term prediction, the most recent CGM values—especially the sharp recent rise and very recent plateau—determine the next hour's glycemic trend. Thus, CGM values outweigh the static characteristics when projecting the next 30-90 minutes.
[Collapse](#)

Chat with your prompt...

Auto-clear

24

prompt 3_explanation:

rawan / Default project

Playground Dashboard Docs API reference

CGM v1

Compare Optimize Evaluate Update

Modelgpt-4.1text.format: text temp: 1.00 tokens: 2048 top_p: 1.00 store: true

VariablesCreate...

ToolsCreate...

System messageDescribe desired model behavior (tone, tool usage, response style)

Add messages to prompt

User

Can you explain how you obtained these results, given that the true glucose values at 30, 60, and 90 minutes are 178.2, 154.8 and 111.6 respectively? Additionally, what had the greatest influence on your predictions: the glucose readings or the static patient features? and why?

Assistant

Let's break down your questions step by step:

1. How were the predictions obtained?

a. Trend Analysis

I examined your most recent CGM values (starting from about 120.6 and climbing quickly to a peak at 196.2, then a dip to 190.8). I calculated 15-minute interval changes, noticing:

- A sharp rise (as high as +28.8 mg/dL per 15 min)
- The last two intervals: a +5.4, then a -5.4 mg/dL (indicating the rise stopped and plateaued, maybe ready to fail)

b. Short-Term Forecasting

Rather than projecting the earlier sharp rise forward, I averaged the recent changes, especially the last two (near zero), and then added a small expected decline post-peak (a typical pattern in CGM data after a

Chat with your prompt...

Auto-clear

rawan / Default project

Playground Dashboard Docs API reference

CGM v1

Compare Optimize Evaluate Update

Modelgpt-4.1text.format: text temp: 1.00 tokens: 2048 top_p: 1.00 store: true

VariablesCreate...

ToolsCreate...

System messageDescribe desired model behavior (tone, tool usage, response style)

Add messages to prompt

(near zero), and then added a small expected decline post-peak (a typical pattern in CGM data after a spike).

- I used the classic approach that, after a sharp peak, glucose either gently plateaus or begins a gentle fall unless further carbohydrate, exercise, or insulin is introduced.

c. Calculation

30, 60, 90 minutes = 2, 4, 6 intervals ahead

- I started with the latest value (190.8 mg/dL).
- I subtracted a small amount per interval (about 2 mg/dL per 15 min), predicting a gentle correction downwards.

Predicted Trajectory:

- 30 min: 186.8
- 60 min: 182.8
- 90 min: 178.8

d. Missing the True Drop

The true pattern (178.2 → 154.8 → 111.6) reveals a much steeper drop than my conservative, gradual estimate. This suggests a strong glucose-lowering process (like delayed insulin action or sharp endogenous correction after a glucose surge), which isn't always obvious just from the last few data points.

Chat with your prompt...

Auto-clear

rawan / Default project

PlaygroundDashboardDocsAPI reference

CGM v1

CompareOptimizeEvaluateUpdate

Modelgpt-4.1
text.format: text temp: 1.00 tokens: 2048 top_p: 1.00 store: true

VariablesCreate...

ToolsCreate...

System message
Describe desired model behavior (tone, tool usage, response style)

Add messages to prompt

2. Why didn't my prediction match the true values?

I essentially underestimated the speed and magnitude of the drop after the peak. The most recent values showed only a minor dip, so I assumed a gentle correction. However, the true trajectory shows a rapid glucose fall, which often happens after a significant peak if insulin (endogenous or exogenous) "kicks in" after the CGM peak, or if the glucose was from a rapidly absorbed meal.

In short: my method, with the limited data, assumed sluggish correction, but the patient's physiology or treatment induced a swift one.

3. What influenced my predictions the most? Glucose readings vs. patient features?

Most Important: Recent CGM Glucose Readings

Reason:

- Short-term forecasts (15–90 min) are overwhelmingly determined by the recent trend shown in CGM data.
- This is because the short-term direction (rising, plateauing, or falling) reflects immediate physiological processes (recent meal, insulin, activity, stress), and predicting several steps ahead depends almost entirely on the recent momentum in the data.
- Patient features (BMI, age, gender, etc.) mostly affect the typical glucose profile over days to years—

Collapse

Chat with your prompt...

Auto-clear

rawan / Default project

PlaygroundDashboardDocsAPI reference

CGM v1

CompareOptimizeEvaluateUpdate

Modelgpt-4.1
text.format: text temp: 1.00 tokens: 2048 top_p: 1.00 store: true

VariablesCreate...

ToolsCreate...

System message
Describe desired model behavior (tone, tool usage, response style)

Add messages to prompt

- Patient features (BMI, age, gender, etc.) mostly affect the typical glucose profile over days to years—how high glucose tends to go, average insulin sensitivity, baseline risk for complications—but contribute very little information about glucose direction/change over the next hour. They are background context, not short-term predictors.

In mathematical/statistical modeling (and in clinical practice):

- Recent data points ("lags") get the biggest weights for predicting the next point(s).
- Static features adjust risk or mean levels but have low or zero weight in immediate forecasting unless integrated into a complex model with multiple interacting terms.

Example (Your Case):

- Even with high BMI and significant diabetes duration, the very rapid CGM rise was the dominating feature, and my forecast (correctly) focused on this, although I misjudged how quickly the glucose would start falling.

Summary Table

Influence	Short-term BG prediction (30-90 min)
CGM trends	Dominant factor (most important)
Background patient features	Minimal influence

Collapse

Chat with your prompt...

Auto-clear

rawan / Default project

CGM v1

Model gpt-4.1
text.format: text temp: 1.00 tokens: 2048 top_p: 1.00 stone: true

Variables Create...

Tools Create...

System message
Describe desired model behavior (tone, tool usage, response style)

Add messages to prompt

Summary Table

Influence	Short-term BG prediction (30-90 min)
CGM trends	Dominant factor (most important)
Patient static features	Minimal influence

In summary:

- I obtained results by extrapolating recent CGM trends, applying a gentle decline after an apparent peak.
- I missed the true sharp drop, as the recent history didn't fully reveal it yet—a common limitation in short-term, point-to-point CGM forecasting.
- Recent CGM readings** influenced the predictions far more than static features, because these readings directly reflect the physiological momentum guiding near-term glucose changes.

Collapse

Chat with your prompt...

Auto-clear

Prompt 4:

rawan / Default project

Playground Dashboard Docs API reference

CGM v1

Model gpt-4.1
text.format: text temp: 1.00 tokens: 2048 top.p: 1.00 store: true

Variables Create...

Tools Create...

System message
Describe desired model behavior (tone, tool usage, response style)

Add messages to prompt

User

You are a medical AI that predicts blood glucose. Given the patient characteristics and CGM history of 20 values recorded each 15 min, predict blood glucose values for the next 30, 60, and 90 minutes. Return predictions in mg/dL. The patient features are: (BMI (kg/m2): 22.9; Age (years): 55.0; Weight (kg): 63.0; Smoking History (pack year): 0.0; Duration of diabetes (years): 4.0; Gender (Female=1, Male=2): 2.0). The CGM recoding of past values each 15 min of the patient 20: 226.8, 226.8, 226.8, 228.6, 230.4, 228.6, 226.8, 226.8, 226.8, 223.2, 230.4, 228.6, 219.6, 219.6, 216.0, 205.2, 199.8, 201.6, 196.2, 190.8

Assistant
Let's analyze your data and predict the next three blood glucose values:

Patient Characteristics

- BMI: 22.9 (normal)
- Age: 55
- Weight: 63 kg
- Smoking: 0
- Diabetes duration: 4 years
- Gender: Male

Chat with your prompt...

rawan / Default project

Playground Dashboard Docs API reference

CGM v1

Model gpt-4.1
text.format: text temp: 1.00 tokens: 2048 top.p: 1.00 store: true

Variables Create...

Tools Create...

System message
Describe desired model behavior (tone, tool usage, response style)

Add messages to prompt

CGM Data (15-min intervals, newest last)

226.8, 226.8, 226.8, 228.6, 230.4, 228.6, 226.8, 226.8, 226.8, 223.2, 230.4, 228.6, 219.6, 219.6, 216.0, 205.2, 199.8, 201.6, 196.2, 190.8

Trend Analysis

- First 10 readings: Fairly flat, 226.8–230.4 mg/dL
- Next 10 readings: Slight dip, small bump, then clear downward trend:
 - 230.4, 228.6, 219.6, 219.6, 216.0, 205.2, 199.8, 201.6, 196.2, 190.8
- Recent 5 readings:
 - 216.0, 205.2, 199.8, 201.6, 196.2, 190.8
- Trend: Gradual decrease, with a minor uptick between 199.8 and 201.6, then resuming the decline.

Recent Changes (per 15 min)

- 216.0 → 205.2: -10.8

Chat with your prompt...

rawan / Default project

PlaygroundDashboardDocsAPI reference

CGM v1

CompareOptimizeEvaluateUpdate

Modelgpt-4.1
text.format: text temp: 1.00 tokens: 2048 top_p: 1.00 store: true

VariablesCreate...

ToolsCreate...

System message
Describe desired model behavior (tone, tool usage, response style)

Add messages to prompt

Recent Changes (per 15 min)

- 216.0 → 205.2: -10.8
- 205.2 → 199.8: -5.4
- 199.8 → 201.6: +1.8 (small uptick)
- 201.6 → 196.2: -5.4
- 196.2 → 190.8: -5.4

Average recent rate:
(-10.8 - 5.4 + 1.8 - 5.4 - 5.4) / 5 = -5.04 mg/dL per 15 min

Prediction Calculation

Starting point (current): 190.8 mg/dL

- 30 min (2 × -5.04): 190.8 + 2 × (-5.04) = 190.8 - 10.08 ≈ 180.7 mg/dL
- 60 min (4 × -5.04): 190.8 + 4 × (-5.04) = 190.8 - 20.16 ≈ 170.6 mg/dL
- 90 min (6 × -5.04): 190.8 + 6 × (-5.04) = 190.8 - 30.24 ≈ 160.6 mg/dL

Collapse

Chat with your prompt...

Auto-clear

rawan / Default project

PlaygroundDashboardDocsAPI reference

CGM v1

CompareOptimizeEvaluateUpdate

Modelgpt-4.1
text.format: text temp: 1.00 tokens: 2048 top_p: 1.00 store: true

VariablesCreate...

ToolsCreate...

System message
Describe desired model behavior (tone, tool usage, response style)

Add messages to prompt

Predicted Glucose (mg/dL):

- 30 min: 180.7
- 60 min: 170.6
- 90 min: 160.6

Explanation

- The recent downward momentum dominates the prediction, though the fall slowed slightly, and a brief uptick is noted.
- Extrapolating the trend, short-term CGM patterns are the strongest predictor; static patient features like BMI or age are far less impactful at this time scale.

Would you like a rationale for these predictions or a comparison to actual values?

Collapse

Chat with your prompt...

Auto-clear

prompt 4_explanation:

rawan / Default project

Playground Dashboard Docs API reference

CGM v1

Model gpt-4.1
text.format: text temp: 1.00 tokens: 2048 top_p: 1.00 store: true

Variables Create...

Tools Create...

System message
Describe desired model behavior (tone, tool usage, response style)

Add messages to prompt

User

Can you explain how you obtained these results, given that the true glucose values at 30, 60, and 90 minutes are 185.4, 171.0 and 172.8 respectively? Additionally, what had the greatest influence on your predictions? the glucose readings or the static patient features? and why?

Assistant

Absolutely, here's a detailed explanation:

How the Results Were Obtained

1. Analyzing the Recent CGM Trend

I focused on the **most recent 5 readings** (covering the past 75 minutes), because short-term future glucose is most strongly influenced by the immediate momentum shown in CGM data:

- 216.0 → 205.2 (change: -10.8)
- 205.2 → 199.8 (change: -5.4)
- 199.8 → 201.6 (change: +1.8)
- 201.6 → 196.2 (change: -5.4)
- 196.2 → 190.8 (change: -5.4)

2. Estimating the Average Rate of Change

Collapse

Chat with your prompt...

Auto-clear

rawan / Default project

Playground Dashboard Docs API reference

CGM v1

Model gpt-4.1
text.format: text temp: 1.00 tokens: 2048 top_p: 1.00 store: true

Variables Create...

Tools Create...

System message
Describe desired model behavior (tone, tool usage, response style)

Add messages to prompt

2. Estimating the Average Rate of Change

I calculated the **average change per 15 min** over the last 5 intervals:

- $(-10.8 - 5.4 + 1.8 - 5.4 - 5.4) / 5 = -5.04$ mg/dL per 15 min

3. Linear Extrapolation

With the last measurement **190.8 mg/dL**, I projected forward:

- 30 minutes (2 intervals):** $190.8 + 2 \times (-5.04) = 180.7$ mg/dL
- 60 minutes (4 intervals):** $190.8 + 4 \times (-5.04) = 170.6$ mg/dL
- 90 minutes (6 intervals):** $190.8 + 6 \times (-5.04) = 160.6$ mg/dL

4. Recap

I assumed the **downward trend** would continue at roughly the same slope. This is a straightforward "momentum-based" projection, which is generally effective for short-term glucose forecasting when no new disruptions (food, insulin, exercise) are reported.

Comparison with True Values

Collapse

Chat with your prompt...

Auto-clear

rawan / Default project

PlaygroundDashboardDocsAPI reference

CGM v1

CompareOptimizeEvaluateUpdate

Modelgpt-4.1
text.format: text temp: 1.00 tokens: 2048 top_p: 1.00 store: true

VariablesCreate...

ToolsCreate...

System message

Describe desired model behavior (tone, tool usage, response style)

Add messages to prompt

Comparison with True Values

Time Ahead	Prediction	True Value
30 min	180.7	185.4
60 min	170.6	171.0
90 min	160.6	172.8

- 30 and 60 minutes: My predictions closely matched the observed values, showing that linear short-term projections can often be quite accurate.
- 90 minutes: I underpredicted the value (160.6 vs 172.8). The actual glucose leveled off and then slightly increased, whereas my model continued to project a decrease. This discrepancy likely results from either physiologic counterregulation or a hidden variable (e.g., small snack, changes in activity, hormonal response) that wasn't captured in the visible CGM history.

What Had the Greatest Influence: Recent Glucose vs. Static Features?

The most significant influence was from the recent CGM glucose readings.

Why?

- Short-term dynamics depend almost entirely on momentum—if glucose is falling, it tends to keep falling unless interrupted; if it is rising, same.

Collapse

Chat with your prompt...

Auto-clear

rawan / Default project

PlaygroundDashboardDocsAPI reference

CGM v1

CompareOptimizeEvaluateUpdate

Modelgpt-4.1
text.format: text temp: 1.00 tokens: 2048 top_p: 1.00 store: true

VariablesCreate...

ToolsCreate...

System message

Describe desired model behavior (tone, tool usage, response style)

Add messages to prompt

Why?

- Short-term dynamics depend almost entirely on momentum—if glucose is falling, it tends to keep falling unless interrupted; if it is rising, same.
- Static patient features (BMI, age, etc.) shape long-term averages and informative context (risk of hypoglycemia, insulin sensitivity ranges, etc.), but have little effect on hour-to-hour or minute-to-minute fluctuations unless used in complex integrated physiologic models, and even then their impact on short-term predictions is usually tiny compared to the recent trend.
- When using established diabetes data science techniques and machine learning, lagged CGM values are nearly always the highest-weighted variables for forecasting the next 1-6 readings.

In summary:

- Recent CGM trends dominated these predictions (most recent ~5 readings, covering the past 75 minutes).
- Static features (BMI, age, duration, etc.) have some background influence, but for 30/60/90 minute forecasts, they are mostly negligible given real-time CGM data.

If you would like to see a different approach (like using a moving average, or emphasizing older trends), let me know!

Collapse

Chat with your prompt...

Auto-clear

Prompt 5:

rawan / Default project

Playground Dashboard Docs API reference

CGM v1

Model gpt-4.1
text.format: text temp: 1.00 tokens: 2048 top.p: 1.00 store: true

Variables Create...

Tools Create...

System message
Describe desired model behavior (tone, tool usage, response style)

Add messages to prompt

User
You are a medical AI that predicts blood glucose. Given the patient characteristics and CGM history of 20 values recorded each 15 min, predict blood glucose values for the next 30, 60, and 90 minutes. Return predictions in mg/dL. The patient features are: (BMI (kg/m2): 24.3; Age (years): 69.0; Weight (kg): 60.0; Smoking History (pack year): 0.0; Duration of diabetes (years): 8.0; Gender (Female=1, Male=2): 2.0). The CGM recoding of past values each 15 min of the patient 20: 259.2, 244.8, 235.8, 219.6, 199.8, 180.0, 167.4, 158.4, 153.0, 158.4, 178.2, 192.6, 208.8, 223.2, 230.4, 230.4, 232.2, 244.8, 257.4, 264.6

Assistant
Let's analyze the CGM trend and predict your next glucose values.

Patient Characteristics

- BMI: 24.3 (normal)
- Age: 69
- Weight: 60 kg
- Smoking History: 0
- Duration of diabetes: 8 years
- Gender: Male

Collapse
Chat with your prompt...
Auto-clear

rawan / Default project

Playground Dashboard Docs API reference

CGM v1

Model gpt-4.1
text.format: text temp: 1.00 tokens: 2048 top.p: 1.00 store: true

Variables Create...

Tools Create...

System message
Describe desired model behavior (tone, tool usage, response style)

Add messages to prompt

CGM Data (every 15 min, most recent last):
259.2, 244.8, 235.8, 219.6, 199.8, 180.0, 167.4, 158.4, 153.0, 158.4, 178.2, 192.6, 208.8, 223.2, 230.4, 230.4, 232.2, 244.8, 257.4, 264.6

Trend Analysis

- Initial values: Rapid decrease from 259.2 → 153.0 (over about 2-3 hours).
- Then: A sharp rebound: 153.0 → 264.6 in about 2.5 hours.
- Recent values (last 5 readings): 230.4, 230.4, 232.2, 244.8, 264.6 (steady and then another sharp rise).

Recent changes (per 15 min, most recent first):

- 232.2 → 244.8: +12.6
- 244.8 → 257.4: +12.6
- 257.4 → 264.6: +7.2

Let's use the last 3 rates, most relevant for short-term direction:
Collapse
Chat with your prompt...
Auto-clear

prompt 5_explanation:

rawan / Default project

Playground Dashboard Docs API reference

CGM v1

Compare Optimize Evaluate Update

Modelgpt-4.1text.format: text temp: 1.00 tokens: 2048 top_p: 1.00 store: true

VariablesCreate...

ToolsCreate...

System messageDescribe desired model behavior (tone, tool usage, response style)

Add messages to prompt

User

Can you explain how you obtained these results, given that the true glucose values at 30, 60, and 90 minutes are 250.2, 223.2 and 181.8 respectively? Additionally, what had the greatest influence on your predictions? the glucose readings or the static patient features? and why?

Assistant

Certainly! Here's a detailed breakdown addressing both your method and the underlying influence of the input data.

1. How Were the Results Obtained?

Step 1: Recent CGM Trend Analysis

I examined the last several CGM readings:

- Recent readings: 230.4, 230.4, 232.2, 244.8, 257.4, 264.6 (oldest to newest)
- Recent changes (last three intervals, all per 15 min):
 - 232.2 → 244.8 = +12.6
 - 244.8 → 257.4 = +12.6
 - 257.4 → 264.6 = +7.2

Step 2: Calculating Average Rate of Change

Chat with your prompt...

rawan / Default project

Playground Dashboard Docs API reference

CGM v1

Compare Optimize Evaluate Update

Modelgpt-4.1text.format: text temp: 1.00 tokens: 2048 top_p: 1.00 store: true

VariablesCreate...

ToolsCreate...

System messageDescribe desired model behavior (tone, tool usage, response style)

Add messages to prompt

Step 2: Calculating Average Rate of Change

- I took the average of the most recent three intervals, which were strongly positive: $(12.6 + 12.6 + 7.2) / 3 = 10.8$ mg/dL per 15 min

Step 3: Linear Extrapolation

- Using the most recent glucose reading (264.6 mg/dL), I projected forward at the average rate above:
 - 30 min (2 intervals): $264.6 + 2 \times 10.8 = 286.2$
 - 60 min (4 intervals): $264.6 + 4 \times 10.8 = 307.8$
 - 90 min (6 intervals): $264.6 + 6 \times 10.8 = 329.4$

Step 4: Assumptions and Pitfalls

- The approach assumes that the most recent, strong upward trend would continue unchanged.

2. Comparison to Actual Outcomes

- True values were:
 - 30 min: 250.2
 - 60 min: 223.2
 - 90 min: 181.8
- My predictions overestimated glucose for all intervals.

Chat with your prompt...

34

rawan / Default project

Playground Dashboard Docs API reference

CGM v1

Modelgpt-4.1text.format: text temp: 1.00 tokens: 2048 top_p: 1.00 store: true

VariablesCreate...

ToolsCreate...

System messageDescribe desired model behavior (tone, tool usage, response style)

Add messages to prompt

CompareOptimizeEvaluateUpdate

Why? Shortly after the last recorded measure, the glucose curve reversed direction and started dropping rapidly. My projection used the latest "momentum" but missed the fact that it was a short-lived spike (possibly due to a meal or a correction bolus about to take effect).

3. Main Influence: Glucose Readings vs. Patient Features

The greatest influence was the recent glucose readings.

Why?

a. How CGM Influences Forecasts:

- Blood glucose in the short term (30–90 minutes) is best predicted by its current trajectory—whether it's rising, falling, or stable in those final few measurements.
- That's why forecasting tools, including AI and manual clinical prediction, overwhelmingly rely on short-term rate or direction (the last few 15-min CGM points).
- My model used the average rate of change from the most recent intervals.

b. Role of Static Patient Features:

- Static features (BMI, age, etc.) define:
 - Overall risk level (risk for highs/lows)

Collapse

Chat with your prompt...

Auto-clear

rawan / Default project

Playground Dashboard Docs API reference

CGM v1

Modelgpt-4.1text.format: text temp: 1.00 tokens: 2048 top_p: 1.00 store: true

VariablesCreate...

ToolsCreate...

System messageDescribe desired model behavior (tone, tool usage, response style)

Add messages to prompt

CompareOptimizeEvaluateUpdate

b. Role of Static Patient Features:

- Static features (BMI, age, etc.) define:
 - Overall risk level (risk for highs/lows)
 - Typical insulin sensitivity or metabolic health
 - They do NOT drive rapid changes in glucose in the same way recent CGM values do.**
- In short-term prediction, static features are minor contextual factors—influencing only by setting the background, not the short-term movement.

4. Why the Predictions Missed the Mark

- The recent readings showed a strong upswing, but in reality there was an imminent drop (perhaps due to delayed insulin onset/action, or end of a meal glucose spike).
- Short-term extrapolation can fail when there's an impending reversal not yet visible in the data.
- You'd need additional info (e.g., insulin just administered, or meal just completed) to anticipate such a change.

Summary Table

Collapse

Chat with your prompt...

Auto-clear

35

rawan / Default project

PlaygroundDashboardDocsAPI reference

CGM v1

CompareOptimizeEvaluateUpdate

Modelgpt-4.1
text.format: text temp: 1.00 tokens: 2048 top_p: 1.00 stane: true

VariablesCreate...

ToolsCreate...

System message
Describe desired model behavior (tone, tool usage, response style)

Add messages to prompt

- Short-term extrapolation can fail when there's an impending reversal not yet visible in the data.
- You'd need additional info (e.g., insulin just administered, or meal just completed) to anticipate such a change.

Summary Table

Factor	Influence	Why?
Recent CGM values	Strongest	Best reflect short-term direction
Static patient features	Minor	Influence baseline, not momentum

In short:
The CGM trend (recent readings) overwhelmingly influenced the prediction because *short-term glucose changes are determined by recent values*, whereas static features are mostly relevant for overall management and not for predicting direction within a 1-hour window. My overprediction happened because the model does not "see" future inflection points—only current momentum.

If you'd like more on mitigating such pitfalls, or how other info (like meal/insulin timing) could improve prediction, just ask!

Collapse

Chat with your prompt...

Auto-clear

36