Advancing Diabetes Management with Conditional Generative Modeling

Matthew Manion & Sathvika Ayyappa Prabhu
Chemical Engineering & Computer Science Engineering
CSE 598: AI For Science
Fall 2025

Diabetes: Process Control for Life or Death

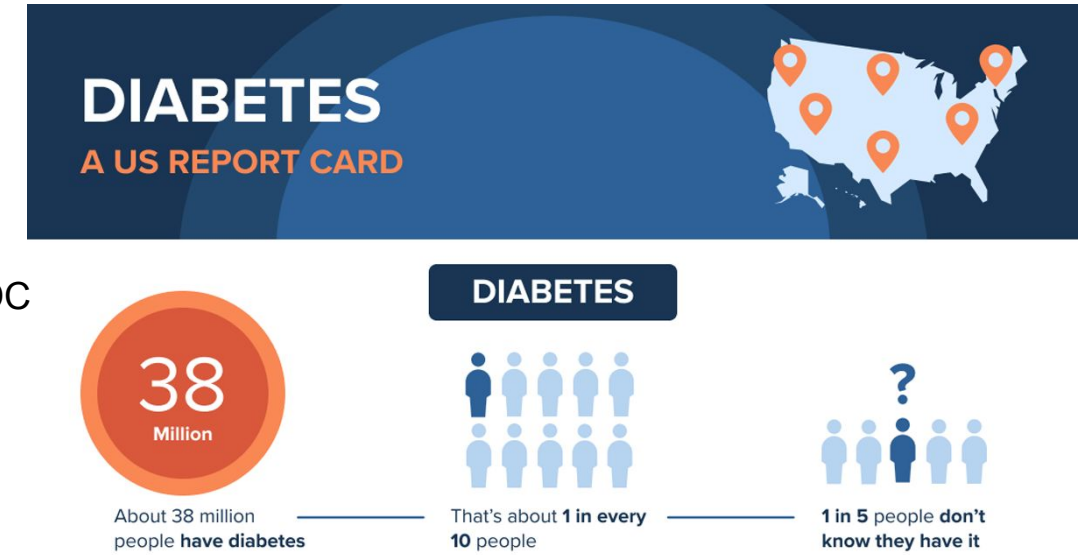


Image: CDC

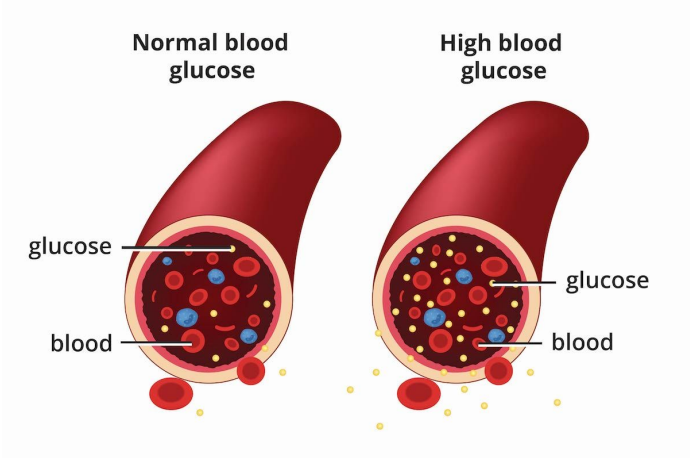


Image: NIDDK

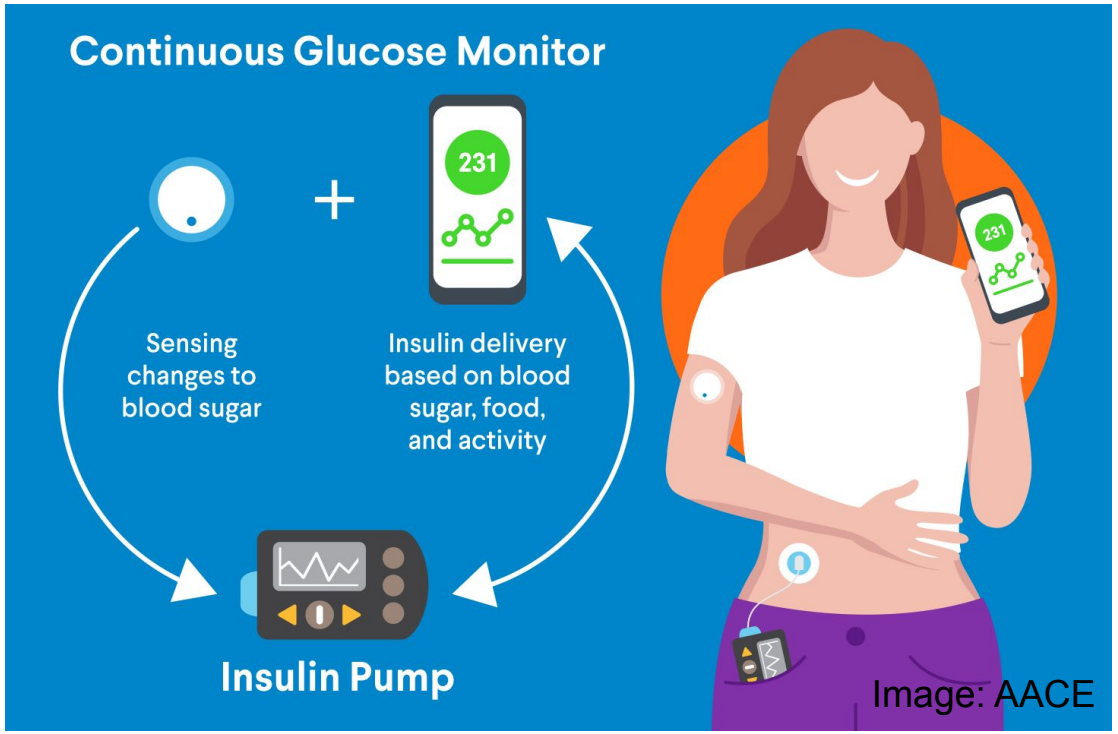


Image: AACE

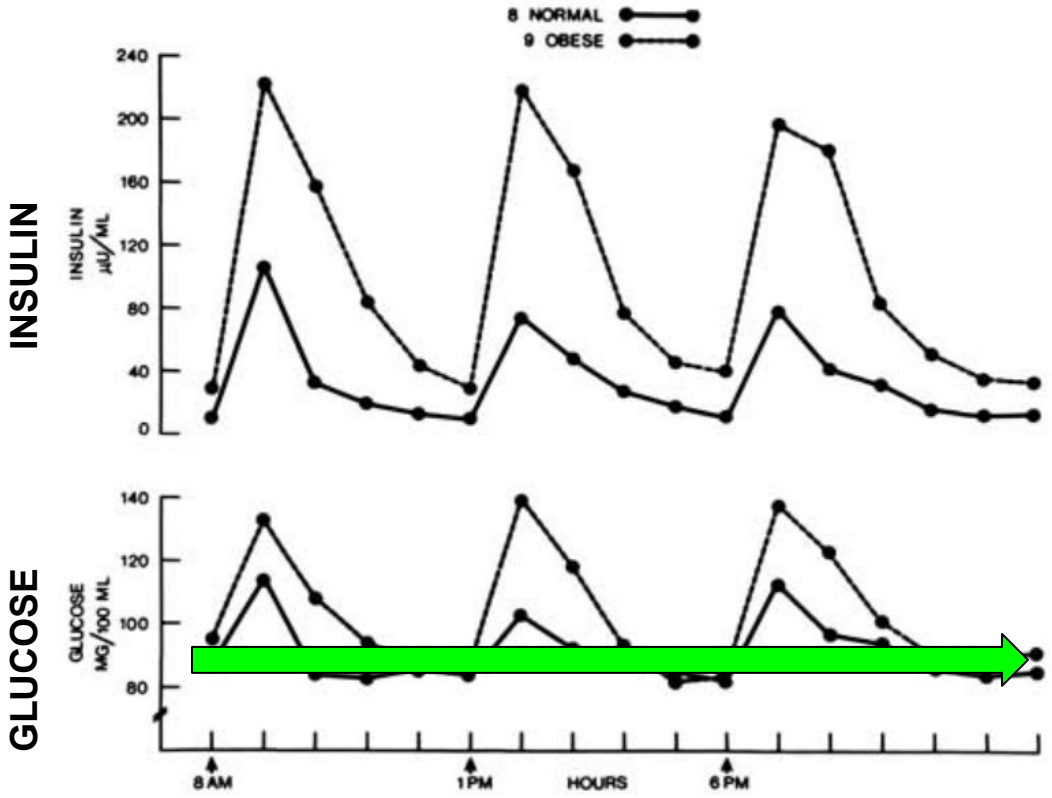
Risks of Pump/Monitor Malfunction:

High Blood Glucose: Coma, Organ Failure

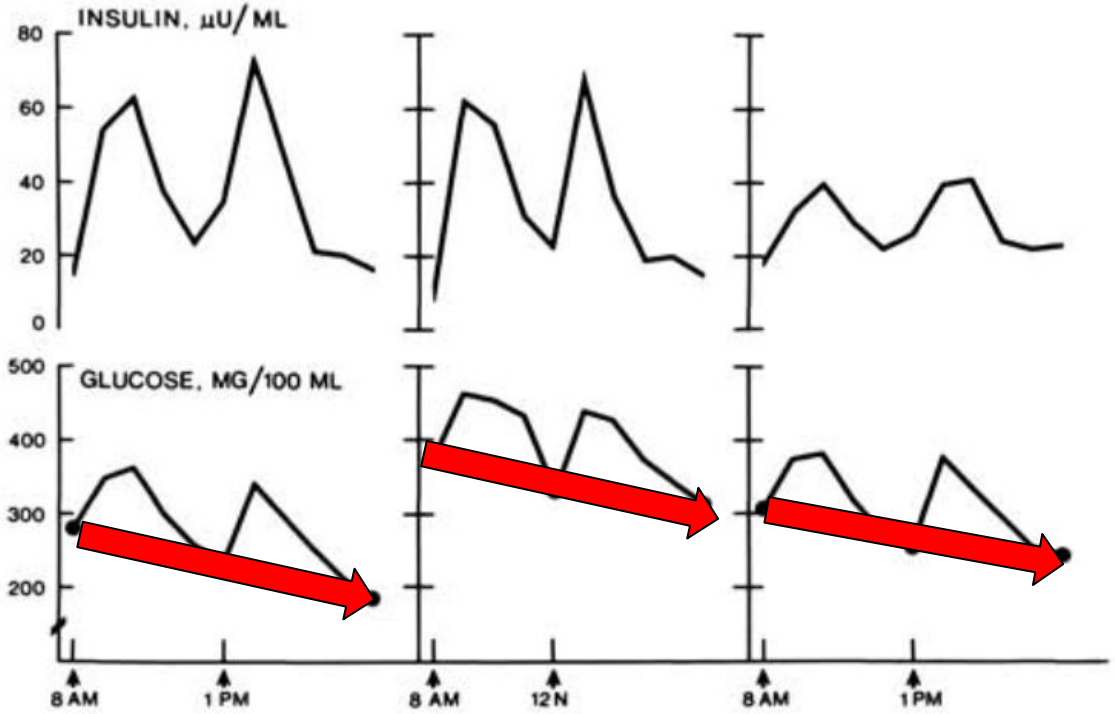
Low Blood Glucose: Coma, Seizure, Death

Problem: Calculate Insulin Dose for Glucose Trajectory

Non-Diabetic Adults

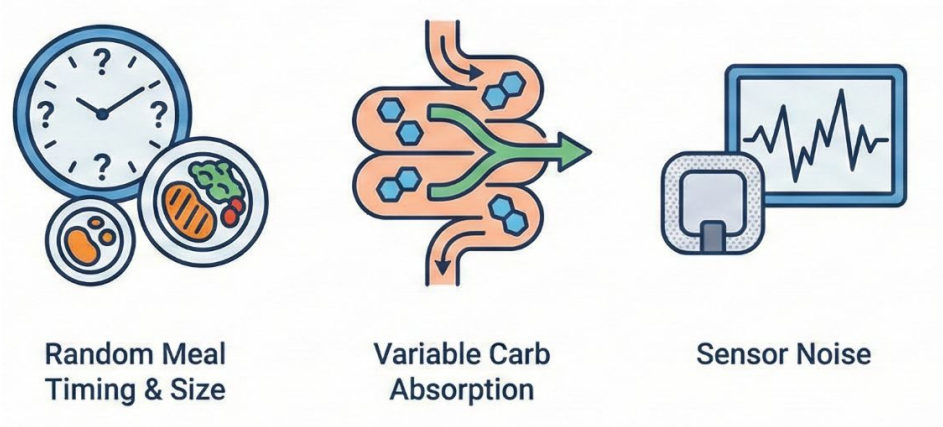


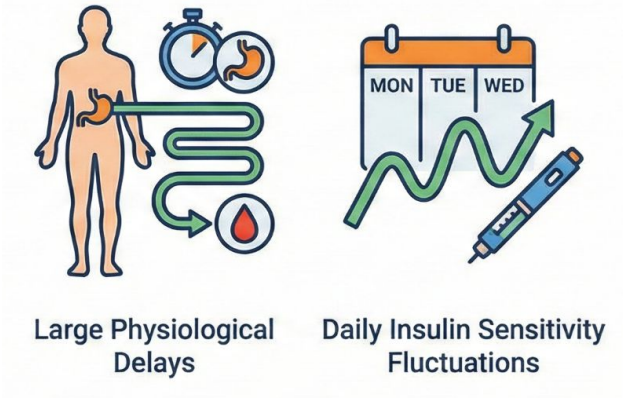
3 Diabetic Adults Given Normal Insulin Dose

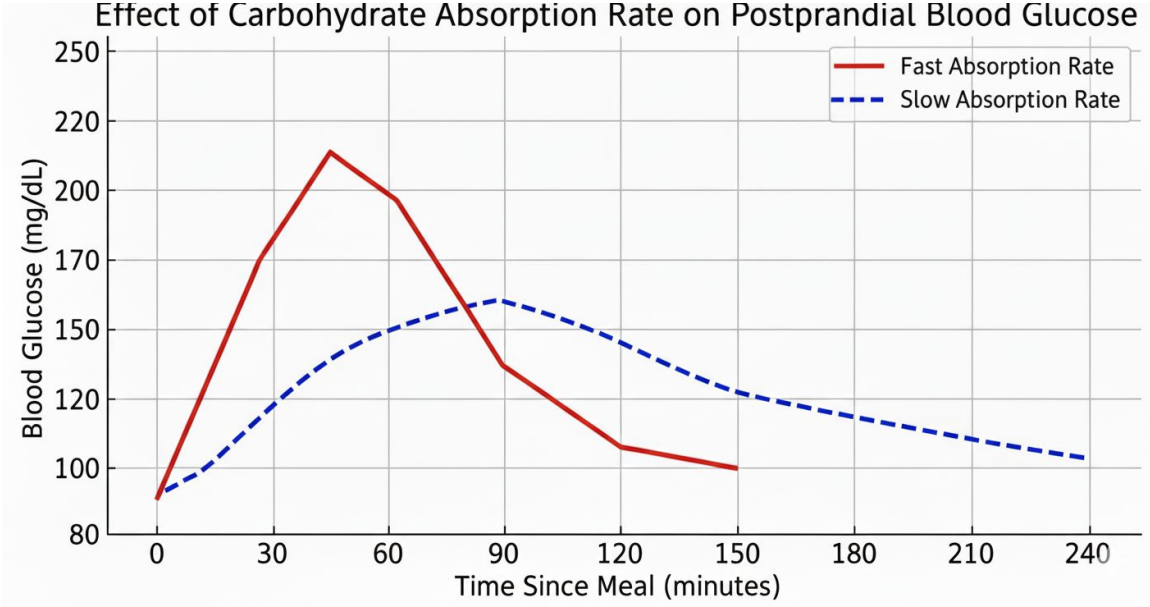


Blood Glucose Dynamics Are Inherently Stochastic

Stochasticity comes from:







Same observed state → **multiple safe insulin doses**

Same action → **different BG outcomes**

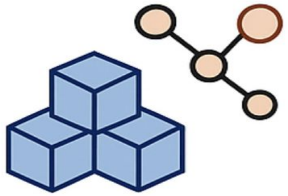
Diffusion Policies in Stochastic Offline RL

- Expressive
- Data-anchored

very successful in deterministic control tasks

However, it is unclear whether they remain:

- Safe
- Stable
- Robust



when transitions are highly stochastic and actions are multimodal

We ask:

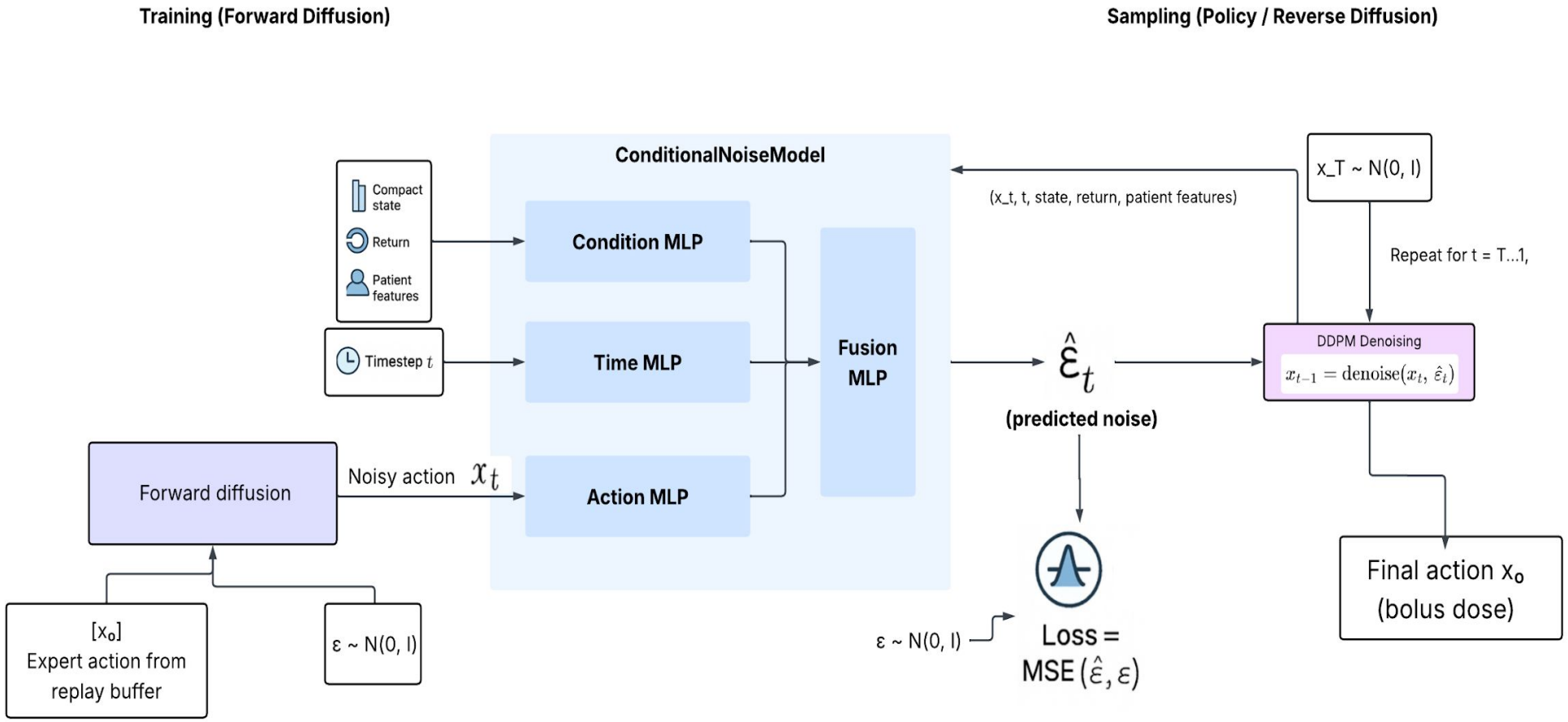
Can conditional diffusion models reliably handle stochasticity in offline RL?

Python implementation of the FDA-approved UVA/Padova Type-1 Diabetes Simulator (2008)

Includes 30 virtual patients
10 adults, 10 adolescents, 10 children

Designed to be “reinforcement-learning-ready”:
OpenAI Gym / Gymnasium API
Each step returns (observation, reward, done, info)

Model Architecture



Problem Definition: Offline RL for Stochastic Blood Glucose Control

We model insulin dosing as a Markov Process Described by:

$$M = (\mathcal{S}, \mathcal{A}, \mathcal{P}, r, \gamma),$$

- **State s_t:** patient glucose/IOB/CHO context
- **Action a_t:** Continuous insulin bolus (administered dose)
- **Stochastic Transitions** (meals, absorption, noise, delays) : $\mathcal{P}(s' | s, a)$
- **Reward r_t** = −RiskIndex(G_t) (penalizes unsafe glucose)
- **Policy :** $\pi_\theta(a|s)$

Reward Function — Clinical Risk Index

We use the clinical Blood Glucose Risk Index, which quantifies how dangerous a glucose value is.

Risk is split into two components:

LBGI (Low BG Risk Index)

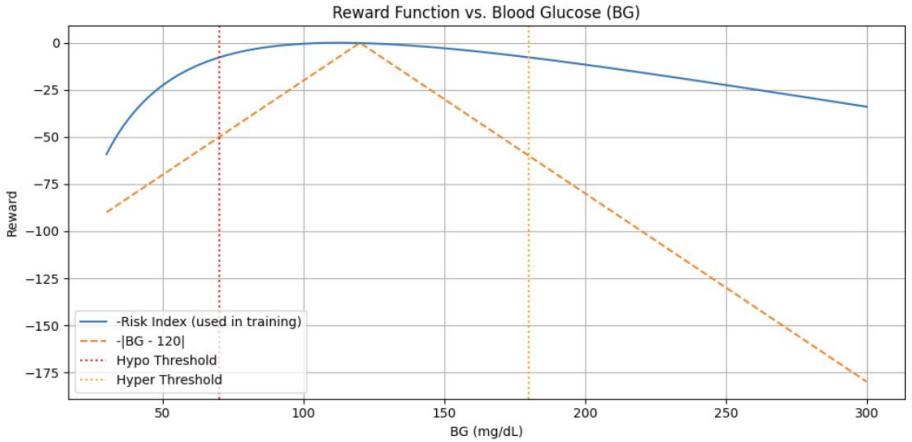
HBGI (High BG Risk Index)

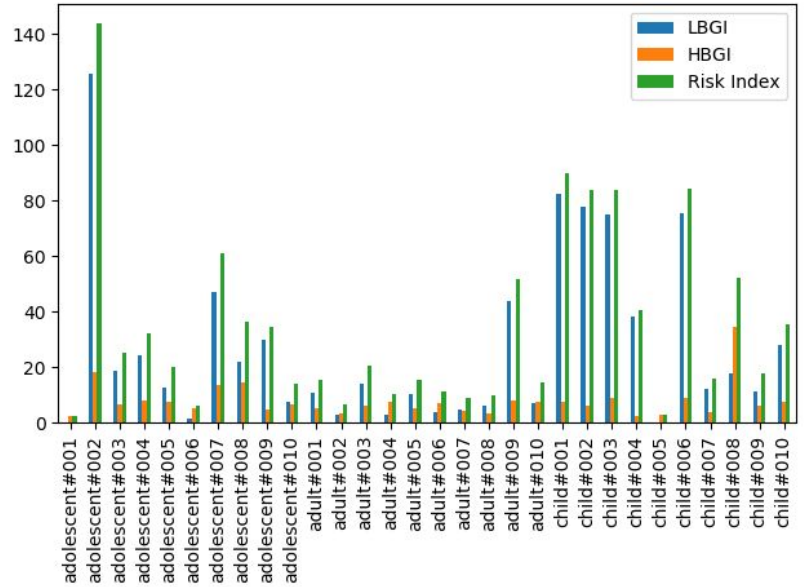
Hypoglycemia carries much higher risk, so LBGI is weighted more heavily.

For each glucose value *Gt* , we compute a risk score:

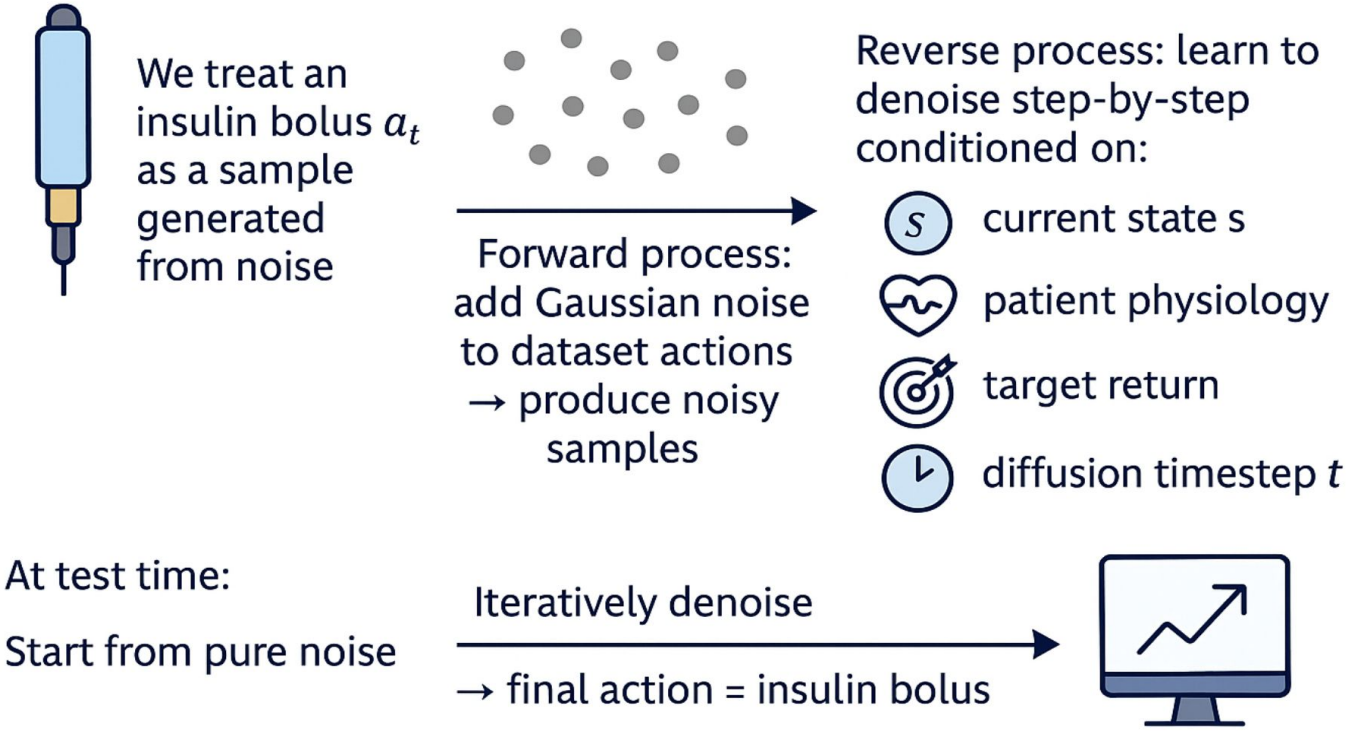
$$Risk_t = \lambda \cdot LBGI_t + (1-\lambda) \cdot HBGI_t \quad , \lambda > 0.5$$

Reward is: $r_t = -Risk_t$





Diffusion Model as a Policy $\pi(a_t \mid s_t)$



Forward (training):

Noise injection to create supervision:

$$x_t = \sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1 - \bar{\alpha}_t}\, \epsilon$$

- $x_0 = a_t$ (true bolus from offline dataset)
- Produces noisy action xtxt at diffusion step t.

Reverse (policy):

Model predicts noise:

$$\hat{\epsilon}_\theta(x_t, s_t, t, \text{patient}, \text{target})$$

Used to denoise:

$$x_{t-1} = \text{Denoise}(x_t, \hat{\epsilon}_\theta)$$

Training objective (Noise Prediction Loss):

$$\mathcal{L}(\theta) = \|\epsilon - \hat{\epsilon}_\theta(x_t, s_t, t, \text{context})\|^2$$

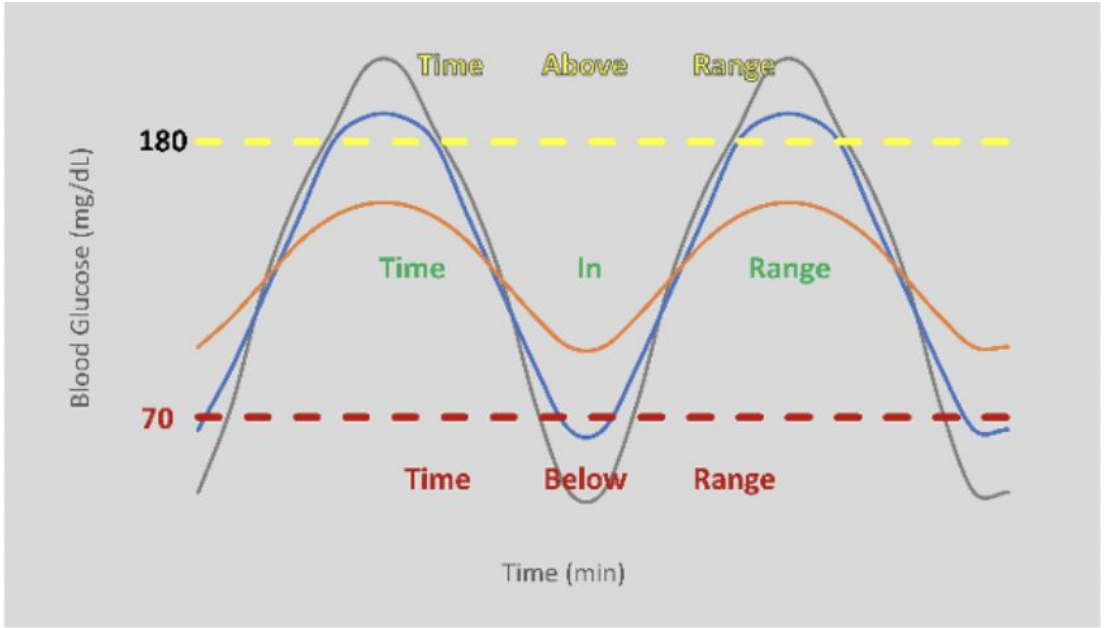
Experimental Setup & Evaluation Protocol

- **Environment:** SimGlucose (T1DSimEnv, adult#001, RandomScenario)
- **Resolution:** 5-minute steps
- **Episode length:** 24 hours → 288 control decisions per episode
- **Training run:** 600 episodes (~600 virtual days)
  - 75 expert-only pretraining episodes (BBController)
  - 525 episodes with diffusion + expert mixing
- **Oracle-style aggregation:**
  - Start with pure Expert trajectories
  - Then mix in BB vs diffusion actions
    i. Model attempts to predict noise
    ii. If TIR <55%, Model decisions not saved
    iii. Expert is given same situation, decisions saved for training
  - All transitions stored in a replay buffer (1M capacity)
- **Optimization:**
  - Batch size = 256
  - 5 diffusion updates per episode
  - Cosine LR schedule (3e-4 → 1e-6 across 600 episodes)

- **Policies evaluated:**
  - **BBController** (Behavior policy / data source)
  - **PID controller**
  - **Tabular Q-learning**
  - **TD3-BC**
  - **Diffusion Policy (ours)**
- **Offline constraint:**
  - No environment interaction during training
  - All models trained only on the Expert & Model replay buffer

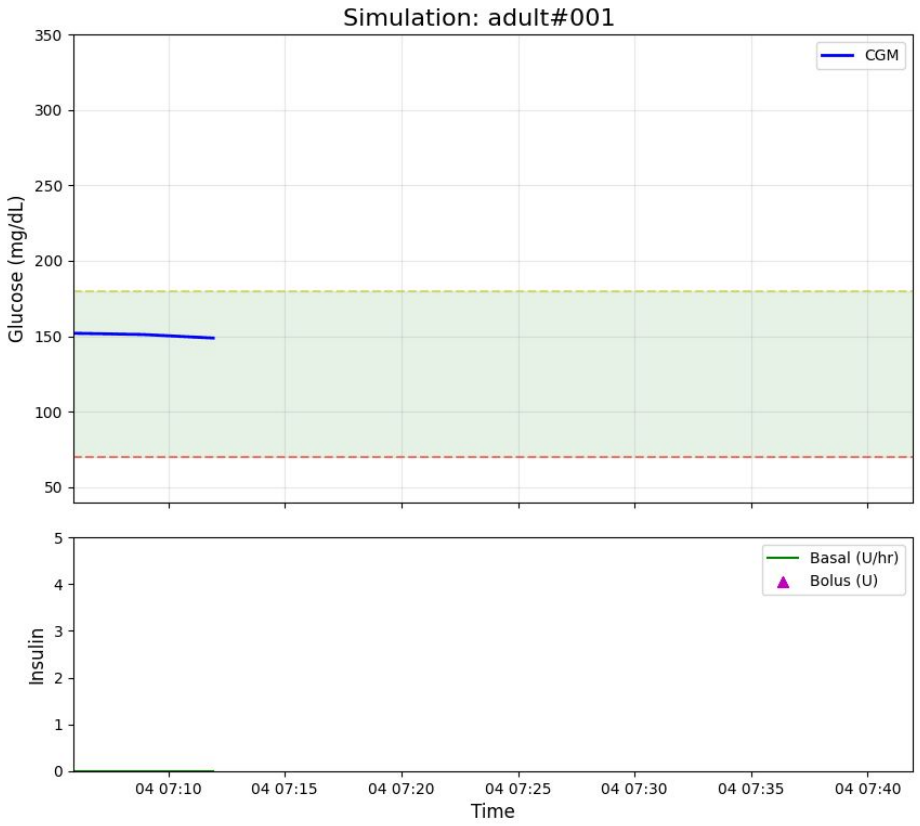
Clinical Evaluation Metrics

Metric	Definition
TIR (%) Time in Range	% of time BG is **70–180 mg/dL**
TBR (%) Time Below Range	% of time BG is **< 70 mg/dL**(hypoglycemia)
TAR (%) Time Above Range	% of time BG is **> 180 mg/dL**(hyperglycemia)
CV (%) Coefficient of Variation	Glucose variability (**SD ÷ mean × 100**)

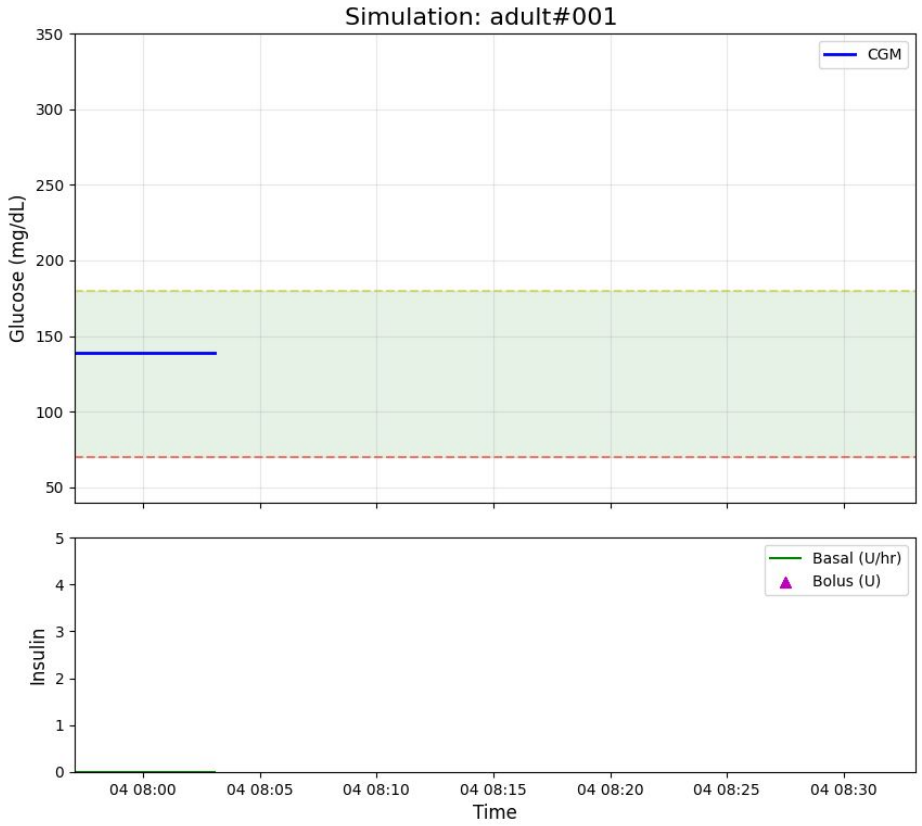


Example Diffusion Training

Early Training (Patient Dies)

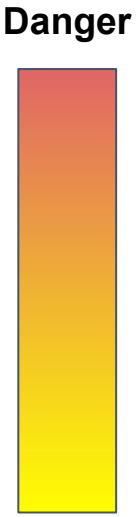


Late Training



Results: Diffusion Trained Only on Body Weight

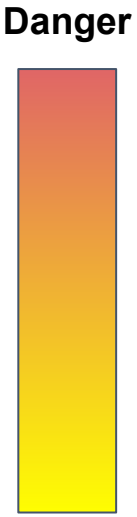
Method	TIR (%)	TBR (%)	CV (%)
PID Controller	68.2	6.1	32.1
Tabular Q-Learning	70.1	5.2	30.3
TD3-BC	71.8	4.5	28.4
Diffusion Policy (ours)*	**100.0**	**0.0**	**13.35**
Clinical Controller*	91.3	0.0	8.83



*Averaged Over 3 Adult Evaluations

Results: BW Generalization to Other Groups

Method	TIR (%)	TBR (%)
Diffusion Policy (ours) (Adolescent 1)	79.5	8.3
Clinical Controller (Adolescent 1)	71.2	0.0
Diffusion Policy (ours) (Child 1)	49.0	51.0
Clinical Controller (Child 1)	76.0	12.8



Results: Diffusion Trained on Multi Params

Method	TIR (%)	TBR (%)	CV (%)
Diffusion Policy (ours)	**58.9**	**0.0**	**16.2**
Clinical Controller	94.2	0.0	14.8

Averaged Over 3 Adult Evaluations

Key Takeaways & Next Steps:

- Diffusion policies are effective for stochastic offline RL in blood glucose control.
- Performs reliably on adult virtual patients, where physiological dynamics are more stable.
- Generalization to adolescents/children remains challenging due to higher variability and noisier glucose–insulin dynamics.
- Training diffusion on multiple patient parameters requires significant compute, especially under stochastic transitions.

Next:

- Scale up compute
- Change Reward Scheme: Emphasize Ideal Glucose Region
- Expand Time Horizon

Questions?

The RL Opportunity

- RL can learn personalized insulin policies.
- It can adapt to:
  - Each patient’s physiology
  - Time-of-day patterns
  - Meal-driven disturbances
- But online RL is risky in healthcare : unsafe, unethical, expensive

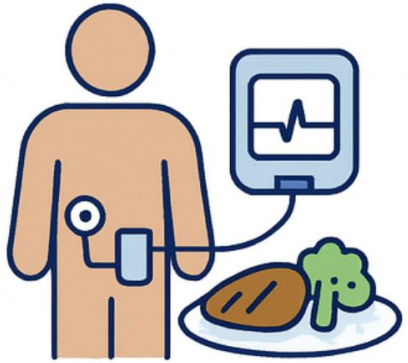
- ➜ **Offline RL** is the only viable option.

Data Collection

Environment: T1DSimEnv
(simglucose, FDA-approved UVA/Padova simulator)

Behavior policy:
Built-in Basal–Bolus (BBController)

Represents standard clinical dosing behavior



Stochastic meal scenarios
Randomized meal timing, meal size, and absorption variability

Physiological variability
Adult virtual patients have diverse glucose/insulin kinetics

- **Dataset generation:**
  - **Run many full-day simulations per adult virtual patient**
  - **Log transitions:**

$$(s_t, a_t, r_t, s_{t+1}, \text{info})$$

  - **(≈ 300–400 episodes → ~7.5×10⁵ transitions)**

Forward & Reverse Processes

Forward (training):
Noise injection to create supervision:

$$x_t = \sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1 - \bar{\alpha}_t}\, \epsilon$$

- $x_0 = a_t$ (true bolus from offline dataset)
- Produces noisy action xtxt at diffusion step tt

Reverse (policy):
Model predicts noise:

$$\hat{\epsilon}_\theta(x_t, s_t, t, \text{patient}, \text{target})$$

Used to denoise:

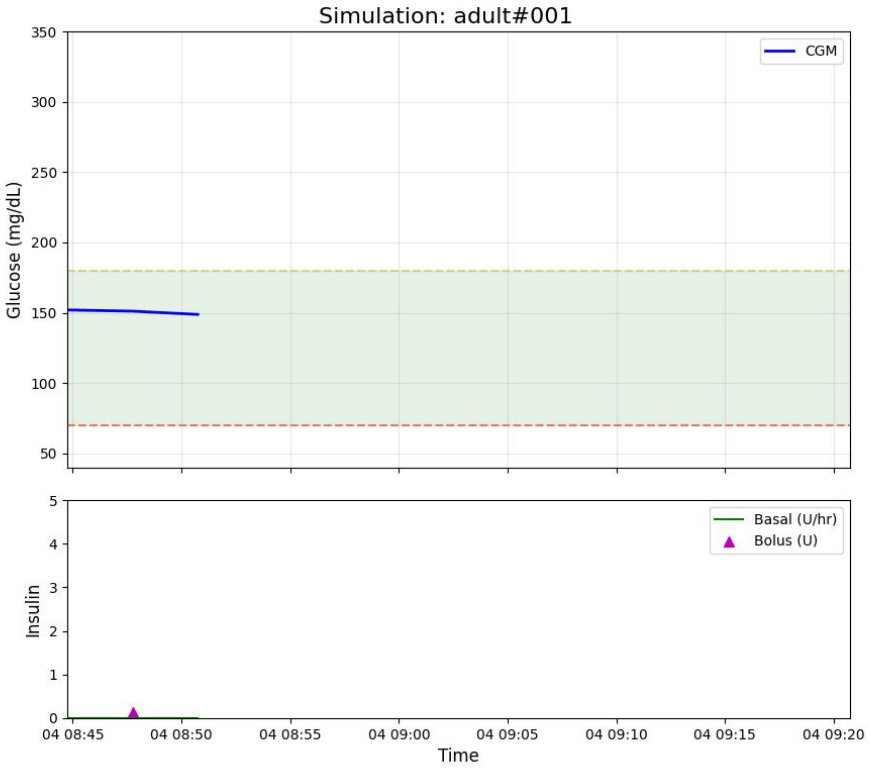
$$x_{t-1} = \text{Denoise}(x_t, \hat{\epsilon}_\theta)$$

Training objective (Noise Prediction Loss):

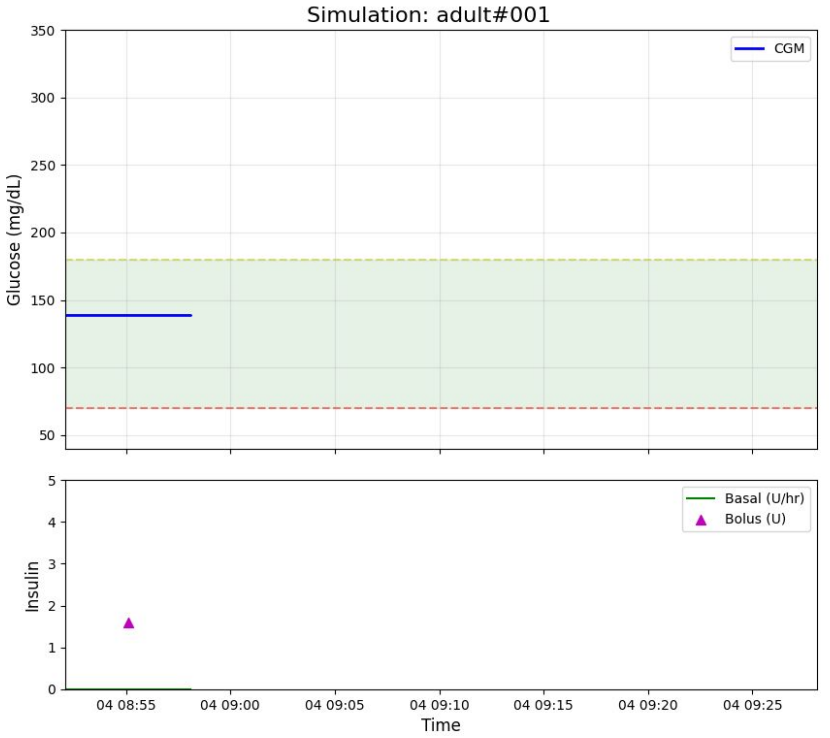
$$\mathcal{L}(\theta) = \|\epsilon - \hat{\epsilon}_\theta(x_t, s_t, t, \text{context})\|^2$$

Example Diffusion Training: Latest

Early Training (Patient Dies)



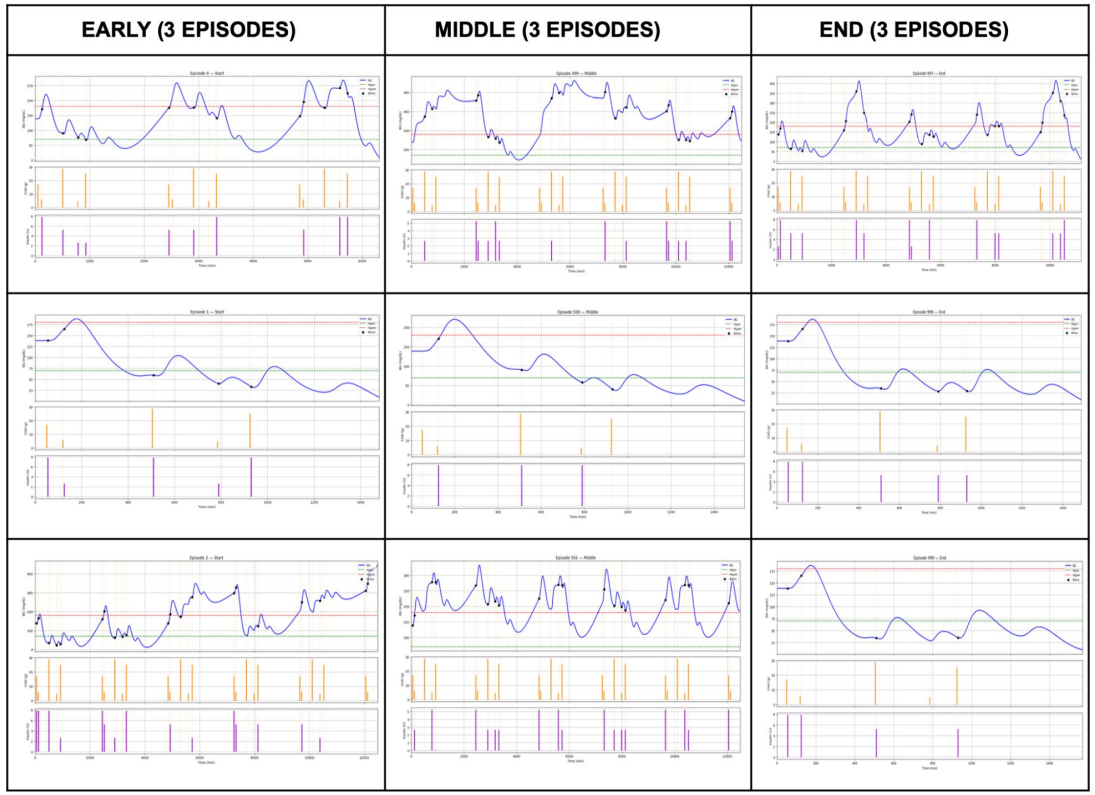
Late Training

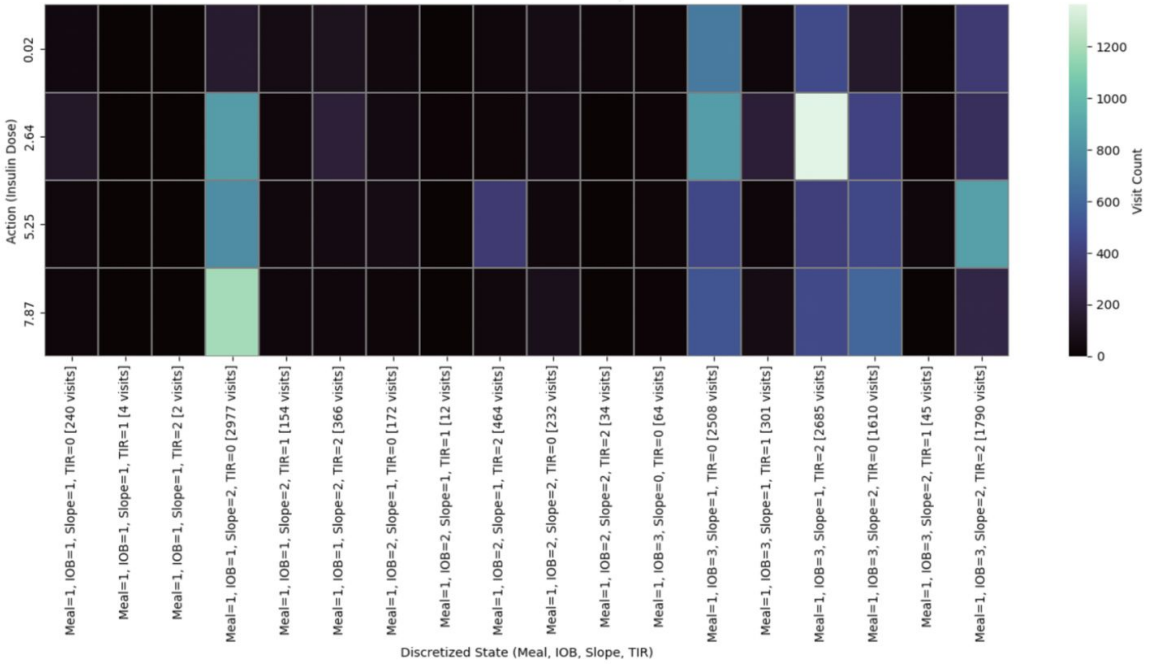


Model Architecture

Placeholder..
Diagram comes here. Will remove/update text after.

Results: Comparison Across Controllers (*TAB RL)*





Stochasticity Creates a General Offline RL Challenge

- Offline RL must learn from a fixed dataset → no corrections

- Under stochasticity:
  a. Q-values become unreliable
  b. Policies collapse multimodal distributions into unsafe averages
  c. Models produce out-of-distribution actions

How do we design policies that remain robust under stochastic offline RL?