Explainable Insulin Pump Control with LLM Controllers for Type 1 Diabetes

Maya Sarkar
Mission San Jose High School
Fremont, CA 94539
mayasarkar010@gmail.com

Abstract

Living with Type 1 Diabetes (T1D) is a constant balancing act, requiring patients to make complex decisions based on endless streams of data. While Artificial Pancreas Systems (APS) powered by Reinforcement Learning (RL) have shown promise in automating insulin delivery, their "black-box" nature makes it hard for patients and doctors to trust them fully. This paper presents LLM-T1D, a groundbreaking approach that combines the precision of RL with the clear, human-like reasoning of Large Language Models (LLMs) to create a more transparent and reliable insulin pump controller. By training an expert RL system and then distilling its knowledge into fine-tuned Llama 3.1 8B and Qwen3 8B models using a LoRA architecture, we developed a controller that not only matches or surpasses the RL system's performance but also explains its decisions in plain, understandable language. Tested on the FDA-approved UVA/Padova T1D simulator, the LLM controllers deliver excellent blood sugar control while giving patients clear, data-driven insights they can trust. This hybrid system transforms a complex algorithm into an approachable "copilot," paving the way for safer, more understandable, and patient-centered AI solutions for managing chronic conditions like T1D.

1 Introduction

Type 1 Diabetes (T1D) is a lifelong condition requiring constant vigilance[1]. Patients must maintain blood glucose between 70–180 mg/dL (3.9–10 mmol/L) by administering insulin[2, 3]. The proportion of time within this range, termed Time in Range (TIR)[4] (see Figure3), is critical, as both hypoglycemia and hyperglycemia can cause severe, even life-threatening, complications. Managing this balance is difficult, as meals, exercise, stress, and hormonal shifts all contribute to a heavy mental burden worldwide[5, 6]. Artificial Pancreas Systems (APS) automate insulin delivery, easing the burden of diabetes care. While PID and MPC controllers have been effective, Reinforcement Learning (RL) methods like Proximal Policy Optimization (PPO) learn personalized dosing directly from data, handling glucose dynamics without manual carb counting [7, 8].

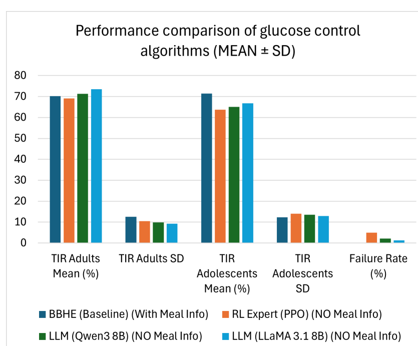

Figure 1: Performance comparison of glucose control algorithms (MEAN ± SD)

A major hurdle for RL-based insulin controllers is **trust**. These opaque "black box" algorithms cannot explain *why* a dose is chosen, making patients and clinicians wary of handing over such a critical therapy. This lack of transparency remains the key barrier to AI adoption in life-critical healthcare.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: 2nd Workshop on Multi-modal Foundation Models and Large Language Models for Life Sciences.

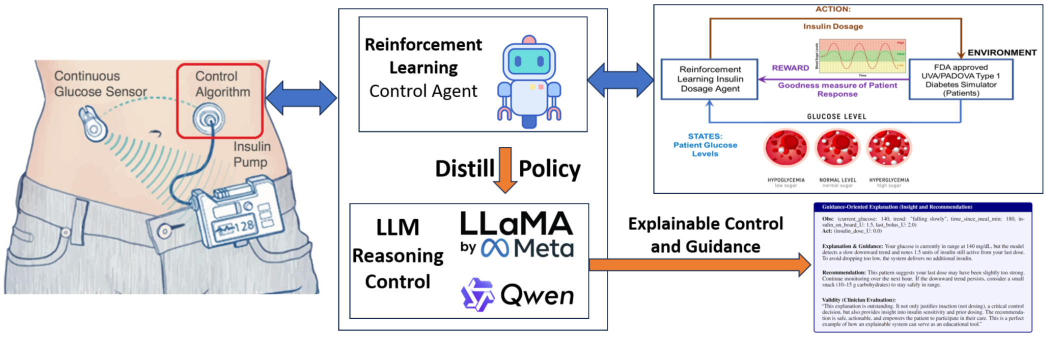

Figure 2: System Disgram for LLM-T1D

We address this challenge with **LLM-T1D**, a framework that distills the expertise of high-performing RL policies into Large Language Models (LLMs) (Figure 2). By training an LLM on RL decisions, we create a controller that retains performance while explaining its choices in plain, human-readable language. Our contributions are:

- **Hybrid RL LLM Control:** A policy distillation method that combines RL's optimization with LLM clarity for T1D insulin management.
- **Trustworthy and Explainable AI:** An LLM controller that explains its insulin dosing, acting as a reliable "copilot" to foster trust and enable safe oversight.
- **Proven Performance and Safety:** Validated on FDA-approved UVA/Padova T1D simulator [9], LLM control matches or surpasses RL baselines while offering decision insights.
- **Patient Empowerment:** Transparency enhances safety and helps patients better understand and engage in their own care.

2 Methodology

Our approach bridges the gap between complex numerical optimization and clear, human-readable reasoning through a multi-stage framework, *LLM-T1D*, designed to optimize blood glucose control in Type 1 Diabetes (T1D) management.

2.1 Generating an Expert Policy with Reinforcement Learning

We begin by training an expert control policy using **Proximal Policy Optimization (PPO)** [10], a state-of-the-art model-free reinforcement learning (RL) algorithm. The PPO agent is trained in a simulated environment using the Simglucose simulator, which implements the FDA-approved UVA/-Padova T1D model [9]. By interacting with simulated T1D patients under diverse and challenging meal scenarios [11], the agent learns a policy $\pi_{\mathrm{RL}}(a_t \mid s_t)$ that maximizes average rewards. This results in a robust, optimized policy that serves as the "expert teacher" for our large language model (LLM) controller.

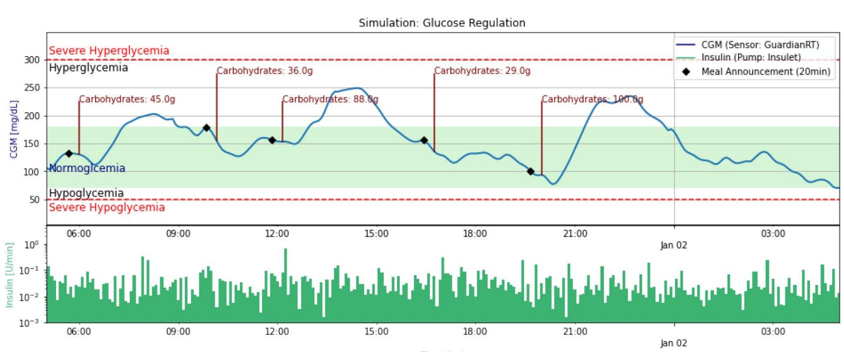

Figure 3: Glucose levels and Time in Range (TIR) [7]

2.2 Distilling RL Expertise into an LLM Controller

The core innovation lies in transferring the RL agent's implicit knowledge into an LLM through supervised fine-tuning, enabling transparent reasoning. This stage involves three key steps:

- **Dataset Generation:** We run the trained expert policy π_{RL} for 20,000 of timesteps in the Simglucose simulator, collecting a large dataset of expert state-action trajectories (s_t, a_t, r_t).
- **JSON Textualization:** A deterministic textualization method converts numerical data into natural language prompts formatted in JSON [12]. Each prompt captures the system's state, including current and historical glucose readings and past insulin doses, paired with the expert action taken by the RL agent. This produces a dataset $\mathcal{D} = (x_t, y_t)$, where x_t is the textualized state and y_t is the textualized expert action.
- **LLM Fine-Tuning:** We fine-tune open-source LLMs (Llama 3 8B and Qwen 8B) using **Parameter-Efficient Fine-Tuning (PEFT) with LoRA** [13]. The LLM is trained via supervised learning to minimize the negative log-likelihood of the expert action y_t given the state prompt x_t:

$$\mathcal{L}_{\text{SFT}}(\theta) = -\sum_{(x_t, y_t) \in \mathcal{D}} \log \pi_\theta(y_t \mid x_t)$$

  This process enables the LLM to emulate the expert RL policy's behavior [14].

To achieve a robust and intelligent system, we combine the strengths of the RL policy and the fine-tuned LLM into a hybrid controller, enhancing both performance and interpretability.

2.3 Explainable Control for Patient Trust

After producing a_t, the LLM is prompted to justify its choice, enabling transparency:

```
{ "context": { "current_glucose": "195 mg/dL",
  "trend": "rising rapidly",
  "glucose_history": [180,165,150,145],
  "last_meal": "30 minutes ago (optional)",
  "insulin_on_board": "2.5U" },
  "decision": { "action": "Deliver 1.2U correction bolus" },
  "instruction": "Explain why this decision was made
  in simple terms for a patient." }
```

The LLM generates a clear rationale (Figure 4,5,6,7), providing the crucial "why" behind each action, thereby transforming an opaque algorithm into a trusted, collaborative copilot.

3 Results

We validated our framework on the Simglucose implementation [15] of the FDA-approved UVA/-Padova T1D simulator [9] across the 10-subject adult and 10-subject adolescent cohorts, which represent the variability of a real T1D population.

Experimental Setup: Models were evaluated over 100 24-hour simulations per subject, using a challenging meal protocol with a significant amount of carbohydrates daily. We compare four controllers based on Time in Range (TIR %) and Failure Rate % (severe hypoglycemia):

1. Basal-Bolus Human Error (BBHE), a clinical baseline requiring manual CHO counting with simulated human error

2. Expert RL (PPO) agent;

3. Fine-tuned Qwen3 8B LLM controller;

4. Fine-tuned LLaMA 3.1 8B controller.

Quantitative Performance: The results, summarized in Table 1 and Figure 1, demonstrate that our approach achieves explainability without compromising on high performance. The **LLaMA 3.1 8B** controller achieves the highest Time in Range (TIR) for the adult cohort (73.5%) and is statistically comparable to the BBHE baseline, despite requiring **no carbohydrate counting**, which is a tedious manual process.. Crucially, it outperforms the standalone RL expert by a small margin, suggesting the LLM provides a beneficial reasoning layer. It also reduces the rate of catastrophic failures (e.g., severe hypoglycemia) compared to the RL expert, highlighting its potential for improved safety.

Table 1: Performance comparison of glucose control algorithms (Mean ± SD).

Controller	**Manual CHO Counting**	**TIR (Adults) (%)**	**TIR (Adolescents) (%)**	**Failure Rate (%)**
BBHE (Baseline)	Yes	70.2 ± 12.5	71.4 ± 12.3	0.00
RL Expert (PPO)	No	69.1 ± 10.5	63.7 ± 14.0	4.93
LLM (Qwen3 8B)	No	71.3 ± 9.8	65.1 ± 13.5	2.15
LLM (LLaMA 3.1 8B)	No	73.5 ± 9.2	66.8 ± 12.9	1.31

Qualitative Performance: Generating Trust through Explanation: The key result of our work is the controller's ability to provide clear, actionable explanations. Figures 4,5,6,7 provide explanations for the decision, patient-friendly rationales, detailed explanations, and insights and recommendations. This transforms the patient's interaction with their device from passive to active understanding.

LLM-Generated Decision & Explanation

Observation: Glucose is 210 mg/dL and rising 60 minutes after a meal. The last insulin dose was 2 hours ago.

Action: Deliver a correction bolus of 1.5 Units.

LLM Explanation: "Your glucose is currently high and still rising after your recent meal. To help bring it back into your target range, I am delivering a small correction dose of 1.5 units. This dose is calculated to be strong enough to lower your glucose without causing a rapid drop or 'crashing' later."

Validity: A human evaluation confirms this rationale is clinically sound, appropriate for the situation, and communicates the controller's intent effectively to the patient.

Figure 4: LLM-Generated Decision / Explanation

Short Explanation (Patient-Friendly Rationale)

Obs: (current_glucose: 195, trend: "rising", time_since_meal_min: 45, glucose_history_mg_dL: [180, 165, 150])
Act: (insulin_dose_U: 1.2)

Explanation: Your glucose is high and still rising after your recent meal. I'm giving you a small correction dose of 1.2 units to help guide your blood sugar back toward your target range.

Validity (Human Evaluation):
This explanation is clear, concise, and uses patient-friendly language. It correctly identifies the reason for the correction (post-meal hyperglycemia) and explains the action without causing alarm. It effectively communicates the system's intent.

Figure 5: Short Explanation (Patient-Friendly Rationale)

Detailed Explanation (Decision Rationale)

Obs: (current_glucose: 220, trend: "stable", time_since_meal_min: 150, insulin_on_board_U: 0.8, glucose_history_mg_dL: [218, 225, 222])
Act: (insulin_dose_U: 0.75)

Explanation: The primary goal is to address persistent hyperglycemia (glucose at 220 mg/dL), which is above the target range. Although the trend is stable, glucose has remained elevated for over two hours post-meal. The system calculates a conservative correction bolus of 0.75 units, balancing the need to reduce high glucose with the 0.8 units of insulin still active (*insulin on board*). This avoids excessive "stacking" that could trigger hypoglycemia. The decision prioritizes safety while nudging glucose down.

Validity (Human Evaluation):
This is an excellent example of the agent explaining a more complex clinical concept—insulin stacking. It correctly identifies the key variables influencing its decision: the stable but high glucose, the time since the meal, and the active insulin. The rationale demonstrates a safe, conservative approach that aligns with best practices in diabetes management.

Figure 6: Detailed Explanation (Decision Rationale)

Guidance-Oriented Explanation (Insight and Recommendation)

Obs: (current_glucose: 140, trend: "falling slowly", time_since_meal_min: 180, insulin_on_board_U: 1.5, last_bolus_U: 2.0)
Act: (insulin_dose_U: 0.0)

Explanation & Guidance: Your glucose is currently in range at 140 mg/dL, but the model detects a slow downward trend and notes 1.5 units of insulin still active from your last dose. To avoid dropping too low, the system delivers no additional insulin.

Recommendation: This pattern suggests your last dose may have been slightly too strong. Continue monitoring over the next hour. If the downward trend persists, consider a small snack (10–15 g carbohydrates) to stay safely in range.

Validity (Human Evaluation):
This explanation is outstanding. It not only justifies inaction (not dosing), a critical control decision, but also provides insight into insulin sensitivity and prior dosing. The recommendation is safe, actionable, and empowers the patient to participate in their care. This is a perfect example of how an explainable system can serve as an educational tool.

Figure 7: Guidance-Oriented Explanation (Insight and Recommendation)

4 Insights

While quantitative results confirm strong performance, the greater impact is shifting insulin control from an opaque black box to a transparent, collaborative partner.

- **From Black Box to Glass Box:** We build an **auditable AI system for healthcare**, allowing patients and clinicians to query model reasoning in real time, enabling safe deployment.
- **Catalyst for Adoption and Safety:** Explainability fosters clinical acceptance, encouraging physician prescription and patient adoption. Transparency also acts as a safeguard, surfacing anomalies hidden in purely numerical controllers.
- **Patient Empowerment:** LLM-T1D serves as an **copilot**, providing not only automated dosing but also clear explanations that educate users and promote active engagement.

5 Conclusion

We introduced a framework for an **explainable, trustworthy AI controller for T1D**, distilling RL expertise into an LLM to overcome the long-standing trust barrier in clinical AI. Our controller delivers strong glycemic control while providing clear, data-driven rationales for its actions. Beyond performance, it marks progress toward safe, autonomous systems that manage physiological complexity. This work offers a concrete path for translating advanced AI into trust-centered clinical practice, empowering patients and clinicians, and extending toward personalized medicine beyond T1D.

References

[1] Linda A DiMeglio, Carmella Evans-Molina, and Richard A Oram. Type 1 diabetes. *The Lancet*, 391(10138):2449–2462, 2018.

[2] Diabetes Control and Complications Trial Research Group. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *New England journal of medicine*, 329(14):977–986, 1993.

[3] Richard M Bergenstal, William V Tamborlane, Andrew Ahmann, John B Buse, George Dailey, Stephen N Davis, Carol Joyce, Tim Peoples, Bruce A Perkins, John B Welsh, et al. Effectiveness of sensor-augmented insulin-pump therapy in type 1 diabetes. *New England Journal of Medicine*, 363(4):311–320, 2010.

[4] Monica Andrade Lima Gabbay, Melanie Rodacki, Luis Eduardo Calliari, Andre Gustavo Daher Vianna, Marcio Krakauer, Mauro Scharf Pinto, Janice Sepúlveda Reis, Marcia Puñales, Leonardo Garcia Miranda, Ana Claudia Ramalho, et al. Time in range: a new parameter to evaluate blood glucose control in patients with diabetes. *Diabetology & metabolic syndrome*, 12(1):22, 2020.

[5] Aria Saunders, Laurel H Messer, and Gregory P Forlenza. Minimed 670g hybrid closed loop artificial pancreas system for the treatment of type 1 diabetes mellitus: overview of its safety and efficacy. *Expert review of medical devices*, 16(10):845–853, 2019.

[6] Marc D Breton and Boris P Kovatchev. One year real-world use of the control-iq advanced hybrid closed-loop technology. *Diabetes technology & therapeutics*, 23(9):601–608, 2021.

[7] Chirath Hettiarachchi, Nicolo Malagutti, Christopher Nolan, Eleni Daskalaki, and Hanna Suominen. A reinforcement learning based system for blood glucose control without carbohydrate estimation in type 1 diabetes: In silico validation. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 950–956. IEEE, 2022.

[8] Chirath Hettiarachchi, Nicolo Malagutti, Christopher J Nolan, Hanna Suominen, and Elena Daskalaki. G2p2c—a modular reinforcement learning algorithm for glucose control by glucose prediction and planning in type 1 diabetes. *Biomedical Signal Processing and Control*, 90:105839, 2024.

[9] Boris P Kovatchev, Marc Breton, Chiara Dalla Man, and Claudio Cobelli. In silico preclinical trials: a proof of concept in closed-loop control of type 1 diabetes, 2009.

[10] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

[11] J Lawton, David Rankin, DD Cooke, M Clark, J Elliot, S Heller, Uk Nihr Dafne Study Group, et al. Dose adjustment for normal eating: a qualitative longitudinal exploration of the food and eating practices of type 1 diabetes patients converted to flexible intensive insulin therapy in the uk. *Diabetes research and clinical practice*, 91(1):87–93, 2011.

[12] Ishup Ali Khan. Xml and json translator. 2025.

[13] Zeyu Han, Chao Gao, Jinyang Liu, Jeff Zhang, and Sai Qian Zhang. Parameter-efficient fine-tuning for large models: A comprehensive survey. *arXiv preprint arXiv:2403.14608*, 2024.

[14] Venkatesh Balavadhani Parthasarathy, Ahtsham Zafar, Aafaq Khan, and Arsalan Shahid. The ultimate guide to fine-tuning llms from basics to breakthroughs: An exhaustive review of technologies, research, best practices, applied research challenges and opportunities. *arXiv preprint arXiv:2408.13296*, 2024.

[15] Chiara Dalla Man, Francesco Micheletto, Dayu Lv, Marc Breton, Boris Kovatchev, and Claudio Cobelli. The uva/padova type 1 diabetes simulator: new features. *Journal of diabetes science and technology*, 8(1):26–34, 2014.

Appendix A: Textualization Engine (TE)

The **Textualization Engine (TE)** is a critical component that bridges the numerical world of Reinforcement Learning (RL) with the language-based world of Large Language Models (LLMs). It acts as a deterministic translator, converting structured numerical state-action data into natural language prompts that LLMs can process and learn from. RL agents operate on arrays of numbers (e.g., glucose levels), whereas LLMs reason over text; the TE enables this alignment.

Key Functions

- **Dataset Construction:** Running an expert RL policy for thousands of steps yields state–action pairs $\{(s_t, a_t)\}$. The TE transforms each s_t into a descriptive text prompt x_t and each a_t into a target response y_t, producing a dataset

$$\mathcal{D} = \{(x_t, y_t)\},$$

  which is then used for LLM fine-tuning.
- **Structured Prompt Formatting:** The TE outputs not just sentences but structured, machine-readable prompts (often JSON-like). A typical prompt includes:
  1. *Instruction Block:* A directive that defines the LLM's role and output format (e.g., "You are an expert insulin dosing assistant. Output must be in Units of insulin.").
  2. *Input/State Block:* Clearly labeled system state (e.g., glucose: 195 mg/dL, trend: rising, insulin on board: 2.5 U).
  3. *Response Block:* A placeholder for the LLM's textualized action.
- **Operational Modes:**
  - *State-to-Action:* Converts s_t into a prompt for LLM-only controllers.
  - *State-and-Action:* Encodes both s_t and a candidate action g_t (e.g., from an RL agent) into a single prompt, enabling the LLM to refine or override RL suggestions in hybrid architectures.

By systematically translating numerical data into structured prompts, the TE enables the LLM to absorb the implicit knowledge of an RL policy while providing natural language outputs. This dual role supports both high-performance control and human-understandable explanations, laying the foundation for safe and trustworthy AI in healthcare and beyond.

Appendix B: Handling of Different Data Types by the TE

The **Textualization Engine (TE)** standardizes diverse data types into labeled, human-readable formats, typically within structured JSON prompts. This ensures that numerical and symbolic information from the simulation environment is presented to the LLM with sufficient context for reasoning and learning.

Continuous Values

For continuous variables (e.g., glucose levels), the TE:

- **Adds Labels and Units:** Each value is paired with descriptive text and its unit. Glucose values include (mg/dL).
- **Provides Contextual Hierarchy:** Values are placed in logical groupings to preserve the system's structure.

Time-Series Data

Temporal context is critical for detecting trends. The TE:

- **Formats as Arrays:** Historical measurements are represented as lists (e.g., `"glucose_history_mg_dL": [180, 165, 150]`).

- **Uses Descriptive Labels:** Arrays are clearly named so the LLM can recognize sequences and reason about trends (e.g., rising vs. falling glucose).

By meticulously labeling, structuring, and contextualizing continuous, categorical, and temporal data, the TE delivers a rich, unambiguous representation of the environment. This allows the LLM to capture the implicit decision logic of the expert RL policy while producing outputs that remain interpretable to humans.