Explainable Insulin Pump Control with LLMs 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 chronic autoimmune disease requiring individuals to perform a continuous, life-long optimization problem [1]: maintaining blood glucose levels within a narrow, healthy range (70 180m~g/dL) (3.9–10 mmol/L) through exogenous insulin administration [2, 3]. Maintaining optimal blood glucose levels is a lifelong challenge, as both low blood glucose (hypoglycemia) and high blood glucose (hyperglycemia) can harm health and, in severe cases, lead to life-threatening conditions. The total time spent within this range is known as Time in Range (TIR) [4] (see Figure 2). The complexity of this task is immense, influenced by meals, exercise, stress, and hormonal changes, creating a significant cognitive load for millions worldwide [5, 6].

Modern Artificial Pancreas Systems (APS) have begun to alleviate this burden by automating basal insulin delivery using traditional controllers like PID and MPC. More recently, Reinforcement Learning (RL) has emerged as a promising frontier, with its ability to learn complex, personalized control policies from data, adeptly handling the non-linear dynamics and delays inherent in the human glucoregulatory system [7, 8]. Studies using state-of-the-art algorithms like Proximal Policy Optimization (PPO) have shown that RL agents can learn effective insulin dosing strategies without manual carbohydrate counting, a major leap towards a fully automated system.

However, this technological advance faces a profound human barrier: **the trust deficit**. RL policies are fundamentally opaque. An RL agent cannot explain *why* it recommended a specific insulin dose—a decision with immediate safety implications. For a patient or a clinician, ceding control

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: The Second Workshop on GenAI for Health: Potential, Trust, and Policy Compliance.

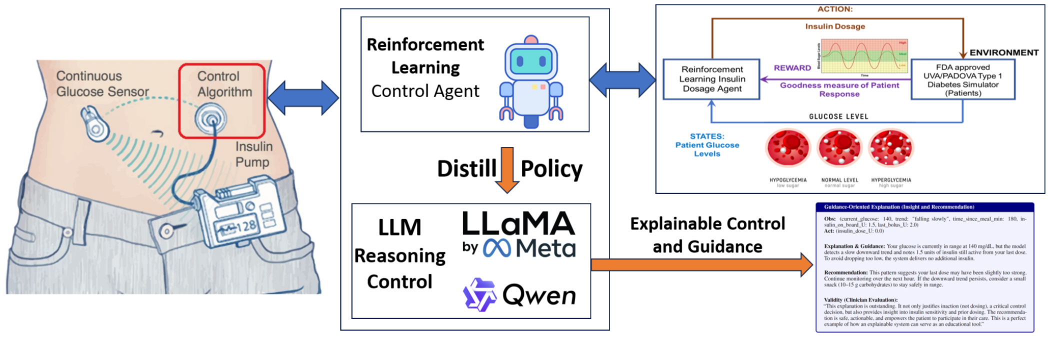

Figure 1: System Disgram for LLM-T1D

of life-sustaining therapy to a black box is an unacceptable risk. This lack of interpretability is the single greatest obstacle to the widespread adoption of advanced AI in critical healthcare applications.

This paper proposes a paradigm shift to solve this challenge. Inspired by recent successes in distilling RL policies for complex physical systems like data center cooling, we introduce a framework to create a transparent, explainable, and trustworthy insulin pump controller. Our core innovation is to **distill the implicit, numerical knowledge of an expert RL policy into the explicit, reasoning framework of a Large Language Model (LLM)** Figure 1. By training an LLM on the decisions of a high-performing RL agent, we create a controller that retains the nuanced performance of its "teacher" while gaining the crucial ability to articulate its decision-making process in natural language.

Our main contributions are:

- **A Hybrid RL LLM Control Framework:** First to adapt and apply a policy distillation methodology to create an explainable LLM-based controller for T1D insulin management, merging the optimization strengths of RL with the interpretability of LLMs.
- **Trustworthy and Explainable AI:** This solution demonstrates that the LLM controller can generate clear, data-driven rationales for its insulin dosing decisions, transforming it from a black box into an intelligent "copilot" that builds patient trust and facilitates safer human-in-the-loop oversight.
- **Demonstrated Performance and Safety:** We validate our approach using the FDA-approved UVA/Padova T1D simulator [9], showing our hybrid controller achieves excellent glycemic control, comparable to or exceeding the expert RL baseline, while providing a new layer of insight.
- **A Pathway to Patient Empowerment:** We argue that this explainable AI approach is a crucial step toward not only safer adoption but also greater patient understanding and engagement in their own care.

2 Problem Definition

Automated glucose control in T1D can be modeled as a *Partially Observable Markov Decision Process (POMDP)*, defined by the tuple:

$$(\mathcal{S}^*, \mathcal{S}, \mathcal{O}, \mathcal{A}, \mathcal{P}, \mathcal{R}),$$

where the true physiological state is not fully observed and must be inferred from noisy sensor data.

2.1 State Space

The environment is defined as a Partially Observable Markov Decision Process (POMDP). Observed states include current and historical glucose measurements, insulin delivery, and meal announcements:

$$s_t = (g_{t-k:t}, i_{t-k:t}, m_{t-k:t})$$

where k is set to one hour. This captures the effect of insulin action and variability in meal timing.

Including history provides essential context on glucose trends and insulin-on-board.

2.2 Action Space

The agent selects an action a_t, which is mapped to an insulin infusion I_{pump} (U) delivered by the pump. The effect can be summarized by:

$$I_{\text{pump}} = I_{\max} \cdot e^{\eta(a_t - 1)}.$$

This exponential mapping allows fine-grained control of small basal doses and rapid delivery of larger boluses.

2.3 Reward Function

The objective is to maximize *Time in Range (TIR)*, defined as the percentage of time spent within 70–180 mg/dL. The reward is derived from the Blood Glucose Risk Index (RI), which assigns exponentially higher penalties to deviations outside the target range, with particularly severe penalties for hypoglycemia ($g_t \leq 39$ mg/dL). The agent maximizes expected average reward over an infinite horizon, reflecting the continuous, life-long nature of diabetes management.

$$R(s_t, a_t) = \begin{cases} -15000, & g_{t+1} \leq 39 \text{ mg/dL} \\ 10 \cdot \big(-RI(g_{t+1}) + 100 \big), & \text{otherwise.} \end{cases}$$

This penalizes hypoglycemia heavily while rewarding safe glucose ranges.

2.4 MDP Formulation

Table 1: MDP Formulation for Glucose Control

States S	Glucose history, insulin history, meal announcements
Actions A	Insulin infusion (basal + bolus)
Reward R	$-RI(g)$ with severe penalty for $g \leq 39$ mg/dL

While the RL agent's objective is to optimize this numerical reward function, the overarching clinical goal is to develop a system that is not only effective but also **safe, reliable, and trustworthy**. An optimal policy from a purely mathematical standpoint is useless if the patient cannot trust it enough to use it. Therefore, our problem extends beyond policy optimization to include the generation of faithful, understandable explanations for the agent's actions.

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.

3 Methodology

3.1 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 $\pi_{\text{RL}}(a_t \mid s_t)$ that maximizes average rewards. This results in a robust, optimized policy that serves as the "expert teacher" for our large language model (LLM) controller.

3.2 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:

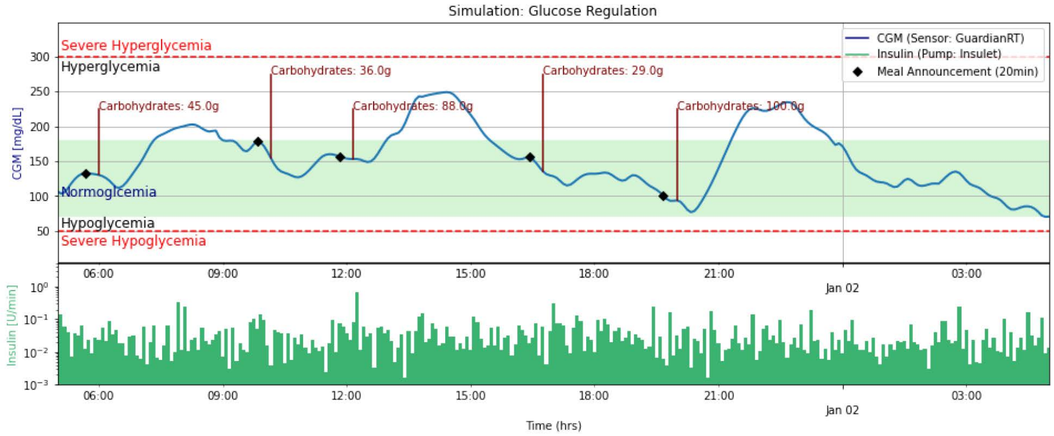

Figure 2: Glucose levels and Time in Range (TIR) [7]

- **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).
- **Textualization:** We developed a deterministic Textualization Engine (TE) that 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 use this textualized dataset to fine-tune open-source LLMs (Llama 3.1 8B and Qwen3 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.

3.3 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 used for LLM fine-tuning.

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

- **Structured Prompt Formatting:** The TE outputs not just sentences but structured prompts (JSON). 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).

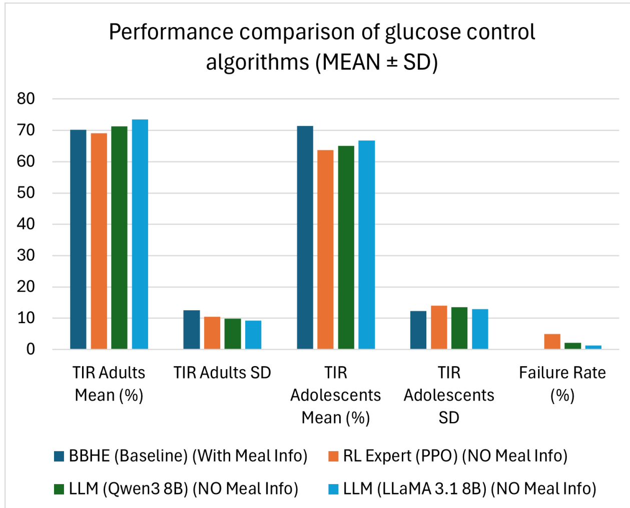

Figure 3: Performance comparison of glucose control algorithms (MEAN ± SD)

3. *Response Block:* For the LLM's textualized action.

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.

3.4 Handling of Different Data Types by the TE

The **Textualization Engine (TE)** standardizes diverse data types into labeled, human-readable formats, 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.

Table 2: 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

3.5 Explainable Control for Patient Trust

After producing a_t, the LLM is prompted to justify its choice in natural language, 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, providing the crucial "why" behind each action, thereby transforming an opaque algorithm into a trusted, collaborative copilot.

4 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 2 and Figure 3, 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. '

Qualitative Performance: Generating Trust through Explanation: The key result of our work is the controller's ability to provide clear, actionable explanations. The following provides 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.

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.

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.

Guidance-Oriented Explanation (Insight and Recommendation)

Obs: (current_glucose: 165, trend: "falling", time_since_meal_min: 90, insulin_on_board_U: 1.5, last_bolus_U: 2.0, glucose_history_mg_dL: [150, 130, 200])
Act: (insulin_dose_U: 0.0)

Explanation & Guidance: Your glucose is currently in range at 165 mg/dL, but the model detects a fast 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 two hours. 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.

LLM-Generated Decision & Explanation Example:

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

5 Insights

While quantitative results confirm strong performance, the deeper impact lies in transforming insulin control from an opaque black box into a transparent, collaborative partner.

- **From Black Box to Glass Box:** We create an **auditable AI system for healthcare**, enabling patients and clinicians to query the model's reasoning in real time. This is not a convenience but a prerequisite for trust and safe deployment in high-stakes care.
- **Catalyst for Adoption and Safety:** Explainability drives clinical acceptance. Physicians are more willing to prescribe, and patients more likely to adopt, a system that can justify its actions. Transparency also acts as a safety net, exposing anomalies that would remain hidden in a purely numerical controller.
- **Patient Empowerment:** LLM-T1D becomes an **intelligent copilot**, offering not just automated dosing but also clear explanations. By articulating the 'why' behind insulin adjustments, the system educates users about their metabolic responses, fostering understanding and active engagement in care.

6 Conclusion

We presented a novel framework for building an **explainable and trustworthy AI controller for T1D**, distilling the expertise of an RL policy into a reasoning LLM. This directly addresses the long-standing barrier of trust in deploying advanced AI for clinical care.

Our LLM controller achieves strong glycemic control while providing clear, data-driven rationales for its actions. Beyond performance, it represents a step toward autonomous systems that safely manage the complexities of the human body. This work offers a concrete pathway for translating advanced AI into trust-building clinical practice, empowering both patients and clinicians, for personalized medicine beyond Type 1 Diabetes.

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.