Ethical Considerations in Automated Insulin Delivery: A Case Study in AI-Driven Diabetes Management

Student Number - 720017170

January 6, 2026

Word Count: Section 1 - 2079, Section 2 - 1099

1 Case Study

1.1 Introduction

The Diabeloop DBLG1 Closed-Loop System (Diabeloop, 2024) is a development in AI-driven diabetes management, but it raises some ethical questions. DBLG1 integrates three core components: a Dexcom G6 CGM sensor transmitting glucose readings every five minutes via Bluetooth, an insulin pump for hormone delivery, and a machine learning algorithm that makes automated treatment decisions (Benhamou et al., 2019).

DBLG1 is an example of a self-learning system where the algorithm adapts to individual patient characteristics over time, requiring patients to only provide four initial settings: body weight, total daily insulin dose, typical meal sizes, and a safety basal rate (Diabeloop, 2024). However, this minimal input design prioritises ease of use over detailed personalisation.

Clinical evidence reports DBLG1's efficacy in improving glycaemic control. Benhamou et al. (2019) found an improved time in range (70-180 mg/dL) of 68.5% compared to 59.4% with conventional therapy, alongside a 53% reduction in time spent in hypoglycemia, both statistically significant improvements. Following its clinical success, DBLG1 was evaluated in a real-world 6-month follow-up trial, reinforcing these clinical results. Twenty-four patients increased time in range from 53% to 69.7% ($p < 0.0001$) and decreased HbA1c from 7.9% to 7.1% ($p < 0.001$), with no serious adverse events reported (Amadou et al., 2021).

When considering the data journey (Leonelli and Tempini, 2020) within DBLG1, there are several stages that can introduce ethical issues. Data is collected at high frequency from the CGM sensor, which causes concerns around data privacy and security. Transmission is over Bluetooth making data interception or loss feasible, compounding these privacy concerns. Once the data is used within the machine learning algorithm, issues surrounding algorithmic bias and fairness may occur, especially if the training data does not represent diverse patient populations. Finally, the self-learning logic within the algorithm may introduce concerns around transparency and accountability.

1.2 Ethical Issues

As stated above, several ethical issues arise during different stages of the data journey within DBLG1. For this case study, focus will be on the two deemed most relevant in current medical AI research, algorithmic bias and fairness (Cross et al., 2024), and accountability and transparency (Shaban-Nejad et al., 2020).

1.2.1 Algorithmic Bias and Fairness

Algorithmic bias occurs when an AI algorithm produces different outputs or outcomes, caused by prejudice or incomplete data (Panch et al., 2019). Differing outputs are worrying in medical applications because biased algorithms may lead to suboptimal or harmful interventions for certain patient groups, worsening health disparities (Cross et al., 2024).

Clinical research on DBLG1 has focused on adult populations with type 1 diabetes within a small sample (Benhamou et al., 2019; Amadou et al., 2021). This creates a bias against minority patient groups, such as children, elderly patients or individuals from different ethnic backgrounds. These individuals may have different physiological responses to insulin or varying lifestyle factors affecting glucose tendencies. Underrepresentation could affect the fairness of DBLG1 when deployed in real-world environments as minority patient groups may experience inaccurate treatment because the algorithm has not been properly trained on their specific characteristics. Inaccurate treatment can reduce fairness in healthcare, posing an ethical dilemma as all patients should be equally benefiting from medical AI regardless of their demographic characteristics (Cross et al., 2024).

Data representation is not the only source of bias within DBLG1. Bias can be introduced across the entire data journey, from data collection to algorithmic design (Cross et al., 2024). For example, the Dexcom CGM used within DBLG1 experiences inaccuracies in glucose readings when behaviour and location vary (Mensh et al.,

2013; Dyess et al., 2024). Inaccuracies lead to systematic errors in glucose data for those patients, influencing the algorithm's recommendations by propagating through the system. Similarly, if the algorithm does not account for variations in insulin sensitivity among different demographics, biased insulin dosing recommendations are likely to occur.

1.2.2 Accountability and Transparency

Another ethical issue within DBLG1 is accountability and transparency. A main consideration within AI-driven medical systems is the 'black box' nature of many algorithms, where the learning and inference logic is not easily interpretable by humans (Shaban-Nejad et al., 2020).

Opacity is exemplified within DBLG1. Diabeloop markets the benefits of their self-learning algorithm, which adapts to individual patient characteristics over time based on limited user characteristics shown in Section 1.1 (Diabeloop, 2024). However, the exact workings of this algorithm and how it uses these inputs are not transparent to users or clinicians. Opacity violates the ethical principle of explicability (Floridi and Cowls, 2019), as AI systems should be understandable and interpretable by humans, particularly in high-stakes domains like healthcare.

The complexity of DBLG1 creates a barrier for establishing accountability. Shaban-Nejad et al. (2020) argue that the 'black box' nature of medical AI makes it unclear who is responsible for unforeseen outcomes. In the case of DBLG1, if the system makes an incorrect insulin dosing decision leading to an adverse event, accountability is unclear. As Leonelli (2016) identifies, this is a common limitation of 'distributed knowledge production systems', where development and application is fragmented across developers, clinicians, and users, making it difficult to determine who is responsible for any errors. Limited accountability leaves the patient to take on the risks where no single entity takes responsibility for the results or consequences.

The lack of transparency introduces further issues with informed consent. Patients who are using DBLG1 may not fully understand how the algorithm functions or the potential risks involved in relying on an unclear AI system for critical health decisions. Reduced understanding undermines the ethical principle of autonomy (Floridi and Cowls, 2019), which outlines that patients should have the right to make decisions about their healthcare based on clear information.

1.3 Ethical Frameworks and Principles

To evaluate the ethical issues presented in Section 1.2, two ethical frameworks will be discussed: consequentialism from a utilitarian perspective and principlism. This section aims to determine the ethicality of DBLG1 whilst identifying which framework is best suited to deal with the case study's ethical issues.

1.3.1 Consequentialism (Utilitarian Approach)

Beaulieu and Leonelli (2021) define utilitarianism as an ethical framework that classifies an action as good, if it maximises a population's welfare by making the majority of individuals who are affected happier than they were before the action occurred. When evaluating the ethicality of DBLG1, utilitarian ethics relies on a social welfare function described by Card and Smith (2020), shown below in Equation 1:

$$v(s) = \sum_{e \in \mathcal{E}} w_e(s) \tag{1}$$

Where $\mathcal{E}$ represents the total population of patients, and $w_e(s)$ represents the individual well-being utility, in this case glycemic control, of patient e using the system.

Applying the function to DBLG1, with the clinical results from Benhamou et al. (2019) and Amadou et al. (2021), confirms that the system maximises $v(s)$ or well-being utility. According to Equation 1, the device is ethically justified as the total sum of benefits outweighs any harms to the patient population. Therefore, in this case study, a utilitarian approach would allow the DBLG1's deployment, despite its uneven distribution of utility, as it improves overall health outcomes for the majority of patients with type 1 diabetes whilst not causing overall harm.

Card and Smith (2020) discuss several limitations of the utilitarian approach that are pertinent for medical AI, especially issues caused by these social welfare functions when considering algorithmic bias and fairness discussed in Section 1.2.1. The function in Equation 1 assumes that all individuals are equally affected and the harms are distributed evenly, though this is likely not the case in practice. DBLG1 is only clinically validated on a small uniform subset of the total diabetes population, indicating that there might be bias within the algorithm that could lead to suboptimal outcomes for underrepresented groups. This means that the overall well-being $v(s)$ is maximised but this might not be the case for all individuals. In addition, the maximisation of $v(s)$ in this case refers only to glycaemic control, ignoring other important psychological factors that contribute to overall

well-being which are harder to quantify: anxiety, trust in the system, sleep disruption and many more that will negatively impact certain individuals' well-being.

These limitations are where a utilitarian approach falls short when addressing issues in medical AI systems like DBLG1 due to sacrificing a few individuals for the 'greater good' of the population. More importantly, it fails to consider what actually constitutes improved well-being for individuals, instead focusing on aggregate metrics that overlook important discrepancies in patient experiences. In this case, as with many medical AI systems, other ethical frameworks are more appropriate, such as principlism.

1.3.2 Principlism

To address the limitations of utilitarianism, principlism offers a holistic ethical framework by ensuring thoughts span multiple ethical perspectives: autonomy, beneficence, non-maleficence, justice, and explicability (Floridi and Cowls, 2019). Principlism is useful when considering the ethical issues presented in Section 1.2, and allows us to evaluate DBLG1 through multiple ethical lenses, revealing tension as DBLG1 excels at delivering glycaemic control (beneficence) and violates the principles designed to protect individual patients (justice, explicability, autonomy).

Justice requires AI development to reduce all types of discrimination (Floridi and Cowls, 2019). DBLG1 contradicts this due to algorithmic bias, resulting from unrepresentative training data, consistently disadvantaging minority patient groups. Explicability, referring to the need to understand and hold to account the decision-making processes of AI (Floridi and Cowls, 2019), is also violated in this case because DBLG1's 'black box' nature creates a lack of transparency and accountability, making it difficult for patients and clinicians to understand how decisions are made. Another important aspect of explicability is informed consent, relating to the principle of autonomy, which states that we must respect individuals' views and rights to make informed decisions about their healthcare (Beauchamp and Childress, 2019). Patients cannot have educated control over their care if they do not understand DBLG1's algorithm or how their data is being used, which Diabeloop fails to provide.

On the other hand, beneficence and non-maleficence (Beauchamp and Childress, 2019) are upheld within DBLG1 although inconsistently across all patient groups. DBLG1 upholds beneficence as clinical evidence discussed within Section 1.1 demonstrates that DBLG1 successfully manages glucose conditions, improving time in range and reducing hypoglycemic occurrences along with limiting adverse events (Benhamou et al., 2019; Amadou et al., 2021). Therefore, DBLG1 provides clear benefits to patients whilst reducing potential harms. Conversely, non-maleficence is inconsistently upheld across all patient groups due to algorithmic bias, which causes DBLG1 to fail to protect specific subgroups from harm, revealing a conflict between aggregate and individual safety.

Principlism offers a clearer evaluation of the ethical issues within DBLG1, due to the individual clarity of its principles and its ability to address differing utility conditions, a key limitation of utilitarianism discussed in Section 1.3.1. It clearly evaluates the strengths and weaknesses of DBLG1 from multiple ethical perspectives, identifying areas where it is ethically responsible and areas that require improvement to maintain responsible and fair development.

Similar to utilitarianism, principlism has its shortcomings. Mittelstadt (2019) discusses some of these limitations, specifically that AI ethics lacks common aims and duties or clear methods to translate these principles into practice. These limitations become clear when considering competing principles. For example, methods to improve transparency and accountability may reduce the clinical efficacy of the algorithm, creating conflicts between explicability and non-maleficence. Alternatively, algorithms trained on more diverse datasets to improve justice may reduce the overall clinical efficacy for the majority group, impacting beneficence. These conflicts show that principlism provides a more comprehensive evaluation than utilitarianism but it lacks a clear route to implementation as there is no clarity on resolving these conflicts. As a result, principlism can function more as a diagnostic framework than a practical tool for responsible data governance (Mittelstadt, 2019).

1.4 Conclusion

DBLG1 shows well-established clinical performance in diabetes management; however, it contains a dispute between the benefits of improved glycemic control and the ethical issues posed by algorithmic bias and fairness, and accountability and transparency. The system's clinical efficacy is clear, yet the ethical issues identified make robust governance and mitigation strategies necessary.

Utilitarianism justifies DBLG1's deployment based on aggregate benefits, however it fails to consider distributed utility, a critical failure in medical contexts where incorrect or biased decisions can have severe consequences for individuals. Consequently, principlism offers a better evaluation for the context of this case study, as it allows discussion of the system's ethical strengths as well as revealing a tension between principles and identifying ethical failures. Without governance that maintains fairness and transparency, DBLG1 risks normalising harm on vulnerable patients when overall clinical success is met.

2 Critical Reflection

2.1 Introduction

Section 1 identified ethical issues within DBLG1: algorithmic bias and fairness, and accountability and transparency, both of which need careful governance. When evaluating these issues through the lenses of utilitarianism and principlism, it became clear that DBLG1 demonstrates a clinical and practical success by benefiting its users, but it also contains ethical shortcomings that require governance to ensure the deployment remains responsible. These issues span across the data journey, meaning governing complex systems like DBLG1 requires a process-based framework that encompasses the entire data journey to maintain ethical integrity.

2.2 Mitigating Algorithmic Bias and Fairness Issues

Bias and fairness issues discussed in Section 1.2.1 were caused by unrepresentative data that propagated throughout the data journey to create a model that learns patterns in homogeneous populations. Those patterns introduce an inability to generalise to underrepresented groups, compounding inequalities in healthcare, and reducing effectiveness (Wang et al., 2022).

Governance methods should first mandate dataset diversity during the collection phase of the data journey. To enforce this, regulatory bodies such as the Food and Drug Administration (FDA) and European Medicines Agency (EMA), along with research ethics committees (RECs), should formulate protocols and standards for dataset composition. Datasets must include a wide range of demographic groups, including children, elderly patients, and people of different races. Diversity means that the patterns learnt accurately reflect the entire patient population (Rajkomar et al., 2018).

Unfortunately, true diversity may not be plausible in practice. Logistical and ethical barriers exist when collecting sensitive health data from some groups, especially children. Agarwal et al. (2023) suggest using bias-mitigation techniques, such as adversarial debiasing and reweighting, during model training to increase representation of minority groups without altering the underlying data distribution. Regulators can enforce this by requiring the submission of a bias reduction report containing evidence of bias-mitigation strategies during model training, although these techniques have been seen to produce instability within the model's performance, and so, as stated by Zhang et al. (2018) they are still being developed in clinical settings, posing further complications for regulators. An alternative avenue of technical research involves fairness-enhancing algorithms, as Friedler et al. (2018) demonstrated that they can improve equity metrics; therefore, they could be included within model development and the bias reduction report. Their study reports that these techniques can be negatively influenced by dataset composition and fairness definitions, suggesting regulators must provide explicit guidelines for their appropriate use. Technical approaches may not be enough in isolation because manufacturers may view the cost of diversifying datasets for minority subgroups or implementing bias-mitigation techniques as financially unjustifiable, given the smaller market size. Regulators could encourage this by introducing financial incentives for manufacturers who demonstrate bias-mitigation, to reduce the financial burden of these techniques.

Finally, governance frameworks must think past initial regulation to include continuous monitoring of algorithmic performance post-deployment, especially given that DBLG1 contains a self-learning algorithm that adapts over time. Regulators should implement frequent audits of the algorithm's performance across different demographic groups to detect and address any biases. In practice, this could be achieved by ensuring manufacturers report demographic subgroup analyses to determine if performance is equal across demographic groups (Wang et al., 2022). If differences are identified, corrective measures must be mandated by regulators, such as retraining the model with more diverse data or adjusting the algorithm to account for identified biases.

2.3 Governing Accountability and Transparency

The vagueness around the algorithm's training, inference and self-learning logic makes it difficult for patients and clinicians to understand how decisions are made. This 'black box' nature of DBLG1 creates barriers interfering with accountability and transparency.

Several strategies can be incorporated throughout the data journey, allowing for these barriers to be reduced. Explainable AI (XAI) techniques can be integrated into model development to enhance transparency. SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations) are two frequently used XAI methods that visualise the contributions of each feature to the model's predictions (Ahmed et al., 2025). Additionally, Zhu et al. (2023) suggest the use of confidence intervals to quantify the uncertainty in model predictions, which can provide regulatory bodies with a more comprehensive understanding of the model's outputs, thus improving reliability. These techniques can help regulatory bodies and RECs understand how the algorithm operates allowing governance decisions to be made. For instance, explanations could be used

to establish accountability mechanisms by providing a trail of the model's decisions and locating where errors may occur.

Only relying on technical solutions limits effective governance. Ahmed et al. (2025) discuss that XAI methods often introduce a trade-off between model interpretability and predictive performance; a conflict between beneficence and explicability. In this case, DBLG1's self-learning algorithm would need to be simplified to allow explanation, reducing clinical accuracy. The complexity of XAI could lead to a false sense of transparency, introducing scenarios where Diabeloop is meeting regulatory requirements without actually enhancing clinician or patient understanding. For example, a patient reading XAI explanations may believe they understand the system, but technical language could have caused misunderstanding, undermining informed consent. Consequently, regulatory bodies must enforce the inclusion of XAI during development and the publication of clear and concise materials that communicate the system's processes, limitations, and potential risks. Yet, Diabeloop may avoid full transparency due to intellectual property concerns and may lack the resources to communicate XAI appropriately, making financial incentives ever more necessary.

Governance of accountability faces further barriers. The complexity of the data journey creates difficulty in establishing clear accountability pathways. Petermann et al. (2022) discuss that RECs often lack the machine learning expertise to evaluate complex AI systems and struggle to assess harms and risks that occur post-deployment. These failures suggest that individual governance is insufficient. Instead, governing accountability in DBLG1 requires collaborative governance (Price et al., 2023) where responsibility for monitoring AI is shared among health systems, regulators, and manufacturers, ensuring that knowledge from multiple perspectives is utilised to govern the entire data journey.

2.4 Conclusion

The ethical issues of algorithmic bias and fairness, and accountability and transparency within the DBLG1 Closed-Loop System can be governed through a combination of technical strategies and robust structural frameworks. Governance must span the entire data journey to allow complete coverage of the ethical bottlenecks.

Mitigating ethical issues within DBLG1 cannot rely on a single regulatory entity. Collaborative governance must be introduced to allow knowledge from multiple entities to provide comprehensive coverage across the data journey. Technical solutions like bias-mitigation and XAI are necessary, but they are insufficient without structural changes to data governance. Regulation should emphasise dataset diversity and alter intellectual property rights that currently prevent 'black box' algorithms from receiving deeper evaluation.

References

Agarwal, R., Bjarnadottir, M., Rhue, L., and et al. (2023). Addressing algorithmic bias and the perpetuation of health inequities: An ai bias aware framework. *Health policy and technology*, 12(1):100702–. doi: https://doi.org/10.1016/j.hlpt.2022.100702.

Ahmed, S., Kaiser, M. S., Shahadat Hossain, M., and Andersson, K. (2025). A comparative analysis of lime and shap interpreters with explainable ml-based diabetes predictions. *IEEE access*, 13:37370–37388. doi: https://doi.org/10.1109/ACCESS.2024.3422319.

Amadou, C., Franc, S., Benhamou, P.-Y., and et al. (2021). Diabeloop dblg1 closed-loop system enables patients with type 1 diabetes to significantly improve their glycemic control in real-life situations without serious adverse events: 6-month follow-up. *Diabetes Care*, 44(3):844–846. doi: https://doi.org/10.2337/dc20-1809.

Beauchamp, T. L. and Childress, J. F. (2019). *Principles of biomedical ethics*. Oxford University Press, New York, eighth edition.

Beaulieu, A. and Leonelli, S. (2021). *Data and society : a critical introduction*. SAGE, Los Angeles. Chapter 9.4 - 9.7.

Benhamou, P.-Y., Franc, S., Reznik, Y., and et al. (2019). Closed-loop insulin delivery in adults with type 1 diabetes in real-life conditions: a 12-week multicentre, open-label randomised controlled crossover trial. *The Lancet. Digital health*, 1:e17–e25. doi: https://doi.org/10.1016/S2589-7500(19)30003-2.

Card, D. and Smith, N. A. (2020). On consequentialism and fairness. *Frontiers in artificial intelligence*, 3:34–. doi: https://doi.org/10.3389/frai.2020.00034.

Cross, J. L., Choma, M. A., and Onofrey, J. A. (2024). Bias in medical ai: Implications for clinical decision-making. *PLOS digital health*, 3(11):e0000651–. doi: https://doi.org/10.1371/journal.pdig.0000651.

Diabeloop (2024). Diabeloop dblg1 system overview. Available at: https://www.dbl-diabetes.com/dblg1system-dana-i (Accessed: 21 November 2025).

Dyess, R. J., McKay, T., Feygin, Y., Wintergerst, K., and Thrasher, B. J. (2024). Factory-calibrated continuous glucose monitoring system accuracy during exercise in adolescents with type 1 diabetes mellitus. *Journal of diabetes science and technology*, 18(3):584–591. doi: https://doi.org/10.1177/19322968221120433.

Floridi, L. and Cowls, J. (2019). A unified framework of five principles for ai in society. *Harvard data science review*. doi: https://doi.org/10.1162/99608f92.8cd550d1.

Friedler, S. A., Scheidegger, C., Venkatasubramanian, S., and et al. (2018). A comparative study of fairness-enhancing interventions in machine learning. doi: https://doi.org/10.48550/arxiv.1802.04422.

Leonelli, S. (2016). Locating ethics in data science: responsibility and accountability in global and distributed knowledge production systems. *Philosophical transactions of the Royal Society of London. Series A: Mathematical, physical, and engineering sciences*, 374(2083):20160122–. doi: https://doi.org/10.1098/rsta.2016.0122.

Leonelli, S. and Tempini, N. (2020). *Data Journeys in the Sciences*. Open Access. Springer Nature, Cham, 1 edition.

Mensh, B. D., Wisniewski, N. A., Neil, B. M., and Burnett, D. R. (2013). Susceptibility of interstitial continuous glucose monitor performance to sleeping position. *Journal of diabetes science and technology*, 7(4):863–870. doi: https://doi.org/10.1177/193229681300700408.

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical ai. *Nature machine intelligence*, 1(11):501–507. doi: https://doi.org/10.1038/s42256-019-0114-4.

Panch, T., Mattie, H., and Atun, R. (2019). Artificial intelligence and algorithmic bias: implications for health systems. *Journal of global health*, 9(2):010318–. doi: https://doi.org/10.7189/jogh.09.020318.

Petermann, M., Tempini, N., Kherroubi-Garcia, I., et al. (2022). Looking before we leap: Expanding ethical review processes for ai and data science research. Ada Lovelace Institute, London. https://www.adalovelaceinstitute.org/report/looking-before-we-leap/.

Price, W. N., Sendak, M., Balu, S., and Singh, K. (2023). Enabling collaborative governance of medical ai. *Nature machine intelligence*, 5(8):821–823. doi: https://doi.org/10.1038/s42256-023-00699-1.

Rajkomar, A., Hardt, M., Howell, M. D., and et al. (2018). Ensuring fairness in machine learning to advance health equity. *Annals of internal medicine*, 169(12):866–872. doi: https://doi.org/10.7326/M18-1990.

Shaban-Nejad, A., Michalowski, M., and Buckeridge, D. L. (2020). Explainability and interpretability: Keys to deep medicine. In *Explainable AI in Healthcare and Medicine*, Studies in Computational Intelligence, pages 1–10. Springer International Publishing, Cham. doi: https://doi.org/10.1007/978-3-030-53352-6_1.

Wang, A., Ramaswamy, V. V., and Russakovsky, O. (2022). Towards intersectionality in machine learning: Including more identities, handling underrepresentation, and performing evaluation. doi: https://doi.org/10.48550/arxiv.2205.04610.

Zhang, B. H., Lemoine, B., and Mitchell, M. (2018). Mitigating unwanted biases with adversarial learning. doi: https://doi.org/10.48550/arxiv.1801.07593.

Zhu, T., Li, K., Herrero, P., and Georgiou, P. (2023). Personalized blood glucose prediction for type 1 diabetes using evidential deep learning and meta-learning. *IEEE transactions on biomedical engineering*, 70(1):193–204. doi: https://doi.org/10.1109/TBME.2022.3187703.