Is an application that prepares your data and prompts for a chatbot a medical device? Insulin settings advice mediated by large language models under the European and British device regulations

Tim Street, MEng BEng (Hons), ORCID 0009-0008-4417-6581. Diabettech Ltd, London, United Kingdom. tim@diabettech.com

Preprint, not peer reviewed. Posted to SSRN and diabettech.com in August 2026.

Abbreviations: AID, automated insulin delivery; IMDRF, International Medical Device Regulators Forum; LLM, large language model; MDCG, Medical Device Coordination Group; MDR, Regulation (EU) 2017/745 on medical devices; MDSW, medical device software; MHRA, Medicines and Healthcare products Regulatory Agency.

Keywords: AndroidAPS; automated insulin delivery; large language models; medical device regulation; software as a medical device

Abstract

A paid Android application prepares a user's AndroidAPS data and four prompts for any large language model, checks the returned file and presents it as a review of insulin pump settings. A companion article shows that eleven models read the record consistently and disagreed about what to change, and that the application's check cannot tell one answer from another. This commentary asks what such a product is in regulatory terms. On the definition in Regulation (EU) 2017/745 and Rule 11 as read in MDCG 2019-11, its stated purpose places it in Class IIb, or at the least IIa; the British, American and Google Play positions are set out, and the obligations that follow are the minimum a product of this kind should meet.

The product and its stated purpose

The application examined in the companion article [1] builds a structured package from a user's Nightscout instance and AndroidAPS profile, supplies four prompts for any large language model (LLM), validates the JavaScript Object Notation file the model returns and renders it as a primary profile recommendation, a decision on each of 23 settings, a meal strategy and a list of recommendations with a progress tracker. The listing describes it as "AI ready Nightscout data for safer AAPS tuning" and as "a tool for organizing data, preparing prompts, and helping users think through possible profile improvements more safely and clearly" [2]. An onboarding screen adds that the user should not "change settings solely based on an app result" and that "you use the app and make decisions at your own risk". The settings in question, basal rates, carbohydrate ratios, sensitivity factors, duration of insulin action, targets and the maximum insulin the algorithm may deliver, determine how much insulin an automated insulin delivery (AID) system gives. A subscription costs £4.49 a month. The application is one of a growing number of listings on the same store that offer advice bearing on insulin dose; a systematic assessment of insulin dose calculator apps a decade ago found that most offered no protection against an inappropriate

recommendation and called for coordinated surveillance by regulators and app stores [3]. What is new here is that the reasoning is delegated to a general-purpose model chosen by the user.

The companion article replayed the application's package and prompts against eleven models on one 14-day record. The models quoted the record accurately and then disagreed: three lowered the maximum insulin on board limit in 73 to 98 per cent of accepted conversations, proposing 40 distinct pairs of limits between them, while four left it alone throughout; one proposed sensitivity factors from 30 to 180 mg/dl per unit against a current 56; and the application's check, which inspects the form of the answer and not its content, accepted all of it. Whether any answer was right was not measured; the question here is what the product is, because that decides who must show that its answers are safe.

Qualification as a medical device

Regulation (EU) 2017/745 (MDR) defines a medical device as any instrument, software or other article "intended by the manufacturer to be used, alone or in combination, for human beings" for a medical purpose, including the "diagnosis, prevention, monitoring, prediction, prognosis, treatment or alleviation of disease" [4]. The intended purpose is what the manufacturer states "on the label, in the instructions for use or in promotional or sales materials or statements" [4]. MDCG 2019-11 sets out five decision steps and makes the central test whether the software has a medical purpose on its own [5]. Software "intended to process, analyse, create or modify medical information may be qualified as a medical device software if the creation or modification of that information is governed by a medical intended purpose", and the guidance gives as an example software that searches data "for findings that support a clinical hypothesis as to the diagnosis or evolution of therapy" [5].

The application performs actions well beyond communication. It aggregates the data, decides which 23 settings are open for review, composes the prompts that tell the model how to weigh the evidence and what form its answer must take, validates the answer and presents it as a review with a recommendation tracker. Each is governed by the stated purpose, to help the user decide whether and how to change the settings that control their insulin delivery: the "evolution of therapy" of the guidance's example and the treatment of disease in the Regulation's definition. The guidance also lists, among its examples of medical device software, software "that provides insulin dose recommendations to a patient regardless of the method of delivery of the prescribed dose" [5]; a recommendation to change the rates, ratios and factors from which an AID system computes every dose is of the same kind.

Three counter-arguments arise. The first is that the application merely organises data and the analysis is done by a third-party model the developer does not control. The guidance answers this directly: the location of the processing does not matter, since software "may be qualified as MDSW regardless of its location (e.g. operating in the cloud, on a computer, on a mobile phone…)" [5], and a manufacturer who builds a product around a component it does not make remains its manufacturer. The prompts, the schema, the check and the presentation are the developer's and turn a chatbot conversation into a settings review. The second is that the user makes the decision. Rule 11, discussed below, "relates to the consequences of indirect harm from failure to provide correct information" [5], that is, information a person acts on. The third is the

disclaimer. The intended purpose is read from the product and its marketing as a whole, and a product marketed for “safer AAPS tuning” and “possible profile improvements”, which tracks whether the user has implemented its recommendations, has a therapeutic purpose whatever its onboarding screen says. The “at your own risk” clause allocates risk between the developer and the user as a matter of contract; it does not change what the software is for.

Classification under Rule 11

Rule 11 of Annex VIII provides that software “intended to provide information which is used to take decisions with diagnosis or therapeutic purposes is classified as class IIa, except if such decisions have an impact that may cause: death or an irreversible deterioration of a person’s state of health, in which case it is in class III; or a serious deterioration of a person’s state of health or a surgical intervention, in which case it is classified as class IIb” [4]. MDCG 2019-11 explains that the first limb applies to practically all medical device software, and that the exceptions turn on the significance of the information to the healthcare decision and on the state of the patient, the two axes of the International Medical Device Regulators Forum (IMDRF) framework [5, 6]. Its table maps information that is used to treat or diagnose in a serious condition to Class IIb, and information that drives clinical management in a serious condition to Class IIa [5].

Type 1 diabetes treated with insulin is a serious condition in the IMDRF sense [6]. The information the application presents is a set of proposed values for the settings from which an AID system computes insulin delivery, and the product’s tracker expects the user to apply them. A proposed value that is wrong for the user, such as a sensitivity factor that cuts corrections by two thirds, can produce severe hypoglycaemia or ketoacidosis, each a serious deterioration of health. Information used to set the parameters of insulin therapy is used to treat rather than merely to drive management, which places the software in IMDRF category III.ii and Class IIb, the class in which bolus calculators and dosing software are customarily found. Either class requires a notified body, a quality management system, a clinical evaluation and post-market surveillance; neither can be self-declared. The application’s listing carried no medical device label when examined on 22 August 2026 [2], and Google Play’s policy requires that “apps that are regulated because they are a medical device must be declared as such” [7].

Great Britain, the United States and the Regulation’s reach

The developer is in Poland and the store reaches every member state, so the MDR governs the European market whether or not any other regime applies. In Great Britain the Medical Devices Regulations 2002 still implement the 1993 directive, under which standalone software is an active device and, unless another rule captures it, falls into Class I under Rule 12; the Medicines and Healthcare products Regulatory Agency (MHRA) guidance on standalone software and apps sets out that position; Class I still requires conformity with the essential requirements, a declaration, registration with the MHRA and evidence for any claim made [8]. The draft Medical Devices (Amendment) Regulations 2026, published on 11 May 2026 with a call for evidence that closed on 19 June, move classification onto an IMDRF-aligned, risk-proportionate footing within the active device rules [9]. In the United States the clinical decision support exemption is confined to software that supports a healthcare professional who can independently review the

basis of its recommendation; software that gives a patient a treatment recommendation remains a device function under the Food and Drug Administration's 2022 guidance [10].

What the classification would have required

The classification matters for the obligations it carries, and the companion article shows what each would have surfaced. A statement of intended purpose would have had to say what the product does when the user chooses a model the developer has not tested, since it recommends none. Risk management would have had to treat the model as a component outside the manufacturer's control and address the hazards observed: 40 distinct pairs of safety limits on one record, judged against an insulin total that omits basal delivery, a sensitivity factor of 180 mg/dl per unit, and a meal step marked completed on a record with no meals. Software lifecycle control would have required a pinned and validated model with a change control plan for when the provider changes it. A clinical evaluation would have had to support the claim of "safer" tuning, since Article 7 forbids claims that the evidence does not bear out [4, 5], and the only evidence available is that the advice depends on which model the user opened.

This is the ordinary content of a conformity assessment for Class IIa or IIb software, and what a bolus calculator built into a pump goes through before it reaches a patient. The decision is as serious in both cases; the calculator's output is determined by its inputs and this application's is not.

The user's position

The Diabetes Technology Network UK has advised that suggested changes to insulin doses or pump settings from a general-purpose LLM must be reviewed carefully and sense checked, and warned against relying on such models for insulin adjustments [11]. That advice was written for a person talking to a chatbot; a product that presents the chatbot's output in the dress of a settings review makes it harder to follow, because the form no longer signals the source. The companion article found that the user is shown one answer from a distribution whose width they are not told, and that the check which appears to stand behind the answer inspects only its shape. A person who follows the Network's advice and sense checks the proposal has nothing to check it against except their own judgement, which is what they were paying to supplement.

Data protection adds a further obligation that sits outside the device regulations. The package is health data about an identifiable person, and the workflow has the user paste it into a consumer chat service whose terms may permit the conversation to be used to improve the provider's models unless the user opts out; the user, not the developer, makes that transfer, and the application's onboarding says nothing about it.

Conclusions

On the Regulation's own definition and the Coordination Group's own guidance, an application whose stated purpose is to help a person with type 1 diabetes decide how to change the settings that control their insulin delivery is medical device software, and the consequences of a wrong change place it in Class IIb, or at the least Class IIa. The choice of an external model as the

engine, the interposition of the user as decision-maker and the presence of a disclaimer do not alter that, because each is addressed in the guidance. Classification is for the manufacturer and then for the competent authority and notified body; the reasoning is set out here so that it can be tested, and applied to the other listings of the same kind. The obligations that follow are the minimum a product of this kind should meet before it is sold, and the companion article's findings are what a manufacturer who met them would have discovered first.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Acknowledgments

Claude (Anthropic, San Francisco, California, USA), a large language model assistant, was used under the author's direction to write the harness and analysis code, to run the analyses, to prepare the figures and to draft and revise the text. The author specified the study, made every decision of design and interpretation, checked each reported number against the stored outputs, and is responsible for the content.

Disclosures

The author runs Diabettech Ltd, is a committee member of the Diabetes Technology Network UK and wrote its statement on large language models. He has no relationship with the developer of the application discussed and paid for its subscription. Three of the eleven models studied are Anthropic's, as is the assistant used to produce this work; Anthropic had no role in the study, was not informed of it, and all model usage was paid for at list price. He is not a lawyer, and this commentary is not legal advice.

References

1. Street T. A paid application standardises what large language models read about insulin pump settings and leaves what they decide to chance: reproducibility of an automated insulin delivery settings review across eleven models. SSRN preprint. 2026.
2. Kowalczyk A. TuneMyBG. Google Play. Version 1.0.5, updated 20 August 2026. https://play.google.com/store/apps/details?id=com.adamkowalczyk.tunemybg. Accessed 22 August 2026.
3. Huckvale K, Adomaviciute S, Prieto JT, Leow MK, Car J. Smartphone apps for calculating insulin dose: a systematic assessment. BMC Med. 2015;13:106. https://doi.org/10.1186/s12916-015-0314-7
4. Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices. Official Journal of the European Union. 2017;L117:1-175. Article 2(1), Article 2(12), Article 7, Annex VIII Rule 11.
5. Medical Device Coordination Group. MDCG 2019-11: Guidance on qualification and classification of software in Regulation (EU) 2017/745 and Regulation (EU) 2017/746.

European Commission. October 2019.

6. International Medical Device Regulators Forum. Software as a medical device: possible framework for risk categorization and corresponding considerations. IMDRF/SaMD WG/N12FINAL:2014. 18 September 2014.
7. Google. Health content and services. Google Play Developer Policy Center. https://support.google.com/googleplay/android-developer/answer/16679511. Accessed 22 August 2026.
8. Medicines and Healthcare products Regulatory Agency. Medical device stand-alone software including apps (including IVDMDs). Guidance. London: MHRA; updated 2023. https://www.gov.uk/government/publications/medical-devices-software-applications-apps. Accessed 22 August 2026.
9. Medicines and Healthcare products Regulatory Agency. Draft Medical Devices (Amendment) Regulations 2026 and call for evidence. 11 May 2026. https://www.gov.uk/government/organisations/medicines-and-healthcare-products-regulatory-agency. Accessed 22 August 2026.
10. US Food and Drug Administration. Clinical decision support software: guidance for industry and Food and Drug Administration staff. 28 September 2022. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/clinical-decision-support-software. Accessed 22 August 2026.
11. Diabetes Technology Network UK. DTN-UK statement on large language models. Association of British Clinical Diabetologists. January 2026. https://abcd.care/dtn/resource/current/dtn-uk-statement-large-language-models. Accessed 22 August 2026.