

## Если ваш ИИ не может объяснить свои решения, сможет ли FDA его одобрить?

**Источник:** MedTech Intelligence

**Дата публикации:** 2025

**Оригинал:** [https://medtechintelligence.com/feature\\_article/if-your-ai-cant-explain-itself-can-fda-authorize-it/](https://medtechintelligence.com/feature_article/if-your-ai-cant-explain-itself-can-fda-authorize-it/)

FDA

SaMD

безопасность

клинические испытания

объяснимый ИИ

регулирование

Рассмотрим гипотетический сценарий, который все чаще описывают производители: подача заявки по процедуре 510(k) на диагностический инструмент на базе ИИ с чувствительностью 99%. Впечатляющие цифры. Солидная валидация. И тем не менее — от FDA (Управление по санитарному надзору за качеством пищевых продуктов и медикаментов США) приходит письмо о несоответствии (deficiency letter).

Причина не в том, что модель работает плохо. А в том, что рецензенты не могут определить, что именно она делает и почему.

Это реальность, с которой сталкиваются производители, когда программное обеспечение как медицинское изделие (**SaMD** — Software as a Medical Device) на основе ИИ/МО (искусственного интеллекта и машинного обучения) выходит на пути регистрации 510(k) и De Novo. Высокая производительность необходима, но сама по себе она недостаточна. FDA уделяет все больше внимания другому вопросу: не просто «Насколько точен ИИ?», а скорее:

- Можете ли вы описать, что делает этот ИИ, и способен ли клиницист полагаться на его результаты?

FDA не просит производителей упрощать сложные алгоритмы. Оно просит производителей сделать их целевое назначение, логику, ограничения, производительность и клиническую роль достаточно понятными, чтобы обеспечить безопасное использование.

## **Производительности больше недостаточно для закрытия вопроса:**

Ранее основные усилия производителей медицинских изделий по разработке технологий ИИ были сосредоточены на создании выдающихся моделей машинного обучения, валидированных с использованием обширных наборов данных с подробными показателями эффективности. Хотя эти подходы по-прежнему важны, они не учитывают в достаточной мере возросшее внимание регуляторов к тому, как устройства используют алгоритмические методы и какие доказательства используются для подтверждения предположения о том, что клиницисты могут доверять этим медицинским изделиям.

Таким образом, клиницист является конечным звеном в процессе принятия решений, когда система искусственного интеллекта (ИИ) ставит диагноз или корректирует лечение. Функция FDA заключается в том, чтобы подтвердить точность модели, но также и в том, чтобы гарантировать, что у клиницистов есть возможность понимать, подвергать сомнению и вмешиваться в отношении полученных результатов, когда это необходимо.

Результат по принципу «черного ящика», даже обладающий очень высокой точностью, не дает клиницистам такой возможности. Письма о несоответствии, связанные с 510(k) для ИИ/МО, все чаще указывают на одни и те же типы недостатков: отсутствие объяснения того, как модель выводит свои результаты, недостаточный анализ производительности в подгруппах, отсутствие четкого понимания того, как устройство работает на границах — например, на представителях меньшинств, при вводе данных, выходящих за пределы распределения (out of distribution), или в условиях реальной клинической практики.

## Что на самом деле требуют руководства FDA?

У FDA нет отдельного документа под названием «Руководство по объяснимости» (Explainability Guidance). Вместо этого существует несколько ключевых документов, требующих прозрачности, и совокупность всех этих документов устанавливает четкие ожидания для систем ИИ и машинного обучения.

Принципы **GMLP** (Good Machine Learning Practice — Надлежащая практика машинного обучения) от 2021 года устанавливают требование прозрачности как основополагающий элемент; системы ИИ и МО должны четко доносить до конечного пользователя цель, допущения и ограничения таких систем. Это требование не является факультативным.

Рекомендации по подаче маркетинговых заявок на 2025 год для **PCCP** (Predetermined Change Control Plan — План predetermined контроле изменений) для устройств с поддержкой ИИ идут еще дальше: чтобы получить возможность модифицировать модель ИИ после того, как она была одобрена FDA, без необходимости проходить процесс повторной подачи, производители должны предоставить документацию о том, как работает алгоритм и каковы пределы его изменений, чтобы FDA могла оценить любые будущие обновления модели. Такой уровень документации будет невозможен без прозрачного, хорошо охарактеризованного алгоритма.

**IMDRF N41** (Clinical Evaluation of Software-as-a-Medical Device (SaMD) — Клиническая оценка программного обеспечения как медицинского изделия) требует, чтобы разработчики продемонстрировали понимание клинической связи между выходными данными ИИ (т. е. результатом вычислений ИИ) и состоянием, которое он предназначен лечить. Чтобы продемонстрировать это с помощью клинических доказательств, разработчику потребуется модель, которую он сможет объяснить.

Если рассматривать в совокупности, принципы GMLP, требования PCCP и структуры клинической оценки создают эффективный стандарт объяснимости — даже без отдельного руководства. Производители, которые упускают это из виду, создают пробел в комплаенсе (соответствии нормативным требованиям).

## Три вещи, которые FDA хочет увидеть:

Когда рецензент проверяет заявку ИИ/МО на прозрачность, обычно возникают три общие области:

1) **Прослеживаемость (Traceability):** Есть ли у вас возможность проследить выходные данные модели до данных, признаков (features) и проектных решений, которые к ним привели? FDA требует документированной цепочки владения, включая состав обучающего набора данных, любую проведенную предобработку и ваши причины использования этих предобработанных данных, решения, связанные с их признаками, историю версий вашего ИИ и известные режимы отказов и т. д. Если вы представляете показатель производительности вашего ИИ без этого контекста, это всего лишь цифра.

2) **Подотчетность (Accountability):** Когда ИИ делает что-то неправильно, кто несет ответственность? На практике это означает определение вашей модели человеческого надзора, чтобы клиницисты знали, как работать с выходными данными ИИ, каков путь, если они решат отменить результат ИИ, и как контролировать производительность ИИ в промежутке между готовностью к выходу на рынок и нахождением на рынке.

Опубликованная FDA структура GMLP требует, чтобы в междисциплинарную команду, разрабатывающую ИИ, входили люди, знающие, как практиковать медицину, непосредственно в процессе создания ИИ, а не только для того, чтобы эти люди проверяли ИИ в самом конце разработки.

3) **Понятность (Comprehensibility):** Поймут ли ваши целевые конечные пользователи (радиолог, медсестра или терапевт), что говорит им ИИ, и смогут ли они предпринять действия на основе этой информации? Это не означает раскрытие внутренних механизмов работы ИИ. Это означает, что выходные данные ИИ должны быть клинически интерпретируемыми, предоставлять клиницистам уровни уверенности, связанные с ИИ, и информировать клиницистов, использующих ИИ, о его ограничениях.

## **Где происходят сбои при подаче заявок?**

Ошибки в заявках на ИИ/МО, по всей видимости, следуют одной и той же общей схеме. Известная техническая команда создает солидную модель с доказательствами предварительной валидации или опыта разработки подобных систем.

Заявка включает подробные описания архитектуры и метрик модели, но содержит лишь несколько строк об объяснимости (одно утверждение об использовании значений **SHAP** и заявление об обучении клиницистов).

Команда заявителя не рассматривала объяснимость как обязательное условие, и в конечном итоге она рассматривалась как второстепенный вопрос. Однако после получения письма о «несоответствии» с запросом именно того, что считалось второстепенным (т. е. как генерируются выходные данные, как клиницисты должны интерпретировать результаты, и насколько хорошо сохраняется производительность модели в различных группах пациентов?), команда заявителя обычно вступает в процесс исправления, который длится от 6 месяцев до 1 года.

Этого процесса исправления можно было бы избежать, если бы команды закладывали прозрачность в свою заявку изначально. Другими словами, использовали прозрачность при создании модели и подаче заявки, а не ждали конца, чтобы добавить её.

## **Как выглядит прозрачная заявка?**

В целом, успешные заявки имеют несколько характеристик, которые способствуют успешному рассмотрению:

### **Документ с описанием алгоритма:**

- Описывает функцию алгоритма (модели), то, как он обучается, валидируется, и ограничения алгоритма простым клиническим языком, а не просто ссылкой на опубликованную статью.
- Документированные обучающие и валидационные наборы данных, которые включают не только количество образцов (количество пациентов): демографический состав выборки, то, как собирались данные и как контролировалось их качество/точность, а также то, как решались проблемы известных предвзятостей (biases).
- Анализ производительности в подгруппах, т. е. разделение показателей на основе: возраста, пола, расы/этнической принадлежности, тяжести заболевания и других соответствующих клинических характеристик, а не отчет только об агрегированной статистике производительности.

План постмаркетингового наблюдения с четко определенными порогами производительности, а также четкими триггерами для внедрения изменений или подачи новой заявки. Документация по человеческим факторам (**human factors**), демонстрирующая способ отображения выходных данных ИИ, то, как сообщается о неопределенности результата, и то, как тестировалась целевая группа пользователей.

При подаче заявок по процедуре **De Novo**, где FDA также создает регуляторную базу для будущих устройств того же класса, стандарты еще более жесткие. Рецензенты должны обладать достаточным пониманием алгоритма, чтобы создать соответствующие специальные меры контроля (**special controls**) для всего класса устройств. Это невозможно достичь, если со стороны заявителя нет прозрачности.

Еще один момент, заслуживающий внимания: с начала 2026 года FDA полностью перешла на новое Положение о системе управления качеством (**QMSR** — Quality Management System Regulation), которое приводит 21 CFR Part 820 в соответствие с ISO 13485. Это имеет практические последствия для заявок на ИИ/МО.

Документация о происхождении данных, контроль версий алгоритма и записи об объяснимости теперь должны вписываться в **QMS** (систему управления качеством), а не существовать только в виде технических файлов. Производителям, создающим инфраструктуру прозрачности, следует подтвердить, что она интегрируется непосредственно с их QMS с самого начала.

## **Заключение:**

Нынешняя эра использования ИИ по принципу «черного ящика», подкрепленного надежными метриками, подошла к концу. Руководства FDA (содержащиеся в нескольких документах и единообразные по своему замыслу) четко указывают на то, что алгоритмическая прозрачность является предварительным условием для одобрения (а не бонусом).

На самом деле, то, о чем на самом деле просит FDA, — это не то, насколько точен ваш ИИ; скорее, это то, насколько ваш ИИ подотчетен. Не существует одобрения точного ИИ без отчета о его работе. Первый шаг к установлению уровня подотчетности — это способность общаться на конкретных клинических терминах, объясняя именно то, как работает алгоритм и почему.

Спросите себя перед подачей:

- Можете ли вы объяснить, что делает ваша модель клиницисту, который никогда не видел нейронную сеть?
- Можете ли вы показать FDA именно те входные данные, которые привели к каждому выходному результату?
- Есть ли у вас план на случай, если модель ошибется?

Если нет — объяснимость — это то, с чего должна начинаться подготовка вашей заявки, а не то, чем она заканчивается.

---

---

Перевод выполнен: 10.06.2026 | ai4med.ru

Машинный перевод. Рекомендуем сверять с оригиналом при клиническом использовании.