

Guardoc Health обрабатывает клиническую документацию с использованием моделей Amazon Nova

Источник: AI News Healthcare

Оригинал: <https://www.artificialintelligence-news.com/news/guardoc-health-processes-clinical-documentation-using-amazon-nova-models/>

NLP

RAG

автоматизация

клиническая документация

управление данными

Компания Guardoc Health сообщает, что ежедневно обрабатывает более одного миллиона клинических документов, используя модели Amazon Nova через платформу Bedrock.

Внедрение искусственного интеллекта (ИИ) в процесс ведения клинической документации сводится к специфическому расчету рисков. Ошибка в этом расчете может привести к каскаду последствий: отказам в выплатах по программам Medicare в рамках модели оплаты на основе состояния пациента (**Patient-Driven Payment Model, PDPM**), штрафам при аудите, судебным искам и, в худших случаях, к пропуску патологии, что меняет тактику лечения пациента.

Однако при правильном подходе результат проявляется в уменьшении количества исправлений, снижении частоты переводов пациентов в стационары и сокращении расходов на обеспечение соответствия нормативным требованиям (комплаенс). Компания Guardoc Health, разрабатывающая программное обеспечение для ведения документации в учреждениях длительного ухода, опубликовала показатели развертывания системы, которые, по её утверждению, подтверждают такие результаты.

Масштаб underlying проблемы

Конвейер обработки данных Guardoc Health должен справляться с документами практически всех форматов, которые могут возникнуть в клинических условиях: многостраничные **PDF**-файлы с рукописными аннотациями врачей, наложенными на печатный текст; формы предварительной авторизации, где решение о покрытии расходов зависит от одного отмеченного пункта; списки лекарственных препаратов, которые в одной карте представлены в виде четких таблиц, а в другой — в виде свободного текста; а также формы приема пациентов, где на одной странице смешаны напечатанные поля, резиновые штампы и рукописный ввод.

Согласно исследованию, опубликованному в журнале **BMJ Quality and Safety**, число американских амбулаторных пациентов, пострадавших от диагностических ошибок, составляет около 12 миллионов в год, при этом фактором, способствующим этому, называются сбои при обработке информации. При объемах, обрабатываемых Guardoc, ошибка всего в один процент при выявлении состояний привела бы к созданию тысяч неверных записей ежедневно. Каждая такая ошибка несет в себе риски для безопасности пациента или нарушения нормативных требований.

Guardoc сообщает о снижении количества ошибок в документации на 46%, сокращении штрафов при аудите на 70% и получении более 400 000 долларов США ежегодной окупаемости инвестиций (**ROI**) для одного учреждения, не публикуя при этом базовый период или методологию, лежащую в основе этих расчетов.

В рамках квартального развертывания, охватившего два учреждения и 200 пациентов, компания заявила, что её система выявила 847 исправлений в документации, отметила 86 проблем, связанных с точностью возмещения по системе **PDPM**, и способствовала 74-процентному снижению частоты переводов в больницы на 100 госпитализаций. Согласно данным Guardoc, отдельное тематическое исследование, охватившее семь учреждений и 1618 резидентов, выявило 10 612 проблем.

Конвейер извлечения данных, ориентированный на стоимость и точность

Архитектура Guardoc осуществляет классификацию состояний с помощью технологии генерации с дополненной выборкой (**Retrieval Augmented Generation, RAG**), извлекая доказательства из собственной документации пациента перед проведением логического анализа для формирования окончательного ответа.

Сначала **Amazon Textract** извлекает текст и структурные метаданные из каждой входящей страницы — на этапе, который компания считает точкой минимальной стоимости обработки одной страницы в конвейере. Полученный результат разбивается на фрагменты (**chunks**) по клиническим границам, чтобы список лекарств или раздел с диагнозом оставались цельными, а не разделялись произвольным количеством символов.

Каждый фрагмент преобразуется в векторное представление с помощью **Amazon Titan Text Embeddings V2** и сохраняется в **Amazon DynamoDB**, сегментируясь по пациентам, чтобы процесс поиска никогда не выходил за пределы данных одного пациента. Специальный предварительный фильтр сужает набор кандидатов по типу документа и актуальности, после чего поиск методом k-ближайших соседей (**k-nearest neighbour search**) извлекает фрагменты, наиболее релевантные конкретному запросу на классификацию; на этом этапе возвращаются только ссылки на страницы, чтобы минимизировать объем передачи данных.

Затем **Amazon Nova 2 Lite** выполняет текстовый проход для удаления очевидных несоответствий. Только те страницы, которые прошли все предыдущие фильтры, попадают в **Amazon Nova Pro**, которая получает необработанные байты **PDF** и анализирует макет, рукописный ввод, подписи и штампы для формирования классификации, на основе которой будут действовать последующие системы.

Дизайн системы на всех этапах следует логике многоуровневой стоимости: дешевые компоненты выполняют высокообъемные задачи, такие как создание эмбеддингов и грубая фильтрация, а более ресурсоемкие мультимодальные рассуждения резервируются для финальной стадии, где они действительно необходимы.

Сложные случаи в клинической документации

По данным Guardoc, два типа документов составляют основную часть того, что пропускали предыдущие версии конвейера. Первый — это поля подтверждения врача в формах предварительной авторизации, где рукописная заметка может отменять значение напечатанного чекбокса. Второй — разделы с симптомами, сообщаемыми пациентами, где рукописный текст часто содержит информацию, которая больше нигде в записи не встречается.

Извлечение данных о лекарственных препаратах представляет собой смежную проблему. Названия препаратов, дозировки, способы введения и частота приема могут встречаться в структурированных таблицах, в прозе, скрытой внутри заметок врача, в рукописных дополнениях к печатным спискам и в сканах, прошедших через множество рук по факсу. Гибридный конвейер Guardoc сначала использует **Amazon Textract** для чистых печатных таблиц, а затем передает как оригинальный **PDF**, так и результат работы **Textract** в **Amazon Nova Pro** для разрешения проблем с переносом столбцов таблиц, рукописными дополнениями и нестандартными форматами, которые невозможно правильно распознать только с помощью **OCR** (оптического распознавания символов).

«С семейством моделей Nova мы облегчаем медицинским организациям задачу раннего выявления высокорисковых случаев и принятия мер до того, как проблемы станут дорогостоящими», — сказал Ассаф Амиаз, директор по продукту Guardoc Health. «Автоматизируя рабочие процессы, которые раньше требовали ручного контроля, семейство Nova помогает командам сокращать пробелы в комплаенсе, предотвращать ошибки и уделять больше времени улучшению результатов лечения пациентов».