

Отчёт OpenAI связывает агентирование кода с ускорением сборки научного ПО

Источник: AI News Healthcare

Оригинал: <https://www.artificialintelligence-news.com/news/openai-report-coding-agents-faster-science-software-builds/>

автоматизация

биоинформатика

геномика

иммунология

научные вычисления

разработка ПО

OpenAI опубликовала новый отчёт, отслеживающий восемь проектов в области научных вычислений, где агенты для написания кода сократили время выполнения.

В отчёте задокументированы проекты, в которых в пяти случаях использовался только Codex, а в трёх других — комбинация Codex и Claude Code от Anthropic. Стоит сразу отметить: это поставщик публикует обзор применения собственного продукта в исследовательских условиях, составленный на основе кейсов, написанных участвовавшими `_contributors`.

Это не делает лежащую в основе закономерность менее заслуживающей изучения. Исследовательское программное обеспечение имеет документированную проблему поддержки. Инструменты, созданные для сопровождения отдельной научной статьи, написанные небольшими академическими командами без выделенной инженерной поддержки, склонны накапливать технический долг, который никто не имеет бюджета или полномочий погасить.

В отчёте OpenAI утверждается, что агенты могут решить эту проблему с долгом, и восемь проектов, которые в нём цитируются, охватывают геномику, иммунологию, статистику и секвенирование РНК.

Какие задачи выполняли агенты

Задачи примерно разделились на три категории: упаковка и очистка системы сборки, оптимизация производительности существующего кода и полный перенос языка или бэкенда.

cyvcf2, библиотека Python для чтения файлов геномных вариантов, заменила свою устаревшую систему сборки и упаковки на новый унифицированный процесс, по словам участника Brent Pedersen, который отметил, что быстро работать с агентами — это одно, но идти далеко в науке всё ещё требует «экспертного руководства, понимания, вкуса и заботы».

HI.SIM, симулятор считываний при секвенировании ДНК, прошёл два в значительной степени автономных цикла оптимизации от GPT-5.2 и GPT-5.6, которые, по словам участника Andrew Ho, сократили время выполнения на 31 процент на репрезентативном тестовом наборе без изменения выходных данных.

Ho, который описывает себя как не специалиста по геномике и не программиста на C, назвал результат «не чем иным, как волшебством» с точки зрения конечного пользователя, ранее терявшего время из-за ошибок производительности и проблем с упаковкой, которые он мог распознать, но не мог исправить лично.

Hifiasm, используемый для сборки генома из показаний PacBio HiFi, получил сокращение времени выполнения на 25 процентов на целевой оптимизации и примерно на 15 процентов на отдельных данных секвенирования человека, по словам участника Suyash Shringarpure.

Shringarpure описал, как агент настроил собственную каркасную систему бенчмаркинга и самостоятельно предложил кандидатов, хотя он подчеркнул, что предоставление результатов профилирования и направление модели в сторону от повторяющихся режимов отказа оставались работой, которую мог выполнить только человек.

MHCflurry, который предсказывает белковые фрагменты, представленные T-клеткам, перенёс свой бэкенд TensorFlow/Keras на PyTorch, сохранив совместимость с ранее выпущенными весами модели — изменение, которое участники Alex Rubinsteyn, Sergey Feldman и Timothy O'Donnell характеризуют как тот вид «негламурной, трудоёмкой поддержки», который поддерживает жизнь открытых научных проектов, а не оставляет их угасать.

bayesm-rs, порт на Rust статистических моделей из пакета bayesm для R, соответствовал оценкам оригинального программного обеспечения в пределах заранее установленной допустимой погрешности и работал в 2,3–2,7 раза быстрее на одном потоке процессора, достигая 4,4–9,5 раз быстрее на восьми потоках. По словам участников Andrew Bai и Andrew Ho, агенты быстро и правильно обрабатывали всё, что имело прямую ссылку для проверки; расширения, требующие статистического суждения, которое оригинальный код никогда не определял, требовали вместо этого прямой человеческой валидации.

Порты на Rust и редизайн GPU продвигают закономерность дальше

Три дополнительных проекта — rustar-aligner, svb и kuva — включали сборки на Rust, выполненные с помощью агентов для написания кода, включая полную реконструкцию STAR, широко используемого инструмента выравнивания последовательностей РНК, который потерял активную поддержку.

Участник James M. Ferguson говорит, что агенты меняют то, что стоит попытаться сделать: переписывание выравнивателя на 20 000 строк вручную не является разумным использованием времени, но с агентом это становится неделями направляемой работы. Верификация, добавил он, — это совершенно отдельный вопрос. Модель может утверждать, что график выглядит нормально, но проверка более 900 из них глазами перед выпуском всё ещё ложилась на человека.

RustQC объединил 15 отдельных инструментов контроля качества секвенирования РНК в единую программу, которая, по словам участника Phil Ewels, сократила время выполнения в 60 раз, а ввод-вывод диска — в 25 раз, при этом сопутствующие пересборки FastQC-Rust и Trim Galore работали в семь и три раза быстрее соответственно, сохраняя поведение оригинальных инструментов.

Ewels также отметил обратную сторону: дешёвые пересборки несут собственный риск, потому что инструменты, расходящиеся в поведении, фрагментируют сообщество и делают результаты из разных лабораторий несравнимыми со временем. «Технология — это лёгкая часть», — сказал он. «Ответственное управление — это открытый вопрос».

HelixForge, нативная для GPU пересборка инструмента симуляции мутаций BAMSurgeon, как сообщается, сократила время выполнения примерно в 60 раз на бенчмарке с участием реальных человеческих данных, по словам участников Mamad Ahangari, Varun Goyal и Hassan Masoudi, которые также говорят, что она произвела частоты мутаций, более близкие к запрошенным целевым значениям, и исправила несколько ошибок, которые генерировали артефакты в оригинальном инструменте.

Верификация, а не генерация кода, теперь является ограничением

То, что проявляется во всех отчётах, — это то, что агенты компетентно обрабатывали хорошо определённые запросы на реализацию, но не могли судить, были ли их собственные выводы научно обоснованными.

Участники описывают, как агенты выражали уверенность в работе, которая содержала явные ошибки, что перекладывало фактическое бремя на людей для создания тестов принятия: точное соответствие выходных данных, проверки паритета с существующим инструментом или ответы, установленные заранее с использованием симулированных данных.

Проекты, как правило, продвигались поэтапно, при этом агенты создавали быстрые первые черновики, а оставшееся время уходило на граничные случаи и небольшие числовые расхождения, которые один только бенчмарк не обнаружил бы.

Снижение инженерных затрат имеет две стороны. Это позволяет команде из двух человек взяться за пересборку, которая когда-то требовала бы нанятого за счёт гранта инженера, и упрощает для трёх разных лабораторий создание трёх несовместимых версий одного и того же инструмента. Изменения в MHCflurry и suvscf2 вернулись в их оригинальные вышестоящие проекты. rustar-aligner перешёл к новому общественному управлению, потому что инструмент, который он заменил, уже был заброшен.

Отчёт OpenAI указывает на конкретный выбор, а не на общее одобрение: решите, кто владеет пересобраным инструментом, и обеспечьте эту приверженность до того, как будет выпущена первая строка сгенерированного агентом кода.

Перевод выполнен: 03.08.2026 | ai4med.ru

Машинный перевод. Рекомендуем сверять с оригиналом при клиническом использовании.