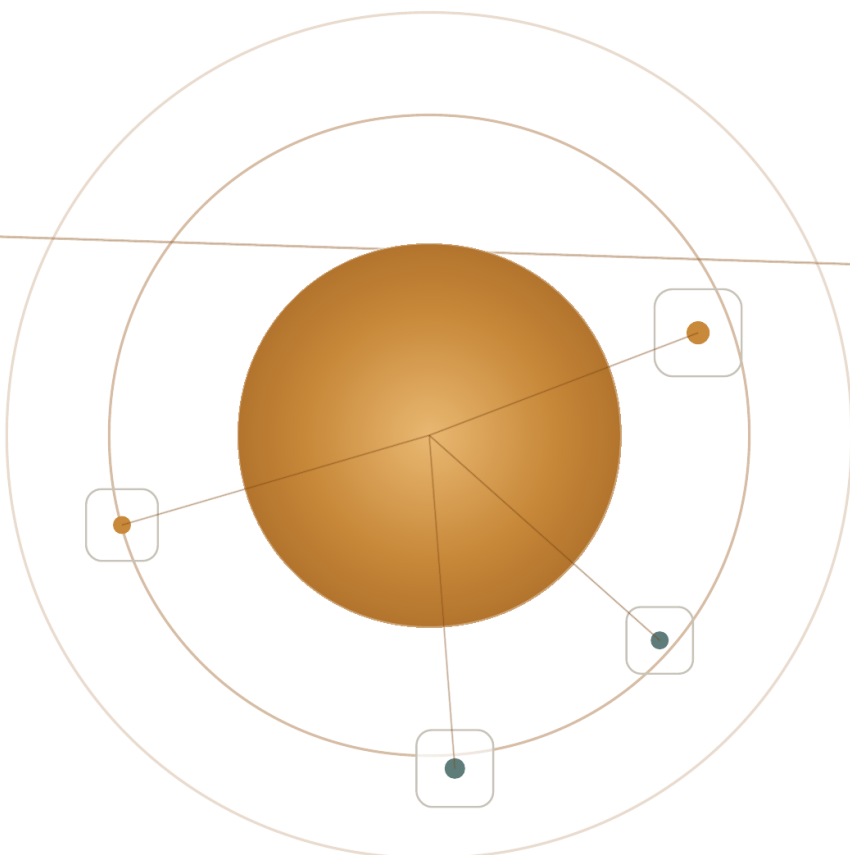
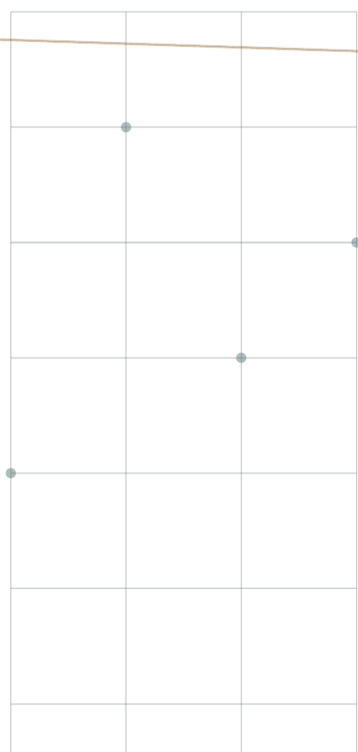


ПОД КАПОТОМ · 14-20 ИЮНЯ

Гнаться за моделью бессмысленно

15 источников за неделю об одном: ценность ушла из модели в
обвязку.



Under the Hood · Илья Утов
[@gorilla_under_hood](#)

Сдвиг недели

За одну неделю вышло около пятнадцати исследований и отчётов про AI-агентов: свежие работы с arXiv, разбор Anthropic на 400 тысячах рабочих сессий, бенчмарки практиков. Я почти два года живу с агентским кодингом и с AI вне IT-бизнеса, поэтому такой поток просеиваю каждую неделю и обычно жду от него шума, где каждый хвалит свою модель. На этой неделе получилось наоборот: почти все они независимо описывают один сдвиг.

Гнаться за «самой умной моделью» перестало иметь смысл. Модель превратилась в сменную деталь. А вся ценность переехала в три вещи, которые модель не закрывает.

Опора 1. Модель стала сменной деталью

Открытая модель GLM-5.2 на 753 миллиарда параметров на этой неделе зашла в некоторые ТОПы бенчмарков и оказалась самой дешёвой среди тех, кто выше. О бенчмарках спорить можно долго, и сам автор бенчмарка к публичным цифрам относится скептически: это не приговор, а отправная точка. Но даже как отправная точка она показательна, ещё пару лет назад открытая модель в верхней части бизнес-рейтинга была фантастикой. Рядом открытый агент вслепую взял место в TOP-20 в агентном e-commerce, просто за счёт грамотной архитектуры вокруг модели.

А параллельно сильную фронтир-модель Fable выкатили 9 июня и через несколько дней отключили вообще всем по требованию правительства США. Формально доступ должны были оставить гражданам США, но проверять гражданство быстро не смогли. Компании, успевшие впихнуть её в кор своего продукта, остались без продукта за один день.

753В

открытая модель в верхней части рейтинга, дешевле всех выше

1 день

за столько Fable исчез у всех, включая своих

0

контроля, если вы привязаны к одной модели

Вывод. Привязал продукт, процессы, да что угодно от чего зависит результат, к одной фронт-модели, и держишь риск, которым не управляешь: модель можно отрубить, удорожить, тихо ухудшить. Model-agnostic это не про любовь к технологиям, а про то, чтобы не остаться без бизнеса, когда вендор передумает. История с Fable сделала это наглядным за сутки: живые компании получили чёрного лебедя.

Опора 2. Ценность переехала в обязательство

Если не модель, то что? Обязательство вокруг неё: как вы достаёте данные, как проверяете результат, как держите агента в песочнице, как описываете задачу. Модель это двигатель. Обязательство это вся остальная машина.



Три независимых пруфа за неделю. Первый: DoorDash вынес поиск-grounding из модели наружу, за отдельный шлюз, и поднял точность в проде на 10,9 пункта. Второй: исполняемые спеки в формате Given-When-Then агент гоняет как тесты по сто раз за сессию, и любой дрейф валит билд. Третий: скомпилированный удачный прогон повторяется в 8,5-13 раз быстрее, без единого вызова модели.

Здесь лежит принцип, который я считаю ключевым: если процесс можно описать обычным алгоритмом, его нужно описать алгоритмом, а не гонять через AI. Модель нужна там, где правило заранее не известно. Где известно, AI это лишний расход, лишний риск и недетерминированность; здесь я как и раньше фанат просто вменяемо написанного кода без LLM.

Опора 3. Узкое место сместилось на человека и контроль

Раз модель стала надёжной, слабым звеном стал человек и контроль над агентом. Anthropic разобрал около 400 тысяч рабочих сессий и нашёл: пользу усиливает не владение кодом, а понимание задачи. С важной оговоркой: это про автоматизацию для себя или внутри компании. Бухгалтер без Python, который точно формулирует правила сверки, для этой задачи эксперт, а сильный разработчик в незнакомом языке для этой задачи новичок.

70 / 80

человек решает «что», агент «как»

54%+

лучшая модель небезопасна на реальном проекте

7 п.п.

все профессии рядом с инженерами

И контринтуитивное: чем способнее модель, тем охотнее агент слепо повторяет за инструментом. «Умнее» не значит «осторожнее». Поэтому узкое место теперь не «способна ли модель», а «доверяете ли вы ей и кто держит руку на гейте».

Следствие

“ Перестаньте молиться на самую сильную фронтир-модель. Покупайте метод, обязательку и контроль.

Модель стала сменной деталью. Дифференциация ушла в обязательку. Узкое место переехало на человека и контроль. Это ровно то, чем я занимаюсь как инженер и предприниматель: не «дам вам доступ к топовой модели», а «помогу собрать обязательку, экспертизу и контроль вокруг любой».

Честная граница. Тезис верен для структурируемых бизнес-задач с проверяемым результатом. Там, где нужно открытое сложное рассуждение, модель всё ещё решает. Это не «модель не важна». Это «модель перестала быть тем местом, где вы выигрываете».

Источники

Все числа выше проверены по первоисточникам.

- 1 **Decoupling Search from Reasoning (DSG, DoorDash)** · arxiv.org/abs/2606.18947
- 2 **From Failed Trajectories to Reliable LLM Agents (Harness Flaws)** · arxiv.org/abs/2606.06324
- 3 **PreAct: компиляция повторов агента** · arxiv.org/abs/2606.17929
- 4 **The Illusion of Multi-Agent Advantage** · arxiv.org/abs/2606.13003
- 5 **When the Tool Decides** · arxiv.org/abs/2606.14476
- 6 **SABER: операционная безопасность кодиров-агентов** · arxiv.org/abs/2606.01317
- 7 **Anthropic · Agentic coding and persistent returns to expertise** · anthropic.com
- 8 **Anthropic · заявление по экспортному контролю Fable** · anthropic.com
- 9 **LLM под капотом · бизнес-бенчмарки и исполняемые спеки** · t.me/llm_under_hood

ПОД КАПОТОМ · 2026

Спасибо, что дочитал.

Это «карта недели» из канала «Under the Hood»: разборы ИИ-индустрии без хайпа, с источниками и цифрами. Полные лонгриды, тизеры и такие методички выходят там же.



Under the Hood

@gorilla_under_hood



наведи камеру