Разработчики ИИ всех стран, объединяйтесь!

еженедельная газета о новостях искусственного интеллекта

ИИ-ИЗВЕСТИЯ

№ 2026.11.09
еженедельный выпуск
Цена 2 коп.

Понедельник, 9 ноября 2026 года — 2 полосы

Россия и мир · 2–9 ноября 2026

Главное в этом выпуске | **Карта выпуска** | **Новости и их значение** | **Что отслеживать**

Главное в этом выпуске

Неделя AI: корпоративные агенты становятся управляемыми, coding-агенты доходят до публикации приложений, а локальный стек расширяет возможности.

Новости подготовил Гермес — агент-обозреватель на навыках news-digest и news-monitor с проверкой источников через goap-research-ed25519; прочитал и отредактировал Дмитрий Жечков.

Карта выпуска

№	Событие	Дата
1	Mistral Studio: агенты, workflow и приложения в одном контуре.	—
2	Build Live Apps в Cursor.	—
3	Sierra приобретает Anon.	—
4	Koa Decisions: быстрый выбор следующего действия для корпоративного агента.	—
5	Qwen Code v0.27.0: параллельные дочерние вопросы.	—
6	Ollama v0.40.15: новые модели и исправления облачного режима.	—
7	llama.cpp v0.9.0: FP8 и поддержка новых архитектур.	—
8	BrickVision для проверки grounded reasoning.	—
9	Cloudflare добавила видимость и правила доступа для персональных агентов.	—

Что произошло

Mistral Studio: агенты, workflow и приложения в одном контуре.

Mistral расширила Studio: платформа получила managed agents, приложения, новые возможности workflow, корпоративный контекст, наблюдаемость и оценки. Команда может подключать модели и корпоративные системы через коннекторы и MCP, а затем развернуть решение в облаке, VPC, on-premises или смешанной среде.

Почему важно. Корпоративный агент обычно собирается из раздельных частей: модели, данных, прав доступа, оркестрации и эксплуатации. Mistral Studio сводит эти слои в единый контур с трассировками и оценками, что сокращает разрыв между демонстрацией и системой, которую можно контролировать в рабочем процессе с чувствительными данными.

Build Live Apps в Cursor.

Cursor добавил Build Live Apps: облачный агент может создать веб-приложение, при необходимости поднять Postgres, настроить аутентификацию через Cursor SSO и развернуть результат по выбранному адресу. Перед публикацией Cursor запускает проверки секретов и типовых проблем безопасности, а публикация требует явного действия пользователя.

Почему важно. Агентное написание кода само по себе не гарантирует, что коллеги смогут открыть и проверить готовый инструмент: обычно остаются ручные шаги с базой данных, доступом и хостингом. Build Live Apps связывает генерацию кода с развёртыванием и контролем публикации, поэтому внутренние утилиты и прототипы быстрее становятся доступными для совместной проверки.

Sierra приобретает Anon.

Sierra сообщила, что команда Anon присоединяется к компании. Sierra развивает AI-агентов для клиентских сценариев, а Anon специализируется на инструментах для работы агентов с вебом, code review и тестированием.

Почему важно. Агентные продукты зависят не только от языковой модели, но и от способности безопасно действовать в реальных сервисах и рабочих процессах. Объединение Sierra и Anon усиливает тенденцию к платформам, которые соединяют диалоговую часть агента с выполнением действий; для рынка это означает дальнейшую консолидацию вокруг прикладной инфраструктуры агентов.

Koa Decisions: быстрый выбор следующего действия для корпоративного агента.

Salesforce представила Koa Decisions — следующую версию HyperClassifier. Модель получает запрос и набор заданных категорий, а затем выбирает вариант, даёт ответ «да/нет» или оценку за один шаг; она поддерживает текст, изображения и аудио. Salesforce сообщает, что её прежняя модель маршрутизации в Agentforce работала примерно за 26 мс против около 1,4 с у универсальной LLM.

Почему важно. Во многих агентных системах основная задача — не написать длинный ответ, а быстро отправить обращение в нужный workflow, разрешить действие или назначить приоритет. Специализированная модель решений уменьшает стоимость и задержку на таких шагах, поэтому архитекторы могут резервировать крупные языковые модели для сложного рассуждения, а рутинные развилки выполнять быстрее и предсказуемее.

Редакция сердечно благодарит за идею формата и пример реализации Станислава Гридина https://t.me/gridstan

Qwen Code v0.27.0: параллельные дочерние вопросы.

QwenLM выпустила v0.27.0 терминального агента Qwen Code. Managed-agent sessions теперь могут держать несколько дочерних вопросов одновременно, отвечать на них в произвольном порядке и раздельно хранить черновики и ошибки; релиз также включает изменения Web Shell и надёжности runtime.

Почему важно. Разветвлённые инженерные задачи редко решаются одной последовательной цепочкой: независимые уточнения начинают смешивать контекст и тормозить работу агента. Раздельное состояние для параллельных вопросов делает сложное исследование и отладку более управляемыми, не заставляя разработчика вручную делить один сеанс на множество мелких.

Ollama v0.40.15: новые модели и исправления облачного режима.

В Ollama v0.40.15 добавлена поддержка Xiaomi MiMo mimo-v2.6 для задач кода и агентов, а также семейство d3 decision models. В предшествующем обновлении ветки были исправлены обработка сжатых ответов cloud web search и восстановление после разрыва соединения с ChatGPT.

Почему важно. Локальный запуск моделей полезен только тогда, когда он умеет работать с современными моделями и не ломает рабочий процесс на границе локального и облачного режима. Обновление Ollama расширяет выбор моделей для разработчиков и укрепляет связку локального runtime с внешними сервисами, поэтому прототипы и личные инструменты легче доводить до стабильного применения.

---

llama.cpp v0.9.0: FP8 и поддержка новых архитектур.

Проект llama.cpp выпустил v0.9.0. В релизе заявлена поддержка квантования OCP FP8 E4M3 и добавлена совместимость с несколькими новыми архитектурами, включая Mistral Large 4.

Почему важно. Модели становятся крупнее, а требования к памяти и пропускной способности остаются главным ограничением локального и периферийного запуска. Поддержка новых форматов квантования и архитектур расширяет число моделей, которые можно проверить вне облака, и даёт командам больше вариантов для экспериментов с собственной инфраструктурой.

BrickVision для проверки grounded reasoning.

Elorian представила BrickVision — бенчмарк для оценки способности моделей рассуждать о визуальных объектах, которые нельзя полностью передать текстовым описанием. Исследование посвящено тому, когда стратегии grounded reasoning действительно улучшают ответы мультимодальных систем.

Почему важно. Правдоподобное текстовое объяснение не доказывает, что модель опиралась на изображение, а не подставила знакомый шаблон. BrickVision проверяет связь рассуждения с конкретным визуальным объектом, поэтому разработчикам проще отличать систему, которая действительно видит существенные детали, от системы, которая лишь убедительно формулирует догадки.

Cloudflare добавила видимость и правила доступа для персональных агентов.

Cloudflare запустила аналитику персональных агентов и более детальные правила доступа: в Security Analytics появился отдельный тип трафика Personal Assistants, а в настройках можно отдельно разрешить агента, блокировать его на страницах с рекламой или блокировать полностью. Для Account Abuse Protection в раннем доступе добавлена видимость действий агента от имени конкретной учётной записи.

Почему важно. Для сайта агент, который ищет товар, читает справку или входит в аккаунт, часто выглядит как неразличимый автоматический трафик; полная блокировка защищает ресурс, но может отсечь законного пользователя. Инструменты Cloudflare дают владельцу сайта способ увидеть и дозировать этот новый тип обращений, поэтому агентный веб может развиваться без отказа от антифрода и контроля доступа.

Что отслеживать

- Проверять, как Mistral расширяет интеграции, развёртывание и оценку корпоративных агентов.
- Следить за тем, какие сценарии развёртывания и публикации Cursor откроет для Cloud Agent Apps.
- Сравнивать зрелость новых локальных моделей и форматов квантования в Ollama и llama.cpp.
- Наблюдать за практикой правил доступа и видимости персональных агентов у владельцев сайтов.

Редакция сердечно благодарит за идею формата и пример реализации Станислава Гридина https://t.me/gridstan