

Новая SLM выведет работу с ПК на новый уровень эффективности.

Microsoft – это американская многопрофильная компания, занимающаяся разработкой программного обеспечения и производством компьютерной техники. Она была основана в 1975 году Биллом Гейтсом и Полом Алленом и на сегодняшний день является одной из самых крупных и известных IT-компаний в мире.
Среди продуктов Microsoft наиболее известными являются операционные системы Windows, пакеты офисных приложений Office, браузер Internet Explorer и поисковая система Bing. Кроме того, компания занимается разработкой программного обеспечения для серверов, баз данных, игровых консолей Xbox и многих других устройств.
Microsoft также предоставляет услуги облачных вычислений и хранения данных через свою платформу Azure, а также занимается разработкой искусственного интеллекта и других инновационных технологий. Компания имеет филиалы по всему миру и сотрудничает с многими крупными корпорациями и организациями." data-html="true" data-original-title="Microsoft" >Microsoft представила Phi Silica — первую малую языковую модель (SLM) для оптимизации производительности искусственного интеллекта на персональных компьютерах. Продукт поможет решить проблему высоких вычислительных затрат, из-за которых часто перегружаются центральный и графические процессоры, снижая общую эффективность системы.

Phi Silica, имея всего 3,3 миллиарда параметров, является самой компактной моделью в семействе Phi от Microsoft. Несмотря на свой небольшой размер, она демонстрирует неплохие показатели — задержка при генерации первого токена составляет 650 токенов в секунду, а энергопотребление всего 1,5 Ватт. Это позволяет освободить ресурсы ЦПУ и ГПУ для других задач, обеспечивая плавную работу остальных приложений.

Phi Silica разработана специально для нейронных процессоров (neural processing units, NPU) новых персональных компьютеров Copilot+. Генерация токенов модели использует кэш KV на NPU и работает на ЦПУ, производя около 27 токенов в секунду. Разработчики получат доступ к API Phi Silica через Windows App SDK, что поможет им создавать инновационные приложения с ИИ в экосистеме Windows.

Функционал Phi Silica будет включать распознавание оптического символьного ввода (optical character recognition, OCR), студийные эффекты, живые субтитры и API для отслеживания активности пользователя. В дальнейшем планируется добавить встраивание векторов, API для обобщения ответов (retrieval-augmented generation, RAG) и автоматического реферирования текста.

Phi Silica входит в серию Phi-3 наряду с моделями Phi-3-mini (3,8 млрд параметров), Phi-3-small (7 млрд параметров), Phi-3-medium (14 млрд параметров) и недавно анонсированной Phi-3-vision (4,2 млрд параметров). Однако Phi Silica станет первой передовой малой языковой моделью, которая будет поставляться вместе с Windows, что означает важный шаг в расширении доступа конечных пользователей к продвинутым возможностям искусственного интеллекта.

Первые компьютеры Copilot+ с поддержкой Phi Silica поступят в продажу в середине июня 2024 года. Они будут оснащены процессорами Snapdragon X Elite и Plus от Qualcomm на базе архитектуры Arm. Позднее в третьем квартале 2024 года ожидается выпуск ноутбуков Copilot+ с процессорами Lunar Lake от Intel.