

Яндекс первым среди российских компаний подтвердил соответствие международному стандарту ISO/IEC 42 001, который оценивает ответственный подход к разработке искусственного интеллекта (ИИ). Независимый аудит установил, что процессы создания и использования YandexGPT соответствуют требованиям в области безопасности и этики.

Стандарт ISO/IEC 42 001 оценивает не сами ИИ-продукты, а систему управления их разработкой. Он учитывает прозрачность процессов, безопасность данных и этические аспекты. YandexGPT — семейство языковых моделей Яндекса, применяемых более чем в 20 сервисах компании, включая поиск, Алису и облачные решения.

В ходе проверки эксперты убедились, что Яндекс обеспечивает защиту пользовательских данных, отслеживает ошибки в работе моделей, контролирует качество ответов и анализирует обратную связь. Компания также учитывает социальные риски, связанные с ИИ, и обучает сотрудников принципам ответственной разработки.

Как отмечают в компании, работа над этичностью и безопасностью ИИ велась задолго до выхода YandexGPT. Яндекс опубликовал свои принципы разработки ИИ в открытом доступе и участвует в отраслевых инициативах. С 2025 года компания привлекает «белых хакеров» для поиска уязвимостей в нейросетях, предлагая вознаграждения до миллиона рублей.