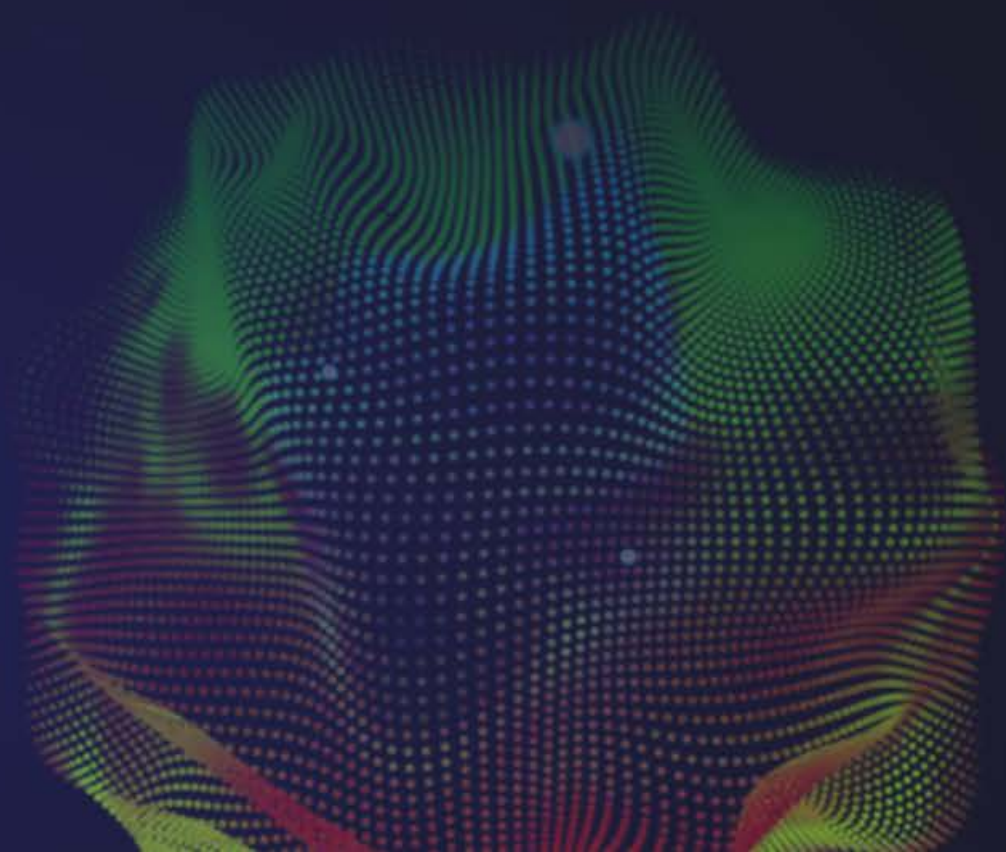


Декларация об ответственном генеративном ИИ

Системы генеративного искусственного интеллекта открывают большие возможности для развития экономики и творческого потенциала. Генеративный искусственный интеллект может существенно улучшить жизнь миллионов человек. При этом развитие и применение технологий генеративного искусственного интеллекта должно быть этическим и ответственным





Яндекс



Сколтех



ИТМО



УНИВЕРСИТЕТ
ИННОПОЛИС

ИСП РАН



УНИВЕРСИТЕТ
ЛОБАЧЕВСКОГО



ИННОПОЛИС 13 марта 2024

**Российские вузы и ИТ-компании
определили этические принципы
создания генеративного ИИ**

Декларация

об ответственной разработке и использовании сервисов в сфере генеративного искусственного интеллекта

Генеративный искусственный интеллект открывает большие возможности для развития экономики и творческого потенциала. Потенциал внедрения генеративного искусственного интеллекта для решения конкретных задач безграничен. Уже сегодня на основе базовых моделей создаются как сервисы для пользователей, позволяющие самостоятельную работу с генеративным искусственным интеллектом, так и сервисы для разработчиков (например, доступ к генеративным моделям через API). Мы верим, что все эти сервисы на основе генеративного искусственного интеллекта смогут внести существенный вклад в улучшение жизни миллионов людей. При этом развитие и применение технологий генеративного искусственного интеллекта должно быть ответственным. Разработчики сервисов в сфере генеративного искусственного интеллекта должны понимать все риски развития этой технологии и принимать необходимые меры для их устранения на основе законодательства и принципов этики искусственного интеллекта. Большую роль в развитии технологий генеративного искусственного интеллекта играют их пользователи. Они - вторая сторона технологического диалога, благодаря которому можно и нужно выявлять возможности и направления для улучшения работы сервисов, основанных на работе генеративного искусственного интеллекта.

Для обеспечения безопасного и ответственного развития технологий генеративного искусственного интеллекта следуйте рекомендациям этой Декларации.

О декларации и ее реализации

1. Эта Декларация устанавливает этические принципы, рекомендации и стандарты поведения, которыми рекомендуется руководствоваться участникам отношений в сфере генеративного искусственного интеллекта.
2. Присоединение к Декларации добровольное.
3. Декларация принята в развитие и для конкретизации положений Кодекса этики в сфере искусственного интеллекта применительно к сервисам на основе генеративного искусственного интеллекта. Под «сервисом на основе генеративного искусственного интеллекта» понимается технологическое решение, обеспечивающее за счёт применения технологий искусственного интеллекта возможность самостоятельного создания голосовой информации, письменного текста, изображений, звуков, видео-или иных электронных сообщений на основании текстового или иного запроса пользователя, а также создание инструментов для разработчиков, позволяющих создать вышеупомянутые продукты.
4. Положения Декларации могут быть расширены и/или конкретизированы для отдельных категорий участников отношений в сфере генеративного искусственного интеллекта путём отдельных дополнений и приложений к Декларации.

Рекомендации для разработчиков

1. Обеспечивайте безопасность и надёжность ваших сервисов на основе генеративного искусственного интеллекта. Принимайте меры для предотвращения злоупотреблений и создания некорректной информации.
2. До запуска сервисов на основе генеративного искусственного интеллекта тщательно тестируйте эти сервисы, а также разрабатывайте и внедряйте процедуры системного мониторинга. После запуска сервисов на основе генеративного искусственного интеллекта следуйте внедрённым процедурам системного мониторинга. По результатам мониторинга принимайте меры для предотвращения создания некорректной информации в будущем, а также для совершенствования внедрённых процедур.
3. Создавайте команду экспертов из разных отраслей для системной проверки контента, создаваемого сервисами на основе генеративного искусственного интеллекта, на соответствие нормам морали и нравственности, а сами сервисы – на взломоустойчивость и информационную безопасность.
4. Тщательно анализируйте сообщения о фактах генерации некорректной и незаконной информации и учитывайте эту информацию при доработке сервисов на основе генеративного искусственного интеллекта.
5. Информировать пользователя о факте использования генеративного искусственного интеллекта в ваших сервисах там, где это возможно и неочевидно.
6. Осуществляйте маркировку (видимую или скрытую в зависимости от целесообразности и технической возможности) информации, создаваемой с помощью сервисов на основе генеративного искусственного интеллекта, там, где это разумно применимо и оправдано с учетом специфики сервиса.
7. Принимайте разумные меры для безопасного использования сервисов на основе генеративного искусственного интеллекта всеми категориями пользователей, включая несовершеннолетних. Принимайте меры по повышению осведомленности пользователей о возможных ошибках при применении искусственного интеллекта.
8. В пользовательском соглашении или другом документе напоминайте пользователям о недопустимости создания информации для причинения вреда и об ответственности за распространение некорректной информации, созданной с помощью сервисов на основе генеративного искусственного интеллекта.
9. Разработайте и разместите в сети Интернет:
 - пользовательское соглашение или другой документ с правилами пользования сервисом на основе генеративного искусственного интеллекта;
 - ссылку на текст этой Декларации;
 - адрес электронной почты или форму обратной связи для направления сообщений о создании некорректной информации сервисом на основе генеративного искусственного интеллекта;

- инструмент для проверки скрытой маркировки информации, созданной с помощью сервисов на основе генеративного искусственного интеллекта, если вы посчитали необходимым осуществлять такую маркировку.
10. В пользовательских соглашениях или других документах определяйте позицию разработчика по вопросу принадлежности прав на контент, созданный с использованием его сервиса на основе генеративного искусственного интеллекта. Указывайте, существуют ли ограничения на использование результатов работы сервиса, например, для коммерческих целей. Описывайте иные условия использования созданных материалов.
 11. Проводите междисциплинарные исследования о последствиях применения генеративного искусственного интеллекта в различных сферах. Принимайте по необходимости меры, направленные на минимизацию возможных негативных эффектов от пользования этими сервисами.
 12. Разрабатывайте сервисы на основе генеративного искусственного интеллекта с учетом особенностей и с уважением к ценностям и традициям различных народов, культур, религий и других социальных общностей.
 13. Обменивайтесь с другими разработчиками опытом обеспечения этичности работы сервисов на основе генеративного искусственного интеллекта. Разрабатывайте совместные стандарты этичности работы сервисов на основе генеративного искусственного интеллекта.
 14. Публикуйте информацию о новых возможностях, функциях ваших сервисов на основе генеративного искусственного интеллекта, а также, где это важно для пользователей, о выявленных проблемах.

Рекомендации для пользователей

1. Не используйте сервисы на основе генеративного искусственного интеллекта для причинения вреда или в иных противоправных целях, не создавайте целенаправленно некорректную информацию, например, противоречащую закону или общепринятым нормам нравственности.
2. Помните: законодательством установлена ответственность за распространение противоречащей закону информации, даже если она была создана с применением генеративного искусственного интеллекта.
3. Если вы создали с помощью сервисов на основе генеративного искусственного интеллекта некорректную информацию, сообщите об этом разработчику. Этим вы сможете сделать сервис лучше и безопаснее.
4. Не используйте сервисы на основе генеративного искусственного интеллекта там, где очевидно требуется проявление личного творческого вклада человека (например, сочинения, экзаменационные и квалификационные испытания и т.д.) или когда использование генеративного искусственного интеллекта способно ввести других в заблуждение.

5. Ответственно относитесь к любым способам использования сгенерированной информации в вашей профессиональной деятельности. Вам следует знать установленные законом или деловой практикой ограничения на применение сервисов на основе генеративного искусственного интеллекта в вашей работе.
6. Помните, что любая технология не совершенна и требует вашего контроля. Особенно внимательно относитесь к применению сервисов на основе генеративного искусственного интеллекта детьми.
7. Знайте, что в конечном счете генеративный искусственный интеллект всегда является лишь одним из инструментов в ваших руках.
8. В случае использования сервисов искусственного интеллекта в профессиональной деятельности, доступно и понятно объясните о существующих в вашей сфере и в вашей организации условиях и ограничениях по применению сервисов на основе генеративного искусственного интеллекта.
9. Используйте сервисы на основе генеративного искусственного интеллекта как источник информации для любого уровня работ ответственно и с осознанием ограничений этого инструмента.