

Методы регулирования генеративного искусственного интеллекта (проект для общественных консультаций)

Опубликован Администрацией киберпространства Китая 11.04.2023 г., замечания и предложения принимаются до 10.05.2023 г.

Источник: http://www.cac.gov.cn/2023-04/11/c_1682854275475410.htm

Статья 1. Настоящие Методы разработаны на основе законов КНР «О кибербезопасности», «О цифровой безопасности», «О защите персональных данных» и др. в целях содействия здоровому развитию и регламентированному применению генеративного искусственного интеллекта (далее — ИИ).

Статья 2. Настоящие Методы применяются к разработке и использованию генеративного ИИ для оказания услуг пользователям на территории Китайской Народной Республики.

В контексте настоящих Методов «генеративный ИИ» обозначает технологии, генерирующие текст, изображения, звук, видеоматериалы, код и иные виды контента при помощи алгоритмов, моделей и правил.

Статья 3. Государство поддерживает собственные разработки, внедрение и применение таких базовых технологий, как алгоритмы и фреймворки для ИИ, а также международное сотрудничество в данной области; поощряет приоритетное использование безопасных и надежных программ, инструментов, вычислительных ресурсов и источников данных.

Статья 4. При предоставлении продуктов или услуг на основе генеративного ИИ необходимо соблюдать законодательные, социальные и моральные нормы, а также выполнять следующие требования:

1) Контент, генерируемый при помощи генеративного ИИ, должен отражать ключевые ценности социализма. Не допускается содержание, связанное с подрывом государственной власти, свержением социалистического режима, агитацией к расколу государства, разрушением национального единства, пропагандой терроризма, экстремизма, межнациональной розни, национальной дискриминации, насилия, эротического и порнографического контента, недостоверной информации, а также контента, способного нарушить экономический или социальный порядок.

2) В процессе дизайна алгоритмов, обучения выбору данных, генерации и оптимизации моделей, предоставлении услуг и т.д. необходимо применять меры по предотвращению дискриминации по расовому, национальному или религиозному признаку, по гражданству, территориальной принадлежности, полу, возрасту, профессии и др.

3) Необходимо соблюдать права на объекты интеллектуальной собственности и нормы деловой этики. Не допускается использование алгоритмов, данных, платформ и иных преимуществ для осуществления недобросовестных конкурентных практик.

4) Контент, генерируемый при помощи генеративного ИИ, должен быть достоверным и точным. Необходимо применять меры по предотвращению генерации недостоверной информации.

5) Необходимо уважать законные интересы других лиц, препятствовать нанесению ущерба физическому и психическому здоровью других лиц, нарушению права на изображение, права на честь и достоинство и право на личную тайну, а также прав на объекты интеллектуальной собственности. Запрещается незаконным образом получать доступ, раскрывать или использовать персональные данные, личную и коммерческую тайну.

Статья 5. Организации и физические лица, предоставляющие услуги генерации чатов, текста, изображений и звука при помощи продуктов на основе генеративного ИИ, включая оказание другим лицам поддержки в самостоятельной генерации текста, изображений, звука и проч. при помощи программируемого интерфейса и др., несут ответственность как производители контента, генерируемого данными продуктами; если используются персональные данные, они также в установленном законом порядке несут ответственность как операторы обработки персональных данных и обязаны обеспечивать их защиту.

Статья 6. Перед оказанием пользователям услуг при помощи продуктов на основе генеративного ИИ необходимо предоставить государственному органу регулирования киберпространства отчет о проверке безопасности в соответствии с «Положениями об оценке безопасности информационных интернет-услуг, способным влиять на общественное мнение или мобилизовать общественность», а также оформить необходимые документы о постановке алгоритмов на учет, изменении сведений и снятию с учета в соответствии с «Положениями о регулировании алгоритмических рекомендаций при оказании информационных интернет-услуг».

Статья 7. Поставщик обязан нести законную ответственность за источник данных, которые применяются для предварительного и дополнительного обучения продуктов на основе генеративного ИИ.

Данные, которые применяются для предварительного и дополнительного обучения продуктов на основе генеративного ИИ, должны отвечать следующим требованиям:

1) Соответствовать положениям закона КНР «О кибербезопасности» и иных применимых законодательных и нормативных актов;

2) Не содержать информацию, нарушающую права на объекты интеллектуальной собственности;

3) Если какие-либо из данных являются персональными, необходимо получить согласие субъекта данных либо соблюсти иные условия, установленные законами, административными актами и др.;

4) Присутствует возможность гарантировать истинность, точность, объективность и разнообразие данных;

5) Соответствовать иным требованиям, установленным государственным органом регулирования киберпространства в отношении услуг, оказываемых при помощи генеративного ИИ.

Статья 8. При осуществлении ручного присвоения тегов в ходе разработки продуктов на основе генеративного ИИ поставщик обязан составить понятные, конкретные и выполнимые правила такого присвоения, соответствующие положениям настоящих Методов, проводить инструктаж и подготовку сотрудников, занимающихся присвоением тегов, а также осуществлять выборочную проверку корректности присвоения.

Статья 9. При предоставлении услуг на основе генеративного ИИ необходимо в соответствии с положениями закона КНР «О кибербезопасности» запрашивать у пользователей достоверную информацию, удостоверяющую личность.

Статья 10. Поставщик обязан обозначить аудиторию, обстоятельства применения и назначение своих услуг и опубликовать эту информацию в открытом доступе, а также применить уместные меры для предотвращения чрезмерной зависимости пользователей от генерируемого контента.

Статья 11. При оказании услуг поставщик обязан обеспечивать защиту вводимой пользователем информации и истории использования. Не допускается незаконным образом сохранять вводимую информацию, на основе которой можно установить личность пользователя; создавать портрет пользователя на основе вводимой им информации и истории использования; предоставлять вводимую информацию третьим лицам. Если применимыми законами и нормативными актами установлено иное, необходимо следовать их положениям.

Статья 12. Поставщик не может заниматься дискриминирующей пользователя генерацией контента на основе расы, гражданства, пола или иных характеристик.

Статья 13. Поставщик обязан создать механизм приема и обработки обращений пользователей, своевременно обрабатывать частные запросы обновить, удалить или скрыть персональные данные. Если поставщик выявляет, что генерируемый текстовый, изобразительный, звуковой или видеоконтент нарушает права третьих лиц на изображение, честь и достоинство, личную и коммерческую тайну или положения настоящих Методов, либо ему становится известно о таком нарушении, он обязан принять необходимые меры для остановки генерации и предотвращения дальнейшего ущерба.

Статья 14. В течение жизненного цикла [программы] поставщик обязан предоставлять безопасные, стабильные и непрерывные услуги, обеспечивая штатный порядок использования.

Статья 15. Если в процессе ведения деятельности или на основе обращения пользователя поставщик выявляет, что генерируемый контент нарушает положения настоящих Методов, помимо принятия таких мер, как фильтрация контента и др., он также обязан в течение 3 месяцев предотвратить повторную генерацию такого контента при помощи дополнительного обучения модели и др.

Статья 16. Поставщик обязан маркировать генерируемые изображения и видеоконтент в соответствии с «Положениями о регулировании технологий глубокого синтеза при оказании информационных интернет-услуг».

Статья 17. Поставщик обязан по запросу государственного органа регулирования киберпространства и иных компетентных ведомств предоставлять необходимую информацию, способную повлиять на доверие и выбор пользователей, включая описание источника, моделей, типов и качества данных для предварительного и дополнительного обучения, правил ручного присвоения тегов, объемы и тип данных, которым вручную присваиваются теги, сведения о базовых алгоритмах и технологических системах и др.

Статья 18. Поставщик обязан информировать пользователей о научном подходе к контенту, генерируемому при помощи генеративного ИИ, и его рациональному использованию, о недопустимости использования генерируемого контента, нарушающего имидж, достоинство и иные законные права и интересы третьих лиц, а также о его применении для коммерческих спекуляций и недобросовестного маркетинга.

При выявлении несоответствия генерируемого контента положениям настоящих Методов пользователь вправе проинформировать об этом органы регулирования киберпространства или иные компетентные ведомства.

Статья 19. Если поставщик обнаруживает, что при использовании продуктов на основе генеративного ИИ пользователь нарушает положения законодательства, нормы деловой этики, социальные и моральные нормы, включая осуществление интернет-спекуляций, отправку сообщений и публикацию комментариев некорректного содержания, рассылку спама, разработку вредоносного ПО, осуществление недобросовестного маркетинга и др., он обязан приостановить или прекратить оказание услуг.

Статья 20. Если поставщик нарушает положения настоящих Методов, органы регулирования киберпространства или иные компетентные ведомства применяют наказание в соответствии с законами КНР «О кибербезопасности», «О безопасности данных», «О защите персональных данных» и другими применимыми законодательными и нормативными актами.

Если наказание не предусмотрено законодательством, органы регулирования киберпространства или иные компетентные ведомства в пределах своих полномочий выдают предупреждение или выносят выговор, а также предписывают устранить нарушение в установленный срок. При отказе устранить нарушение или наличии отягчающих обстоятельств допускается предписание приостановить или прекратить оказание услуг, предоставляемых при помощи генеративного ИИ, а также применение штрафа в размере от 10 000 до 100 000 китайских юаней. Если действия квалифицируются как нарушение общественного порядка, применяется соответствующее наказание; если совершено уголовное преступление, преследуется уголовная ответственность.

Статья 21. Настоящие Методы вступают в силу ____2023 г.