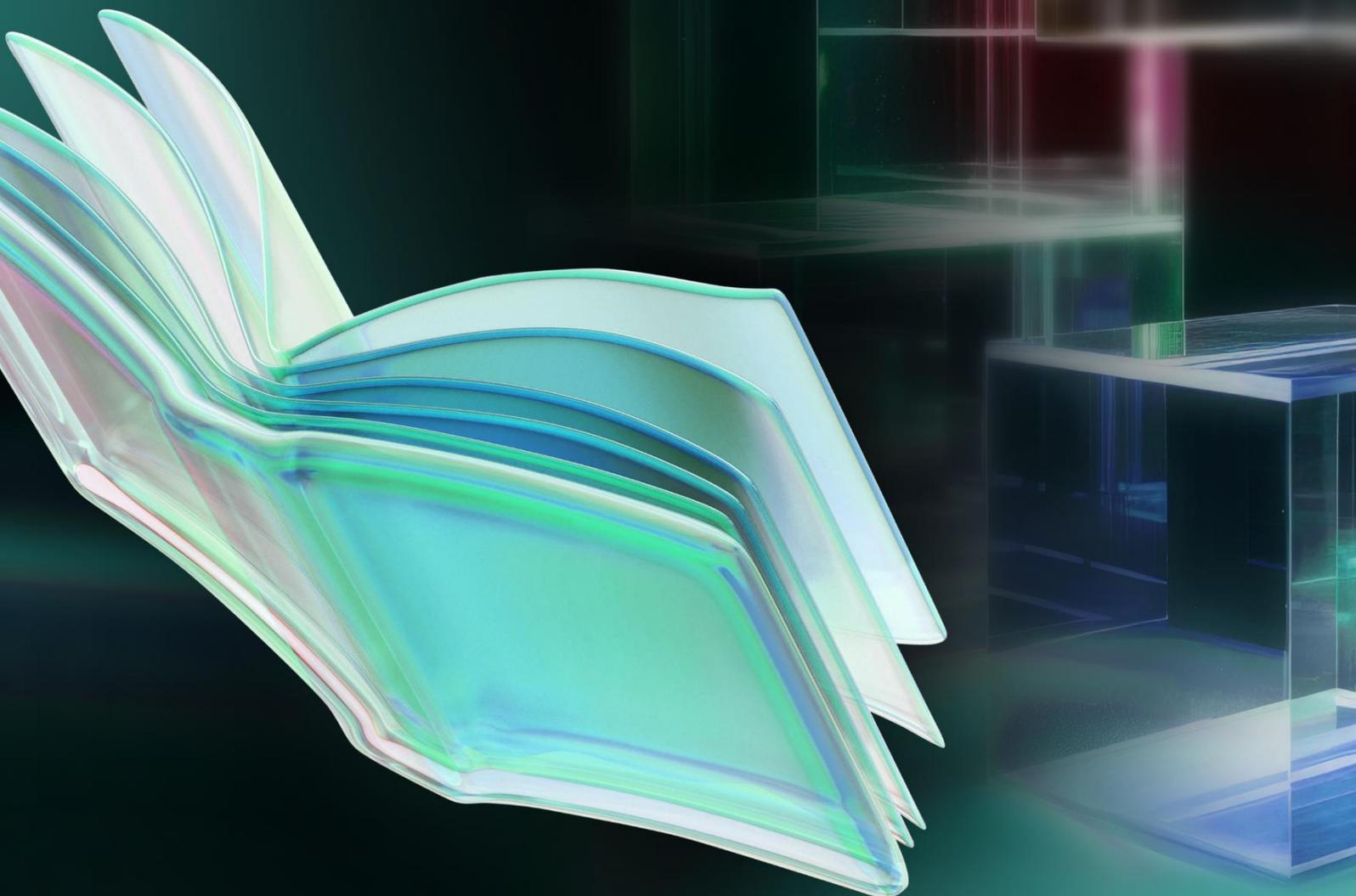


ЭТИЧЕСКИЙ СВЕТОФОР

рабочей группы по этике
в сфере ИИ в образовании



ОБЩИЕ ПОЛОЖЕНИЯ

«Этический светофор» представляет собой набор рекомендаций, уточняющих отдельные положения Кодекса этики в сфере искусственного интеллекта (далее – ИИ) в образовании на конкретных примерах использования ИИ. Эти рекомендации, как и конкретные практики использования ИИ в образовательных и исследовательских целях в этой области, могут претерпевать существенные изменения в процессе накопления опыта применения и быстрого совершенствования самих моделей ИИ.

Отнесение той или иной практики использования ИИ к категории запрещенной («красная зона»), разрешенной с учетом определенных ограничений («желтая зона»), разрешенной («зеленая зона») является контекстуально зависимым. «Этический светофор» представляет собой пример анализа различных практик использования ИИ, значимый на момент его публикации. Обновление «Этического светофора» должно осуществляться на постоянной основе.

В настоящее время различные приложения и инструменты на основе ИИ применяются в образовании для широкого спектра задач и практик. С развитием технологий происходит быстрое внедрение передовых ИИ-решений в сферы, где раньше инструменты интеллектуальной автоматизации никогда не применялись, а потому отсутствовал опыт квалифицированной и надежной оценки воздействия систем ИИ на человека, образовательный и воспитательный процессы в целом.

В ГОСТ Р 59895—2021 «Технологии искусственного интеллекта в образовании» приводятся следующие основные технологии, которые используются в образовательном процессе (отдельно или в совокупности):

- технологии компьютерного зрения;
- технологии обработки естественного языка, распознавания и синтеза речи;
- технологии интеллектуальной поддержки принятия решений.

Ниже приведены некоторые возможные практики использования ИИ в образовании (перечень не является исчерпывающим и должен обновляться с учетом развития как самих технологий, так и способов их использования):

- приложения и ассистенты для обучения в учебном классе и для самостоятельной подготовки учащихся;
- инструменты для удобного и продуктивного взаимодействия в процессе обучения;
- инструменты для быстрого поиска материалов и полезной информации, оформления решений в заданном стандарте, в т. ч. для подготовки библиографии в нужном формате, создания различных списков и перечней, др.;
- обучающие чат-боты, голосовые ассистенты и роботы (универсальные или предназначенные для решения конкретных задач);
- общедоступные генеративные системы, в т. ч. на основе предобучения и использования больших языковых моделей для широкого круга образовательных и исследовательских задач, практик учебного моделирования, для поиска

информации, обработки текста, видео и аудиоматериалов, выдачи по целевому запросу формализованных решений; а также различные корпоративные системы и большие языковые модели с ограниченным пользовательским доступом (в т. ч. на платной основе) и с повышенным уровнем верификации выходных решений;

- системы обучения на основе игровых моделей и различных средств виртуализации, включая адаптивные метавселенные, используемые для усвоения материала и приобретения навыков в командной игре или в индивидуальном интерактивном формате;
- системы для тестирования обучающихся, педагогов, административного персонала, в т. ч. с целью выявления талантов, оценки способностей и наклонностей, уровня знаний, стрессоустойчивости, коммуникативных качеств, способности к быстрой социализации и т. п.;
- программы-ассистенты для автоматизированной оценки общей успеваемости, проверки выполнения учебных заданий и их оценки;
- приложения для разработки персонализированных программ/траекторий обучения, в т. ч. рекомендательные системы по формам и методам работы с конкретным учеником или категорией обучающихся;
- инструменты для разработки учебных программ, методик с учетом конкретных задач образовательной организации, содержания и объема тематических материалов;
- средства визуального наблюдения за аудиторией, в т. ч. с целью контроля психоэмоционального состояния, концентрации внимания, уровня интереса слушателей к материалу, ситуативным учебным сценам и событиям в классе, а также предиктивного выявления различных инцидентов, в т. ч. связанных с безопасностью, например, проблем со здоровьем участников образовательного процесса, конфликтных ситуаций, неадекватного поведения, других угроз и рисков в образовательной организации;
- приложения и различные инструменты для управленческих задач, планирования и подготовки отчетности в учебном заведении;
- различные инструменты, в т. ч. на основе перспективных технологий ИИ, для автоматической генерации заданий с заданной психометрической сложностью;
- экспериментальные модели/устройства (инвазивные и неинвазивные) для быстрого усвоения материала и концентрации познавательных функций мозга, тренировки и расширения памяти и других способностей (относятся к категории высокого риска) и др.

ВОЗМОЖНОСТИ И ОГРАНИЧЕНИЯ ИСПОЛЬЗОВАНИЯ ИИ В ОБРАЗОВАТЕЛЬНОМ ПРОЦЕССЕ

Зеленая зона

Сферы деятельности и практики, где использование ИИ может повысить продуктивность

Для обучающегося

1. Самостоятельная проверка уровня знаний и умений (навыков) вне учебной аудитории;
2. Использование инструментов на основе ИИ как тренажера отдельных умений. Например: аудирование, письмо, счет, рисование, программирование и т. д.;
3. Тьюторская поддержка со стороны ИИ при самостоятельной работе учащегося с учебными материалами или заданиям;
4. Использование сервисов на основе ИИ для перевода учебных материалов с и на иностранные языки;
5. Озвучка учебных материалов;
6. Использование ИИ для оформления письменных учебных и научных работ в соответствии с действующими в образовательной организации (или профессиональной области) стандартами оформления работы;
7. Использование ИИ для генерации изображений с целью оформления учебных работ (при условии отсутствия в образовательной организации запрета на использование генеративных инструментов на основе ИИ при выполнении данного вида работ и при условии указания источника, в котором было сгенерировано соответствующее изображение).

Для преподавателя

8. Генерация с помощью ИИ оценочных средств закрытого типа на основе выверенного учебного материала и генерация открытых тестовых заданий на основе примера, заложенного преподавателем или разработчиком образовательного содержания;
9. Проверка тестовых и текстовых заданий при наличии дескрипторов от преподавателя и разработчика образовательного содержания;
10. Создание преподавателем-экспертом справочных материалов по учебному курсу/ дисциплине с использованием ИИ;
11. Разработка преподавателем-экспертом учебных материалов с помощью инструментов генеративного ИИ;

12. Использование ИИ-сервисов для перевода учебных материалов с на иностранные языки;
13. Озвучка учебных материалов;
14. Использование цифровых аватаров (в т. ч. известных исторических личностей, ученых и пр.), позволяющих представить содержание дисциплины, научной работы и т. п., при условии обязательного предупреждения аудитории о факте общения с цифровой имитацией;
15. Использование систем прокторинга с ИИ при условии доступности прозрачных апелляционных процедур для обучающихся.

Для администратора

16. Решение с помощью сервисов на основе ИИ ряда рутинных административных задач (учебных офисов);
17. Мониторинг загруженности аудиторного фонда и формирование схем оптимального его использования, составление расписания занятий;
18. Использование чат-ботов для автоматизации ответов на организационные вопросы обучающихся при условии прозрачности ИИ в этом процессе (не-введение участников процесса в заблуждение относительно того, что общение происходит не с живым человеком, а с системой ИИ);
19. Использование генеративных инструментов для оформления отчетов, аналитики, дэшбордов, в том числе, предиктивной аналитики образовательных результатов обучающихся.

Желтая зона

Сферы деятельности и практики, где использование ИИ может приводить как к позитивным, так и к негативным последствиям, и потому требует особого внимания со стороны участников образовательного процесса

Для обучающегося

1. Использование систем адаптивного обучения на основе ИИ, которые учитывают возможности, интересы, потребности и индивидуальные особенности обучающегося, обеспечивая персонализацию образования путем составления индивидуальных учебных планов;
2. Генерация решения или алгоритма правильного решения для учебных заданий (без выдачи правильного готового ответа и при условии наличия экспертной оценки, подтверждающей, что данные ИИ-решения не препятствуют формированию критически значимого самостоятельного навыка у обучающегося);
3. Экспериментальное использование ИИ для решения задач структурирования: классификация, суммаризация, выделение ключевых тезисов, - при условии

наличия у обучающегося навыка самостоятельного выполнения данных задач и критической оценки сгенерированного материала¹;

4. Экспериментальное использование ИИ для поисково-аналитических задач: сбор первичной литературы и иных материалов по теме, сравнение материалов, подготовка аннотаций и т. п., - при условии наличия у обучающегося навыка самостоятельного выполнения подобных задач и навыка критической оценки сгенерированного содержания.
5. Ограничение использования генеративных моделей связано с тем, что им на данный момент свойственна непрозрачность (ситуация так называемого «черного ящика»), допущение фактологических ошибок, сложность обеспечения надежных проверок «плохого» вопроса системой ИИ, принятие ею «на веру» информации в вопросе, зависимость от больших высококачественных наборов данных для проведения точного анализа, фокус на самом «частотном» термине, сложность контроля за соответствием генерируемого содержания этическим нормам.

Для преподавателя

6. ИИ-поддержка преподавателя в режиме корректора и редактора при формулировании образовательных результатов, тематических разделов учебного плана (определение целей и содержания осуществляется преподавателем или разработчиком);
7. Поддержка преподавателя при проверке письменных и устных заданий;
8. Поддержка преподавателя при проведении код-ревью на технических и IT направлениях;
9. Проверка работ на плагиат и степени самостоятельности выполненной работы (при условии реализации многофакторной оценки учебной работы);

Для администратора

10. Мониторинг уровня вовлеченности и эмоционального состояния обучающихся с учетом соблюдения требований приватности, т. е. без идентификации личности только при условии согласия преподавателя и обучающихся, а также в соответствии с целями повышения качества образовательного процесса, определенными преподавателем. Недопустимо использование данных сервисов с целью контроля и оценки работы преподавателя;
11. Диагностика и анализ профиля обучающегося для формирования рекомендаций по возможным профессиональным траекториям;
12. Проектирование траекторий развития преподавателей на основе собранного цифрового следа в профессиональной деятельности, интересов преподавателя, результата диагностик и пр.

1. Ограничение использования генеративных моделей связано с тем, что им на данный момент свойственна непрозрачность (ситуация так называемого «черного ящика»), допущение фактологических ошибок, сложность обеспечения надежных проверок «плохого» вопроса системой ИИ, принятие ею «на веру» информации в вопросе, зависимость от больших высококачественных наборов данных для проведения точного анализа, фокус на самом «частотном» термине, сложность контроля за соответствием генерируемого содержания этическим нормам.

Сферы деятельности и практики, где использование ИИ имеет предсказуемо негативное влияние на результат и потому должно быть исключено

Для обучающегося

1. Выполнение заданий вместо обучающихся инструментами ИИ, включая генерацию содержания текстов, решение задач, тестов и т. п.

Для преподавателя

2. Полная генерация содержания исследовательских работ и образовательных программ ИИ-инструментами.

Для администратора

3. Составление с использованием ИИ-инструментов различных рейтингов обучающихся и преподавателей, которые могут нанести психологическую травму или способствовать разжиганию атмосферы ненависти, вражды или нездоровой конкуренции в учебных и профессиональных коллективах;
4. Делегирование ИИ полномочий по принятию административных решений, которые могут создать риски или нанести ущерб людям, образовательному процессу, репутации образовательного учреждения;
5. Использование в учебном процессе систем ИИ, которые изначально не разрабатывались или не тестировались для использования в образовании, в т. ч. с учетом действующих в этой сфере возрастных и психофизиологических особенностей обучающихся;
6. Сбор данных, не имеющих прямой связи с качеством образовательного процесса и (в явной или неявной форме), относящихся к личной жизни, поведению, привычкам и другим индивидуальным особенностям обучающихся и преподавателей.