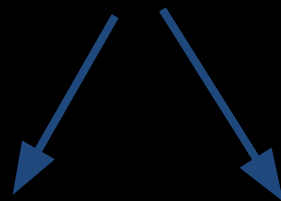




1. ЭТИКА И ФИЛОСОФИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Подготовила
Левичева Н.Б.,
гр. ПМИОZ-811

Этика искусственного интеллекта является частью этики технологий, характерной для роботов и других искусственно интеллектуальных существ

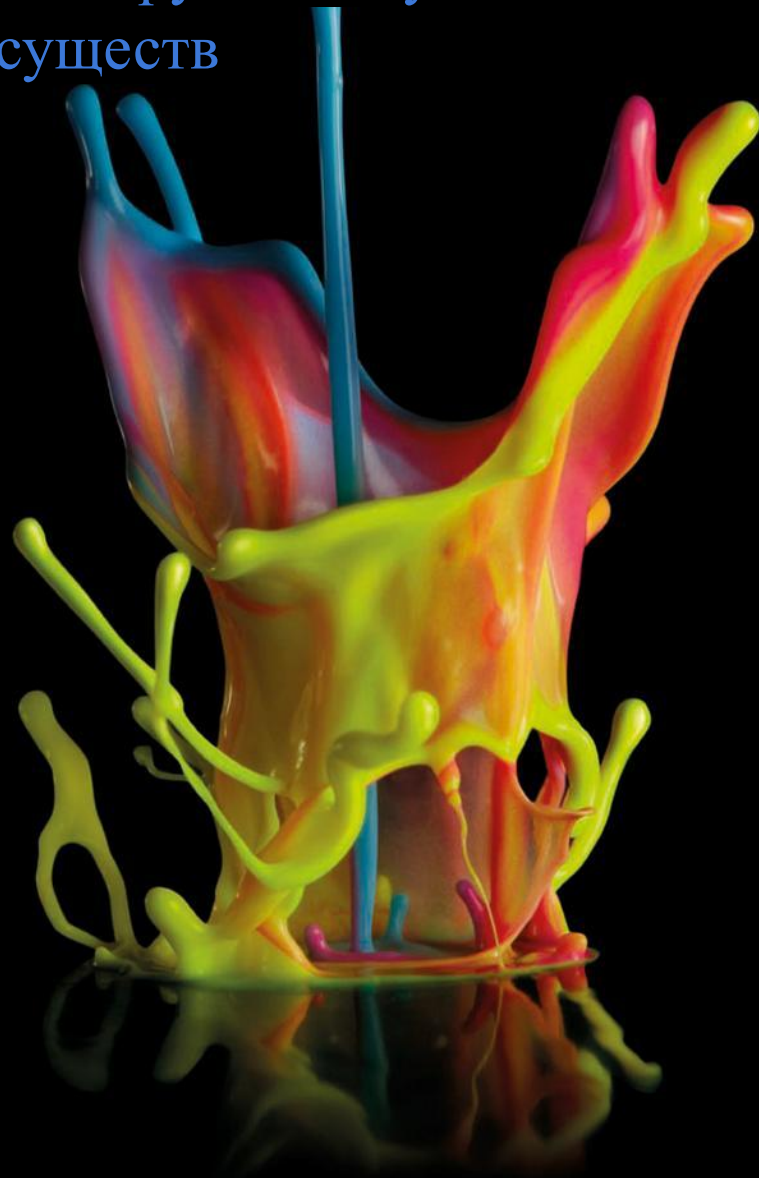


РОБОЭТИКА

решает вопросы морального поведения людей при проектировании, конструировании, использовании и лечении искусственно разумных существ

МАШИНАЯ ЭТИКА

затрагивает проблемы морального поведения искусственных моральных агентов (ИМА)



Этика искусственного интеллекта

рассматривается в двух основных аспектах:

- 1) этические принципы, лежащие в основе принимаемых ИИ решений,
- 2) этичное поведение ИИ в ситуации, напрямую касающейся людей.

Система искусственного интеллекта способна:

- самостоятельно принимать решения, касающиеся человека,
- анализировать данные в таких объемах и с такой скоростью, как человек делать не в состоянии (следовательно, человек не может проверить верность решений).

Основная проблема — определение того, насколько решения, принимаемые интеллектуальной автономной системой (ИАС), соответствуют этическим нормам, то есть насколько она этична.



Два аспекта этики искусственного интеллекта: этичность решения и этичность применения

Этика ИИ



Этическая ситуация, в которой ИИ принимает решения.

Этичность применения ИИ и связанные с этим социальные вызовы.

Этика разработчика

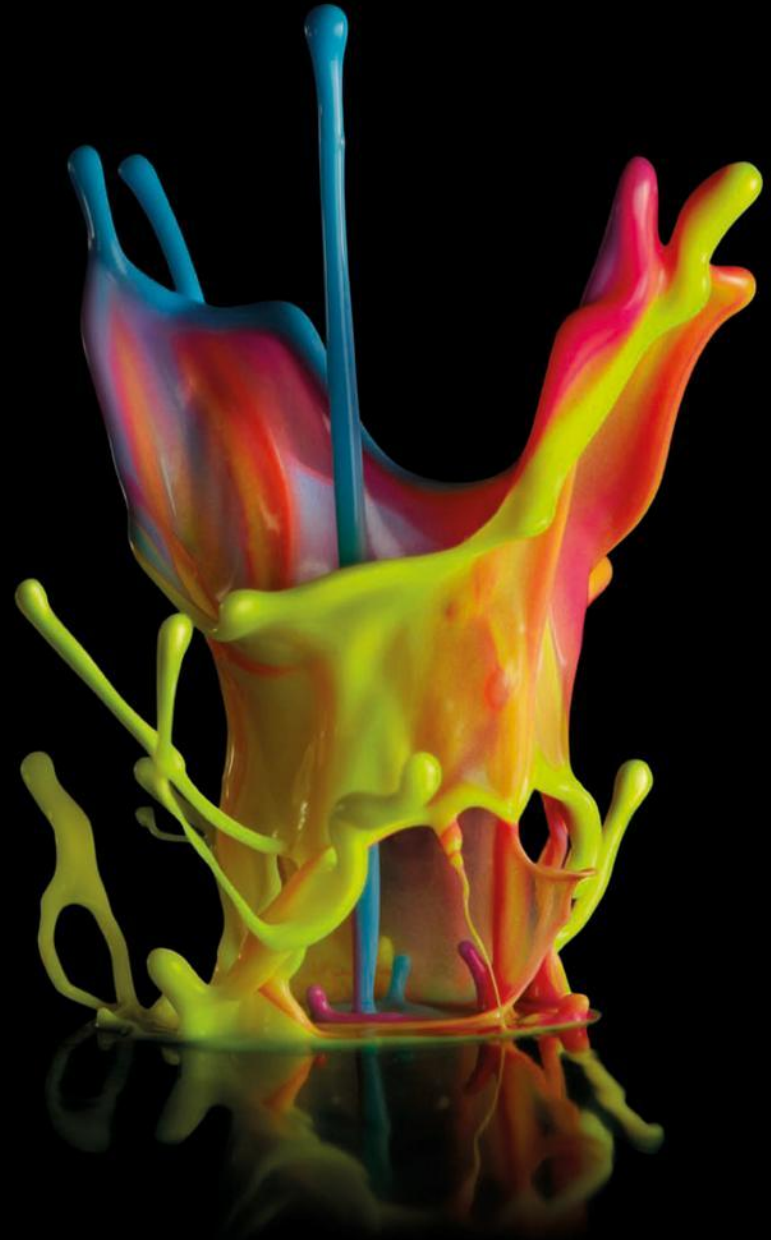


Философские размышления можно свести к двум глобальным вопросам:

1) что такое искусственный интеллект, возможно ли его создание и каким образом;

2) каковы возможные последствия его возникновения в жизни человечества.

Основная проблема философского осмысления искусственного интеллекта:
реальность создания действующей модели мышления живого человека.



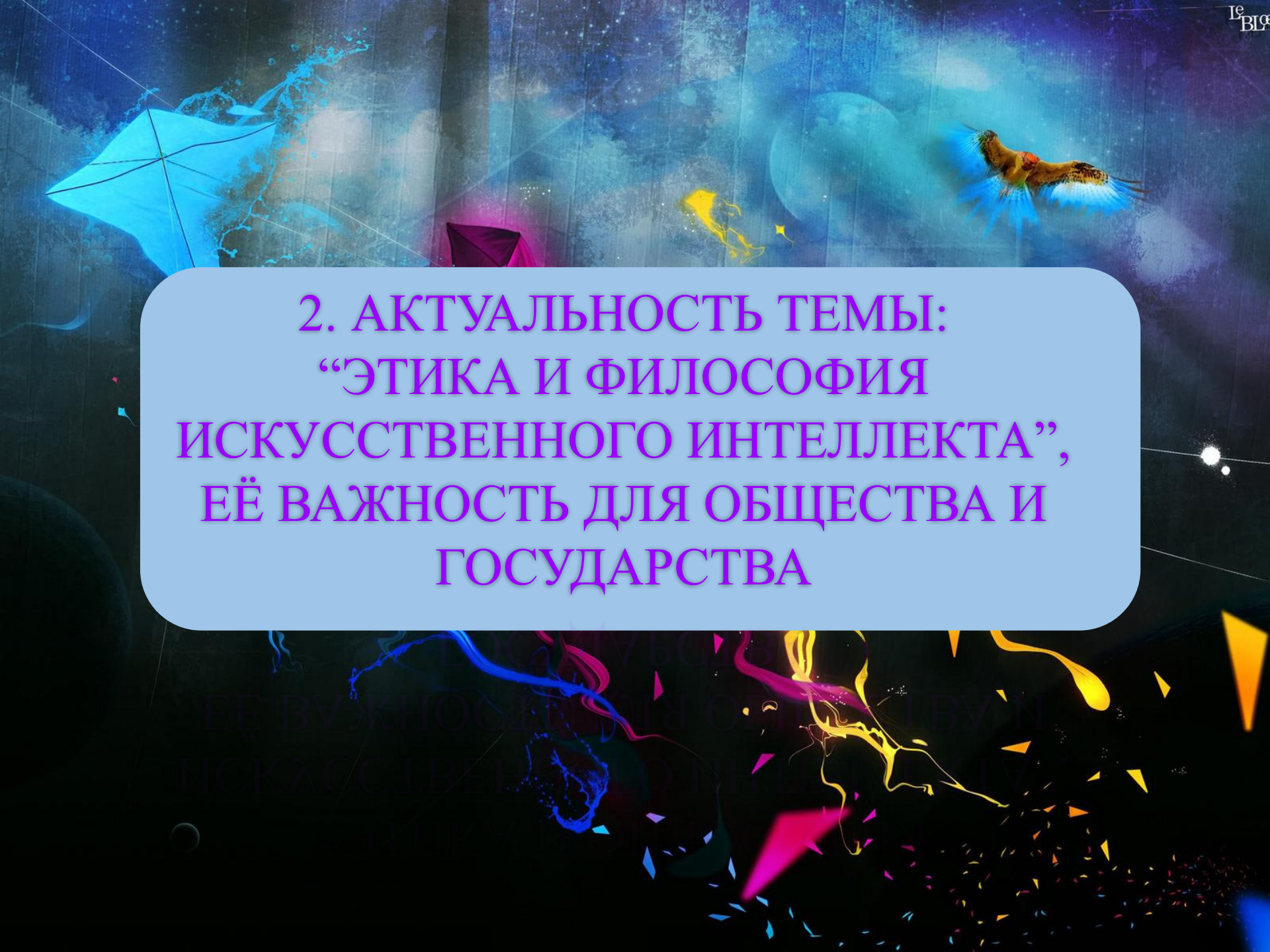
Мнение исследователей по поводу искусственного интеллекта:

- человек создан по образу и подобию бога и он, в свою очередь может создавать подобных себе;
- разум ребенка создается биологически, связан с генетикой, но обновление, углубление, расширение разума чаще связано с накоплением знания, обучением подрастающего поколения;
- пик творчества, где ранее считались главными талант, одаренность, интуиция человека, теперь связывают с нахождением наиболее оптимальных способов и алгоритмов, что можно заменить автоматическим перебором вариантов при традиционно умственно развивающих играх, в шахматы, например, или при нахождении технических и экономических решений;
- о возможностях воспроизведения мышления свидетельствует наличие компьютерных вирусов, которые нарушают существование целостных систем;
- автоматизация разумного решения интеллектуальных задач связывается с работой ЭВМ, которые представляют универсальные алгоритмы и позволяют создать многообразие программ для преобразования информации.

Философию искусственного интеллекта интересуют возможности мышления машин:

- сможет ли она решать проблемы, сознательно размышляя?
- сможет ли она проявить сознание, и даже ощутить психическое состояние, как человек?
- способна ли машина чувствовать?
- насколько мозг человека – компьютер?
- одинакова ли природа естественного и искусственного интеллекта?





2. АКТУАЛЬНОСТЬ ТЕМЫ:
“ЭТИКА И ФИЛОСОФИЯ
ИСКУССТВЕННОГО ИНТЕЛЛЕКТА”,
ЕЁ ВАЖНОСТЬ ДЛЯ ОБЩЕСТВА И
ГОСУДАРСТВА

Вопрос доверия к искусственному интеллекту

48% доверяют

42% не доверяют

Положительно воспринимают:

- 74% из сферы науки,
- 78% из промышленности

Причины недоверия:

- недостаток изученности (18%),
- ошибки и сбои (15%),
- неготовность заменить человека «машиной» (14%).



Почему искусственный интеллект несправедлив?



Предсказательные алгоритмы, которые используют полиция и суды, — популярный сюжет про конфликт технологий и прав человека. Примеры компаний: [PredPol](#), [COMPAS](#).

Почему IBM отказался делать системы распознавания лиц?

Системы распознавания лиц — одна из технологий, вызывающих самые большие споры. С каждым годом становится все больше требований общества и правозащитных организаций отказаться от таких алгоритмов.

Примеры компаний: IBM, Google, Amazon.



Почему программа в Amazon отказывала женщинам в работе?



Как алгоритмы стали расистскими? Почему они начали дискриминировать женщин при приеме на работу? Решения искусственного интеллекта хороши настолько, насколько хороши данные, на которых он обучался. Примеры компаний: Amazon.

Почему сайты знакомств устроены неправильно?

Tinder присваивает пользователям секретный индекс привлекательности, на основе которого показывает потенциальных кандидатов и тех, кто географически ближе. Алгоритм вычислял уровень доходов (ради этого искал разную информацию о пользователях в других соцсетях), уровень интеллекта (насколько пользователь умный, он выяснял на основании лексики, которую он использует в переписке).

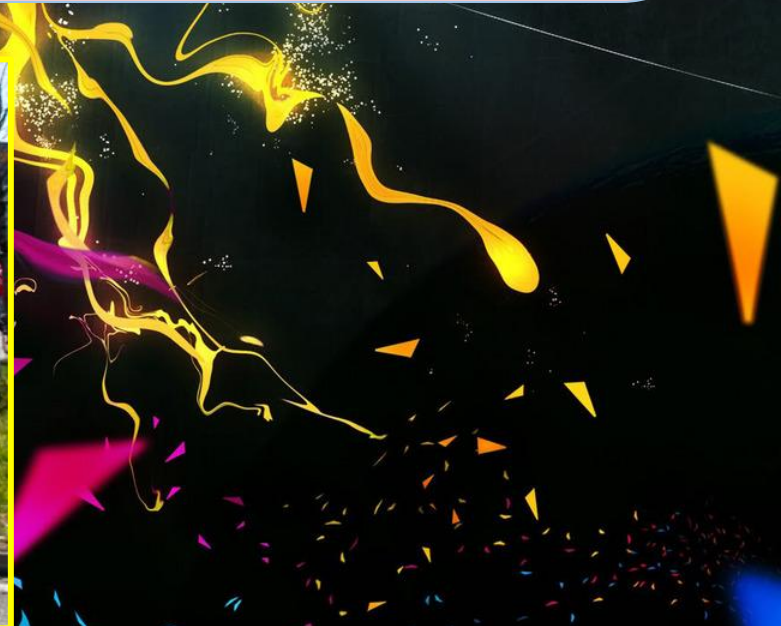
Алгоритмы-менеджеры

Часто люди боятся, что искусственный интеллект отберет у них работу. Как насчет того, что вы будете трудиться больше под присмотром программ?


Алгоритмы-менеджеры захватывают рабочие места в Китае. Правительство выдает бизнесу субсидии для перехода на цифровые решения для управления.

Технологии помогают загружать сотрудников работой по максимуму.

Примеры компаний: Amazon, IT-компании в Ханчжоу.



В чем проблема рекомендательных алгоритмов



Рекомендательные алгоритмы разносят проблемный контент (от фейков до псевдонаучной и противоправной информации)

Рекомендательные алгоритмы помещают пользователя в информационный пузырь

3. Существующее правовое регулирование

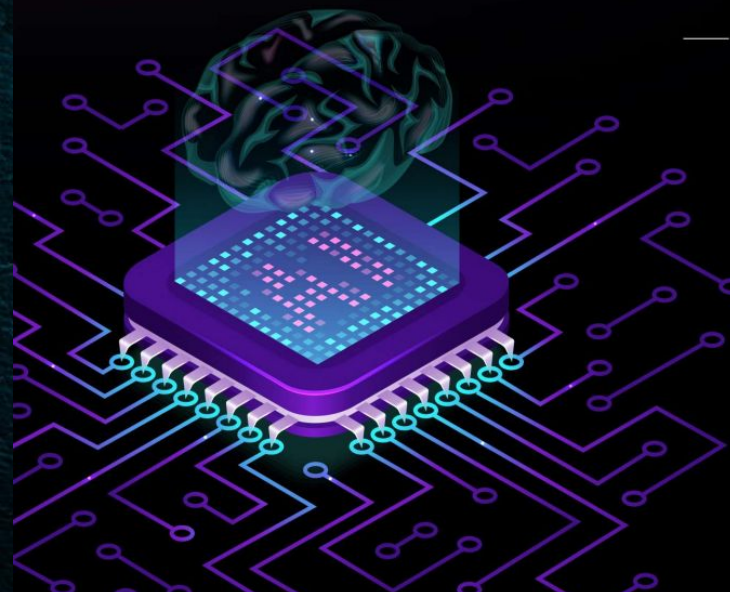
Что такое «Кодекс этики в сфере ИИ»?

— это единая система рекомендательных принципов и правил, предназначенных для создания среды доверенного развития технологий искусственного интеллекта в нашей стране

Важно!

Кодекс не определяет этику ИИ. Он помогает организовать взаимоотношение людей и компаний в связи с развитием ИИ

- Государственное регулирование уравнивается инструментами **мягкого права**
- Носит **рекомендательный характер**
- Присоединение осуществляется на **добровольной** основе
- Распространяется только на **гражданские** разработки



Из чего он состоит?

Содержание Кодекса базируется на **6 принципах**, положенных в основу **детальных рекомендаций**:



- 01** **Главный приоритет развития технологий ИИ** – защита интересов людей, отдельных групп, каждого человека.
- 02** **Необходимость осознания ответственности** при создании и использовании ИИ.
- 03** **Ответственность** за последствия применения ИИ всегда лежит на человеке.
- 04** Технологии ИИ внедрять там, где это принесёт **пользу людям**.
- 05** **Интересы развития технологий ИИ** выше интересов конкуренции.
- 06** Важна **максимальная прозрачность и правдивость** в информировании об уровне развитии технологий ИИ, их возможностях и рисках



Зачем он нужен?

01

Предоставить заинтересованным сторонам рекомендации для принятия этического решения относительно создания и использования ИИ

02

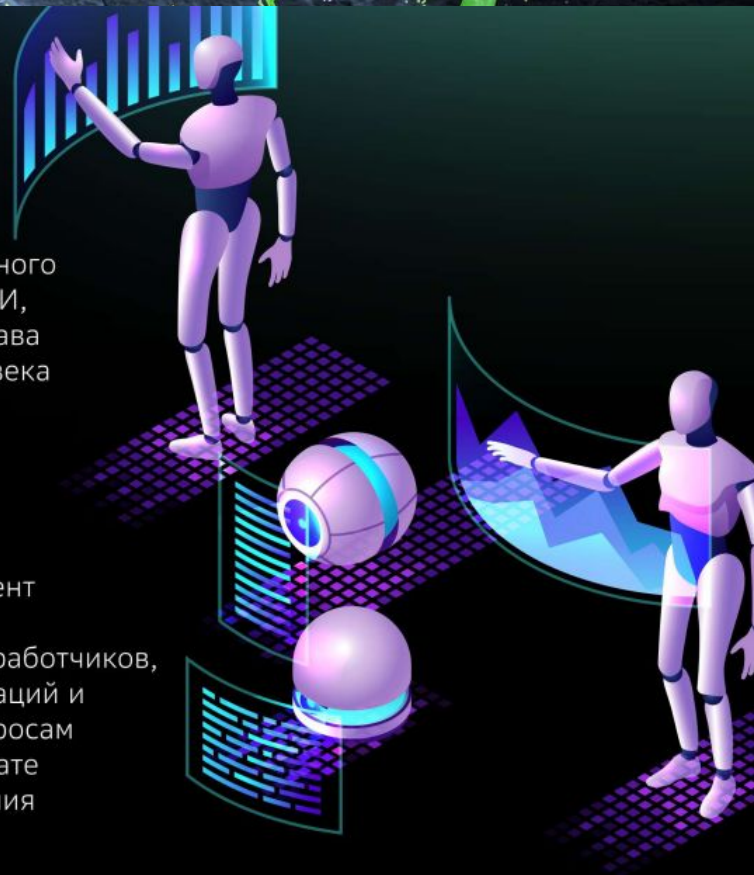
Избежать неэтичного использования ИИ, нарушающего права и интересы человека

03

Установить «мягкое» регулирование применения технологий ИИ

04

Создать инструмент взаимодействия государства, разработчиков, научных организаций и общества по вопросам этики ИИ в формате саморегулирования





Кто его создал?

Национальный кодекс этики ИИ разработан при поддержке государства крупнейшими ведущими компаниями России из Альянса в сфере ИИ (Газпромнефть, МТС, ВК, РФПИ, Сбер, Яндекс) совместно с научным сообществом и общественными институтами

Кодекс обсуждался с **более 1000 экспертов** на площадках:

- Общественной Палаты
- Совета Федерации
- АНО «Цифровая экономика»
- Аналитического центра при Правительстве РФ

**26 октября
2021 года**

Кодекс открыт к подписанию на Всероссийском форуме по этике ИИ, организованном Аналитическим Центром при Правительстве РФ



Как он реализуется?

- Акторам ИИ рекомендуется назначить **Уполномоченного по этике**, ответственного за реализацию Кодекса
- Подписанты разрабатывают **методик и руководства**, обеспечивающих соблюдение положений Кодекса
- Акторы ИИ могут создавать **Комиссии по этике** в сфере ИИ и участвуют в ее работе
- На сайте Комиссии по этике ИИ ведется **публичный Реестр** Акторов ИИ
- Акторы ИИ могут создавать публичный свод **наилучших и/или наихудших практик** решения возникающих этических вопросов в жизненном цикле ИИ



Как это работает за рубежом?



Сегодня весь мир занимается вопросами изучения и внедрения принципов этики ИИ

Принципы этики ИИ утверждаются на уровне:

Международного сообщества



ЮНЕСКО

В 2021 году 193 страны мира при активном участии России приняли первую мировую Рекомендацию об этике ИИ

Государства



и другие

Более 10 государств опубликовали национальные принципы этики в сфере ИИ

Отдельных организаций



Более 30 ведущих мировых разработчиков опубликовали корпоративные принципы этики ИИ

НКО и частных групп



Более 1000 инициатив в сфере этики ИИ создано экспертами и НКО

Российский Кодекс создавался с учетом всех мировых тенденций



Зачем присоединяться разработчику и бизнесу ИИ?



- Присоединиться к профессиональному сообществу акторов ИИ
- Получить ответы на этические дилеммы, которые возникают при создании систем ИИ
- Подтвердить ответственное отношение к разработкам в сфере ИИ и усилить репутацию надежного разработчика
- Заранее узнать мнение и опасения пользователей, общества и государства
- Увеличить степень доверия клиентов к разрабатываемым продуктам
- Донести свою позицию государству через инструменты саморегулирования



4. Сравнение с международной практикой применения правового регулирования по этике ИИ

Восток



Доверие ИИ

Запад



Недоверие ИИ



Китайский подход

White Paper on Trustworthy Artificial Intelligence

Свойства ИИ:

- он надежный и управляемый;
- его решения прозрачны и объяснимы;
- его данные защищены;
- его ответственность четко регламентирована;
- его действия справедливы и толерантны по отношению к любым сообществам.

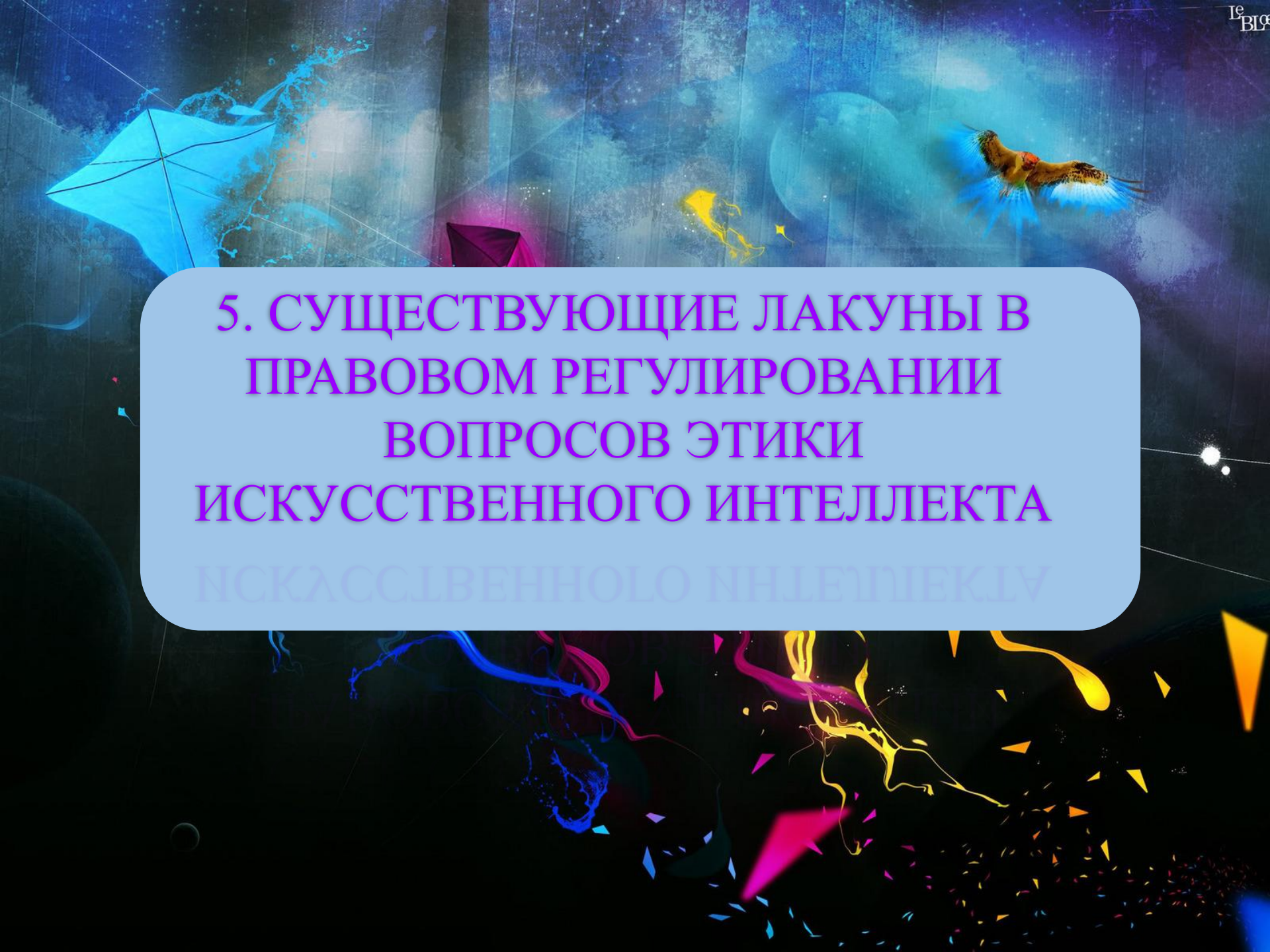
Российский подход в корне отличается от китайского, он построен по принципу “от противного”

- «Акторы ИИ должны принимать необходимые меры, направленные на сохранение автономии и свободы воли человека в принятии им решений, права выбора»;
- «Акторы ИИ должны удостовериться, что алгоритмы не влекут умышленную дискриминацию по признакам расовой, национальной, половой принадлежности, политических взглядов, религиозных убеждений, возраста, социального и экономического статуса или сведений о частной жизни»;
- «Акторам ИИ рекомендуется проводить оценку потенциальных рисков применения СИИ, включая социальные последствия для человека, общества и государства»;
- «Характер действий Акторов ИИ должен быть пропорционален оценке уровня рисков, создаваемых ИИ для интересов человека и общества»

Национальная стратегия США по лидерству в области искусственного интеллекта

Этические принципы ИИ:

- Ответственный (Responsible);
- Беспристрастный (Equitable);
- Отслеживаемый (Traceable);
- Надежный (Reliable);
- Управляемый (Governable).



5. СУЩЕСТВУЮЩИЕ ЛАКУНЫ В
ПРАВОВОМ РЕГУЛИРОВАНИИ
ВОПРОСОВ ЭТИКИ
ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Вопросы практической реализации этической компоненты ИИ становятся все более и более значимыми, среди них:

- реализация машинной этики;
- формализация этических понятий;
- верификация и валидация этической компоненты;
- стандартизация машинной этики;
- стандартизация этических аспектов ИИ.

Этика И/АС

Этичность поведения И/АС

Формализация правил и норм морали

Реализации правил морального поведения

Верификация этичности поведения

~~Проблемы и опасности применения И/АС~~

~~Вопросы профессиональной этики~~



Проблема формализации этических норм включает в себя две основные задачи:

- 1) создание форм представлений этических норм (критериев, признаков и т. п.);
- 1) выбор соответствующего математического аппарата для работы с ними: сопоставления, измерения, анализа и т. д.



Какую этику заложить в машину?

Как мы решим, что именно этично для искусственной системы в том или ином случае, а что — нет?

И по каким критериям мы будем выбирать этичные поступки для ИИ? Будет ли это мнение большинства людей, или мнение государства, например правящей партии, или мнение особых людей — моральных философов?

**В связи с культурными различиями
возникает целый ряд вопросов:**

Нормы какой культуры целесообразно закладывать в ИИ?

Нужно ли предусматривать работу ИИ в разных этических рамках в зависимости от региона применения (этическая локализация)?

Нужно ли сначала разным странам договориться о едином этическом кодексе (если вообще принципиально возможно договориться об этом)?

Основными принципами развития и использования технологий искусственного интеллекта, соблюдение которых обязательно при реализации Российской национальной стратегия ИИ, являются:

- **защита прав и свобод человека:** обеспечение защиты гарантированных российским и международным законодательством прав и свобод человека, в том числе права на труд, и предоставление гражданам возможности получать знания и приобретать навыки для успешной адаптации к условиям цифровой экономики;
- **безопасность:** недопустимость использования искусственного интеллекта в целях умышленного причинения вреда гражданам... а также предупреждение и минимизация рисков возникновения негативных последствий использования технологий искусственного интеллекта;
- **прозрачность:** объяснимость работы искусственного интеллекта и процесса достижения им результатов, недискриминационный доступ пользователей продуктов, которые созданы с использованием технологий искусственного интеллекта, к информации о применяемых в этих продуктах алгоритмах работы искусственного интеллекта...

Лакуны в правовом регулировании вопроса этики искусственного интеллекта

Существующий в РФ кодекс этики искусственного интеллекта учитывает принципы Российской национальной стратегии ИИ.

Однако, по моему мнению, он должен постоянно дорабатываться с учётом практики применения ИИ.

Это следует из того, что на сегодняшний день существует много примеров неэтичного поведения ИИ, а также неприятных ситуаций, которые происходят с конечными пользователями. До тех пор, пока будут возникать случаи неэтичного поведения ИИ, проблему этики ИИ можно считать открытой.



6. ПРЕДЛОЖЕНИЯ ПО СОВЕРШЕНСТВОВАНИЮ ЭТИКИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

ИСКУССТВЕННОГО ИНТЕЛЛЕКТА
СОВЕРШЕНСТВОВАНИЮ ЭТИКИ

Перед учеными, разработчиками, предпринимателями и государственными служащими встает вопрос: как совместить все разумные этические принципы (особенно жесткий контроль над самообучающимися системами ИИ) и не затормозить их развитие? Ответов пока нет. Сейчас, когда начинают говорить об этике, на самом деле речь идет о более масштабных вещах, чем просто этика, — о том, каким путем развивать ИИ, чтобы избежать всевозможных рисков.

Поэтому этика ИИ должна граничить с регуляторикой и быть направленной прежде всего на коммерческие компании. Этические сдвиги в их работе и отношении к пользователям не произойдут завтра — должно поменяться поколение, чтобы, например, воровство данных стало также осуждаться обществом, как воровство кошельков и женских сумок в метро.

Возможно, могут быть действенными этические принципы ИИ на уровне отрасли или компании. Например, могут быть созданы саморегулируемые организации, внутренние департаменты компаний, которые будут контролировать соблюдение разработчиками этических норм ИИ.

В целом, для разработчиков ИИ необходимо присутствие стимула для создания этического ИИ. А в этом должны быть заинтересованы как руководители компаний, занимающихся разработкой ИИ, так и общество в целом.

Спасибо за внимание!

