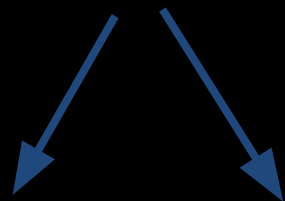




1. ЭТИКА И ФИЛОСОФИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

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

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

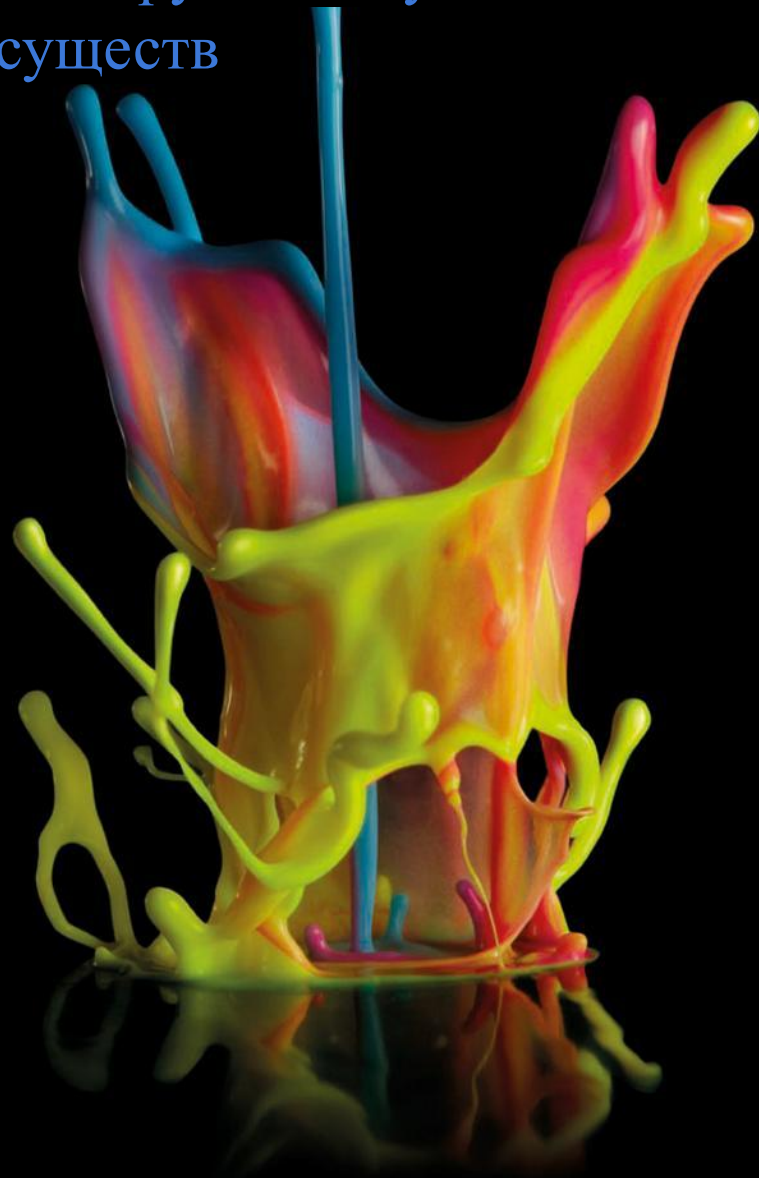


РОБОЭТИКА

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

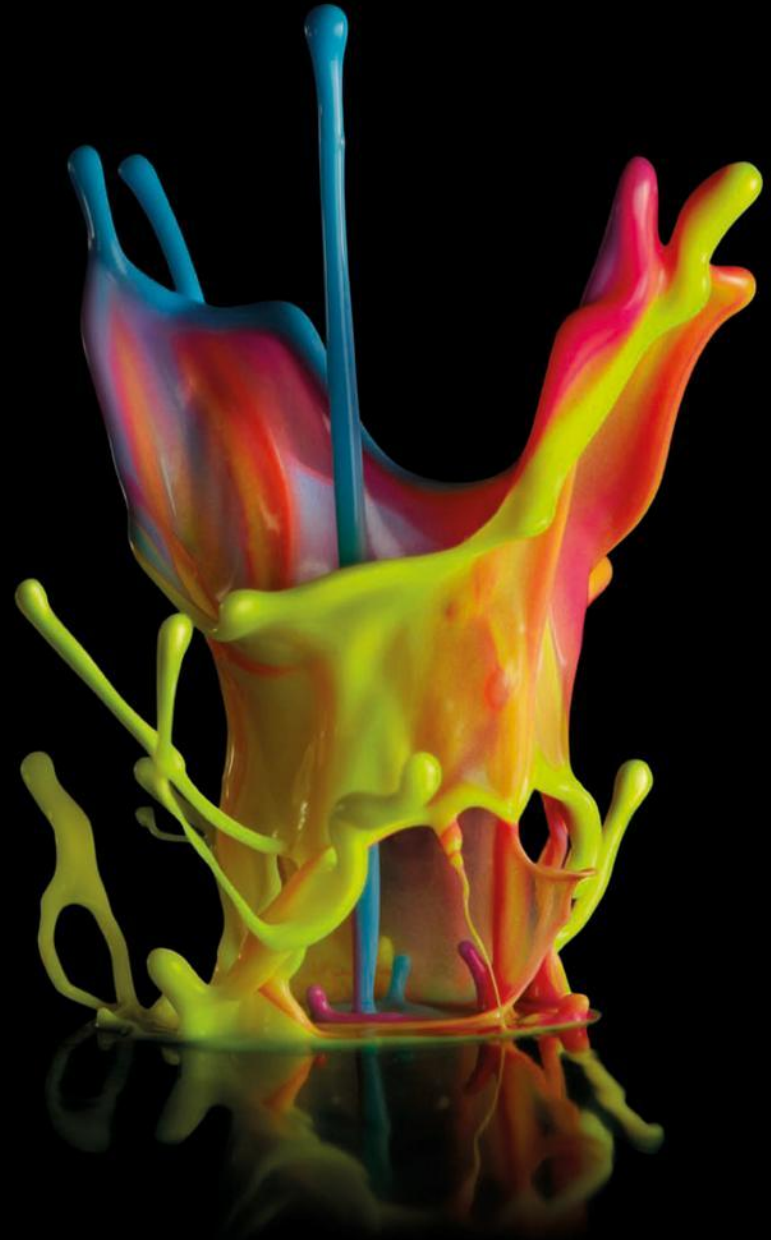


Философские размышления можно свести к двум глобальным вопросам:

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.



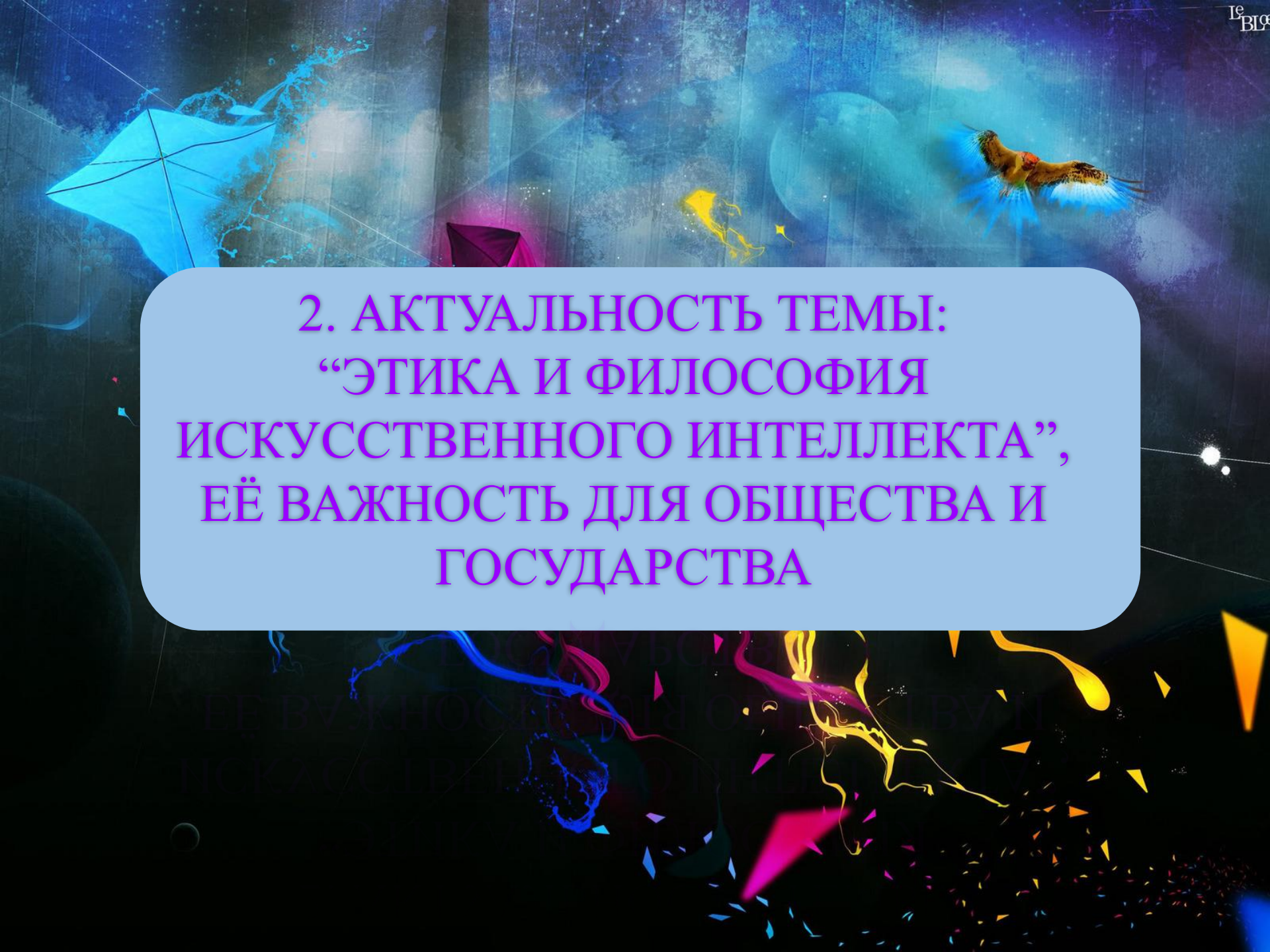
Мнение исследователей по поводу искусственного интеллекта:

- 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

Положительно воспринимают:

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

42% не доверяют

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).

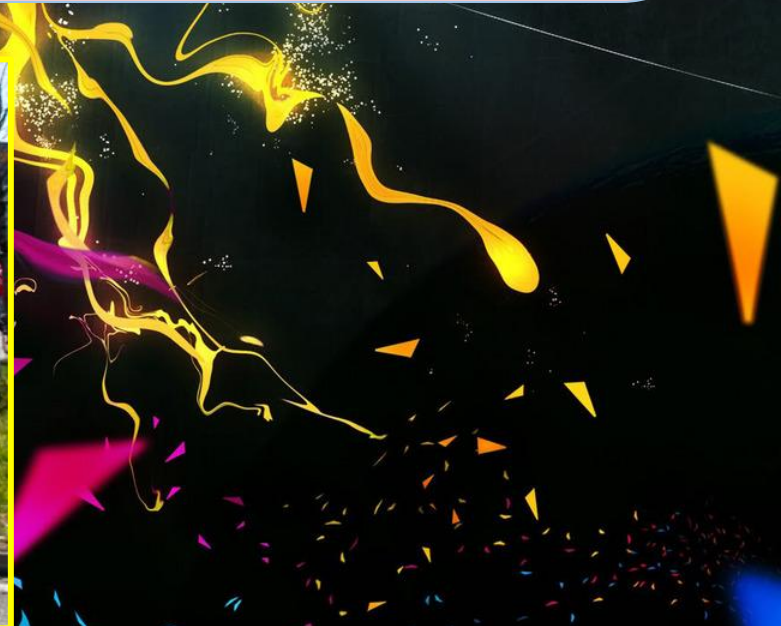
Алгоритмы-менеджеры

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

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”

- «Актеры ИИ должны принимать необходимые меры, направленные на сохранение автономии и свободы воли человека в принятии им решений, права выбора»;
- «Актеры 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).

Спасибо за внимание!

