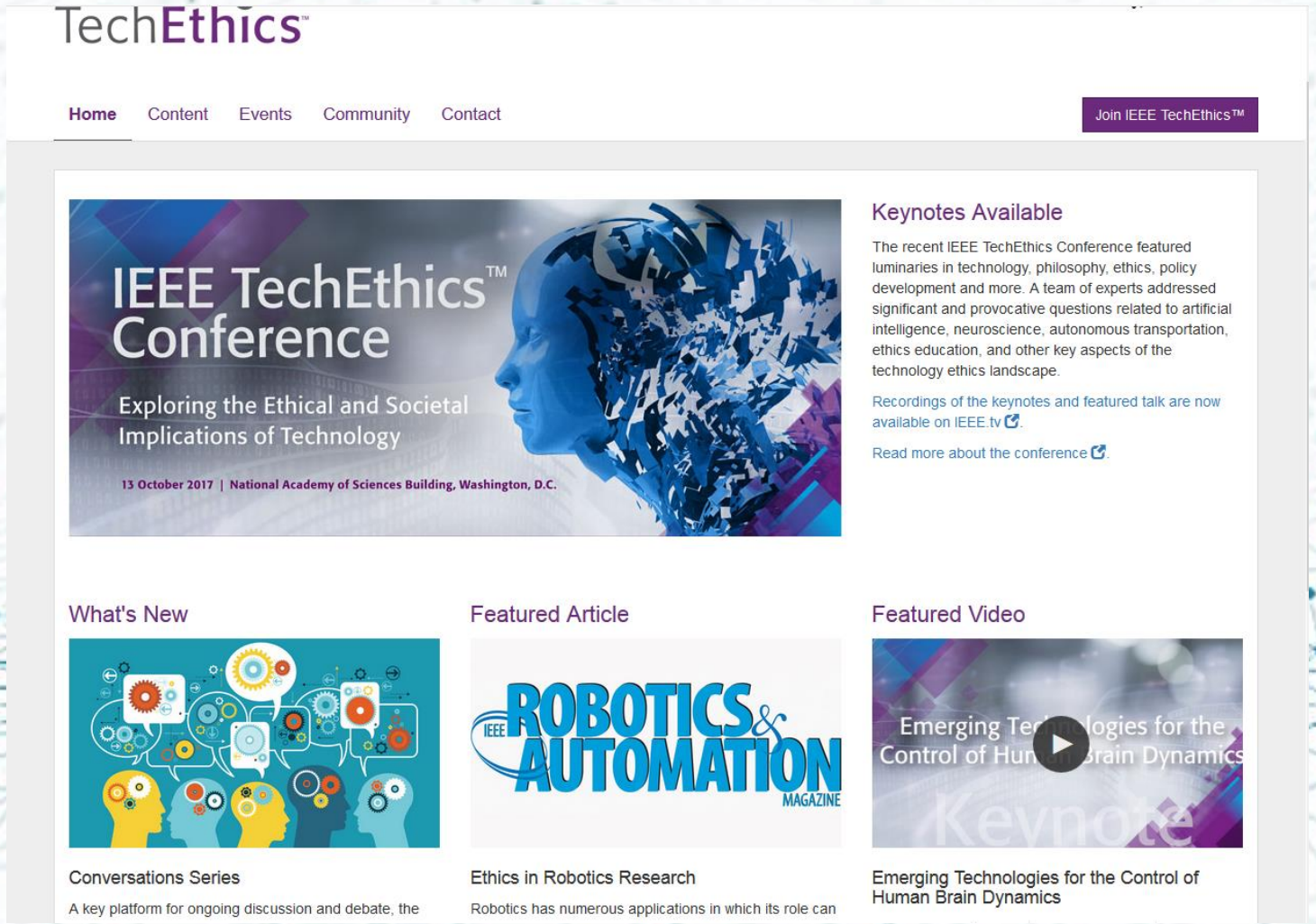
Этические аспекты применения и регулирования интеллектуальных и автономных систем

Почему этика интеллектуальных и автономных систем?

Почему этика ИИ?

Почему этика ИИ?

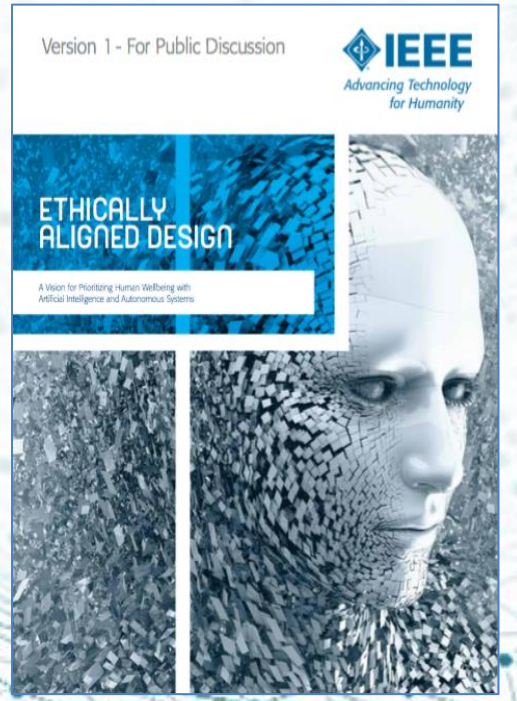
https://techethics.ieee.org/



Ethically Aligned Design.

A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems

Версия 1



- Выпущена в декабре, 2016 под лицензией Creative Commons для общего доступа;
- Разработан более чем 100 экспертами со всего мира в ходе абсолютно открытого проекта;
- Вся работа велась в рамках восьми сообществ;
- Содержит основные вызовы связанные с ИИ и автономными системами;
- Разработан по принципам технических рекомендация, для простоты использования при разработке регулирования ИИ и автономных систем;

Ethically Aligned Design.

A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems

Nation of Stated Affiliation (EAV,v1)
Australia: 4 (2.5%)
Austria: 2 (1.3%)
Canada: 5 (3.2%)
Ethiopia: 1 (0.6%)
France: 8 (5.1%)
Germany: 2 (1.3%)
Global: 2 (1.3%)
Greece: 1 (0.6%)
Hong Kong: 1 (0.6%)
India: 2 (1.3%)
Ireland: 1 (0.6%)
Italy: 3 (1.9%)
Netherlands: 6 (3.8%)
Poland: 1 (0.6%)
South Africa: 1 (0.6%)
Sweden: 1 (0.6%)
Taiwan: 1 (0.6%)
UK: 26 (16.5%)
US: 25 women, 58 men (15.9%+ 36.9%= 52.9%)

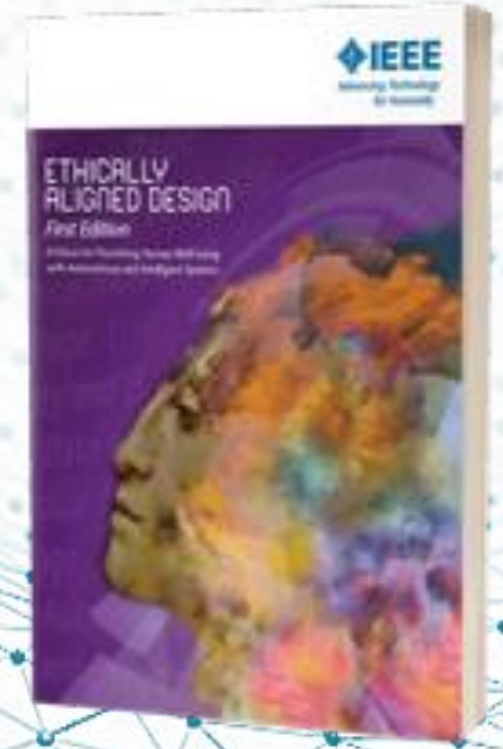
Demographics of EAD v1 Committees
Gender
Men: 108/157 (68.7%)
Women: 49/157 (31.2%)

Organization Type
Corporation: 27/157
Government: 3/157
IEEE: 11/157
NGO: 20/157
R&D: 7/157
Other: 4/ 157
University: 85/ 157

Ethically Aligned Design.

A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems

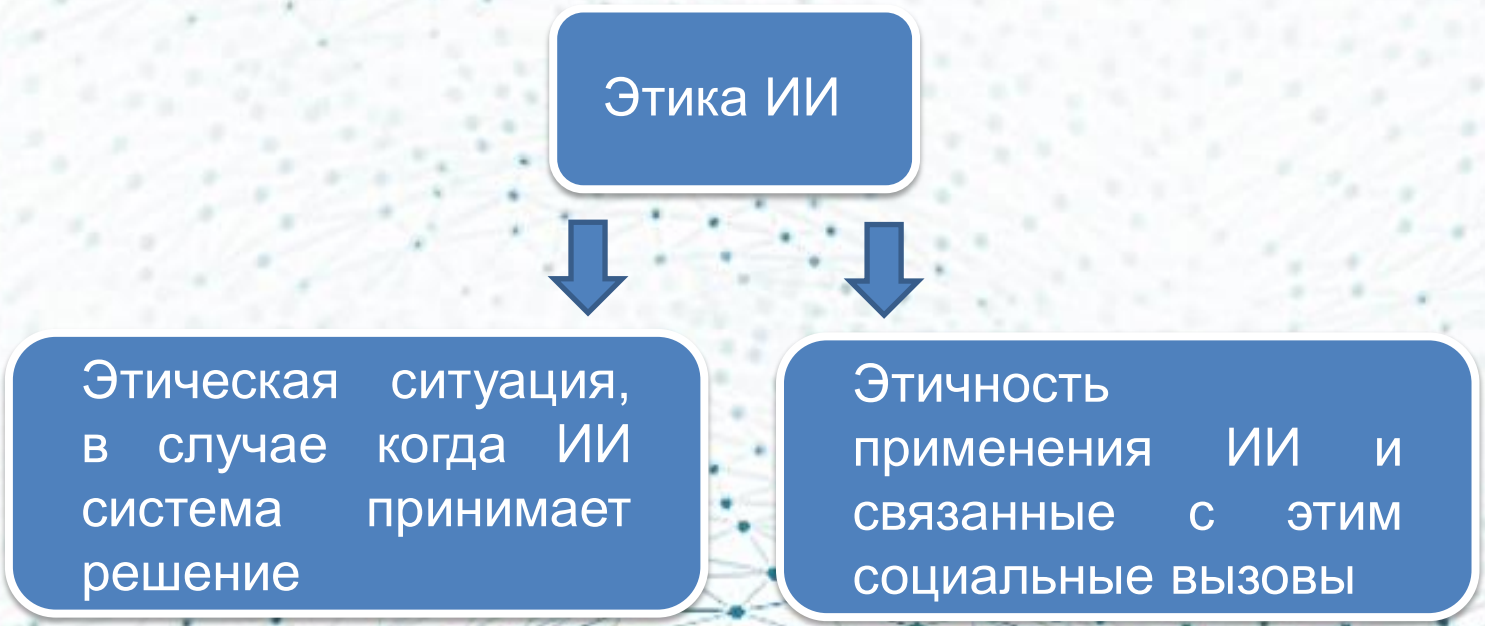
Издание 1 https://ethicsinaction.ieee.org/



- Выпущена в начале 2019 г. под лицензией Creative Commons для общего доступа;
- Разработан более чем 700 экспертами со всего мира в ходе абсолютно открытого проекта;
- Содержит основные вызовы связанные с ИИ и автономными системами;
- Разработан по принципам технических рекомендация, для простоты использования при разработке регулирования ИИ и автономных систем;

II Зачем это нужно?

Этика И/АС


Этика ИИ
Этическая ситуация, в случае когда ИИ система принимает решение
Этичность применения ИИ и связанные с этим социальные вызовы

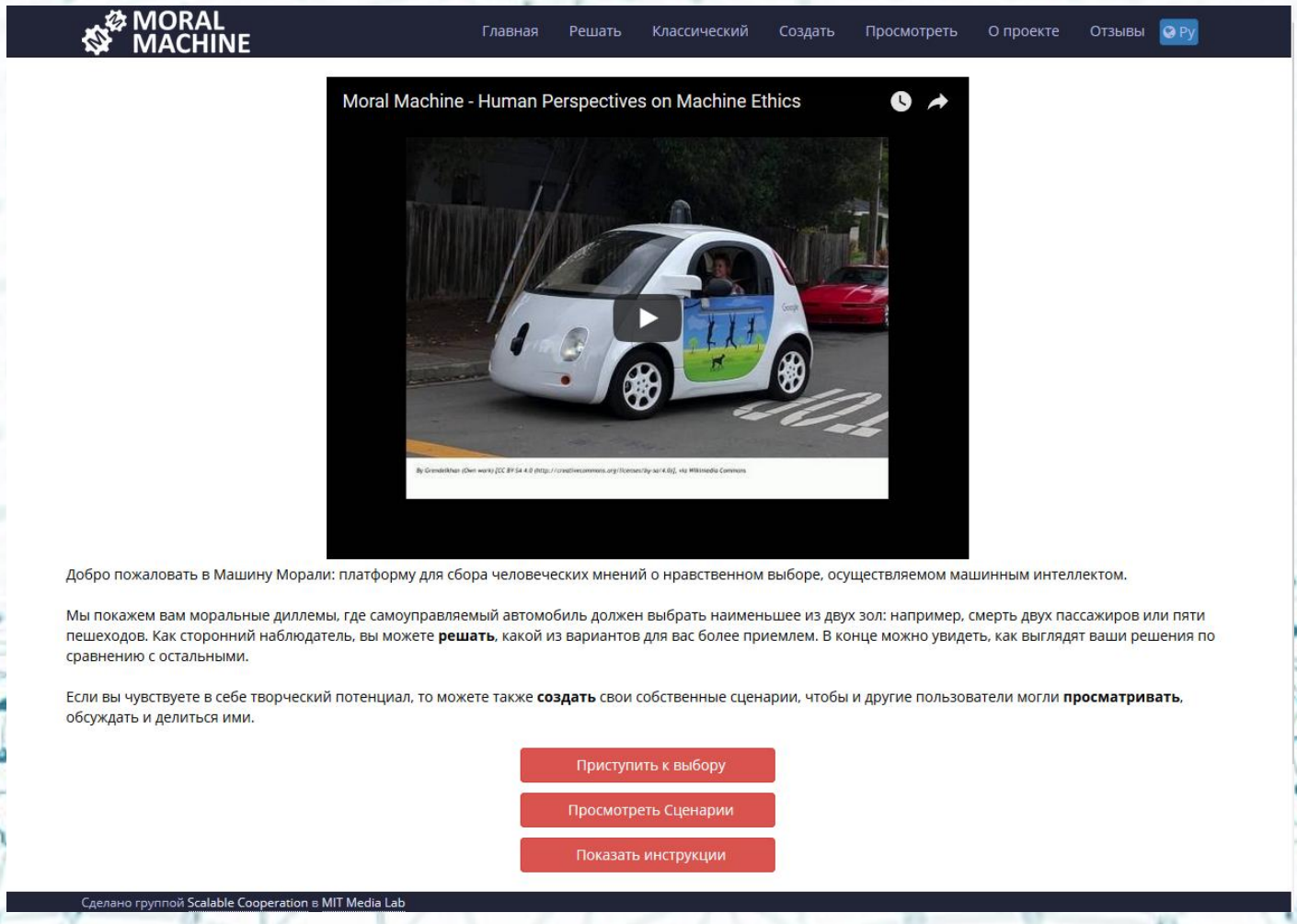
ИИ как система «принимающая решение»

Некоторые технологии на основе ИИ, которые сейчас находят применение:

- Автономные автомобили;
- Программы-советчики, поддержка операторов сложных технических систем, поддержка в принятии финансовых решений и т.д.;
- Системы извлечения и анализа метаданных в интернете;
- Системы автоматического управления на основе подходов, относящихся к ИИ, и автономные системы.

ИИ как система «принимающая решение» - автономные автомобили

http://moralmachine.mit.edu/hl/ru



ИИ как система «принимающая решение» - автономные автомобили

В каких ситуациях автономный автомобиль остановится, если на дороге случилось ДТП?

- Если среди пассажиров есть врач;
- Если рядом нет машин скорой помощи;
- Если спасателям нужна помощь, что бы извлечь пострадавших;
- ….

Нарушит ли автономная машина ПДД если:

- Рядом с дорогой происходит правонарушение;
- Если пассажиры видят, что могут оказать помощь в случае какого-либо события рядом с дорогой:
- ….

Программы - советчики

- Независимость программ советчиков?
- Как программы советчики влияют на способность пользователей принимать решения самостоятельно?
- Программа советчик vs опыт оператора?
- Кто несет ответственность если программа советчик ошиблась?
- Может ли, например, программа, анализирующая состояние потенциального заемщика, учесть особые факторы, связанные со сложной жизненной ситуацией?

Извлечение и анализ метаданных

- Метаданные могут дать информацию о человеке, которую он предпочел бы не афишировать. Насколько системы извлекающие и анализирующие их способны учитывать это?

- Как понять, этично машине ли использовать полученные метаданные без согласия человека, даже не разглашая их?

Системы автоматического управления на основе подходов, относящихся к ИИ и автономные системы

- Как этические аспекты могут быть включены в систему принимающую решение?
- Автономные системы вооружений;
- Продукция производимая машинами (от рассказов, до продукции получаемой с помощью 3D-печати).

http://shelley.ai/

Что нужно сделать?

Как отразить этические нормы в цифровых системах?

- Можно ли выделить базовые универсальные этические нормы, как обязательные для всех, и переменные, которые меняются в течении срока работы ИИ/АС или зависят от культурных традиций?
- В каких случаях какие нормы должны соблюдаться?
- Обратное влияние людей на поведение ИИ/АС?

Как отразить этические нормы в цифровых системах?

- Какие математические методы могут найти применение?

Многие работы сегодня основаны на использовании машинного обучения, т.е. попытках «научить» этике. В таких случаях ключевым вопросом является репрезентативность обучающего набора данных или говоря проще «кто и чему учит?»

Можно ли использовать логический вывод для описания универсальных этических норм?

Как отразить этические нормы в цифровых системах?

Необходимо рассмотреть возможность включения технической этики в курсы подготовки специалистов в области ИИ/АС.

«Этическое регулирование» не должно ограничивать прогресс в области разработки ИИ/АС.

Стандартизация процессов тестирования ИИ/АС, взаимодействующих с человеком с учетом этических норм.

IV Этичность применения ИИ

Рабочие места

В период бурного внедрения автоматизации и интеллектуальных систем с 1990 по 2015 гг. в США выросло число рабочих мест. Прогнозируется дальнейший рост в горизонте до 2022 г.

Изменяется структура занятости, сокращаются низкооплачиваемые рабочие места и места офисных служащих. Рост занятости в специальностях требующих творческой и интеллектуальной деятельности.

Необходима программа взаимодействия государства и частного сектора для переобучения людей, сокращенных при внедрении ИИ/АС. При этом эта программа не должна приводить к росту стоимости внедрения ИИ/АС.

Рабочие места

При анализе рабочих мест, необходимо учитывать структурные изменения связанные не только с внедрением ИИ/АС, но и с развитием других технологий, которые создают новый рабочие места. Например:

- Аддитивное производство;
- Биотехнологии, включая принципиально новые производства;
- Новые сельскохозяйственные технологии;
- Новые технологии в энергетике;
- И т.д.

V Стандарты

Стандарты в разработке

IEEE P7006: Standard for Personal Data Artificial Intelligence (AI) Agent

Проект стандарта, определяющего обращение с персональными данными с использованием ИИ, включая получение метаданных. Создается в координации со стандартами IEEE P7005 и IEEE P7004 регламентирующими обращение с персональными данными сотрудников организаций и детей и студентов соответственно, в том числе с использованием автоматизированных систем.

IEEE P7007: Ontological Standard for Ethically Driven Robotics and Automation Systems

Проект стандарта, который задает набор базовых онтологий для использования при разработке АС.

IEEE P7008: Standard for Ethically Driven Nudging for Robotic, Intelligent and Autonomous Systems

Проект стандарта, который регламентирует подходы к разработке этически обоснованных взаимодействий между человеком и ИИ/АС, в части повседневного взаимодействия.

IEEE P7009: Standard for Fail-Safe Design of Autonomous and Semi-Autonomous Systems

Стандарт регламентирует стадии разработки и тестирования АС, направленные на исключение неполадок, в том числе и ведущих к негативным последствиям при взаимодействии с людьми.

IEEE P7010 : Wellbeing Metrics Standard for Ethical Artificial Intelligence and Autonomous Systems

Определяют метрики и базовые уровни описывающие уровень этически обоснованного взаимодействия между человеком и ИИ/АС.

Спасибо за внимание!