

Этические рекомендации для этически обоснованного проектирования

Концепция приоритезации жизнедеятельности людей при взаимодействии с и автономными и интеллектуальными системами

## Введение

Поскольку использование и влияние автономных и интеллектуальных систем (А/ИС) становятся все более распространенными, нам необходимо установить общественные и политические руководящие принципы, чтобы такие системы оставались человеко-направленными, соответствуя ценностям и этическим принципам человечества. Эти системы должны вести себя так, что это выгодно для людей помимо соответствия функциональным требованиям и решения технических проблем. Это позволит повысить уровень доверия между людьми и технологиями, который необходим для их плодотворного, повсеместного использования в нашей повседневной жизни.

Чтобы иметь возможность внести свой вклад позитивным и недогматическим путем, мы, научно-научные сообщества, должны усилить нашу самокритику, нам нужно иметь открытую и честную дискуссию вокруг наших мнимых, наших наборов явных или неявных ценностей, наших институтов, символов и представлений.

Эвдаймония, как объяснял Аристотель, - это практика, которая определяет благополучие людей как высшей добродетели для общества. Примерно переведенная как «процветание», польза эвдаймонии начинается с осознанного созерцания, где этические соображения помогают нам определить, как мы хотим жить.

Являются ли наши этические практики западными (Аристотелевскими, Кантианскими), восточными (Синтоистскими, Конфуцианскими), африканскими (Убунту) или других традиций, при создании автономных и интеллектуальных систем, которые в явной форме уважают неотъемлемые права человека и выгодные ценности пользователей, мы можем приоритезировать повышение благосостояния людей как показатель прогресса для алгоритмической эпохи. Измерение и признание потенциала целостного экономического процветания должно стать более важным, чем достижение одномерных целей, таких как повышение производительности или рост ВВП.

## **Миссия Глобальной Инициативы IEEE по Этике Автономных и интеллектуальных Систем**

**Для обеспечения того, чтобы все заинтересованные лица, участвующие в проектировании и развитии автономных и интеллектуальных систем, получили образование, подготовку и имели право приоритизировать этические соображения, чтобы эти технологии были продвинуты на благо человечества.**

Под «заинтересованным лицом» мы подразумеваем всех, кто участвует в исследованиях, проектировании, изготовлении или обмене сообщениями с интеллектуальными и автономными системами, включая университеты, организации, правительства и корпорации, которые делают эти технологии реальностью для общества.

Наша цель состоит в том, чтобы Ethically Aligned Design предоставил информацию и рекомендации, которые будут служить ключевым источником для работы технологов в смежных областях науки и техники в ближайшие годы. Для достижения этой цели в текущей версии Ethically Aligned Design (EAD2v2) мы определяем соответствующие «Проблемы» и «Рекомендации кандидатов», которые, как мы надеемся, будут способствовать формированию национальной и глобальной политики, согласующейся с этими принципами.

Глобальная Инициатива IEEE объединяет несколько сотен участников с шести континентов, которые являются лидерами из академических кругов, промышленности, гражданского общества, политики и правительства в соответствующих технических и гуманитарных дисциплинах для выявления и нахождения консенсуса по актуальным вопросам.

Вторая цель Глобальной инициативы IEEE заключается в предоставлении рекомендаций для стандартов IEEE на основе этически обоснованного дизайна. Этически обоснованный дизайн (v1 и v2) и члены Глобальной инициативы IEEE являются вдохновителями набора рабочих групп стандартов IEEE P7000™, которые являются бесплатными и открытыми для всех, кто присоединится к ним.

## О нас

Глобальная инициатива IEEE по этическим рекомендациям для искусственного интеллекта и автономных систем («IEEE Global Initiative») является программой Института Электротехники и Электроники («IEEE»), крупнейшей в мире технической профессиональной организацией, посвященной продвижению технологий на благо человечества с более чем 420 000 членами в более чем 160 странах.

Глобальная инициатива IEEE обеспечивает возможность объединить многочисленные мнения сообществ искусственного интеллекта и автономных систем для выявления и нахождения консенсуса по актуальным вопросам.

IEEE разработает рекомендации для этически обоснованного проектирования (Ethically Aligned Design EAD), доступны под некоммерческой лицензией Соединенных Штатов Америки Creative Commons Attribution-Non-Commercial 3.0.

В соответствии с условиями этой лицензии организации или отдельные лица могут принять аспекты этой работы на их усмотрение в любое время. Также ожидается, что содержание и предмет проблемы Этически обоснованного проектирования будут выбраны для ввода в формальные процессы IEEE, в том числе для разработки стандартов.

Глобальная инициатива IEEE и этические рекомендации способствуют значительным усилиям, предпринимаемым в IEEE для открытого, широкого и содержательного разговора об этике в технологиях, известной как программа IEEE TechEthics™.

## Миссия Глобальной инициативы IEEE

Каждый научно-технический работник должен быть обучен, подготовлен и мотивирован делать приоритетными этические соображения при разработке и развитии автономных и интеллектуальных систем.

Под «научно-техническим работником» мы подразумеваем любого, кто участвует в исследованиях, разработке, производстве или обмене сообщениями в области ИИ/АС, включая университеты, организации и корпорации, претворяющие эти технологии в реальность для общества.

Этот документ представляет собой коллективный труд более ста человек с мировым именем в областях искусственного интеллекта, права и этики, философии и политики в сфере научных сообществ, науки, правительственных и корпоративных секторов. Наша цель состоит в том, чтобы рекомендации для этически обоснованного проектирования могли предоставить информацию и предложения от этих специалистов, обеспечивающих ключевые компетенции для работы научно-технических сотрудников ИИ/АС в ближайшие годы. Для достижения этой цели в текущей версии рекомендаций для этически обоснованного проектирования (ЭОП версия1) мы определяем проблемы и рекомендации для специалистов в областях, охватывающих искусственный интеллект и автономные системы.

Вторая цель Глобальной инициативы IEEE - предоставлять рекомендации для стандартов IEEE, основанных на ЭОП. IEEE P7000™ - процесс моделирования для решения этических проблем в процессе проектирования системы был первым проектом стандарта IEEE (утвержден и находится в разработке), вдохновленным инициативой. Два других проекта по стандартизации, IEEE P7001™ – информационная открытость автономных систем и IEEE P7002™ – процесс сохранения конфиденциальности данных, были одобрены, что демонстрирует общественное влияние инициативы на вопросы этики ИИ/АС.

## Этически обоснованное проектирование v2 – Обзор

### 1. Намерение

Интеллектуальные и автономные технические системы специально разработаны для сокращения участия человека в повседневной жизни. При этом новые области вызывают озабоченность по поводу их воздействия на отдельных лиц и общества. Текущие обсуждения включают пропаганду позитивного воздействия, а также предупреждения, основанные на потенциальном ущербе, наносимом конфиденциальности, дискриминации, утрате навыков, экономических последствиях, безопасности критических инфраструктур и долгосрочных эффектов для социального благосостояния. В силу своей природы, полная польза от этих технологий будет достигнута только в том случае, если они будут соответствовать нашим определенным ценностям и этическим принципам. Поэтому мы должны установить парадигмы, которые будут направлять развитие и информировать, вести диалог и обсуждать нетехнические последствия этих технологий.

### 2. Цели

Этическое проектирование, разработка и внедрение этих технологий должны руководствоваться следующими Общими принципами:

- Права человека: обеспечить, чтобы они не ущемляли международно-признанные права человека
- Благополучие: определение приоритетов показателей благосостояния при разработке и использовании
- Подотчетность: обеспечить, чтобы проектировщики и операторы были ответственны и подотчетны
- Прозрачность: обеспечить прозрачность работы
- Осознание неправильного использования: свести к минимуму риски неправильного использования

### 3. Требования

#### **Права личных данных и индивидуальный контроль доступа**

Основное требование заключается в том, что люди имеют право определять доступ и предоставлять информированное согласие в отношении использования своих персональных цифровых данных. Отдельные лица нуждаются в механизмах, помогающих выявлять их уникальную личность и личные данные в сочетании с политикой и практикой, которые делают их явным образом осведомленными о последствиях, возникающих в результате объединения или перепродажи их личной информации.