

СОГЛАСОВАНИЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА И ГЛАВНЫЙ НЕ ЗАДАННЫЙ ВОПРОС

Ради чего должно развиваться человечество?

За последние годы проблема согласования искусственного интеллекта стала одной из центральных тем научных исследований.

Как сделать ИИ безопасным?

Как исключить нежелательные последствия?

Как обеспечить соответствие его действий интересам человека?

Эти вопросы чрезвычайно важны.

Однако все они опираются на предположение, которое почти никогда не обсуждается.

Предполагается, что человечество уже знает, каким должно быть его будущее.

Но так ли это?

Что означает «согласованный искусственный интеллект»?

В большинстве современных исследований согласование понимается как способность интеллектуальной системы действовать в соответствии с человеческими целями и ценностями.

Но возникает следующий вопрос.

1. Каким именно человеческим целям?
2. Целям отдельного человека?
3. Компании?
4. Государства?
5. Военного союза?
6. Экономической системы?
7. Религиозной традиции?
8. Всего человечества?

Ответа на этот вопрос пока не существует.

Потому что современный мир не обладает единым представлением о собственном будущем.

Искусственный интеллект не способен выбрать цель вместо человека

Любая интеллектуальная система может эффективно искать решения.

Но она не способна самостоятельно определить, какие цели являются правильными.

Даже самый совершенный интеллект остается инструментом по отношению к тем смыслам, которые ему задает человек.

Поэтому вопрос согласования ИИ неизбежно превращается в вопрос согласования самого человечества.

Если люди расходятся в понимании будущего, невозможно ожидать, что искусственный интеллект сможет устранить это противоречие.

Он лишь будет работать в рамках поставленных ему целей.

Главный не заданный вопрос

Сегодня обсуждается множество вопросов.

1. Как сделать ИИ безопасным?
2. Как избежать потери контроля?
3. Как распределить ответственность?
4. Как предотвратить злоупотребления?

Но существует вопрос, без ответа на который все остальные остаются неполными.

Ради какого будущего должно развиваться человечество?

Пока этот вопрос остается без общего обсуждения, невозможно окончательно ответить и на вопрос о целях искусственного интеллекта.

Почему это становится критически важным именно сейчас

На протяжении большей части истории последствия человеческих ошибок были ограничены масштабами отдельных государств или поколений.

Сегодня ситуация изменилась.

Искусственный интеллект, биотехнологии, глобальные информационные сети, автономные системы и другие технологии способны влиять на судьбу всей цивилизации.

Чем мощнее становятся инструменты, тем важнее становится понимание целей, ради которых они используются.

Технологическое могущество больше не может развиваться отдельно от зрелости человеческого мышления.

От согласования машин — к согласованию смыслов

Современная дискуссия в основном сосредоточена на согласовании поведения искусственного интеллекта.

Но, возможно, перед человечеством стоит более широкая задача.

Не только согласовать машины с человеком.

Но и согласовать человеческие представления о будущем.

Это не означает необходимость полного единомыслия.

Люди всегда будут отличаться культурами, языками, традициями и убеждениями.

Однако без минимального согласия относительно фундаментальных целей развития невозможно создать устойчивую цивилизацию.

Общая цель как условие безопасного развития технологий

Любая сложная система нуждается в направлении.

Это относится к государствам.

К экономике.

К науке.

И тем более — к искусственному интеллекту.

Поэтому вопрос о будущем человечества перестает быть исключительно философским.

Он становится практической задачей, определяющей развитие науки, технологий и международного сотрудничества.

Искусственный интеллект как участник нового этапа цивилизации

ИИ способен помочь человечеству увидеть взаимосвязи, которые невозможно охватить отдельному человеку.

1. Он способен ускорить научные открытия.
2. Повысить качество управления.
3. Расширить доступ к знаниям.
4. Содействовать поиску решений сложнейших глобальных задач.

Но для реализации этого потенциала необходимо, чтобы развитие ИИ происходило в контексте более широкого проекта — проекта развития самого человечества.

Не искусственный интеллект должен определять направление цивилизации.

Человечество должно осознанно определить направление своего развития и сделать искусственный интеллект одним из инструментов его достижения.

Приглашение к новой постановке вопроса

Прежде чем спрашивать, каким должен быть искусственный интеллект, необходимо задать вопрос:

Каким должно стать человечество?

Именно этот вопрос, на наш взгляд, способен стать общей точкой пересечения науки, философии, технологий, образования, культуры и международного сотрудничества.

Ответ на него невозможно найти усилиями одной дисциплины, одной страны или одной мировоззренческой традиции.

Он требует открытого глобального диалога, в котором развитие технологий будет рассматриваться как часть более широкого процесса — развития человеческой цивилизации.