

Татьяна Ивановна Врублевская-Токер

канд. филос. наук, доц. каф. управления и социальной работы

Филиала Российского государственного социального университета в г. Минске

Tatiana Vroublevskaya-Toker

Candidate of Philosophical Sciences,

Associate Professor of the Department of Management and Social Work

of Minsk Filial Branch of Russian State Social University

e-mail: VroublevskaiaTI@rgsu.net

ТЕОРЕТИЧЕСКИЕ ОСНОВАНИЯ ЭТИКИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В РАБОТАХ Н. ВИНЕРА

Рассматривается кибернетический подход американского ученого и философа, основателя кибернетики Норберта Винера как источник теоретических оснований этики искусственного интеллекта. Изучается проблема определения теоретических оснований этики искусственного интеллекта, которая складывается как прикладная наука и ориентирована на отбор лучших практик использования современных технологий. Цель статьи – представить кибернетический подход Н. Винера как объяснительную этическую модель, одно из теоретических оснований этики искусственного интеллекта. Показано, что система понятий кибернетического подхода (информация, обратная связь, неопределенность, энтропия) определяет содержание этических понятий в рамках этики искусственного интеллекта, задает поле нравственных отношений между человеком и технологией как коммуникацию, направленную против беспорядка и дезорганизации, участники которой стремятся наладить ее наилучшим образом.

Ключевые слова: этика, искусственный интеллект, прикладная этика, кибернетический подход, Н. Винер.

Theoretical Foundations of the Ethics of Artificial Intelligence in the Works of N. Wiener

The article examines the cybernetic approach of the American scientist and philosopher, the founder of cybernetics, Norbert Wiener, as a source of theoretical foundations for the ethics of artificial intelligence. The problem of defining the theoretical foundations of the ethics of artificial intelligence, which is developing as an applied science and is focused on selecting the best practices for using modern technologies, is studied. The purpose of the article is to present the cybernetic approach of N. Wiener as an explanatory ethical model, one of the theoretical foundations of the ethics of artificial intelligence. It is shown that the system of concepts of the cybernetic approach (information, feedback, uncertainty, entropy) determines the content of ethical concepts within the framework of the ethics of artificial intelligence, sets the field of moral relations between man and technology as communication directed against disorder and disorganization, the participants of which strive to establish it in the best possible way.

Key words: ethics, artificial intelligence, applied ethics, cybernetic approach, N. Wiener.

Введение

Предмет настоящей статьи представляет теоретический и прикладной интерес для этики и других наук, решающих проблему связи между технологией искусственного интеллекта и процессами в жизни общества. Стремительное освоение технологий все новых сфер общественного производства и деятельности выглядит поразному с точки зрения их практики и теоретического осмысления в философии, этике, психологии, социологии и других науках, изучающих «человеческое в человеке». В первом случае проблема сущего и должного рассматривается в категориях совершенствования качества, сокращения

времени совершения операций, расширения технических и творческих возможностей человека. Предполагается, что новшества ведут к улучшению результатов общественного производства и деятельности, а также жизни человека в целом. Во втором случае гуманитарное знание не может вынести оценку результатам прогресса технологий, пока не будет установлено, какая именно реальность при этом возникает, в каких новых понятиях осмысливается то, что создано при помощи технологий. Речь идет не только о продукте, но и о новых видах связей, возникающих по ходу и в результате этого процесса.

Исследования проблемы связи между технологией искусственного интеллекта и общественными процессами сегодня масштабны, многосторонни и наполнены идеями. Гуманитарные науки сегодня видят искусственный интеллект предметом трансдисциплинарным, входящим в понятийное поле философии познания, прикладной этики, философии науки и техники, кибернетики [1, с. 72]. Дискуссия вокруг искусственного интеллекта является продолжением исследований компьютерной и инженерной этики, начатых в середине XX в. работах основателя кибернетики Норберта Винера (1894–1964). Он предполагал развитие технологии машинного обучения, теоретически возможного уже в то время, и гуманитарные проблемы, связанные с ним.

Цель статьи – представить кибернетический подход Н. Винера как объяснительную этическую модель, одно из теоретических оснований этики искусственного интеллекта.

Задачи статьи – охарактеризовать систему понятий, составляющих кибернетический подход Н. Винера, определить его этические доминанты, выявить значение кибернетического подхода в ряду теоретических оснований этики искусственного интеллекта.

Материалы и методы исследования

Перед нами три самые крупные работы Н. Винера, в которых излагается предмет его научного и практического интереса. Это книги «Кибернетика, или Управление и связь в животном и машине» [2], «Кибернетика и общество» [3], а также изданная намного позже книга «Творец и робот» [4] с подзаголовком «Обсуждение некоторых проблем, в которых кибернетика сталкивается с религией».

В первой представлены теоретические основы созданного Н. Винером направления мысли – кибернетики, задуманной как наука, чей предмет – коммуникация и управление в системах с обратной связью, т. е. в машинах и живых организмах [4, с. 16], и соответствующая методология. Две другие работы содержат авторское развитие многих проблем, затронутых в первой работе, их обсуждение в контексте нового опыта, полученного в ходе практического развития кибернетики.

Методами исследования являются контент-анализ, компаративный метод, метод мысленного моделирования.

Ход исследования

Свой подход к науке Н. Винер называл не догматичным и чуждым универсальных объяснительных моделей. Более того, появление всякой научной теории он связывал с развитием человеческого опыта и мысли. «Проблема роли науки в обществе представляется мне тесно связанной с проблемой роли чувственного опыта и мысли в жизни индивида... В основе своей она подобна той роли, какую выполняет гомеостазис в поддержании определенного рода динамического равновесия между индивидом и окружающим миром» [5, с. 117]. В статье «Наука и общество» он отмечал, что все, в т. ч. и научные обобщения, производится для того, чтобы возникло «определенного рода динамическое равновесие, которое благоприятствует продолжению нашего существования как отдельных людей и как человеческой расы в условиях постоянно изменяющейся среды» [5, с. 118]. С одной стороны, этот подход методологически продолжает традицию сенсуализма. Винер пишет о том, что главным средством познания окружающей среды являются органы чувств, так что продолжение человеческого существования определяется массой впечатлений, которые формируют отношение индивида к окружающей среде и комбинируются в соответствующих действиях [5, с. 118]. С другой стороны, этот подход ориентирован надысторически. Винер пишет: «Цель науки в обществе состоит в том, чтобы позволить нам гомеостатично реагировать на превратности будущего... которое движется вперед во времени вместе с прогрессом нашего опыта» [5, с. 120]. Он также полагает, что «внутренняя жизнь науки не должна находиться в слишком прямой зависимости от политики момента или от официального образа мысли... В качестве условий эффективной деятельности у ученого должно оставаться нечто – не слишком много – от позиции “башни из слоновой кости”, которую сейчас так модно разоблачать» [5, с. 120].

Две эти позиции – сенсуалистичность и надысторичность – отражают сущность кибернетического подхода Винера, который

он применяет весьма широко и к области естественно-научного, и гуманитарного знания. Сенсуалистская методология придает подходу динамичность, она позволяет рассмотреть явления и процессы любой реальности как такие, в которых возникающие связи определяют буквально все.

Надисторическая ориентация позволяет кибернетическому подходу быть принципиально незамкнутым и объяснять любые типы реальности, которые можно помыслить в его базовых категориях: «автоматы и физиологические системы можно охватить одной теорией» [2, с. 96].

Соответственно, как Винер сформулировал комплекс идеалов, принципов и норм компьютерной этики, так с помощью кибернетического подхода можно сформулировать и теоретические основания этики искусственного интеллекта. В настоящее время этика искусственного интеллекта существует как вид прикладной этики, создавая моральные практики для комфортного взаимодействия человека и высоких технологий. Ее теоретическими основаниями являются понятия человека и машины (технологии), представление о связи между ними, о должном и правильном в этой связи и ее результатах.

Система категорий кибернетического подхода включает понятия неопределенности, энтропии, информации, обратной связи. Их содержание составляют предельно обобщенные идеи, что позволяет понимать в данных терминах и физические, и духовные явления, и процессы.

Понятие неопределенности поддерживает идею незамкнутости, надисторичности кибернетического подхода, а также является основой объединения отдельных направлений и исторических периодов знания: Винер (с рядом оговорок) представляет разные версии индетерминизма как стороны общего признания преобладания случайности в процессах и явлениях. «Это признание наличия в мире элемента неполного детерминизма, почти иррациональность, в известной степени аналогично фрейдовскому допущению глубоко иррационального компонента в поведении и мышлении человека. В современном мире политической, а также интеллектуальной неразберихи естественно объединить в одну группу Гиббса, Зигмунда Фрейда и сторон-

ников современной теории вероятностей как представителей единой тенденции... Пропасть между образом мышления Гиббса – Лебега и интуитивным, однако в некотором отношении дискурсивным методом Фрейда слишком велика. Все же в своем признании случая как основного элемента в строении самой Вселенной эти ученые очень близки друг к другу, а также традиции св. Августина. Ибо этот элемент случайности, это органическое несовершенство можно рассматривать, не прибегая к сильным выражениям, как зло – негативное зло, которое св. Августин охарактеризовал как несовершенство, а не как позитивное предумышленное зло манихейцев» [3, с. 23].

Понятие энтропии в кибернетическом подходе тесно связано с понятием неопределенности. Винер берет его в том значении, которым оно обладает в физике: «Нововведение Гиббса состояло в том, что он стал рассматривать не один мир, а все те миры, где можно найти ответы на ограниченный круг вопросов о нашей среде. В центре его внимания стоял вопрос о степени, до которой ответы относительно одного ряда миров будут вероятны по отношению к другому, более широкому ряду миров. Кроме того, Гиббс выдвигал теорию, что эта вероятность, по мере того как стареет Вселенная, естественно стремится к увеличению. Мера этой вероятности называется энтропией, характерная тенденция энтропии заключается в ее возрастании» [3, с. 24]. Энтропия как свойство, присущее всем системам, связана с уменьшением определенности и организации, разнообразия и множества форм в них и, соответственно, увеличением хаоса и наступлением состояния стертости границ, единообразия. «Однако в то время, как Вселенной в целом, если действительно существует Вселенная как целое, присуща тенденция к гибели, то в локальных мирах направление развития, по видимому, противоположно направлению развития Вселенной в целом, и в них наличествует ограниченная и временная тенденция к росту организованности. Жизнь находит себе приют в некоторых из этих миров. Именно исходя из этих позиций начала свое развитие наука кибернетика» [3, с. 24].

Данное противопоставление (энтропии в целом и обратной ей тенденции к организации в локальных мирах) Винер пере-

носит на вопросы организации нравственной жизни, и с определенных позиций может сложиться впечатление, что для него нет ничего важнее и проблемнее, чем «энтропия морали». В книге «Кибернетика, или Управление и связь в животном и машине» читаем: «Средний человек достаточно разумен по отношению к вещам, представляющим его непосредственному вниманию, и достаточно альтруистичен там, где дело касается общественного блага или индивидуальных страданий, которые он видит собственными глазами. В небольшой сельской общине, существующей достаточно долго, чтобы в ней сложились более или менее одинаковые уровни понимания и поведения, существуют вполне достойные уважения нормы попечения об обездоленных, управления дорогами и другими общественными средствами, терпимости к тем, кто лишь один-два раза нарушил общественные законы. Как бы то ни было, нарушители находятся здесь, и остальная община должна жить с ними и впредь» [2, с. 233]. Соответственно, и моральный порядок, и этика локальны, привязаны к непосредственным переживаниям людей; основой самоорганизации сообщества является его опыт добра и зла, а моральные практики более развиты, чем идеи.

Можно сказать, что с позиции кибернетического подхода мораль представляет собой один из видов связи, объединяющий людей, если эта связь хорошо организована, и разделяющий, если она является предметом манипулирования и искажения в интересах какой-либо общественной группы. Этот комплекс идей является теоретическим основанием того, что в этике искусственного интеллекта значительное внимание уделяется нормам, предотвращающим недостойное использование информации, а также поиску способов выявления подложной информации, или фейков, и защиты от нее.

Понятие информации и связанное с ним понятие обратной связи являются ключевыми для кибернетического подхода. Винер определяет информацию как «обозначение содержания, полученного из внешнего мира в процессе нашего приспособления к нему и приспособления к нему наших чувств. Процесс получения и использования информации является процессом нашего приспособления к случайно-

стям внешней среды и нашей жизнедеятельности в этой среде» [3, с. 27]. Подобное представление о мире задает этике коммуникативную направленность. Сущность нравственной жизни понимается как правильная организация системы межличностных и межгрупповых связей, для которой важно, как и какой информацией распоряжаются моральные субъекты. «Действенно жить – это значит жить, располагая правильной информацией. Таким образом, общение и управление точно так же связаны с самой сущностью человеческого существования, как и с жизнью человека в обществе» [3, с. 27]. Статус информации как нравственной ценности обосновывается тем, что она способствует организации и разнообразию и тем самым противостоит энтропии, разладу, стиранию различий, догматизации [3, с. 30]. Нравственная жизнь не может быть связана только с одним образом добра, приверженцы которого стремятся сделать его единственно возможным. Свободный доступ к информации, ее прозрачность и возможность ее критического освоения становятся ценностями в компьютерной этике и перенимаются этикой искусственного интеллекта постольку, поскольку она также строится вокруг коммуникативных связей.

Понятие обратной связи в кибернетическом подходе также является универсальным как в логическом смысле, так и содержательно. Винер показывает, что обратной связью обладает любая ситуация коммуникации, независимо от того, человек в ней участвует или машина. «Когда я отдаю приказ машине, то возникающая в данном случае ситуация не отличается от ситуации, возникающей в том случае, когда я отдаю приказ какому-либо лицу. Иначе говоря, что касается моего сознания, то я осознаю отданный приказ и возвратившийся сигнал повиновения. Лично для меня тот факт, что сигнал в своих промежуточных стадиях прошел через машину, а не через какое-либо лицо, не имеет никакого значения и ни в коей мере существенно не изменяет моего отношения к сигналу. Таким образом, теория управления в человеческой, животной или механической технике является частью теории информации» [3, с. 26].

Обратная связь – это необходимый элемент динамической коммуникативной

системы. Винер рассматривает многие виды обратной связи, выделяя их в соответствии с системой, в которой она возникает: гомеостатическая, обратная связь произвольного движения и позы, зрительно-мышечная, упреждающая, информирующая и др. Обратная связь служит индикатором того, как происходит обмен информацией.

В паре человека и машины контролирующим субъектом является человек: «Управление машиной на основе действительного выполнения ею приказов, а не ожидаемого их выполнения, называется обратной связью и включает в себя чувствительные элементы, которые приводят в движение моторными элементами и выполняют функцию предупреждающих сигнальных... устройств, показывающих выполнение приказов. Функция этих механизмов состоит в управлении механической тенденцией к дезорганизации, иначе говоря, в том, чтобы вызывать временную и местную перемену обычного направления энтропии» [3, с. 32]. Технология искусственного интеллекта обучается контролю обратной связи, поскольку именно с технологической возможностью адекватных до тонкостей ответов в коммуникации связывается успех ее развития теперь. В этическом смысле это усиливает иллюзию очеловечивания машины, с которой столкнулся еще Дж. Вейценбаум в результате эксперимента «Eliza».

Результаты и их обсуждение

Дискурс неопределенности и случайности в кибернетическом подходе уводит его в сторону от проблемы свободы воли, в известном смысле заменяет ее, предлагая другой взгляд на возможность выбора зла и моральное оправдание поступка. Так, неопределенность характеризует моральные коллизии, колебания, душевные терзания, нравственные противоречия и многие другие, неоднозначные с точки зрения этики ситуации. Их возникновение часто случайно и не зависит от воли личности. В самом деле, развитие технологии искусственного интеллекта не является результатом злонамеренной деятельности, а, скорее, вызвано стремлением к прогрессу, к освоению недоступных ранее областей деятельности — к новым пределам совершенства человека.

Реальные и возможные безнравственные последствия использования кибернети-

ческих технологий Винер относит на счет «технологической наивности» людей, полагающих, что в безоблачном будущем «включают роботы — счастлив человек» и машине можно полностью доверить ответственность за результаты деятельности. Теоретически это можно сделать, поскольку программные алгоритмы уже на заре кибернетики были способны достигать цели запрограммированным путем. Но нет гарантии, что нас устроит результат «самостоятельности» машин, пишет Винер, вспоминая притчу об обезьяньей лапе [4, с. 68–69]. Винер подытоживает свою концепцию деятельного участия личности и ключевой роли экспертов в управлении цифровыми технологиями знаменитой фразой: «Отдайте же человеку — человеческое, а вычислительной машине — машинное. В этом и должна, по-видимому, заключаться разумная линия совместных действий людей и машин» [4, с. 82–83].

С увеличением объема компьютерных данных решения машин все больше будут выглядеть нравственно свободными, но разделять их на допустимые и недопустимые, т. е. давать или не давать им моральное оправдание, можно только с позиции сложившихся в сообществе нравственных практик. Иными словами, предполагается, что нормативная часть этики искусственного интеллекта регулирует отношения людей по поводу цифровых технологий, препятствуя несправедливости, дискриминации и другим видам неравенства, которые закрепляют превосходство одних людей перед другими [6, с. 231].

Идея сохранения морального равновесия в небольших локальных сообществах, где определены нормы, признаваемые всеми как очевидные, необходимые для самосохранения, и где выработаны повседневные нравственные практики, поддерживающие, по словам Винера, «гомеостаз системы», и противоположная ей — возрастание морального хаоса, неопределенности, дезорганизации в мире в целом — являются одной из характерных моральных доминант кибернетического подхода. Мораль не может быть всеобщей, это превращает ее в выхолощенный протокол разрешенных и запрещенных действий, но разнообразие морали локальных сообществ удивительно солидарно в своем содержании относительно

но того, что их существование невозможно без моральных ценностей: справедливости, смысла жизни, счастья, свободы, при этом их содержанию присуще известное разнообразие.

Обобщая всю философию Н. Винера, можно назвать учением о правильной жизни в условиях энтропии и неопределенности, где сущностными характеристиками любой системы является информационный обмен и обратная связь [7]. Кибернетический подход объясняет, как в едином нравственном поле могут оказаться человек и машина, обходя при этом проблему морального субъекта. Моральное преимущество в нравственных отношениях с технологией принадлежит человеку, но идея информации как меры организации обратной связи, как способа изменять направление дезорганизации системы, наряду с техническими особенностями развития искусственного интеллекта, обосновывает способность данной технологии к принятию моральных решений, самоорганизации, ставит вопрос о статусе морального субъекта для технологии искусственного интеллекта в этой системе отношений.

Заключение

Кибернетический подход Н. Винера характеризуется как сенсуалистический и надысторический. Он складывается на ос-

нове четырех понятий: «неопределенность», «энтропия», «информация» и «обратная связь». Содержание каждого из них задается предельно широко, чтобы с их помощью можно было объяснить наибольшее количество процессов и явлений. Ни одно из них не является этическим в строгом смысле, но вместе они моделируют реальность морального порядка. Данный подход предполагает, что этика искусственного интеллекта объединяет человека и машину в одной системе связей, где все является коммуникацией, направленной против зла, т. е. беспорядка и дезорганизации (неопределенности и энтропии), участники которой стремятся наладить ее наилучшим образом (информационный обмен и обратная связь). В этике искусственного интеллекта данный подход, с одной стороны, уравнивает человека и машину как участников информационного обмена, с другой – обосновывает моральное преимущество эксперта, закрепляя за ним ответственность в виде обязанности принимать решения и контролировать их последствия. Достоинно внимания замечание Винера об организованности локальных сообществ. Можно сделать вывод, что этике искусственного интеллекта не присуща универсальность, она существует как локальная, по-разному воплощающая общие моральные ценности.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Алексеева, И. Ю. Этика искусственного интеллекта как прикладная этика / И. Ю. Алексеева // *Философия и общество*. – 2024. – № 3. – С. 69–85.
2. Винер, Н. Кибернетика, или Управление и связь в животном и машине // Н. Винер. – М. : Совет. радио, 1968. – 325 с. – URL: <https://www.twirpx.com/file/3140740/> (дата обращения: 24.03.2025).
3. Винер, Н. Кибернетика и общество / Н. Винер. – М. : Тайдекс Ко, 2002. – 184 с.
4. Винер, Н. Творец и робот / Н. Винер. – М. : Прогресс, 1966. – 109 с. – URL: <https://www.twirpx.com/file/3697218/> (дата обращения: 24.03.2025).
5. Винер, Н. Наука и общество / Н. Винер // *Вопросы философии*. – 1961. – № 7. – С. 117–122. – URL: <https://www.twirpx.com/file/1603715/> (дата обращения: 24.03.2025).
6. Минаев, С. В. Норберт Винер: у истоков компьютерной этики / С. В. Минаев // *Вестник ТвГУ. Серия «Философия»*. – 2018. – № 1. – С. 227–232.
7. Bynum, Terrell Ward. Norbert Wiener's Foundation of Computer Ethics / Terrell Ward Bynum // *The Research Center on Values in Emerging Science and Technology*. – URL: <https://rcvest.southernct.edu/norbert-wieners-foundation-of-computer-ethics> (date of access: 20.03.2025).

REFERENCES

1. Alieksiejeva, I. Ju. Etika iskusstviennogo intielliekta kak prikladnaja etika / I. Ju. Alieksiejeva // *Filosofija i obshchestvo*. – 2024. – № 3. – S. 69–85.
2. Viner, N. Kibiernetika, ili Upravlenije i svjaz' v zhivotnom i mashinie // N. Viner. – M. : Soviet. radio, 1968. – 325 s. – URL: <https://www.twirpx.com/file/3140740/> (data obrashchijenija: 24.03.2025).
3. Viner, N. Kibiernetika i obshchestvo / N. Viner. – M. : Tajdeks Ko, 2002. – 184 s.
4. Viner, N. Tvoriec i robot / N. Viner. – M. : Progress, 1966. – 109 s. – URL: <https://www.twirpx.com/file/3697218/> (data obrashchenija: 24.03.2025).
5. Viner, N. Nauka i obshchestvo / N. Viner // *Voprosy filosofii*. – 1961. – № 7. – S. 117–122. – URL: <https://www.twirpx.com/file/1603715/> (data obrashchijenija: 24.03.2025).
6. Minajev, S. V. Norbert Viner: u istokov komp'juternoj etiki / S. V. Minajev // *Viestnik TvGU. Sierija «Filosofija»*. – 2018. – № 1. – S. 227–232.
7. Bynum, Terrell Ward. Norbert Wiener's Foundation of Computer Ethics / Terrell Ward Bynum // *The Research Center on Values in Emerging Science and Technology*. – URL: <https://rcvest.southernct.edu/norbert-wieners-foundation-of-computer-ethics> (date of access: 20.03.2025).

Рукапіс наступіў у рэдакцыю 08.05.2025