

РАЗМЫШЛЕНИЯ НАД КНИГОЙ

АНГЕЛИНА ВИКТОРОВНА БАЕВА

кандидат философских наук,
старший преподаватель философского факультета
Московского государственного университета
имени М.В. Ломоносова,
Москва, Россия;
e-mail: baeva@philos.msu.ru



О чем мы говорим, когда говорим об «этике искусственного интеллекта»?

УДК: 004.896:172

DOI: 10.24412/2079-0910-2024-4-196-205

Статья посвящена обзору книги современного философа Лучано Флориди «Этика искусственного интеллекта: принципы, вызовы и возможности», вышедшей в 2023 г. и ставшей частью более масштабного исследовательского проекта автора. В статье рассматривается концепция «этики искусственного интеллекта» Л. Флориди и предпринимается попытка реконструировать ее основные положения.

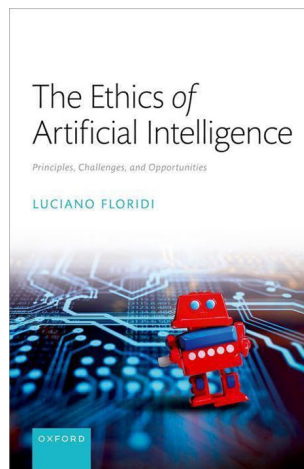
Ключевые слова: искусственный интеллект (ИИ), этика ИИ, технологии, агентность, регулирование, ответственность, эпистемология, реонтологизация.

Введение

Сегодня в философии тема этики искусственного интеллекта (ИИ) несомненно выходит на первый план в связи с ускоренным развитием технологий в целом и искусственных интеллектуальных систем в частности. Л. Флориди — один из самых влиятельных современных итальянских философов в области философии техники и этики ИИ. В 2023 г. была опубликована его книга «Этика ИИ: принципы, вызовы и возможности», которая, по заявлению самого автора, «стала частью масштабного исследовательского проекта, посвященного трансформации свободы воли (способности успешно взаимодействовать с окружающим миром и учиться у него в соответствии с поставленной целью), вызванной цифровой революцией» [Floridi, 2023, p. xii].

В этой книге Флориди преследует две основных цели. Первая, по его собственному заявлению, носит метатеоретический характер и реализуется в первой части книги, состоящей из трех глав. В ней автор предлагает интерпретацию прошлого, настоящего и будущего ИИ. Первая часть представляет собой не введение в ИИ в техническом смысле, а скорее философскую интерпретацию ИИ как технологии, и здесь Флориди обозначает свою позицию, которой будет придерживаться и далее. Основной тезис, который он отстаивает, заключается в том, что ИИ — «это беспрецедентный разрыв между агентностью и интеллектом» [*Ibid.*, p. xiv]. На протяжении всей книги Флориди настаивает на понимании ИИ не как интеллекта, а как особой новой формы агентности, что само по себе тоже является революцией, но не в когнитивном смысле, а в прагматическом. В этом смысле ИИ (или, по крайней мере, философия ИИ) слишком много внимания уделяет разработке какой-то биологической формы интеллекта, в то время как, по мнению автора, следовало бы заниматься разработкой артефактов, которые могут успешно функционировать без какой-либо необходимости в интеллекте.

Вследствие такого понимания ИИ вторая цель книги состоит в теоретическом исследовании последствий того разрыва, о котором говорит Флориди. Во второй части им развивается идея о том, что ИИ — это новая форма агентности, которую можно использовать этично или неэтично. Здесь Флориди более подробно предлагает единую точку зрения на многие принципы, которые уже были предложены ранее для формирования этики ИИ. Он отмечает еще одну специфическую особенность, связанную с таким пониманием ИИ: те конкретные насущные проблемы, которые связаны с ИИ, возникают именно из-за разрыва между этой новой формой агентности и интеллектом. И более того, этот разрыв будет продолжать расширяться по мере того, как ИИ будет становиться все более успешным и сильным.



Прошлое, настоящее и будущее ИИ

Итак, в первой части Флориди, говоря о прошлом ИИ, останавливается на основных изменениях в нашей жизни, произошедших вследствие цифровой революции. Во-первых, скорость и объем памяти наших цифровых технологий растут не такими темпами, как объем данных. Вычислительная мощность и скорость, объемы данных, эффективность алгоритмов и статистических инструментов, а также количество наших онлайн-взаимодействий — все это растет невероятно быстро. А поскольку даже те данные, которые были созданы нынешним поколением, стремительно устаревают с точки зрения способов поддержки технологий, которые также стремительно устаревают, мы уже неизбежно сталкиваемся с проблемой хранения данных. И эта проблема будет становиться все более актуальной. Таким образом, мы совершаем переход от практики записи к практике удаления: вопрос состоит уже не в том, что сохранить, а в том, что удалить, чтобы освободить место для новых данных [*Ibid.*, p. 5].

Цифровая революция также влияет и на то, как мы концептуализируем и понимаем нашу реальность, которая все чаще интерпретируется в компьютерных и цифровых терминах. Флориди предлагает: «...просто подумайте о “старой” аналогии между вашей ДНК и вашим “кодом”, которую мы сейчас воспринимаем как должное» [*Ibid.*, p. 6]. Революция способствовала развитию ИИ, поскольку мы делимся своим жизненным опытом и нашей информационной средой с искусственными агентами (ИА), будь то алгоритмы, боты или роботизированная среда. Но Флориди интересует не столько вопрос «как», сколько вопрос «почему»: почему цифровые технологии обладают такой способностью к расщеплению, позволяющему соединять, разъединять и изменять мир и наше представление о нем? И почему другие технологические инновации не оказывают, по всей видимости, аналогичного влияния? Автор предполагает, что ответ кроется в сочетании двух факторов. С одной стороны, Флориди понимает цифровые технологии как технологии третьего порядка [*Floridi*, 2014]. Это не просто технологии, которые связывают нас с природой, как топор (первый порядок), или с другой технологией, как двигатель (второй порядок). Это скорее промежуточные технологии между двумя другими, как компьютеризированная система, управляющая роботом, красящим автомобиль (третий порядок).

С другой стороны, цифровые технологии не просто улучшают или приумножают реальность: они радикально преобразуют ее, поскольку создают новые условия нашей жизни и новые способы нашего взаимодействия между собой и с этой реальностью. Однако для этого глубокого вида трансформации все еще нет термина. В более ранней своей работе Флориди использовал термин «реонтологизация» [*Floridi*, 2010], как наиболее удачно описывающий суть происходящих изменений: этот вид трансформации не столько проектирует, конструирует или структурирует какую-либо систему, сколько фундаментально преобразует внутреннюю природу самой системы, то есть ее онтологию. В этом смысле нано- и биотехнологии не просто перестраивают, но и реонтологизируют наш мир. Таким образом, цифровые технологии наделяются беспрецедентной силой благодаря главным образом наличию этих двух характеристик: это реонтологизирующие технологии третьего порядка.

Как мы отмечали выше, общая тема исследовательского проекта Флориди — это трансформация свободы воли, которая изменяется в том числе под влиянием цифровой революции: цифровые технологии изменили природу агентности, но мы все еще интерпретируем результаты таких изменений с точки зрения нашего современного мышления, что порождает их глубокое непонимание. Некоторые (а возможно, многие) люди, похоже, считают, что ИА — это новые артефакты, сочетающие искусственную деятельность с разумным поведением. Однако Флориди убежден, что это недоразумение и на самом деле верно обратное: «цифровая революция сделала ИИ не только возможным, но и все более полезным» [*Floridi*, 2023, p. 12]. Это достигается за счет того, что ИИ отделяет способность успешно решать проблему или выполнять задание от необходимости обладать для этого интеллектом. Только когда достигается такое разделение, ИИ становится успешным. И в этом главная отличительная особенность агентности ИИ в настоящее время: в нем важен результат, а не то, является ли агент или его поведение разумным. Значит, ИИ — это не воспроизведение какого-либо биологического интеллекта. Напротив, речь идет о том, чтобы обойтись вообще без него. Что еще более важно, так это то, что артефакты могут

быть интеллектуальными, не обладая интеллектом. И это достижение ИИ, который является успешным продолжением человеческого интеллекта другими средствами.

Переходя к разговору о будущем ИИ, Флориди отмечает, что в последнее время ИИ настолько усовершенствовался, что в некоторых случаях мы переходим от акцента на количестве больших массивов данных (иногда неправильно называемых «большими данными» [Floridi, 2012]), к акценту на качестве хорошо обработанных наборов данных. В качестве примера Флориди описывает кейс *DeepMind* в партнерстве с офтальмологической больницей в Великобритании, где в 2018 г. систему ИИ обучили выявлять признаки глазных заболеваний с помощью данных оптической когерентной томографии (ОКТ). В итоге команде удалось продемонстрировать эффективность в выдаче рекомендаций по направлению к врачу, которая соответствует или даже превосходит рекомендации экспертов по целому ряду угрожающих зрению заболеваний сетчатки. И это после обучения ИИ, проводимого всего на 14 884 снимках. Флориди отмечает, что такой массив данных вовсе не является *Big Data*: как раз-таки это «небольшие данные» высокого качества, и это одно из будущих достижений ИИ, поскольку у ИИ будет больше шансов на успех, когда появятся тщательно отобранные, обновленные и полностью надежные наборы данных, которые позволят обучить систему работе в конкретной области применения.

В связи с этим решающее значение приобретает происхождение данных. В случае с офтальмологической больницей данные были получены аутентичным способом, то есть из реальной жизни — в терминологии Флориди это «исторические» данные. Но мы также знаем, что ИИ может и самостоятельно генерировать свои собственные данные: такие данные Флориди предлагает называть «полностью синтетическими». Однако он сам отмечает неоднозначность этого термина, в силу того что его начали использовать еще в 1990-х гг. для обозначения исторических данных, которые были обезличены до их использования (зачастую для защиты конфиденциальности). Эти данные являются синтетическими только потому, что они были синтезированы на основе исторических данных, например, путем «замаскировки»: они имеют более низкое разрешение, но их происхождение не является искусственным. Флориди же подчеркивает значимость именно «полностью синтетических» данных, которые имеют полностью искусственное происхождение. Здесь важно именно онтологическое различие между «историческими» и «синтетическими» данными, которое может иметь существенные последствия с точки зрения эпистемологии, особенно когда речь заходит о нашей способности объяснить полученные синтетические данные, а также результаты обучения, проводимого ИИ с их использованием. Ведь очевидно, что будущее ИИ будет связано не только с «небольшими данными», но и (возможно, главным образом) с его растущей способностью генерировать свои собственные данные.

Этика ИИ: вопросы регулирования и ответственности

В первой части своей работы Флориди показал, что ИИ — это новая форма агентности, которая может успешно решать задачи и проблемы исходя из поставленной цели и без какой-либо потребности в интеллекте: другими словами, это беспрецедентная форма агентности. И любой успех любого приложения с ИИ не повышает планку того, что значит быть интеллектуальным агентом. Напротив, по

утверждению Флориды, он вообще превосходит эту планку. А успеху такой искусственной агентности, как ИИ, все больше способствует превращение среды, в которой ИИ работает, в благоприятные для этой работы условия [Floridi, 2023, p. 57].

Во второй части работы Флориды представляет результаты сравнительного анализа нескольких наиболее авторитетных наборов этических принципов для ИИ и утверждает, что анализируемые наборы принципов в значительной степени совпадают в отношении того, что следует считать «этичным ИИ». Эту согласованность можно представить в виде пяти основных принципов этичного ИИ. Четыре из них — это основные принципы, обычно используемые в биоэтике: принцип «делай благо», принцип «не навреди», принцип автономии личности и принцип справедливости. Неудивительно, что ИИ интерпретируется как форма агентности: согласованность между биоэтикой и этикой ИИ можно рассматривать как свидетельство того, что подход к ИИ как к новой форме агентности является плодотворным. При этом, с точки зрения Флориды, перспектива подходить к этике ИИ и цифровой этике в целом с позиций этики добродетелей является не вполне убедительной, но она все еще остается жизнеспособной, несмотря на критические возражения Флориды, которые он выдвигал на этот счет ранее [Floridi, 2013].

Основываясь на проведенном сравнительном анализе, Флориды настаивает на необходимости добавить новый (пятый) принцип — объяснимость. Объяснимость понимается им как в эпистемологическом смысле — как понятность (то есть как ответ на вопрос «как это работает?»), так и в этическом смысле — как ответственность (то есть как ответ на вопрос «кто несет ответственность за то, как это работает?»). Необходимость в этом дополнительном принципе обусловлена новым пониманием ИИ как новой беспрецедентной формы агентности. При этом четыре биоэтических принципа на удивление хорошо адаптируются к новым этическим вызовам, которые ставит ИИ [Floridi, 2023, p. 60]. Внедряя ИИ, мы тем самым добровольно уступаем часть своих полномочий по принятию решений технологическим артефактам. Таким образом, утверждение принципа автономии в контексте ИИ означает установление баланса между полномочиями по принятию решений, которые мы сохраняем за собой, и теми полномочиями, которые мы делегируем ИИ. Риск же, как отмечает Флориды, заключается в том, что рост искусственной автономии может подорвать процветание человеческой автономии [Ibid., p. 62].

Во избежание этических рисков или, по крайней мере, для сведения их к минимуму при воплощении принципов цифровой этики на практике Флориды говорит о необходимости обсуждения вопросов регулирования ИИ (и в более широком смысле — цифровых технологий). Прежде всего он проводит различие между «мягкой» (soft) и «жесткой» (hard) этикой: «жесткая этика сначала предшествует формированию законодательства, а затем вносит свой вклад в его дальнейшее развитие. Напротив, мягкая этика применяется после соблюдения законодательства (то есть после его соблюдения этических норм)» [Ibid., p. 77]. И это один из важных моментов для обсуждения вопросов регулирования. Очевидно, что реальная проблема заключается уже не в хороших цифровых инновациях, а в хорошем регулировании цифровых технологий. Доказательством того, что это так, по мнению Флориды, является растущее число инициатив, направленных на изучение влияния цифровых технологий на повседневную жизнь и способов их регулирования. В том же контексте, где люди жалуются на скорость цифровых инноваций и невыполнимую задачу догнать их с помощью какой-либо нормативной базы, можно найти такую же

уверенность в серьезности риска, связанного с неправильным законодательством и соответствующей нормативной базой. Плохое законодательство может полностью уничтожить цифровые инновации или даже целые технологические секторы и разработки. Флориди видит выход в переходе от попытки догнать технологии законодательно к лидерству в этой сфере: именно из-за отсутствия четкого представления о социально-политическом направлении, в котором нам следует двигаться, мы обеспокоены скоростью нашего технологического прогресса. Однако решение состоит не в том, чтобы замедляться, а в том, чтобы сообща решить, куда именно мы хотим двигаться. Чтобы это произошло, Флориди считает, что общество должно перестать защищаться и пойти в атаку. Вопрос уже состоит не в том, стоит ли это делать, а в том, как это делать, чему и посвящена оставшаяся часть книги, где речь идет о регулировании цифровых технологий в целом и ИИ в частности.

Однако Флориди поясняет, что не следует путать друг с другом разные нормативные подходы, которые дополняют друг друга: 1) регулирование цифровых технологий (цифровое регулирование), 2) этику цифровых технологий (цифровую этику, которая также известна как компьютерная, информационная или информационная этика данных) [Floridi, Taddeo, 2016], и 3) регулирование цифровой сферы. Что касается вопросов регулирования и ответственности, то отсутствие прозрачности также является неотъемлемой характеристикой самообучающихся алгоритмов, которые изменяют логику принятия решений (для создания новых наборов правил) в процессе обучения. Это затрудняет разработчикам детальное понимание того, почему были внесены некоторые конкретные изменения. Однако это не обязательно приводит к непрозрачным результатам: даже не понимая каждый логический шаг, разработчики могут настроить гиперпараметры (параметры, управляющие процессом обучения) для тестирования различных конечных результатов. Принятие прозрачности за самоцель — в отличие от этического фактора, который должен быть должным образом разработан для обеспечения соблюдения важнейших этических норм, — может не решить существующие этические проблемы, связанные с использованием алгоритмов. И тогда действительно путаница может привести к появлению новых проблем [Floridi, 2023, p. 99].

Флориди уже ранее отмечал, что, рассматривая алгоритмы и их влияние, мы имеем дело с распределенными моральными действиями, возникающими в результате морально нейтральных взаимодействий потенциально гибридных сетей агентов [Floridi, 2016]. Другими словами, вопрос заключается в том, кто несет распределенную моральную ответственность за распределенные моральные действия. В этом контексте Флориди предлагает возложить полную моральную ответственность (по умолчанию и безоговорочно) на всех моральных агентов сети (например, людей или компании, состоящие из людей), которые причинно связаны с данным действием сети. Этика алгоритмов стала центральной темой обсуждения среди ученых, поставщиков технологий и политиков с 2016 г. И это неудивительно: этический анализ необходим как для снижения рисков, так и для использования потенциала технологий во благо. Результаты такого двустороннего анализа воплощаются в обоснованных практических рекомендациях по регулированию разработки и использования цифровых артефактов. Говоря об угрозах, которые несут в себе «плохие практики» использования ИИ, Флориди сосредоточил внимание на понимании ИИ как средства автономной интеллектуальной агентности и на конечной ответственности человека за неэтичное и незаконное использование такого ИА. В книге Флориди достаточно

подробно описывает угрозы в различных конкретных областях применения ИИ с точки зрения определенных преступлений, а также в целом с точки зрения характеристик ИИ и вопросов регулирования и ответственности. Кроме того, автор анализирует, какие решения доступны или могут быть разработаны для устранения этих рисков, сосредоточив внимание как на общих, так и на сквозных темах и представив актуальную картину доступных социальных, технологических и юридических решений, а также их ограничений. И, конечно, говоря о практиках (как «плохих», так и «хороших») использования ИИ, необходимо учитывать проблему объективности использованных данных во избежание предвзятости в полученных результатах.

Подводя некоторый итог, Флорида говорит о том, что ИИ предоставляет все больше возможностей для «умной агентности». Будучи поставленным на службу человеческому интеллекту, такой ресурс может значительно расширить возможности человека. Мы уже можем делать больше, лучше и быстрее благодаря той помощи, которую предоставляет ИИ. В этом смысле «[человеческий] расширенный интеллект» ИИ можно сравнить с тем влиянием, которое оказал двигатель на нашу жизнь. И чем больше людей будут пользоваться возможностями и преимуществами такого «доступного» источника интеллектуальных ресурсов, тем лучше будет и наше общество. Поэтому этические вопросы регулирования и ответственности крайне важны и должны учитывать, какой тип ИА мы разрабатываем, как мы его используем и делимся ли мы со всеми его преимуществами. Очевидно, что соответствующим риском является отсутствие такой ответственности. И это, по мнению Флориды, связано не только с социально-политическими ограничениями, но и с тем, что тот способ принятия решений, которым наделен ИИ, представляет собой для нас «черный ящик», поскольку выходит за рамки нашего человеческого понимания и, следовательно, оказывается неподконтрольным человеку. И эти опасения касаются не только резонансных случаев (например, смертельных), вызванных автономными транспортными средствами. Они также применимы и к более обыденным, но все же важным областям применения ИИ, таким как автоматическое принятие решений об условно-досрочном освобождении или проверка кредитоспособности. «Однако взаимосвязь между степенью и качеством свободы воли, которой пользуются люди, и тем, какую долю полномочий они делегируют автономным системам, не сводится к нулю, ни с прагматической, ни с этической точки зрения. На самом деле при продуманной разработке ИИ предоставляет возможность улучшить и приумножить возможности человеческой агентности» [Floridi, 2023, p. 171]. Самый эффективный способ гарантировать, что ИИ будет отвечать потребностям общества, позволяя разработчикам, пользователям и законодателям работать сообща с самого начала, по мнению Флориды, может дать как раз-таки многосторонний подход, подразумевающий, помимо прочего, внимание к различным культурным ограничениям, оказывающим влияние на отношение к новым технологиям.

Заключение

Итак, ИИ — это не новая форма интеллекта, а новая форма агентности, и его успех обусловлен отделением агентности от интеллекта, а также созданием среды, которая становится все более благоприятной для ИИ. Таким образом, будущее ИИ заключается в развитии этих двух факторов. В своей книге, посвященной этике ИИ,

Флориди проанализировал природу ИИ, его положительное и отрицательное влияние, возникающие в связи с этим этические проблемы и то, какие меры могут быть приняты для обеспечения большей вероятности развития «хорошего общества», основанного на ИИ.

В основной части книги Флориди предложил единую систему этических принципов для ИИ на основе сравнительного анализа наиболее влиятельных предложений начиная с 2017 г., когда начали публиковаться этические принципы для ИИ. Далее Флориди проанализировал некоторые риски, которые мешают воплощению этих принципов на практике. Кроме того, центральная часть этого раздела посвящена этическим основам и этическим рискам: здесь Флориди рассмотрел вопросы этического регулирования ИИ и представил концепцию «мягкой этики». Далее он перешел к более прикладным аспектам: представил обновленную карту основных этических проблем, возникающих в связи с алгоритмами. Им была предпринята попытка найти баланс между общим подходом, касающимся всех видов алгоритмов, а не только тех, которые управляют системами или приложениями ИИ, и конкретным подходом к решению проблем. Так, например, конкретными и неотложными являются проблемы, связанные с преступным использованием ИИ в целях причинения «социального зла». Однако гораздо больше ИИ, по мнению Флориди, в основном связан с возможностями, действующими на благо общества.

В заключение Флориди возвращается к вопросам разработки качественного ИИ, сосредоточив внимание на том, как обеспечить социально значимый «хороший ИИ», а также рассмотрев вопросы положительного и отрицательного влияния ИИ на изменение климата и применимость ИИ для поддержания целей устойчивого развития (ЦУР) ООН. Завершая книгу, Флориди напоминает о том, что агентность, включая ее новую форму, представленную ИИ, нуждается в регулировании, и что регулирование — это вопрос социально-политической деятельности, поскольку цифровые технологии также изменили социально-политическую агентность. Наряду с инструментами прогнозирования, которые могут превосходить человеческие решения и манипулировать ими, ИИ предоставляет уникальную историческую возможность переосмыслить нашу человеческую исключительность как не ошибочную, а скорее направленную не по назначению. Наше разумное поведение столкнется с проблемой разумного поведения ИА, которое, вероятно, может более успешно адаптироваться в инфосфере. Наше автономное поведение столкнется с проблемой предсказуемости и управляемости нашего рационального выбора, а также с развитием искусственной автономии. И нашей социальности будет противостоять ее искусственный аналог, представленный ИА. Эти аналоги внешне, возможно, даже будут неотличимы от людей. И чем все это закончится, неизвестно. Но одно кажется Флориди предсказуемым: разработка ИА не приведет к пугающей реализации некоего антиутопического научно-фантастического сценария. В этой книге Флориди неоднократно утверждает, что ИИ — это оксюморон: «умные» технологии будут такими же «глупыми», как и наши старые технологии. Но ИИ побудит нас все же более серьезно задуматься о том, кем мы являемся, кем мы могли бы быть и кем мы хотели бы стать, а следовательно, задуматься о нашей ответственности и о понимании самих себя. Это бросит серьезный вызов нашей идентичности и «исключительности» с точки зрения того, что мы имеем в виду, считая себя в некотором роде «особенными» даже после четвертой революции [Floridi, 2014], в результате которой мы уже не находимся в центре ни Вселенной (Коперник), ни биосферы (Дарвин),

ни ментального пространства (Фрейд), ни инфосферы, наконец (Тьюринг). Однако Флориди не утверждает, что наша исключительность неверна: он полагает, что ИИ поможет нам осознать, что «наша исключительность заключается в уникальном и, возможно, невозпроизводимом способе нашей успешной недееспособности. <...> Используя более цифровую и современную метафору, мы прекрасный сбой в великом программном обеспечении Вселенной, а не конечное приложение. И мы останемся этой уникальной удачной ошибкой, в то время как ИИ будет занимать все большее место в математической книге природы Галилея» [Floridi, 2023, p. 208].

Литература

Floridi L. Information: A Very Short Introduction. Oxford University Press, 2010. 106 p.

Floridi L. Big Data and Their Epistemological Challenge // *Philosophy & Technology*. 2012. Vol. 25. No. 4. P. 435–437.

Floridi L. The Ethics of Information. Oxford University Press, 2013. 357 p.

Floridi L. The Fourth Revolution: How the Infosphere Is Reshaping Human Reality. Oxford University Press, 2014. 248 p.

Floridi L. Faultless Responsibility: On the Nature and Allocation of Moral Responsibility for Distributed Moral Actions // *Philosophical Transactions of the Royal Society. A: Mathematical, Physical and Engineering Sciences*. 2016. Vol. 374 (2083). No. 20160112. P. 1–13. DOI: 10.1098/rsta.2016.0112.

Floridi L., Taddeo M. What Is Data Ethics? // *Philosophical Transactions of the Royal Society. A: Mathematical, Physical and Engineering Sciences*. 2016. Vol. 374 (2083). No. 20160360. P. 1–5. DOI:10.1098/rsta.2016.0360.

Floridi L. The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities. Oxford University Press, 2023. 243 p.

What Are We Talking About, When We Talk About “The Ethics of Artificial Intelligence”?

ANGELINA V. BAEVA

Lomonosov Moscow State University,
Moscow, Russia;
e-mail: baeva@philos.msu.ru

The article is devoted to the review of the book by the modern philosopher Luciano Floridi “Ethics of Artificial Intelligence: principles, challenges and opportunities,” published in 2023. It became part of a larger research project of the author. The author of this article considers the concept of “Ethics of Artificial Intelligence” by L. Floridi and attempts to reconstruct its basic provisions.

Keywords: artificial intelligence (AI), ethics of AI, agency, regulation, responsibility, epistemology, re-ontologization.

References

Floridi, L. (2010). *Information: A Very Short Introduction*, Oxford University Press.

Floridi, L. (2012). Big Data and Their Epistemological Challenge, *Philosophy & Technology*, 25 (4), 435–437.

Floridi, L. (2013). *The Ethics of Information*, Oxford University Press.

Floridi, L. (2014). *The Fourth Revolution: How the Infosphere Is Reshaping Human Reality*, Oxford University Press.

Floridi, L. (2016). Faultless Responsibility: On the Nature and Allocation of Moral Responsibility for Distributed Moral Actions, *Philosophical Transactions of the Royal Society. A: Mathematical, Physical and Engineering Sciences*, no. 374 (2083), 20160112, 1–13.

Floridi, L., Taddeo, M. (2016). What Is Data Ethics?, *Philosophical Transactions of the Royal Society. A: Mathematical, Physical and Engineering Sciences*, no. 374 (2083), 20160360, 1–5.

Floridi, L. (2023). *The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities*, Oxford University Press.