

Проблемы выбора ценностей для искусственных интеллектуальных систем

© В.П. Майкова¹, Э.М. Молчан², А.И. Майков³

¹Мытищинский филиал МГТУ им. Н.Э. Баумана, Мытищи, 141005, Россия

²Московский государственный областной университет, Мытищи, 141014, Россия

³МГТУ им. Н.Э. Баумана, Москва, 105005, Россия

Рассмотрены проблемы выбора ценностей для искусственных интеллектуальных систем (ИИС), которые в скором будущем станут самостоятельными агентами объективной реальности, войдут в жизнедеятельность человека. Показано, что в контексте актуальных проблем меняющейся реальности перед современным обществом стоят задачи как выбора ценностей для ИИС, так и реализации технического и нормативного этапов их решения. Понятие «ценность» заменяется на предпочтение, выступающее в качестве информационно-идеализированной формы, внедренной в программное обеспечение киборг-систем. Сделан вывод: выбор и включение ценностей для ИИС киборг-машин связан с приоритетами заказчика, однако в силу самообучаемости ИИС возможны получение непредсказуемых синергетических эффектов или переход интеллектуальных систем на суперинтеллект, что может спровоцировать неуправляемость ими. Для этого требуется создание новой методологии, определяющей базис научно-технического прогресса современности, когнитивного (эпистемологического), информационного развития в целях формирования нормативно-правовой и этической базы в ответ на вызовы цифровой глобализации, трансформирующейся в глобализационный индивидуализм.

Ключевые слова: системы искусственного интеллекта, ценности, этика, моральные ценности, моральные нормы, моральные принципы, суперинтеллект, киборг-машина

Введение. Основная идея современного миропорядка связана с содержательно-сущностным преобразованием метафизических основ личности, расщеплением целостности человека посредством трансформации духовно-нравственных ценностей, созданием искусственных интеллектуальных систем (ИИС) и интеграцией их в виртуальное и социальное бытие. Быстрое развитие ИИС под предлогом блага для человечества предоставило широкий спектр полезных приложений в различном контексте: от здравоохранения до безопасности и управления. Во многих сферах деятельности человека сейчас используются интеллектуальные роботы. Растущие возможности искусственных интеллектуальных систем, а также разработка бесчисленного множества новых приложений, от которых все больше зависит человек, вызывают тревогу [1, с. 451]. ИИС могут представлять угрозу для человечества из-за злонамеренного использования людьми или

автономных действий искусственной системы. По данным аналитического отчета *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation* (2018), выпущенного ведущими исследовательскими центрами искусственного интеллекта в Великобритании и США, возникает острая необходимость в разработке нормативно-правовых и ценностно-нравственных средств регулирования искусственных ИИС с точки зрения цифровой, физической и политической безопасности [2]. Исследователи прогнозируют, что ИИС превзойдут людей в переводе с одного языка на другой к 2024 г., написании школьных сочинений — к 2026 г., вождении грузовика — к 2027 г., в работе хирургов — к 2053 г. Данные отчета свидетельствуют о 50%-ной вероятности того, что ИИС опередят людей в решении любых задач в течение 45 лет и будут автоматизированы все рабочие места за 120 лет. Поскольку роботы становятся все более автономными и способны принимать решения самостоятельно, возрастает риск того, что люди потеряют контроль над ИИС. Поэтому есть необходимость изучения ценностной доминанты для ИИС, как основного ядра с целью принятия моральных решений во благо человечества.

Искусственные интеллектуальные системы: техническая и нормативная сторона. ИИС — это разработанные людьми программные (аппаратные) системы, которые, имея сложную цель, действуют в физическом или цифровом измерении, воспринимая свою среду через сбор данных, интерпретируя собранные структурированные или неструктурированные данные, рассуждая на основе знаний или результатов обработки информации, полученной из этих данных. ИИС могут принимать решения о наиболее эффективных действиях для достижения поставленной цели, использовать символические правила или изучать числовую модель, а также адаптировать свое поведение, анализируя, как на среду влияют их предыдущие действия [3].

Научно-технический прогресс предполагает вероятность того, что ИИС, соответствующие человеческому мышлению или даже превосходящие его, будут созданы в ближайшем будущем. Не обремененные биологическими ограничениями, такими как узкая память и низкая скорость биохимической обработки, киборг-машины будут более интеллектуальными, чем человек, что вызовет потрясения цивилизации. Согласно мнению экспертов, к концу XXI столетия ИИС перейдут на суперинтеллект, что может спровоцировать неуправляемость ими. Сами разработчики ИИС беспокоятся о том, чтобы данные системы не вызвали разрушительных эффектов. Поскольку потенциальные риски велики, современное научное сообщество отмечает необходимость разработки ИИС, ориентированных на ценностное основание общества. Связь ИИС и традиционных ценностей рассмотрена в отчете [4].

В данном отчете осуществляется попытка выработать общечеловеческие ценности, которые будут интегрированы в ядро ИИС и составят моральный фундамент искусственного интеллекта. Определение ценностного основания интеллектуальной системы заложено в целях и действиях, которые реализует искусственный интеллект на пользу людям. Рассогласованность и соотношение ценностей необходимо рассмотреть с двух точек зрения, сосредоточив внимание на действии искусственного интеллекта (индивидуальный уровень) и на функционировании социотехнических систем (системный уровень).

Исследования проблемы рассогласованности и соотношения ценностей ИИС определяют техническую и нормативную стороны вопроса. Техническая сторона состоит в решении проблемы внедрения человеческих ценностей в ИИС. Нормативная проблема связана с вопросом: с какими ценностями соотнести системы искусственного интеллекта? Необходимо выяснить, какие цели или ценности внедрять в ИИС, учитывая, что существует множество ценностей, моральных принципов и противоречий.

Ясон Габриэль, сотрудник Лондонского представительства «Этика и общество», отмечает, что понятие «ценность» выступает как заполнитель для ИИС, как предполагаемая инструкция, осознанное предпочтение или желание. Человек является главным определителем ценностного основания киборг-машины, гарантируя полезность искусственного интеллекта. Выделяются два подхода морального апгрейда ИИС: программирование по принципам алгоритмизации с внесением морального информационного ядра в структуру искусственного интеллекта машины и обучение системы на данных (контролируемых или неконтролируемых). Гэбриэл предлагает проанализировать ценностный потенциал цивилизации и выделить «истинные» моральные ценности для ИИС, которые будут выражать интересы всего мирового сообщества, несмотря на широко распространенные различия в моральных убеждениях людей [5].

Проблема выбора ценностей искусственных интеллектуальных систем. Идея обучения заключается в том, что желательность любой последовательности состояний выражается как сумма немедленных вознаграждений, связанных с каждым состоянием в последовательности. Искусственный интеллект может читать и понимать истории, может усвоить ценности, негласно сохраняемые культурой. Выбор ценностей в системе искусственного интеллекта обучения теоретически может быть достигнут путем предоставления ИИС сигнала вознаграждения, побуждающего решить данную задачу и препятствующего выполнению действий, которые могут быть сочтены вредными для людей.

Однако идею выбора ценностей с помощью обучения можно расценивать как этически неадекватное и недостаточное для руководства искусственными системами при принятии решений, связано со смещением данных и проблемой обобщения, а также поскольку адекватностью функций вознаграждения для представления сложных во времени норм. В качестве альтернативы выбор ценностей ИИС можно осуществить посредством имитационного обучения. В статье «Добродетельная машина — старая этика для новых технологий» Николас Берберих и Клаус Дипольд, основываясь на трудах Аристотеля, демонстрируют, что решение проблемы выбора ценностей может обеспечить этика добродетели. Они предлагают использовать этику добродетели, заложив ее основные принципы в моральное ядро интеллектуальной системы. Этика добродетели прекрасно сочетается с современными исследованиями искусственного интеллекта и является многообещающей теорией морали; выбор пути этики добродетели к построению ценностного основания киборг-машин позволяет использовать гораздо более широкий подход, чем простое теоретическое суждение о возможных действиях [6, с. 23].

Основная идея руководителя Центра совмещения человека с искусственным интеллектом С. Рассела состоит в том, что для сохранения контроля необходимо разработать полезные ИИС, которые будут «скромными, альтруистическими и приверженными достижению наших целей» [7]. Он формулирует три принципа:

- 1) единственная цель машины — максимально реализовать человеческие предпочтения;
- 2) машина не в состоянии определить основополагающие предпочтения без человека;
- 3) конечным источником информации о человеческих предпочтениях является человеческое поведение [7, с. 172, 173].

Рассел считает, что существует множество данных о человеческих действиях и об отношении к этим действиям. Таким образом, вместо того чтобы запрограммировать подробную этическую таксономию, ИИС должны делать выводы о человеческих ценностях, наблюдая поведение людей и подражая им. В ранних работах Рассел заменяет понятие «человеческие ценности» на «человеческие предпочтения». Он отказался от слова «ценность», поскольку оно часто понимается как «моральная ценность», а использование этого термина не понятно машине. Рассел настаивает на том, что не следует приравнивать соотношения ценностей к желанию привить машинам единую идеальную систему ценностей, которая может вызвать вопрос, значения следует кодировать или кому было дано право согласовывать ИИС с такими ценностями. Он использует слово «ценность» как технический термин, «синоним полезности которого измеряется сте-

пенью желательности чего-либо, от пищи до рая» [7]. Рассел объясняет понятие «предпочтения»: они охватывают все, что может волновать человека, приносить пользу. Альтруистическая машина должна стремиться к максимальной реализации человеческих предпочтений. Вместо того чтобы «вводить ценности» или предпочтения, машины «должны научиться лучше предсказывать для каждого человека, какую жизнь он предпочел бы, при этом осознавая, что предсказания весьма неопределенны и неполны» [7, с. 178]. Рассел считает, что ИИС могут изучить миллиарды различных моделей прогнозных предпочтений. Машина должна учитывать предпочтения множества людей, в связи с чем Рассел ссылается на утилитарный принцип «величайшего счастья для наибольшего числа людей». Ключевой вопрос: как машина может узнать о предпочтениях человека? Ответ на него очень прост: путем наблюдения за человеческим выбором, который раскрывает информацию о человеческих предпочтениях [7, с. 177].

Оказывается, Рассел заменяет понятие «ценность» на «предпочтение». Однако ценность — объективная категория, а предпочтение — субъективная. Ценности руководят действиями; они учитываются при принятии решений и при планировании деятельности. В самом общем смысле предпочтения — это оценки; они касаются вопросов, представляющих ценность. Но здесь возникает проблема. Можно представить ситуацию, когда разработчик беспилотного автомобиля спрашивает потенциальных клиентов, предпочли бы они автомобиль, который защитит людей, едущих в нем, или тех, кто находится на улице. Или, что еще хуже, хотят ли они ехать в машине, которая нарушает правила дорожного движения, игнорирует светофор или ту, которая быстро и безопасно доставляет владельца к месту назначения. Должен ли искусственный интеллект удовлетворять такие предпочтения? Если оставить в стороне моральные ценности и сосредоточиться только на человеческих предпочтениях, то трудно понять, как можно избежать ситуаций, в которых ИИС могли бы выполнять предпочтения аморальных систем. Поскольку в мире больше проявлений аморальных человеческих предпочтений, необходимо быть осторожными, чтобы не запрограммировать ИИС на что-то деструктивное. Однако всегда будут предприниматься попытки создать разрушительные ИИС, например для военных целей. Некоторые эксперты ставят под сомнение использование роботов в боевых действиях, особенно если такие машины будут иметь некоторую степень автономности, например, возможность самостоятельно выбирать цели для атаки с помощью оружия.

Таким образом, вопрос о программировании ИИС с позиций игнорирования аморальных предпочтений и отказа от аморальных действий становится злободневным. ИИС должны уметь различать мо-

ральные и аморальные действия, поэтому, помимо включения ценностей или предпочтений, необходимо предоставить ИИС возможность морального выбора, чтобы решать, когда применяется определенный моральный принцип, и уметь ранжировать ценности в различных ситуациях, а для этого потребуется гораздо больше критериев. Для того чтобы ИИС обладали способностью морального выбора, необходимо самосознание. Если агент не осознает себя в той или иной ситуации и не знает альтернативных путей, которыми можно было бы следовать, тогда никакие моральные решения не могут быть приняты [8].

Эта аргументация указывает на то, что проблему морального выбора ИИС необходимо решать, объединив усилия не только ученых-информатиков, но и философов, юристов, экономистов, ученых-когнитивистов, психологов, ученых-педагогов и др. Успех морального выбора ИИС будет зависеть от того, насколько разработчики смогут интегрировать знания, полученные в этих дисциплинах. Анализ документов по этическому руководству ИИС [4] показал, что наиболее распространенным режимом управления, предлагаемым для работы с ИИС, является международное право в области прав человека — 64 % всех документов содержат ссылки на права человека.

Какими ценностям должны обладать системы искусственного интеллекта? Ответ на данный вопрос находим в современных документах, разработанных Европейской комиссией. В «Руководстве по этике для надежности искусственного интеллекта» утверждается, что «системы искусственного интеллекта должны уважать множественность ценностей и выбор людей, основные из них: уважение прав человека, демократия и верховенство закона» [9, с. 4]. В документе «Повседневная этика искусственного интеллекта» (IBM) указано: «Требуется осторожность, чтобы обеспечить чувствительность к широкому спектру культурных норм и ценностей. Каким бы пугающим ни казалось принятие во внимание систем ценностей, общая суть универсальных принципов заключается в том, что они представляют собой совместное явление» [10].

Международная некоммерческая ассоциация специалистов в области техники, мировой лидер в области разработки стандартов по радиоэлектронике, электротехнике и аппаратному обеспечению вычислительных систем и сетей разработала «Концепцию благополучия человека» [11], реализуемую с помощью автономных и интеллектуальных систем. В данном документе формулируется вопрос: если машины будут участвовать в человеческих сообществах в качестве самостоятельных систем, следуя социальным и моральным нормам сообщества, то чьи нормы взять за основу? С учетом разнообразия моральных норм и ценностей необходимо выработать универсальные этические принципы и признанные на международном уровне права человека, которые будут заложены в ценностное ядро ИИС.

В документе обращаются к этическим нормам и принципам практики западной, восточной, африканской культур или традиций. Вопросы духовного мира человека, ценности культуры и других народов мирового сообщества не учтены или просто растворяются в ценностном многообразии мирового сообщества. Внимание сосредоточивается только в материальной плоскости, все духовное пренебрегается. Автономным и интеллектуальным системам следует уделять первоочередное внимание, главной их целью должны являться уважение неотъемлемых основных прав и достоинства, а также повышение духовного и физического здоровья, благосостояния человека и экологической устойчивости.

Данный перечень основных документов указывает на то, что необходимо учитывать множественность ценностей, в то же время уделяя приоритетное внимание конкретным ценностям, прописанным в международном праве. Однако свобода выбора ценностей может породить хаос в иерархической структуре ценностей, поскольку существует несводимость ценностей разных сообществ. Как показала Сьюзан Вольф в статье «Два уровня плюрализма», эта несводимость может происходить на двух уровнях принятия решений: на уровне ценностей и на уровне моральных систем. Сначала следует определить иерархическую структуру ценностей, их приоритет для ценностного основания ядра интеллектуальной системы, выявить, что первично, а что вторично.

Кроме того, необходимо понять, какая концепция более предпочтительна для ИИС (например, кантианство, утилитаризм этика добродетели или этика благоговения перед жизнью) [12]. По мнению И. Берлина, «...у нас действительно есть общие ценности, которые образуют мосты между людьми и делают возможным взаимодействие между культурами во времени и пространстве. Но наши ценности — наши, а их ценности — их» [13, с. 11]. Можно критиковать ценности других культур, но нужно понимать динамику их появления и развития. Здесь Берлин рассматривает процесс трансформации культурно-ценностных доминант разных систем, поддерживая плюрализм мира объективных ценностей или целей, которые люди преследуют ради самих себя. Цели и моральные принципы множественны, но их не бесконечно много, и они должны оставаться «в пределах человеческого горизонта» [13, с. 12]. Следует уметь различать объективно хорошие и плохие формы жизни, осуждать рабство, расизм, нацизм и другие негативные проявления. Вопрос о включении в ИИС ценностных предпочтений требует самого широкого обсуждения не только в верхних эшелонах власти, в среде ученых и философов, но и общественности в целом. С одной стороны, ИИС могут стать таким оружием для всего человечества, аналогов которому нет, могут изменить саму природу человека, о чем тревожится сегодня весь

мир [14, с. 54], а с другой — могут улучшить качество жизни человека и способствовать ее продолжительности.

Заключение. Научно-технический прогресс вывел человечество на новый виток развития: были созданы ИИС, которые, по мнению ученых, приносят пользу и, следовательно, должны соответствовать человеческим нормам поведения и ценностным доминантам. Однако по словам западных представителей научной мысли, термин «ценность» для ИИС выступает как заполнитель программного кода машины, поэтому его необходимо заменить предпочтением. В действительности можно ограничиться тем, чтобы заставить ИИС выполнять человеческие предпочтения вместо морального выбора. Но здесь возникает проблема, заключающаяся в том, что ИИС необязательно будут действовать морально, а потенциально смогут удовлетворять и аморальные предпочтения человека. Если ценностное ядро искусственного интеллекта можно будет программировать или обучать, то данная система будет выступать как положительная для человека, который программирует под себя, и как отрицательная для всего человечества, неся угрозу безопасности существования. Таким образом, проблема морального выбора ИИС остается открытой, а вопрос внесения моральных принципов и моральных ценностей в программную среду — нерешенным.

Разработчики ИИС сталкиваются с важными проблемами, такими как существование универсальных ценностей для всего человечества и необходимость для ИИС учитывать культурные различия стран. Эти проблемы будут решены без участия мирового сообщества, так как кто платит, тот и заказывает соответствующие ИИС с набором тех или иных ценностей в аксиологическом ядре искусственного интеллекта. С учетом того что ИИС в основном создаются для военной промышленности, а затем для общества, становится понятно, каким ценностям будет отдан приоритет (например 26 июня 2020 г. на канале Яндекс.Эфир знаменитый робот от Boston Dynamics предстал с оружием в руках, его обучали убивать). Таким образом, согласование ценностей должно начинаться с поиска согласия по отрицательным ценностям или с ответа на вопросы: что не хотим терять и что не хотим создать? Следует иметь в виду, что функция морали состоит в удовлетворении основных человеческих ценностей, которые разделяли бы все сообщества.

ЛИТЕРАТУРА

- [1] Sutrop M. Should we trust artificial intelligence? *Trames*, 2019, vol. 23, no. 4, pp. 499–522.
- [2] Brundage M., Avin S. et al. *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*. URL: <https://maliciousaireport.com> (дата обращения 05.03.2021).

- [3] *EU Commission. A Definition of AI: Main Capabilities and Disciplines*. URL: <https://www.aepd.es/media/docs/ai-definition.pdf> (дата обращения 05.03.2021).
- [4] Fjeld J., Achten N., Hilligoss H., Nagy A.C., Srikumar M. Principled artificial intelligence: mapping consensus in ethical and rights-based approaches to principles for AI. *Berkman Klein Center Research Publication*, 2020, no. 1, pp. 1–71.
- [5] Gabriel I. Artificial intelligence, values, and alignment. *Minds and Machines*, 2020, vol. 30, pp. 411–437.
- [6] Berberich N., Diepold K. The virtuous machine — old ethics for new technology? *ArXiv*, 2018, pp. 1–25. URL: <https://deepai.org/publication/the-virtuous-machine-old-ethics-for-new-technology> (дата обращения 05.03.2021).
- [7] Russell S. *Human Compatible. AI and the Problem of Control*. London, Allen Lane, Penguin Books, 2019, 376 p.
- [8] Etzioni A. Incorporating ethics into artificial intelligence. In: Etzioni A., ed. *Happiness is the Wrong Metric: A Liberal Communitarian Response to Populism*. Luxemburg, Springer, 2018, pp. 235–252. URL: <https://www.springer.com/gp/book/9783319696225> (дата обращения 05.03.2021).
- [9] *EU Commission Ethics Guidelines for Trustworthy AI*. URL: <https://ec.europa.eu/futurium/en/ai-alliance-consultation> (дата обращения 05.03.2021).
- [10] *Everyday Ethics for Artificial Intelligence*. URL: <https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf> (дата обращения 05.03.2021).
- [11] *Ethically Aligned Design. A Vision for Prioritizing Human well-being with autonomous and intelligent systems*. URL: <https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead1e.pdf> (дата обращения 05.03.2021).
- [12] Wolf S.M. Two levels of pluralism. *Ethics*, 1992, vol. 102, no. 4, pp. 785–798.
- [13] Berlin I. The Pursuit of the Ideal. In: Hardy H., ed. *The Crooked Timber of Humanity*. Princeton & Oxford, Princeton University Press, 2013, 345 p.
- [14] Майкова В.П., Молчан Э.М. *Предначала философии виртуальной реальности*. Москва, Спутник+, 2020, 65 с.

Статья поступила в редакцию 07.04.2021

Ссылку на эту статью просим оформлять следующим образом:

Майкова В.П., Молчан Э.М., Майков А.И. Проблемы выбора ценностей для искусственных интеллектуальных систем. *Гуманитарный вестник*, 2021, вып. 2. <http://dx.doi.org/10.18698/2306-8477-2021-2-715>

Майкова Валентина Петровна — д-р филос. наук, профессор кафедры «Педагогика, психология, право, история и философия» Мытищинского филиала МГТУ им. Н.Э. Баумана. e-mail: valmaykova@mail.ru

Молчан Эдуард Михайлович — канд. пед. наук, доцент кафедры «Философия» Московского государственного областного университета. e-mail: ed.molchan2015@yandex.ru

Майков Андрей Игоревич — канд. техн. наук, доцент кафедры «Защита информации» МГТУ им. Н.Э. Баумана. e-mail: shtuder@mail.ru

The problems of choosing values for artificial intelligence systems

© V.P. Maykova¹, E.M. Molchan², A.I. Maykov³

¹Bauman Moscow State Technical University,
Mytischki branch, Mytischki, 141005, Russia

²Moscow Region State University, Mytischki, 141014, Russia

³Bauman Moscow State Technical University, Moscow, 105005, Russia

The paper considers the problems of choosing values for artificial intelligence systems, which in the near future will become independent agents of objective reality and will be included in human life activity. The study shows that in the context of urgent problems of the changing reality, modern society faces the task of both the choice of values for artificial intelligence systems and the technical and regulatory stages of their implementation. The concept of “value” is replaced by a preference that acts as an information-idealized form embedded in the software of cyborg systems. Findings of research show that the choice and inclusion of values in cyborg machines is associated with the customer’s priorities. However, due to the self-learning of artificial intelligence systems, unpredictable synergistic effects are possible, or intelligence systems will switch to superintelligence, which can result in their uncontrollability. This requires a new methodology that determines the basis of modern scientific and technological progress, cognitive, i.e. epistemological, information development in order to form a legal and ethical framework in response to the challenges of digital globalization, which is transforming into globalizing individualism.

Keywords: artificial intelligence systems, values, ethics, moral values, moral norms, moral principles, superintelligence, cyborg machine

REFERENCES

- [1] Sutrop M. Should we trust artificial intelligence? *Trames*, 2019, vol. 23, no. 4, pp. 499–522.
- [2] Brundage M., Avin S., et al. *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*. Available at: <https://maliciousaireport.com> (accessed March 5, 2021).
- [3] EU Commission. *A Definition of AI: Main Capabilities and Disciplines*. Available at: <https://www.aepd.es/media/docs/ai-definition.pdf> (accessed March 5, 2021).
- [4] Fjeld J., Achten N., Hilligoss H., Nagy A.C., Srikumar M. Principled artificial intelligence: mapping consensus in ethical and rights-based approaches to principles for AI. *Berkman Klein Center Research Publication*, 2020, no. 1, pp. 1–71.
- [5] Gabriel I. Artificial intelligence, values, and alignment. *Minds and Machines*, 2020, vol. 30, pp. 411–437.
- [6] Berberich N., Diepold K. The virtuous machine — old ethics for new technology? *ArXiv*, 2018, pp. 1–25. Available at: <https://deepai.org/publication/the-virtuous-machine-old-ethics-for-new-technology> (accessed March 5, 2021).
- [7] Russell S. *Human Compatible. AI and the Problem of Control*. London, Allen Lane, Penguin Books, 2019, 376 p.
- [8] Etzioni A. Incorporating ethics into artificial intelligence. In: Etzioni A., ed. *Happiness is the Wrong Metric: A Liberal Communitarian Response to Populism*.

- Luxemburg, Springer, 2018, pp. 235–252. Available at: <https://www.springer.com/gp/book/9783319696225> (accessed March 5, 2021).
- [9] *EU Commission Ethics Guidelines for Trustworthy AI*. Available at: <https://ec.europa.eu/futurium/en/ai-alliance-consultation> (accessed March 5, 2021).
- [10] *Everyday Ethics for Artificial Intelligence*. Available at: <https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf> (accessed March 5, 2021).
- [11] *Ethically Aligned Design. A Vision for Prioritizing Human well-being with autonomous and intelligent systems*. Available at: <https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead1e.pdf> (accessed March 5, 2021).
- [12] Susan M. Wolf. Two levels of pluralism. *Ethics*, 1992, vol. 102, no. 4, pp. 785–798.
- [13] Berlin I. The Pursuit of the Ideal. In: Hardy H., ed. *The Crooked Timber of Humanity*. Princeton & Oxford, Princeton University Press, 2013, 345 p.
- [14] Maykova V.P., Molchan E.M. *Prednachala filosofii virtualnoy realnosti* [The beginning of the philosophy of virtual reality]. Moscow, Sputnik+ Publ., 2020, 65 p.

Maykova V.P., Dr. Sc. (Philos.), Professor, Department of Pedagogy, Psychology, Law, History and Philosophy, Mytischki branch, Bauman Moscow State Technical University. e-mail: valmaykova@mail.ru

Molchan E.M., Cand. Sc. (Ped.), Assoc. Professor, Department of Philosophy, Moscow Region State University. e-mail: ed.molchan2015@yandex.ru

Maykov A.I., Cand. Sc. (Eng.), Assoc. Professor, Department of Information Security, Bauman Moscow State Technical University. e-mail: shtuder@mail.ru