

КУЛЬТУРА ВИНЫ И ОТВЕТСТВЕННОСТЬ ВРАЧА ПРИ ИСПОЛЬЗОВАНИИ СИСТЕМ ПОДДЕРЖКИ ПРИНЯТИЯ ВРАЧЕБНЫХ РЕШЕНИЙ



Углева Анастасия Валерьевна — кандидат философских наук, PhD, доцент Школы философии и культурологии.
Национальный исследовательский университет «Высшая школа экономики».
Российская Федерация, 101000 Москва, ул. Мясницкая, д. 20.
ORCID: 0000-0002-9146-1026
augleva@hse.ru

Аннотация. Цель данной статьи заключается в обосновании необходимости пересмотра структуры моральной и уголовной ответственности врача в общественной системе, в которой ему навязывается культура вины, то есть единоличная персональная ответственность за любую медицинскую ошибку, совершенную в том числе с участием системы поддержки принятия врачебных решений. От патерналистской модели взаимоотношений «врач–пациент» начиная с 1970-х годов совершен переход к модели совместного ответственного поведения, в которой каждой стороне вменяется в обязанность знание и понимание природы выносимых клинических решений, что должно быть подкреплено в случае включения в процедуры диагностики и лечения технологий искусственного интеллекта особой формой информированного согласия как для пациента, так и для врача. При этом фактически отсутствует четкая регламентация процесса сообщения и работы с информацией и персональная ответственность высшего менеджмента за формирование системы медицинского обеспечения. В силу принципиальной непрозрачности

работы постоянно развивающихся технологий искусственного интеллекта окончательное решение должно оставаться за врачом. Однако обоснование его ответственности в актуальных технологических условиях требует нового общественного договора внутри профессиональной среды относительно условий внедрения систем поддержки принятия врачебных решений в широкую медицинскую практику.

Ключевые слова: искусственный интеллект, доверие, медицинская ошибка, ответственность врача, система поддержки принятия врачебных решений, культура вины, автономия.

Ссылка для цитирования: Углева А.В. Культура вины и ответственность врача при использовании систем поддержки принятия врачебных решений // Человек. 2023. Т. 34, № 3. С. 9–23. DOI: 10.31857/S023620070026423-2

Высокая степень ответственности, как моральной, так и юридической, которую несет врач в ситуации принятия сложного клинического решения, обуславливает необходимость осмысления обстоятельств этого выбора и поиска возможностей для эффективной помощи ему. Разработка систем поддержки принятия врачебных решений (СППВР) на основе новейших технологий искусственного интеллекта (ИИ) с машинным обучением вселяет естественную надежду на совершенствование и масштабирование их применения в широкой медицинской практике при условии ответственного проектирования. Эта надежда связана, с одной стороны, с минимизацией рутинной части работы врача и повышением маржинальности, а с другой — с ускорением процесса поиска оптимальных решений на основе обработки масштабных баз данных о пациентах. Более сложной и менее предсказуемой формой таких систем является нейронная сеть, позволяющая осуществлять процедуры прогнозирования заболевания с высокой степенью вероятности, однако порождающая ряд существенных этических вопросов, связанных с принципиальной непрозрачностью механизма, поскольку основана она на принципах глубинного обучения с различного рода предвзятостями [Nebeker et al., 2019] и погрешностями, которые она до сих пор допускает в силу низкой робастности: небольшое изменение входных данных приводит к полному циклу переобучения системы и минимальной возможности интерпретации полученных результатов.

Нетривиальными и крайне актуальными этическими проблемами в условиях все более интенсивного внедрения технологий ИИ являются проблемы доверия к ним, справедливости распределения ответственности в случае допущения медицинских ошибок, а также предвзятостей, которые, как правило, сопровождают

человеческие решения, но тем более будут сопровождать решения, принимаемые ИИ, потому как пишущие алгоритмы люди будут автоматически прописывать эти предвзятости в цифровых кодах. Под предвзятостью имеется в виду свойство системы ИИ (СИИ), заключающееся в принятии ошибочных решений, связанных со статистической смещенностью обучающей выборки исходных данных. Доверие к СИИ — важнейшее условие, определяющее возможность применения этих систем при решении ответственных задач обработки данных. Примером таких задач является поддержка принятия врачебных решений с помощью технологий ИИ, которые представляют собой комплекс технологических решений, позволяющих имитировать когнитивные функции человека (включая самообучение и поиск решений без заранее заданного алгоритма) и получать результаты, сопоставимые с результатами интеллектуальной деятельности человека.

Объектами доверия могут выступать как выводы или итоги работы СИИ, действующие как «черный ящик», так и сами процессы, которые приводят к этим выводам. Возникает необходимость поиска этической модели для обоснования человекоцентричной СИИ. В качестве таковой можно рассматривать релейабелизм как этико-эпистемическую теорию, предлагающую релевантный способ обоснования доверенного ИИ, выносящего суждения, достоверность которых непосредственно зависит от надежности их источника. Если в случае человека таковыми являются его когнитивные способности (память, восприятие, интроспективные состояния и процессы), то в случае технологий ИИ такие когнитивные способности весьма ограничены, хотя именно их создание составляет амбициозную задачу для разработчика. Пока сомнительной видится всякая попытка воспроизведения этики добродетели применительно к *нечеловеческим агентам*. Но процесс принятия СИИ-решений может рассматриваться как надежный настолько, что мы можем ему доверять, как и *человеческому актору*. Это доверие требует строго соблюдения пяти принципов этичности ИИ, среди которых: 1) непричинение вреда; 2) благодеяние; 3) автономия, которая, с одной стороны, предполагает право пациента на принятие самостоятельного решения в отношении своего медицинского обслуживания и лечения на основе своих собственных ценностей и убеждений (этот принцип подчеркивает важность информированного согласия), а с другой — интеллектуальную автономию врача, которая становится весьма уязвимой в случае внедрения ИИ в клиническую практику прежде всего потому, что есть риск своего рода укоренения зависимости, то есть постепенного снижения его когнитивных и моральных навыков; 4) справедливость (в том числе противодействие разнообразных

А.В. Углева

Культура вины и ответственность врача при использовании систем поддержки принятия врачебных решений

дискриминационных эффектов от решений СИИ по причине, как правило, необъективных данных обучения, которые могут использоваться для подготовки моделей и оценки их прогностической эффективности [Barocas et al., 2020]. Это тем более важно в условиях отсутствия единой базы данных о пациентах. Упомянутые первые четыре принципа отсылают к четырем биоэтическим принципам (так называемая джорджтаунская мантра); 5) принцип объяснимости, включающий в себя и эпистемологический смысл понятности (прозрачности механизма), и этический смысл подотчетности (принцип ответственности) [Floridi, Cowls, 2022].

Таким образом, для медицины огромное значение имеет проблема предвзятости алгоритма, которая составляет основную трудность при определении этичности СИИ в двух аспектах. Первый связан с конфликтом между разными этическими принципами (например, благодеяния и автономии), второй — с проблемой выбора конкретной реализации отдельного принципа в некоторой реальной ситуации (например, не всегда ясно, что значит действовать справедливо) [Krijger, 2021]. Вторая проблема, которую иногда связывают с недоопределенностью нормативных высказываний [Feenberg, 2017], осложняется тем, что предвзятости или иные неприемлемые ценностные предпочтения могут быть плотно и глубоко «вшиты» в программный код ИИ [Feenberg, 2005]. В итоге неоднозначным и слишком подвижным метрикам этической валидации медицинского ИИ могут противостоять плотно укорененные предвзятости, например, в отношении качества жизни людей с ограниченными возможностями, понимания медицинской нормы и т.п. Многие из этих важных этических аспектов выливаются в теоретические и практические вопросы, касающиеся ответственности и ее распределения [см., напр., Bleher, Braun M., 2021].

СППВР призвана помогать специалистам путем предоставления им медицинских рекомендаций, подкрепленных доказательствами, которые в противном случае не были бы доступны в разумные сроки [Di Nucci, 2019]. Учитывая растущие технические возможности и эпистемическую зависимость специалиста, возникает вопрос, является ли и в какой степени профессионал по-прежнему полностью ответственным за принятие клинических решений? Несмотря на обширную теоретическую литературу по СППВР в медицине, в настоящее время практически отсутствуют сколько-нибудь значимые исследования пусть даже прогностического характера моральной и юридической ответственности врача за решения, принятые с помощью этой системы, хотя речь идет о радикальной трансформации сущностных представлений о его моральной агентности при наличии персональной ответственности, в том числе уголовной.

Технологии ИИ не обладают самостоятельной когнитивной способностью, вследствие этого постоянно увеличивается разрыв между решением, принятым ИИ, и способностью конкретного специалиста (врача) понять, как это решение было принято и почему оно такое, какое есть. Это ставит вопрос о доверии к решениям, принимаемым СИИ [Ну, 2020]. Доверие заключается в понятности работы ИИ пользователю, так как должно быть сформировано ощущение правильности выбора варианта решения проблемы программой-советчиком.

Возможные негативные последствия, связанные с пограничными уровнями доверия к системе, могут приводить к чрезмерному доверию к неидеальной системе, что чревато медицинскими ошибками¹, или, напротив, к чрезмерному недоверию к системе и, как следствие, к отказу от ее использования. Пока ясно одно: почти все знания, производимые сегодня системами ИИ, имеют исключительно информационный характер. Почти нет инструментов, которые обобщали бы пользовательский опыт использования интеллектуальной системы, основанной на нейросетях. Сегодня СИИ в некотором смысле имитирует продукт человеческих действий. Чтобы приблизиться к возможностям человеческого интеллекта, ИИ должен обладать способностью обобщать знания [Chollet, 2019], осмысливать свои действия и вербализировать свой опыт.

¹ Под медицинской ошибкой понимается «неспособность запланированного действия завершиться согласно изначальному плану (ошибка исполнения), использование неправильного алгоритма для достижения цели (ошибка планирования)» [Reason, 1990]. Медицинские ошибки — третья причина смертности в мире [Makary, Daniel, 2016: 2] и одно из ключевых условий быстрого профессионального выгорания врачей, сопровождающегося тревогой по поводу возможного повторения ошибки в будущем [Waterman et al., 2007], и потери профессионального авторитета. Эта тревога во многом подкрепляется культурой вины, в соответствии с которой, как правило, именно лечащий врач несет единоличную персональную ответственность за допущенные ошибки. Об этом свидетельствует и экспоненциальный рост уголовного судопроизводства и количества обвинительных приговоров. Культура вины влияет на стиль менеджмента, в который входит не только совокупность дисциплинарных взысканий, но и регулирование межличностных отношений (внутри профессиональных коллективов, в отношении «врач—пациент» и пр.). Культура вины поддерживается практикой «закрытых дверей», когда медицинские ошибки не являются предметом регулярного коллективного обсуждения. Уникальным в этом отношении случаем является Кодекс Американского колледжа врачей, вменяющий врачам в обязанность раскрывать пациентам информацию о процедурных ошибках, допущенных в ходе оказания ими медицинской помощи, если такая информация является существенной для лечения и благополучия пациента [American College of Physicians, 1998]. При тенденции к широкому внедрению ИИ в клиническую практику тем более возрастает ответственность врача, в том числе за ошибку ИИ. Во избежание существенного оттока кадров требуется жесткая регуляция процесса со стороны контролирующих органов.

А.В. Углева

Культура вины
и ответствен-
ность врача
при использова-
нии систем под-
держки приня-
тия врачебных
решений

Возникает парадокс: человек может создать инструмент, сложность которого превышает его когнитивные способности, и поэтому он не может эффективно его использовать. Чтобы разрешить этот парадокс, необходимо переосмыслить роль и возможности человека в созданной им технологической реальности, а также разработать новую концепцию совместного развития машинного и человеческого интеллекта, действующую на эволюционной основе [Krinkin et al., 2022]. В некоторых случаях люди воспринимают ИИ как инородное существо, непонятное и порой представляющее некую опасность для человека. Преодолеть это ограничение возможно, если обучение интеллектуальных систем будет происходить в процессе их эксплуатации. В этом случае разрыв в общении между врачом и ИИ будет сокращаться.

Исследования [Lepri et al., 2021] показали, что способ объяснения влияет на восприятие пользователями справедливости алгоритма. Существуют индивидуальные различия, которые определяют, как люди реагируют на различные виды объяснений. Следовательно, восприятие СИИ зависит от способа обучения, а значит, антропоцентрическая система отсчета в развитии ИИ не только законна, но и необходима, и человек не может быть исключен из схемы развития ИИ².

До тех пор, пока системы ИИ будут конкурировать с человеческими ресурсами и пытаться заменить их, человечество будет сталкиваться с проблемой непринятия ИИ [Hu, 2020]. Доверие человека к инструменту (технологии) может быть основано на нескольких принципах: понимании алгоритма его работы и способа получения результатов; успешном опыте его использования коллегами; доверие к его разработчику. Индивидуальный характер использования персонализированных инструментов (например, интеллектуальных помощников) не позволяет ссылаться на других пользователей. Единственное возможное решение — принятие пользователем непосредственного участия в создании и обучении технологии. Именно в этом случае пользователю может быть понятна его собственная ответственность за принимаемые совместно с СППВР решения, а их эффективность напрямую будет зависеть от достаточности клинического опыта у конкретного специалиста.

В контексте взаимодополняющих возможностей человеческого и искусственного интеллекта люди доказали свое превосходство в различных ситуациях, требующих свободы мышления наряду со скоростью, эмоциональностью, стереотипностью и

² См., к примеру, результаты качественного исследования [Funer et al., 2022].

так называемой человеческой интуицией. Машины эффективны в решении повторяющихся задач, требующих быстрой обработки огромных объемов данных. Подход гибридного интеллекта позволяет человеческим экспертам применять предсказательную мощь ИИ, одновременно используя собственную интуицию и эмпатию для принятия решений на основе прогнозов ИИ, что требует непрерывного параллельного обучения как ИИ, так и естественного интеллекта [Dellermann et al., 2019], исходя из главного этического принципа врача — ненанесения вреда. Однако эта категория тоже составляет некоторую проблему в силу своей недоопределенности, несмотря на огромный пласт посвященной ей профессиональной литературы. Вред здоровью является исключительно правовым понятием. С судебно-медицинской точки зрения для определения понятия «вред здоровью» используются медицинские критерии, на основании которых устанавливается степень его тяжести. В самой же медицине «вред» используется в случае оценки показаний и противопоказаний к операции, то есть в связке «вред–польза–риск», в которой ключевым оказывается понятие «риск». Выявление основных рисков и вреда (для человека и общества) составляет основу этического осмысления проблемы внедрения технологий ИИ в клиническую практику. К числу таких рисков можно отнести следующие: деградацию фундаментальных эмпирических знаний и навыков, непосредственно связанных с элементарными рутинными навыками и процедурами, лежащими в основе тех или иных видов научно-технической деятельности; размывание ответственности за решения, принятые с участием экспертных систем на основе ИИ; десубъективацию преимущественно человеческих видов деятельности, таких как врачебная деятельность и др.

Доступность ИИ в перспективе может снизить критичность, понимание ошибки того, работает ИИ или нет. ИИ может быть использован для системного аудита качества диагностики в ходе ретроспективного анализа (автоматизированного пересмотра исследований с целью выявления пропущенных патологий). И это безусловное благо. Одновременно с этим, например, оценка качества рентгенологических снимков и гистологических срезов — по-прежнему сфера ответственности врача, пренебрежение которой может приводить к неожиданным решениям со стороны СИИ.

Известно, что если ошибся морфолог, то клиницист-онколог может принять неверное решение о проведении операции. Однако жаловаться в судебные инстанции потерпевшая сторона будет не на морфолога, а именно на хирурга. Поэтому хирурги доверяют очень ограниченному кругу диагностов, понимая, что ответственность за себя в первую очередь примет именно лечащий врач.

А.В. Углева

Культура вины и ответственность врача при использовании систем поддержки принятия врачебных решений

И большой вопрос: готовы ли они роль посредника между ними передать ИИ? Тем более в условиях российского судопроизводства. Система в нашем случае, как правило, винит врача. В других цивилизованных странах такого фокуса на эти проблемы почти нет, потому что в них медицинские ошибки разбираются в гражданском судопроизводстве. Ответчик — это клиника, которая выбрала себе программное обеспечение и в случае предъявления к ней гражданского иска имеет право обратиться к разработчику с требованием компенсации. Такова справедливая и понятная схема распределенной ответственности. В России же преобладает уголовное судопроизводство, поддерживающее культуру вины лечащего врача.

Дискуссии об этичности технологий ИИ в последние несколько лет, происходящие на разных площадках, убеждают в том, что разработчик в нынешних условиях постановки вопроса наделяется исключительным статусом неприкосновенного субъекта, исключенного из дискурса ответственности за медицинские ошибки, совершенные в том числе при участии разработанных им алгоритмов. Таким образом, есть все основания предполагать, что с массовым внедрением таких технологий в клиническую практику будет возрастать количество обвинительных приговоров против врачей и, как следствие, произойдет существенный отток кадров из отрасли. Поэтому особого внимания заслуживает феномен *deskilling*'а, то есть устаревания определенных трудовых навыков, а также морального разучения (*moral deskilling*). Поскольку моральные навыки являются важными предпосылками для эффективного развития практической мудрости и добродетельного характера практикующего врача, понимание неоднозначного и открытого горизонта влияния новых технологий как на трудовые, так и на моральные навыки может помочь нам не только лучше определить моральные риски и преимущества таких технологий, но и предвидеть новые возможности разработки и ответственного проектирования, которые конкретным образом реализуют потенциальные преимущества этих технологий, сводя к минимуму вред, который они могли бы нанести нашим собственным моральным способностям, в том числе и способности быть ответственными за принимаемые нами решения.

Этика ответственности — пример реляционной этической модели, основанной на кооперативном соглашении. Существует дискуссия о точном количестве реляций, составляющих основу этого соглашения [Langanke et al., 2016]. В то время как более ранние работы, посвященные этике ответственности, в основном обсуждают вопрос отношения субъекта и объекта ответственности и страдают от недостатка ясности в понимании места и роли

нормативного стандарта [Jonas, 1985], более новые концепции расширяют структуру отношения ответственности с целью отразить целостный взгляд на проблему и прояснить ее нормативные последствия.

Общим местом в современной литературе является выделение четырех существенных элементов в структуре ответственности: субъект ответственности (А), объект ответственности (В), адресаты ответственности (С), нормативные стандарты (D) [Coeckelbergh, 2020; Werner, 2021]. Отношения ответственности могут быть выражены так: А несет ответственность перед С за В на основании D. Если субъектом и адресатом ответственности могут быть как человек, так и учреждение, то объектом являются действие и/или его последствия. Нормативное основание для ответственного поведения, или нормативный стандарт, представляет собой набор явных (например, правовых) или неявных (например, этических) правил. Исходя из этой реляционной концепции идея ответственности может быть рассмотрена либо ретроспективно³, либо проспективно. Во втором случае ответственность определяется в контексте заранее сформулированных и обоснованных моральных требований, предвещающих внедрение новой технологии в практику с целью предупреждения еще не нанесенного вреда. И тогда должны быть строго регламентированы действия врача, который окажется не согласен с рекомендацией СППВР в силу несовместимости их эпистемических характеристик (в случае системы — ее принципиальной непрозрачности), что подчеркивает проблематичность атрибуции ответственности [Grote, Berens, 2020; Braun et al., 2020; Funer, 2022]. В таких случаях может потребоваться обучение специалистов соответствующим компетенциям в качестве меры предосторожности для снижения риска причинения вреда в ситуации когнитивного несоответствия между мнениями врача и экспертной системы, когда вывод ИИ не может быть верифицирован или фальсифицирован.

И тогда в ситуации принципиальной неопределенности субъекта ответственности возникает возможность трех различных

А.В. Углева

Культура вины и ответственность врача при использовании систем поддержки принятия врачебных решений

³ То есть в случае оценки последствий уже причиненного вреда и определения фигуры ответственного. Начиная с 1970-х годов обнаруживается переход в общей практике отношений «врач–пациент» от сугубо патерналистского понимания предмета ответственности к концепции совместного принятия решений, которая подразумевает со-ответственность врачей и пациентов, а также совместную ответственность разработчиков и производителей [Braun et al., 2020: 47]. Перспективность такого подхода в том, что он позволяет перейти от культуры вины в отношении врача к поиску эффективных решений для предотвращения рецидивов.

сценариев решения проблемы, таких как: а) (само)распределение ответственности; б) автономия в принятии решений, в) потребность в расширении профессионального опыта.

В первом случае имеет место сомнение в том, кто является субъектом ответственности — врач, совершивший ошибку, разработчик конкретного алгоритма, поставщики, клиника, приобретающая и внедряющая соответствующую технологию, или инстанции, осуществляющие регуляторный контроль, например учреждения здравоохранения / надзорные органы? Сама СППВР такой ответственностью обладать не может в силу отсутствия у нее самостоятельного интеллекта, субъектности и, как следствие, моральной агентности. Напротив, разработчики и провайдеры несут ответственность за надежное функционирование системы и, следовательно, должны нести ответственность в ситуациях, когда причиной ущерба является ошибочное программирование или неточный прогноз, основанный на неверном взвешивании статистических данных. Регулирующие инстанции и закупочные учреждения, такие как больницы или службы ухода за больными людьми, должны обеспечивать дополнительные гарантии надежности. Однако институциональная ответственность существенно зависит от конкретной директивы в отношении использования рекомендательной системы: например, если сотрудники обязаны использовать ее или система используется для решения клинических проблем в условиях нехватки медицинского персонала, справедливым кажется то, что учреждение должно нести основную ответственность. При этом стоит обратить внимание и на риск принуждения пользования услугами СППВР, что открывает пространство для манипулирования ответственностью со стороны учреждения по отношению к врачам, которые заведомо оказываются, таким образом, в уязвимом положении. Иначе говоря, важнейшей характеристикой профессионализма врача была, есть и остается его автономность как морального агента, вмещающая ему в обязанность несение ответственности за принимаемые решения, какими бы СППВР он ни пользовался. Несогласие с рекомендацией СППВР должно для самого специалиста поставить под сомнение его собственные суждения. Требуются и понимание работы алгоритма, и сущностная оценка данных, на которых системой было вынесено противоположное мнению специалиста решение.

В поиске правильного решения особое значение приобретает консолидированное мнение медицинского сообщества внутри конкретного учреждения, которое побуждает к взаимодействию и накоплению коллективного опыта обращения с конкретной СППВР. Причем между объяснимостью ИИ и доверием к ИИ можно установить определенную взаимосвязь через связку обоснования

надежности ИИ с необходимостью контроля за ним во время его использования. Прозрачность механизма способствует доверию к ИИ тогда и только тогда, когда она способствует обоснованному и гарантированному парадигматическому доверию к ИИ, то есть доверию при наличии обоснованного убеждения в том, что ИИ заслуживает доверия. Такое доверие, в свою очередь, каузально способствует тому, что на ИИ можно положиться при отсутствии наблюдения. И здесь можно провести различие между доверием индивидуального пользователя к ИИ, где объяснимость не способствует доверию, и общественного доверия к ИИ. Такой подход объясняет кажущийся парадокс в отношениях ИИ и человека: чтобы доверять ИИ, нужно позволить индивидуальным пользователям ИИ не доверять ему полностью. Что и требует критического взгляда отдельного профессионала на решение, предлагаемое СППВР, само существование которой вполне может быть одобряемое профессиональным медицинским сообществом.

Подводя итог, можно сделать следующие выводы.

Во-первых, в целях эффективного решения медицинских задач с использованием технологий ИИ необходима и неизбежна гибридизация человека и машины, ключевым фактором развития которой является способность к коэволюции⁴. Для искусственных интеллектуальных агентов «мотивацию» к совместному развитию обеспечивает целевая функция. Для сложных объектов постоянное вовлечение человека в среду интеллектуальных агентов неизбежно, и именно его присутствие обеспечивает мотивацию для развития их интеллектуальных возможностей.

Во-вторых, поскольку в медицине каждое решение врача сопровождается риском причинения вреда, внедрение высоких технологий в данной сфере должно быть жестко регламентировано и обосновано с точки зрения ответственности за допущение ошибки. Соответствующие нормативные стандарты должны быть оперативно разработаны и корректироваться с учетом выявления новых рисков.

В-третьих, стандартизированное использование СППВР ставит под угрозу как приобретение, так и поддержание требуемого уровня компетентности конкретного врача. Существует опасность потенциального «недостатка опыта», если система не будет функционировать правильно. И если врач обучен работать только при поддержке системы, возникает угроза его когнитивным и моральным навыкам, а, значит, его автономии. В таком случае «ответственность» должна рассматриваться в двух различных оптиках:

А.В. Углева

Культура вины и ответственность врача при использовании систем поддержки принятия врачебных решений

⁴ Подробнее см., напр., [Krinkin et al., 2022].

как позитивное допущение индивидуальной моральной ответственности (включая виновность) и как юридическая ответственность, обусловленная в том числе вменением врачу в обязанность принятия обоснованного решения, что требует от него интеллектуальной добродетели. Имеющиеся немногочисленные исследования показывают, что врачи в целом рассматривают собственную «конечную ответственность» как условие профессиональной состоятельности [Van Cauwenberge et al., 2022: 50]. Делегирование же ответственности или ее распределение воспринимаются подчас как передача контролирующей функции третьей стороне, что возможно лишь при соблюдении трех следующих условий.

1. Должна существовать структурная свобода или институциональная добровольность, то есть отсутствие институционального принуждения, что позволило бы использовать в работе отдельные рекомендации СППВР или отказаться от них. В этом смысле директивы руководства организации или вышестоящих органов должны рассматриваться как потенциальные ограничения для несения персональной ответственности врача. Иные виды принуждения рассматриваются как вредные для ответственности, например, когда врача обязывают использовать СППВР в условиях нехватки человеческих ресурсов или экономической рентабельности [Langanke et al., 2016].

2. Врачу должны быть гарантированы эпистемическая свобода, доступ к базам данных, на которых обучается СППВР, у него должен быть достаточный практический опыт принятия клинических решений без помощи СППВР. В случае отсутствия надежного синтеза рекомендации системы с мнением врача у последнего должно быть неотъемлемое право обратиться к мнению коллег. В качестве возможной альтернативы (если профессиональное консенсусное мнение не может быть достигнуто) полномочия по принятию решений относительно предпочтительной рекомендации могут быть переданы пациенту (ср. «безответственный аутсорсинг ответственности» [Di Nucci, 2019]). Если профессионалы структурно и эпистемически способны действовать автономно, они считают себя ответственными за принятие решений. Как только одно или оба условия не выполняются, ответственность врача за совершенные медицинские ошибки не может быть определена с достаточной долей достоверности в рамках реляционной модели.

3. Интеграция имеющихся баз данных о пациентах и обеспечение технической возможности доступа врача к единой информационной сети, а также введение персональной ответственности пациента за качество предоставляемой им информации о состоянии своего здоровья и произведенных манипулятивных воздействиях

на его организм в течение всей жизни, что позволит уравновесить ответственность врача и пациента.

The Culture of Guilt and the Responsibility of the Physician when Using Clinical Decision Support System

Anastasia V. Ugleva

PhD, Associate Professor School of Philosophy and Cultural Studies.

National Research University Higher School of Economics.

20, Myasnitskaya Str., Moscow 101000, Russian Federation.

augleva@hse.ru

ORCID: 0000-0002-9146-1026

Abstract. The purpose of this article is to substantiate the need to revise the structure of moral and criminal responsibility of the physician in a social system in which a culture of guilt is imposed on him, that is, sole personal responsibility for any medical error committed, including with the participation of the clinical decision support system. From the paternalistic model of the «doctor-patient» relationship, since the 1970s, there has been a transition to a model of collaborative responsible behavior, in which each party is obligated to know and understand the nature of clinical decisions, which must also be supported by a special form of informed consent for both patient and physician, if AI technologies are included in diagnostic and treatment procedures. At the same time, in fact, there is no clear regulation of the process of communication and work with information, as well as the personal responsibility of top management for the formation of a medical support system. Because of the fundamental opaqueness of the operation of constantly developing artificial intelligence technologies (using neural networks as an example), the final decision must remain with the physician. However, justification of his responsibility in the new technological conditions requires a new social contract within the professional environment regarding the conditions for implementing clinical decision support system into broad medical practice.

Keywords: artificial intelligence, trust, medical error, physician responsibility, clinical decision support system, culture of guilt, autonomy.

For citation: Ugleva A.V. The Culture of Guilt and the Responsibility of the Physician when Using Clinical Decision Support System // *Chelovek*. 2023. Vol. 34, N 3. P. 9–23. DOI: 10.31857/S023620070026423-2

Литература/References

American College of Physicians. Ethics manual. 4th ed. *Annals of Internal Medicine*. 1998. N 128. P. 576–594.

А.В. Углева

Культура вины и ответственность врача при использовании систем поддержки принятия врачебных решений

- Barocas S., Selbst A., Raghavan M. The hidden assumptions behind counterfactual explanations and principal reasons. *Proceedings of the 2020 International Conference on Fairness, Accountability, and Transparency*. 2020. P. 80–89.
- Bleher H., Braun M. Diffused responsibility: attributions of responsibility in the use of AI-driven clinical decision support systems. *AI Ethics*. 2022. N 2. P. 747–761.
- Braun M., Hummel P., Beck S. et al. Primer on an ethics of AI-based decision support systems in the clinic. *Journal of Medical Ethics*. 2021. N 47. P. e3.
- Chollet F. On the measure of intelligence, 2019. Available at: <https://arxiv.org/abs/1911.01547>
- Coeckelbergh M. Artificial intelligence, responsibility Attribution, and a relational Justification of explainability. *Sci Eng Ethics*. 2020. N 26. P. 2051–2068.
- Dellermann D., Calma A., Lipusch N. et al. The future of human-ai collaboration: a taxonomy of design knowledge for hybrid intelligence systems. *Hawaii international Conference on System Sciences (HICSS)*. Hawaii, USA, 2019.
- Di Nucci E. Should we be afraid of medical AI? *J Med Ethics*. 2019. N 45. P. 556–558.
- Feenberg A. Critical theory of technology: An overview. *Information technology in librarianship: New critical approaches*. 2005. Vol. 1. Iss. 1. P. 47–64.
- Feenberg A. *Technosystem: the social life of reason*. Harvard University Press, 2017.
- Floridi L., Cows J. A unified framework of five principles for AI in society. *Machine Learning and the City: Applications in Architecture and Urban Design*. May 2022. N 2. P. 535–545.
- Ford E., Edelman N., Somers L. et al. Barriers and facilitators to the adoption of electronic clinical decision support systems: a qualitative interview study with UK general practitioners. *BMC Med Inform Decis Mak*. 2021. P. 21.
- Funer F. The deception of certainty: how non-interpretable machine learning outcomes challenge the epistemic authority of physicians. A deliberative-relational approach. *Med Health Care Philos*. 2022. N 25. P. 167–78.
- Funer F., Liedtke W., Tinnemeyer S. et al. *J Med Ethics*. 2022.
- Funer F., Liedtke W., Tinnemeyer S. et al. Responsibility and decision-making authority in using clinical decision support systems: an empirical-ethical exploration of German prospective professionals' preferences and concerns. *J Med Ethics*. 2022.
- Grote T., Berens P. On the ethics of algorithmic decision-making in healthcare. *J Med Ethics*. 2020. N 46. P. 205–211.
- Hu X.-B. A methodological framework of human-machine Co-evolutionary intelligence for decision-making Support of ATM. *Integrated Communications Navigation and Surveillance Conference (ICNS)*. 2020. P. 5C3-1–5C3-8.
- Jonas H. *The Imperative of Responsibility // Search of an Ethics for the Technological Age*. Chicago: University of Chicago Press, 1985.
- Krijger J. Enter the metrics: critical theory and organizational operationalization of AI ethics. *Ai & Society*. 2021. P. 1–11.
- Krinkin K., Shichkina Y., Ignatyev A. Co-evolutionary hybrid intelligence is a key concept for the world intellectualization. *Kybernetes*. 2022. Vol. ahead-of-print. N ahead-of-print.
- Langanke M., Liedtke W., Buyx A. Patients' responsibility for their health // Schramme T., Edwards S., eds. *Handbook of the Philosophy of Medicine*. Cham; Heidelberg; New York; Dordrecht; London: Springer, 2016. P. 619–640.

- Lepri B., Oliver N., Pentland A. Ethical machines: the human-centric use of artificial intelligence. *Science*. 2021. Vol. 24, N. 3. P. 102–249.
- Leslie D. Understanding Artificial Intelligence Ethics and Safety: A Guide for the Responsible Design and Implementation of AI Systems in the Public Sector. The Alan Turing Institute, 2019.
- Makary M.A., Daniel M. Medical error — the third leading cause of death in the US *BMJ*. 2016. P. 353.
- Nebeker C., Torous J., Bartlett E.R.J. Building the case for actionable ethics in digital health research supported by artificial intelligence. *BMC Med*. 2019. N 17. P. 137.
- Reason J.T. Human error. Cambridge University Press, 1990.
- Van Cauwenberge D., Van Biesen W., Decruyenaere J. et al. Many roads lead to Rome and the artificial intelligence only shows me one road: an interview study on physician attitudes regarding the implementation of computerised clinical decision support systems. *BMC Med Ethics*. 2022. P. 23.
- Waterman A.D., Garbutt J., Hazel E. et al. The emotional impact of medical errors on practicing physicians in the United States and Canada. *Joint Commission journal on quality and patient safety*. 2007. N 33(8). P. 467–476.
- Werner M.H. Verantwortung // Grunwald A., Hillerbrand R., eds. *Handbuch Technikethik*. Stuttgart: J.B. Metzler, 2021. P. 44–48.

А.В. Углева
Культура вины
и ответствен-
ность врача
при использо-
вании систем под-
держки приня-
тия врачебных
решений