Научная статья **https://doi.org/10.21202/2782-2923.2024.1.246-263**
УДК УДК 347.9:340.13:004.8

А. Файн[1],
С. Ли[1],
М. К. Миллер[1]

[1] *Университет Невады в Рино, г. Рино, США*

Контент-анализ мнений судей об инструментах оценки рисков с использованием искусственного интеллекта

Переводчик *Е. Н. Беляева*

Контактное лицо:
Анна Файн, соискатель степени PhD по междисциплинарной социальной психологии, Университет Невады в Рино
Адрес: 1300 1664, ул. С. Виргинии, Рино, NV 89557, США
E-mail: afine@nevada.unr.edu

Стефани Ли, магистрант, Университет Невады в Рино

Моника К. Миллер, доктор права, PhD, профессор, Университет Невады в Рино

Аннотация

Цель: анализ правовых позиций судейских работников об инструментах оценки рисков с использованием искусственного интеллекта.

Методы: диалектический подход к познанию социальных явлений, позволяющий проанализировать их в историческом развитии и функционировании в контексте совокупности объективных и субъективных факторов, который определил выбор следующих методов исследования: формально-логический и социологический.

Результаты: искусственный интеллект (ИИ) делает прогнозы (например, при вынесении решения об освобождении под поручительство) с использованием компьютерного программного кода и потенциально может положительно влиять на систему правосудия (например, экономить время и снижать уровень предвзятости). Данная работа содержит анализ вторичных данных и оценивает ответы 381 судьи на вопрос «Считаете ли Вы, что искусственный интеллект (использующий компьютерные программы и алгоритмы) способен устранить предвзятость при вынесении решений об освобождении под поручительство и приговоров?».

Научная новизна: на основе опубликованной литературы авторы выделили такие категории, как оценка и неприятие судьями компьютерных алгоритмов, локус контроля, нормы процессуального права, легитимность. Результаты показывают, что судьи испытывают неприятие компьютерных алгоритмов, опасаются усиления предвзятости при использовании ИИ и возможности потери работы из-за развития технологий. Судьи считают, что ИИ может помочь им при вынесении решений об освобождении под поручительство и приговоров, однако он должен пройти эмпирические проверки и следовать разработанным правилам. На основе изученных мнений судей об ИИ авторы обсуждают вопросы интеграции ИИ в правовую систему и направления будущих исследований.

Цитирование оригинала статьи на английском: Fine, A., Le, S., & Millera, M. K. (2023). Content Analysis of Judges' Sentiments Toward Artificial Intelligence Risk Assessment Tools. *Criminology, Criminal Justice, Law & Society*, *24*(2), 31–46.

URL публикации: https://ccjls.scholasticahq.com/article/84869-content-analysis-of-judges-sentiments-toward-artificial-intelligence-risk-assessment-tools

Практическая значимость: основные положения и выводы статьи могут быть использованы в научной, педагогической и правоприменительной деятельности при рассмотрении вопросов, связанных с правовыми рисками использования искусственного интеллекта.

Ключевые слова:

искусственный интеллект, судьи, инструменты принятия решений, общественное мнение, неприятие компьютерных алгоритмов, принятие компьютерных алгоритмов, нормы процессуального права, легитимность

Как цитировать русскоязычную версию статьи: Файн, А., Ли, С., Миллер, М. К. (2024). Контент-анализ мнений судей об инструментах оценки рисков с использованием искусственного интеллекта. *Russian Journal of Economics and Law*, *18*(1), 246–263. https://doi.org/10.21202/2782-2923.2024.1.246-263

Scientific article

A. Fine[1],

S. Le[1],

M. K. Miller[1]

[1] *University of Nevada, Reno, USA*

Content Analysis of Judges' Sentiments Toward Artificial Intelligence Risk Assessment Tools

Translator *E. N. Belyaeva*

Contact:

Anna Fine, Interdisciplinary Social Psychology Ph.D. Program, Mailstop, University of Nevada
Address: 1300 1664 N. Virginia St., Reno, NV, 89557, USA
E-mail: afine@nevada.unr.edu

Stephanie Le, a first-year Master of Sociology student, University of Nevada

Monica K. Miller, J.D., Ph.D., is a Foundation Professor, University of Nevada

Abstract

Objective: to analyze the positions of judges on risk assessment tools using artificial intelligence.

Methods: dialectical approach to cognition of social phenomena, allowing to analyze them in historical development and functioning in the context of the totality of objective and subjective factors, which predetermined the following research methods: formal-logical and sociological.

Results: Artificial intelligence (AI) uses computer programming to make predictions (e.g., bail decisions) and has the potential to benefit the justice system (e.g., save time and reduce bias). This secondary data analysis assessed 381 judges' responses to the question, "Do you feel that artificial intelligence (using computer programs and algorithms) holds promise to remove bias from bail and sentencing decisions?"

The article was first published in English language by Criminology, Criminal Justice, Law & Society and The Western Society of Criminology Hosting by Scholastica. For more information please contact: CCJLS@WesternCriminology.org

For original publication: Fine, A., Le, S., & Millera, M. K. (2023). Content Analysis of Judges' Sentiments Toward Artificial Intelligence Risk Assessment Tools. *Criminology, Criminal Justice, Law & Society*, *24*(2), 31–46.

Publication URL: https://ccjls.scholasticahq.com/article/84869-content-analysis-of-judges-sentiments-toward-artificial-intelligence-risk-assessment-tools

Scientific novelty: The authors created apriori themes based on the literature, which included judges' algorithm aversion and appreciation, locus of control, procedural justice, and legitimacy. Results suggest that judges experience algorithm aversion, have significant concerns about bias being exacerbated by AI, and worry about being replaced by computers. Judges believe that AI has the potential to inform their decisions about bail and sentencing; however, it must be empirically tested and follow guidelines. Using the data gathered about judges' sentiments toward AI, the authors discuss the integration of AI into the legal system and future research.
Practical significance: the main provisions and conclusions of the article can be used in scientific, pedagogical and law enforcement activities when considering the issues related to the legal risks of using artificial intelligence.

Keywords:
artificial intelligence, judges, decision-making tools, public opinion, algorithm aversion, algorithm appreciation, procedural justice, legitimacy

For citation of Russian version: Fine, A., Le, S., & Miller, M. K. (2024). Content Analysis of Judges' Sentiments Toward Artificial Intelligence Risk Assessment Tools. *Russian Journal of Economics and Law, 18(1)*, 246–263. https://doi.org/10.21202/2782-2923.2024.1.246-263

Введение

Искусственный интеллект (далее – ИИ) решает новые задачи, используя достижения информатики и имеющиеся данные (IBM Cloud Education, 2020). В частности, для классификации, анализа данных и прогнозирования ИИ использует алгоритмы, созданные на основе больших массивов информации (Shroff, 2019). Эти алгоритмы включают наборы автоматизированных инструкций, выполняемых после запуска (Scott, 2021).

ИИ – это технологический феномен, имеющий социальные, культурные и политические последствия. Алгоритмы облегчают выполнение множества повседневных задач, относящихся к электронной почте, социальным сетям, поиску в сети, покупкам онлайн (Victor, 2021). Они помогают спасать жизни. К примеру, в Facebook[1] ИИ используется для выявления постов людей с суицидальными намерениями (Cassata, 2019). ИИ потенциально может принести обществу огромную пользу, но на проблему взаимодействия человека и ИИ нужно смотреть с точки зрения социальной психологии (Lindgren & Holmstrom, 2020). Социальное измерение данной проблемы отражено и в настоящем исследовании, поскольку отношение к технологии ИИ является социальным конструктом и связано с надеждами и страхами, которые влияют на поведение людей при взаимодействии с этой технологией (Pinch & Bijker, 1984).

ИИ уже разрабатывается и используется в различных областях системы правосудия (Rigano, 2019), включая инструменты принятия решений, используемые в судах. С помощью этой технологии можно выявить закономерности, сложные для человеческого восприятия; так, прогностическое моделирование помогает предсказать риск рецидивизма среди подсудимых и более эффективно распределять административные ресурсы (Henman, 2020). ИИ может извлекать информацию из объемных юридических документов, сокращая рутинную работу персонала (Zadgaonkar & Agrawal, 2021). В полиции эта технология используется для выявления районов с высоким уровнем преступности и распределения ресурсов на основе полученной информации (Perry, 2013). Государственные органы, например Управление национальной безопасности США, используют ИИ для автоматического распознавания лиц, чтобы выявлять потенциально опасных людей в «списках наблюдения» (Ritchie et al., 2021). Это лишь несколько примеров применения технологий искусственного интеллекта в системе правосудия. Наиболее распространенная область его применения – оценка рисков при освобождении под поручительство и вынесении приговора (McKay, 2020).

[1] Соцсеть принадлежит *Meta* – организации, признанной экстремистской, ее деятельность запрещена на территории Российской Федерации.

ИИ может помочь снизить предвзятость в системе правосудия, поскольку он основывает свои решения на данных, а не на субъективном мнении лица, принимающего решение. Однако этот процесс полностью зависит от характера и качества данных. Сложность использования ИИ в системе правосудия вызывает опасения, что алгоритмические данные могут отражать предвзятость (примером могут служить аресты за наркопреступления в зависимости от района, когда расовая принадлежность и социально-экономический статус оказывали существенное влияние на решение об аресте).

ИИ все еще является относительно новым явлением, и правительство США работает над принятием законов, регулирующих его деятельность (Lee & Lai, 2022). В настоящее время не существует общих федеральных норм, но некоторые штаты уже приняли соответствующее законодательство (National Conference of State Legislatures, 2022). Например, в Калифорнии на рассмотрении находится законопроект (A 331), который обязывает создателей и пользователей любого автоматизированного инструмента принятия решений проводить анализ его воздействия. В частности, такой анализ должен показывать, почему используется данное программное обеспечение, каковы ожидаемые положительные результаты, а также как и где оно будет использоваться. Этот анализ подается в Отдел гражданских прав Министерства юстиции США. Затем, если нормы не соблюдаются, на создателей или пользователей программного обеспечения можно подать в суд. В штате Вашингтон также находится на рассмотрении законопроект (S 5356), который устанавливает правила приобретения и использования автоматизированных инструментов принятия решений для защиты общественности и повышения прозрачности. Это лишь два из целого ряда штатов, рассматривающих введение законопроектов, касающихся автоматизированных инструментов принятия решений (National Conference of State Legislatures, 2022).

Таким образом, необходимо понять, как судьи относятся к использованию технологии искусственного интеллекта в зале суда, ведь именно они будут принимать решение о ее внедрении. В данной работе мы провели контент-анализ вторичных данных, касающихся мнения судей по поводу неприятия и одобрения алгоритмов, локуса контроля, процессуальной справедливости и легитимности. В частности, мы оценивали, считают ли судьи, что искусственный интеллект может уменьшить предвзятость при принятии решений об освобождении под поручительство и вынесении приговора. Данное исследование может послужить основой для изучения мнений судей в отношении применения искусственного интеллекта в судебной практике.

Обзор литературы

Инструменты оценки рисков в судах

Судьи используют инструменты оценки рисков при освобождении под поручительство и вынесении приговора. Цель инструментов оценки риска – помочь судьям быстро классифицировать информацию о подсудимом, чтобы определить, можно ли освободить его под поручительство или какой приговор следует вынести. Хотя судьи, как правило, связаны нормами закона, определяющими приговор, они также имеют право по своему усмотрению назначить наказание в рамках этих норм (Bushway & Piehl, 2001). Для оценки вероятности рецидивизма обычно используются два метода оценки риска – клинический и актуарный (Buskey & Woods, 2018). Эти методы различаются по своей точности прогнозирования рецидивизма; таким образом, выбор метода влияет на принятие решений об освобождении под поручительство и вынесении приговора.

Сравнение клинических и актуарных моделей

При оценке рисков используются клинические и актуарные методы оценки вероятности рецидивизма у подсудимого. В рамках клинического метода специалисты (например, судебные психологи и клиницисты) используют свой личный опыт, знания и интуицию, чтобы отличить преступников с низким и высоким риском рецидивизма (Mossman, 1994). В актуарном методе для прогнозирования уровня рецидивизма применяются статистические инструменты, алгоритмы или методы искусственного интеллекта, что является новшеством в области подобных оценок (Barabas et al., 2018).

Одна из основных причин высокой точности моделей ИИ состоит в их способности быстро и правильно обрабатывать огромные объемы данных (Helm et al., 2020). Эти модели могут выявлять закономерности, анализировать большие объемы данных и делать прогнозы на их основе. Кроме того, искусственный интеллект обучается и совершенствуется с помощью машинного обучения, что со временем повышает точность его оценок (Helm et al., 2020). Точность ИИ – это степень соответствия прогнозов системы искусственного

интеллекта фактическим результатам (Rueda et al., 2022). Точность ИИ важна для оценки его производительности и эффективности. ИИ может превзойти человеческие способности и показывает более высокую предсказательную достоверность, чем прогнозисты-люди (Dietvorst et al., 2015).

ИИ более точен при составлении прогнозов в условиях неопределенности, чем люди, принимающие решения (Dawes, 1979; Grove et al., 2000). Кроме того, опытные прогнозисты, которые в большей степени полагаются на советы человека, чем на алгоритм, показывают более низкую точность (Logg et al., 2019). Несмотря на это, люди не решаются доверять ИИ и предпочитают, чтобы решения принимал человек, даже если знают о высокой точности применяемого ИИ (Diab et al., 2011; Eastwood et al., 2012). Хотя точность ИИ очень важна (Hoff & Bashir, 2015), аспект справедливости также имеет решающее значение для доверия к ИИ (Knowles et al., 2022).

Актуарные методы, использующие технологию искусственного интеллекта, обещают стать эффективным способом снижения предвзятости при освобождении под поручительство и вынесении приговора. Однако искусственный интеллект хорош лишь настолько, насколько хороши данные, на которых его обучали; но в течение многих лет мы систематически наблюдаем расовую предвзятость при освобождении под поручительство (Demuth & Steffensmeier, 2004b; Schlesinger, 2005; Turner & Johnson, 2005) и вынесении приговора (Demuth & Steffensmeier, 2004a; Spohn & Holleran, 2000; Western, 2006).

Поэтому важно понимать ограничения ИИ. Например, компания *Pro Publica* проанализировала оценки рисков совершения уголовных преступлений, предоставляемые алгоритмом *COMPAS*, на предмет расовой дискриминации и обнаружила, что чернокожие обвиняемые в два раза чаще попадают в категорию высокого риска, чем белые (Angwin et al., 2016). Более того, белые обвиняемые в два раза чаще попадали в категорию низкого риска, чем чернокожие. Поэтому судьи должны осознавать, что искусственный интеллект может закреплять предвзятость, если алгоритм обучен на расово предвзятых данных. Указанный анализ привлек внимание к изменчивости и недостаточной ясности роли ИИ в оценке рисков; однако его методология подвергалась критике. Так, Джонс (Jones, 2020), проведя математическое моделирование анализа данных *COMPAS*, обнаружил, что показатели риска для чернокожих были рассчитаны верно в пределах 2 % от фактических показателей, в то время как для белых и латиноамериканцев прогнозировался более низкий уровень рецидивизма, чем показывают реальные данные.

Если для судей существуют нормы морали и юридической этики, то для ИИ также необходимо разработать специальные этические рекомендации. Как правило, этические нормы ничего не говорят о прозрачности инструмента, его справедливости и точности, а также не обязывают пользоваться рекомендованным инструментом. Таким образом, решения судей об использовании ИИ могут определяться их представлениями (и статистическими данными) о его точности и справедливости. Эти критерии, вероятно, не единственные, которые судьи используют для формирования своих решений и представлений. О восприятии искусственного интеллекта судьями можно узнать на основании социально-психологической теории, связанной с изучением общественных настроений.

Представления общественности об искусственном интеллекте

Общественные настроения – это общие убеждения, взгляды и мнения группы людей по определенной теме или вопросу (Miller & Chamberlain, 2015). На общественные настроения влияют различные социальные, культурные и экологические факторы, а настроения, в свою очередь, оказывают значительное воздействие на правовую систему и законы. Общественные настроения часто играют роль в формировании законодательства (Burstein, 2003) и определении приоритетности правовых вопросов, имеющих особое значение для общества (Burstein, 2006). Когда речь идет о конкретном политическом вопросе, настроения сообщества могут основываться на мнении широкой общественности или учитывать только мнения соответствующих экспертов и инсайдеров. Изменения в политике могут быть связаны с настроениями общества и определяться его приоритетами (Miller & Chamberlain, 2015). Понимание общественных настроений может помочь определить, как судьи воспринимают и используют инструменты ИИ в системе правосудия.

Судьи как элемент законной власти оказывают существенное влияние на судебную систему, поэтому важно понимать их настроения. Судьи, негативно относящиеся к ИИ, могут с меньшей частотой использовать его или отменять решения ИИ. И наоборот, судьи с позитивным отношением к ИИ могут охотнее внедрять ИИ в процесс принятия решений.

Разработка и внедрение в систему правосудия успешных инструментов ИИ (т. е. таких, которые дают положительные результаты, например в форме уменьшения предвзятости, и имеют мало отрицательных последствий) требуют понимания общественных настроений. Учет мнений заинтересованных сторон (в частности, судей) позволяет разрабатывать и внедрять инструменты ИИ таким образом, чтобы повысить вероятность их принятия и использования. Это, в свою очередь, должно привести к увеличению числа разработок и испытаний и в конечном счете способствовать принятию более эффективных решений в системе уголовного правосудия. Аспекты общественных настроений включают, в частности, неприятие и одобрение алгоритмов, локус контроля, восприятие процессуальной справедливости и легитимности.

Неприятие и одобрение алгоритмов

Общественные настроения по поводу ИИ можно частично объяснить в рамках концепции неприятия и одобрения алгоритмов. Термин «неприятие алгоритмов» был введен Dietvorst с соавт. (2015) для описания ситуаций, когда люди не желают или противятся использованию алгоритма вместо человека, даже если алгоритм обладает более высокой точностью. Однако в ходе серии исследований Logg с соавт. (2019) обнаружили, что действительно существуют ситуации (например, числовые оценки, прогнозы популярности песен, романтическая привлекательность), когда пользователи высоко оценивают работу алгоритма, т. е. предпочитают использовать алгоритм, а не человека-прогнозиста. Эти результаты свидетельствуют о том, что на решение людей доверять ИИ влияют не только точность, но и другие факторы.

Общественные настроения не только связаны с доверием к ИИ, но и зависят от уровня неопределенности типа решения. В тех областях, где нет неизбежной случайности или непредсказуемости (например, при решении арифметических задач или запоминании известных фактов), лучшим методом принятия решений будет тот, который может дать идеальный, безошибочный ответ (Dietvorst & Bharti, 2020). Однако в областях, где принятие решений связано с неизбежной неопределенностью, ошибки случаются даже при применении самых лучших инструментов (например, при использовании беспилотных автомобилей, при оценке криминальных рисков, в медицинской диагностике). Поэтому вероятность, что судьи будут доверять ИИ свои решения, снижается из-за неопределенного характера решений в области уголовного правосудия. Понятие локуса контроля показывает, как неопределенность может влиять на принятие решений.

Локус контроля

Локус контроля – это мнение человека о том, насколько он контролирует свою жизнь и события, которые на нее влияют (Rotter, 1966). Это понятие описывает степень веры людей в то, что они имеют власть над результатами своих действий, или же в то, что на них в первую очередь влияют внешние факторы, такие как удача, судьба или действия других людей (Rotter et al., 1972). Локус контроля может существенно влиять на мировоззрение, поведение и благополучие человека (Wallston & Wallston, 1978).

Люди с преимущественно внутренней ориентацией локуса контроля считают, что результаты зависят от их поведения (Sherman, 1973), и склонны доверять своим суждениям больше, чем суждениям искусственного интеллекта или других людей, принимающих решения (Sharan & Romano, 2020). Люди, которые подчеркивают автономию, личные достижения, индивидуальную ответственность и компетентность, имеют более сильный внутренний локус контроля. Люди с преимущественно внешней ориентацией считают, что результаты зависят от внешних сил (например, окружающей среды, удачи, других людей), и признают, что внешние силы могут повлиять на их решения (Sherman, 1973).

Ориентация локуса контроля судей влияет на их отношение к искусственному интеллекту и готовность полагаться на него для снижения предвзятости при принятии решений об освобождении под поручительство и вынесении приговора. Вероятно, судьи могут оказаться близко на шкале локуса контроля, что связано с их выбором профессии. Высокостатусные профессии ассоциируются с внутренним локусом контроля (Smith et al., 1997). Учитывая престижность профессии судьи, они могут естественным образом склоняться к тому, чтобы иметь внутренний локус контроля, а значит, с меньшей вероятностью рассматривать ИИ как достойный инструмент принятия решений по сравнению со своим собственным суждением. Кроме того, на их мнение могут влиять также опасения по поводу соблюдения процессуальных норм при использовании искусственного интеллекта. Этот аспект рассмотрен ниже.

Процессуальная справедливость

Процессуальная справедливость – это воспринимаемая человеком степень справедливости принимаемых решений (Lemons & Jones, 2001), которая зависит от процедуры принятия этих решений (Leventhal, 1980). В правовой системе процессуальная справедливость имеет решающее значение для обеспечения защиты прав всех сторон, а также беспристрастности и прозрачности решений. Если процедура принятия решений справедлива, то даже неблагоприятные результаты считаются процессуально справедливыми (Thibaut & Walker, 1975).

Одним из важнейших факторов процессуальной справедливости является наличие права голоса у тех, кого затрагивает решение (Lind & Tyler, 1988). Что касается ИИ и процессуальной справедливости, то одной из ключевых проблем является обеспечение того, чтобы те, кого эти решения затрагивают, имели возможность поставить под сомнение процесс принятия решений. Поскольку процесс принятия решений искусственным интеллектом имеет характер «черного ящика», его трудно подвергнуть сомнению. Обвиняемые могут не иметь возможности высказать свое мнение, поскольку ИИ опирается на статистические данные и не позволяет обвиняемому рассказать свою версию событий. Отсутствие прозрачности может привести к чувству несправедливости и подорвать доверие к процессу принятия решений. Для соблюдения процессуальной справедливости необходимо, чтобы используемые инструменты были прозрачными и объяснимыми. Кроме того, должна быть предусмотрена процедура подачи апелляции, если субъекты права сочтут решение несправедливым. В рамках правовой системы искусственный интеллект может использоваться для помощи в принятии судебных решений. Однако ИИ не должен подрывать процессуальное правосудие.

Использование инструментов ИИ в суде может снизить восприятие справедливости решений. Например, в сфере трудоустройства сотрудники воспринимают собеседования на основе ИИ как менее справедливые с процедурной точки зрения, чем собеседования, проводимые человеком (Parasuraman et al., 2000). Независимо от точности, которую демонстрирует ИИ, люди склонны меньше доверять ему в различных сферах и контекстах. Однако исследования показывают, что включение человеческих элементов в процесс принятия решений может повысить процессуальную справедливость (Lee et al., 2019). Поэтому крайне важно, чтобы система правосудия внедряла системы искусственного интеллекта для совместной работы с людьми, принимающими решения, а не заменяла их. Неспособность поддерживать процессуальную справедливость может привести к чувству нелегитимности процесса принятия решений, подрывая настроения в обществе и негативно влияя на социальную сплоченность и сотрудничество. При использовании инструментов искусственного интеллекта в зале суда общественность может воспринять процесс как несправедливый или скептически отнестись к решению, принятому искусственным интеллектом. Социальная сплоченность опирается на общую веру в справедливость и эффективность системы правосудия. Поэтому, если процесс принятия решений будет восприниматься как несправедливый, это может привести к недостаточной сплоченности в обществе.

Легитимность

Легитимность – это убежденность в том, что власти, организации и социальные соглашения являются правильными и непредвзятыми (Tyler, 2006); ее поддержание зависит от соблюдения властями норм процессуальной справедливости (Clawson et al., 2001; Farnsworth, 2003). Власти воспринимаются как легитимные, если они соответствуют нормам, убеждениям и ценностям группы (Zelditch, 2018), а граждане считают, что они заслуживают поддержки (Gurr, 1974). Судебная система должна обладать высоким уровнем легитимности, чтобы поддерживать верховенство закона, поскольку легитимность – это ее главный политический капитал (Gibson, 2006), а отсутствие легитимности означает отсутствие власти (Zelditch, 2018).

Чтобы правительство было легитимным, решения должны быть справедливыми и каждый должен иметь возможность в какой-то момент оказаться в выигрыше (Buhlmann & Kunz, 2011). Судебные органы должны обладать институциональной легитимностью, которая заключается в том, что общественность будет поддерживать их решения и полномочия (Audette & Weaver, 2015) независимо от степени удовлетворенности решениями (Buhlmann & Kunz, 2011). Без легитимности суды не обладают достаточной властью, чтобы в случае необходимости принимать решения вопреки общественному мнению. Чувство предвзятости судебной системы снижает легитимность судов (Ramirez, 2008), поэтому судьи должны принимать меры для поддержания легитимности.

Участие ИИ в принятии решений может снизить легитимность судебной власти. На восприятие легитимности систем ИИ влияют три актуальные проблемы: 1) легитимность на входе, когда граждане считают, что не могут контролировать собираемые данные; 2) легитимность в процессе, когда граждане не понимают процессы ИИ и сталкиваются с проблемой «черного ящика» и 3) легитимность на выходе, когда граждане сомневаются, что ИИ может принимать более верные решения, чем человек (Starke & Lunich, 2020). Легитимность требует прозрачности, демократичности и подотчетности. Применительно к искусственному интеллекту судебные органы должны использовать надежные системы ИИ, регулируемые с помощью норм закона и независимого надзора (Andrews, 2022).

Поддержание процессуальной справедливости и легитимности необходимо для успешной работы системы правосудия. Поэтому крайне важно понять, как технологии искусственного интеллекта в системе правосудия влияют на восприятие легитимности и процессуальной справедливости. Технологии ИИ уже показали свою перспективность в плане снижения предвзятости при принятии решений об освобождении под поручительство и вынесении приговора. Однако эти инструменты не лишены недостатков. Для того чтобы они служили поставленным целям, им необходима поддержка со стороны пользователей. Очень важно понять, как пользователи относятся к их внедрению. Поскольку исследований, изучающих отношение судей к ИИ, мало, качественный подход позволит понять некоторые из наиболее острых проблем.

Описание исследования

Существует очень мало исследований, изучающих отношение судей к использованию ИИ при принятии решений об освобождении под поручительство и вынесении приговора. Хотя инструменты ИИ более точны, чем прогнозы людей (Dietvorst et al., 2015), эти инструменты все еще нуждаются в совершенствовании, поскольку они могут отражать существующие системные предубеждения в системе правосудия. Однако по мере совершенствования этих инструментов и уменьшения ошибок судьям будет полезно научиться работать с ИИ для обоснования своих решений. В данной работе мы изучили отношение судейского сообщества к искусственному интеллекту, используемому при освобождении под поручительство и вынесении приговора, и выяснили, может ли искусственный интеллект минимизировать предвзятость.

В данном исследовании рассматриваются следующие вопросы:

1. Относятся ли судьи в целом негативно или позитивно к использованию искусственного интеллекта?
2. Отражают ли ответы судей внутренний или внешний локус контроля?
3. Считают ли судьи, что использование искусственного интеллекта создаст угрозу процессуальному правосудию и легитимности?
4. О каких преимуществах и опасениях по поводу ИИ как инструмента принятия решений говорят судьи?

Методология исследования

Участники и процедуры

Сбор данных осуществлялся Национальным юридическим колледжем (*NJC*) в Рино, штат Невада, в январе 2020 г. среди выпускников этого учебного заведения, работающих судьями. Им был задан вопрос: «Считаете ли вы, что искусственный интеллект (использующий компьютерные программы и алгоритмы) способен устранить предвзятость при принятии решений об освобождении под поручительство и вынесении приговора?» Ответы поступили от 381 судьи. Предполагался ответ «да» или «нет» либо развернутый комментарий, который оставили 168 судей[1].

Кодирование данных

Мы провели контент-анализ комментариев судей, ответивших на вопрос анкеты. Авторы составили схему кодирования содержания, которая отражала ключевые темы в комментариях судей на основе научной литературы. Коды обозначали как вопросы исследования, так и общие темы, возникшие в ответах. Единицами анализа были высказывания, т. е. мы выделили из 168 комментариев судей 325 отдельных высказываний, каждое из которых отражало одну идею. Если комментарий содержал три идеи (например, упомянуты три отдельные проблемы, связанные с ИИ), то его делили на три отдельных высказывания. Это позволило определить, сколько раз каждый судья упоминал ту или иную тему. Каждый комментарий проанализировали на предмет наличия каждой темы.

Отношение к проблеме в целом кодировалось как 0, если оно было негативным, 1 – если оно было нейтральным и 2 – позитивным. Темы неприятия и одобрения алгоритмов (например, полезность, этика и желание научиться) получили код 0 (тема отсутствует), 1 (неприятие алгоритмов) и 2 (одобрение алгоритмов). Например, если судья упомянул о сомнениях в полезности, то присваивался код 1. Напротив, если судья упомянул о пользе, присваивался код 2. Остальные темы получили код 0 (тема отсутствует) или 1 (тема присутствует). Коды не являются взаимоисключающими, т. е. авторы оценивали каждое высказывание на предмет присутствия той или иной темы. Например, для комментария «Применение алгоритмов может быть полезно для экономии времени» тема общего настроения была закодирована как 2 (положительная), а тема полезности – как 1.

Соавторы вместе закодировали 20 ответов, а затем по отдельности – 40 ответов. Затем они поделили оставшиеся утверждения и кодировали их независимо друг от друга. При этом удалось добиться согласованности по методу Холсти на уровне 0,89, т. е. кодировщики воспринимали коды одинаково в 89 % случаев, что считается приемлемым показателем (Belur et al., 2021). В Приложении приведены каждая тема и ее определения.

Результаты исследования

ВИ 1: ***Относятся ли судьи в целом негативно или позитивно к использованию искусственного интеллекта?***

Большинство судей (n = 243; 64 %) ответили на этот вопрос отрицательно, т. е. они не считают, что использование ИИ может устранить предвзятость при принятии решений об освобождении под поручительство и вынесении приговора. В развернутых ответах судьи в целом негативно отнеслись к использованию ИИ для освобождения под поручительство и вынесения приговоров: 229 ответов были отрицательными, 33 – нейтральными и 63 – положительными. Мнение часто выражалось в форме одобрения или неприятия алгоритмов, которые входят в тему «Мнение сообщества» как две отдельные категории (см. Приложение). В 48 высказываниях судьи положительно оценили алгоритм, упомянув его полезность в той или иной форме (например, «Предоставление данных и моделей может быть полезным») и готовность обучиться его применению. Некоторые судьи отметили свое неприятие алгоритмов, особенно в отношении этических недостатков использования ИИ при принятии решений о поручительстве и вынесении приговора (n = 15; например, «Поскольку представители меньшинств чаще фигурируют в криминальной истории, учитывая практику системы правосудия в прошлом, мы будем сталкиваться с предвзятостью еще как минимум пару поколений»). Кроме того, некоторые судьи отметили, что эти инструменты не принесут пользы (n = 10; например, «Они не могут учесть тонкие аспекты взаимодействия между людьми, более показательные, чем любой тест»).

ВИ 2: ***Отражают ли ответы судей внутренний или внешний локус контроля?***

Высказывания судей отражают высокую степень внутреннего локуса контроля. Идея использования ИИ для принятия решений об освобождении под поручительство и вынесении приговора вызывает у судей чувство нехватки контроля (n = 12; например, «Процесс [принятия решений] скрыт и постоянно меняется, поэтому возникает риск того, что судья не сможет вынести полностью обоснованное решение, а адвокат – полноценно защищать своих клиентов»). В целом судьи демонстрировали высокий уровень неопределенности (n = 30; например, «Учитывая эти факты или их отсутствие, как судья может оценить обоснованность риска, если он не может понять процесс принятия решений?»), но иногда и низкий уровень неопределенности (n = 10; например, «Да, учитывая последнюю работу Малкольма Гладуэлла»). Кроме того, судьи отмечали угрозу их автономии со стороны ИИ и активно выступали за то, чтобы самым важным аспектом при принятии решений об освобождении под поручительство и вынесении приговора было судейское усмотрение (n = 61; например, «Нам необходимо право на независимое судейское усмотрение»).

ВИ 3: ***Считают ли судьи, что использование искусственного интеллекта создаст угрозу процессуальному правосудию и легитимности?***

Судьи выражали обеспокоенность тем, что использование ИИ может угрожать процессуальному правосудию. Например, отмечалось, что к подсудимым следует относиться справедливо и они должны знать, что решения были приняты после тщательного рассмотрения (n = 31; например, «ИИ ни сейчас, ни в будущем

не сможет обеспечить должную справедливость или милосердие»). Кроме того, многие судьи согласились с тем, что вынесение справедливого приговора требует от судей рассмотрения всех аспектов дела, а не только того, что запрограммировано в коде; это является важной частью процессуального правосудия (n = 39; например, «Мы выносим приговор людям, а не машинам. Каждое дело индивидуально»). Судьи также отметили, что процедура требует учета качественных факторов, которые могут понять только люди (n = 84; например, «Это будет еще один инструмент, но он не сможет заменить человеческий фактор, который должен включать воздействие на жертву. ИИ не сможет заменить его»).

Что касается легитимности, то в 16 ответах утверждалось, что общественность может оспорить авторитет системы, если судебные органы делегируют свои решения компьютерам (например, «Наличие квалифицированных судей, разбирающихся в законах и людях, имеет решающее значение»). Кроме того, 19 человек отметили, что перед использованием ИИ необходимо разработать руководящие принципы для обеспечения легитимности (например, «В настоящее время не существует федерального закона, устанавливающего стандарты или требующего проверки этих инструментов, как это делает Управление по контролю за продуктами питания и лекарственными средствами в отношении новых лекарств»). Судьи высоко оценивают необходимость легитимности и считают, что ИИ может потенциально угрожать ей.

ВИ 4: ***О каких преимуществах и опасениях по поводу ИИ как инструмента принятия решений говорят судьи?***

Помимо пользы, упомянутой в ВИ 1, судьи отметили и другие преимущества, связанные с использованием ИИ при освобождении под поручительство и вынесении приговора. Так, некоторые судьи выразили заинтересованность и желание узнать больше о разработке технологий для снижения предвзятости в судебном процессе (n = 11; например, «Это нужно изучить. Потребуются эмпирические данные, чтобы оценить преимущества ИИ и его роль при рассмотрении вопроса об освобождении под поручительство»). Другие признавали, что у них самих может присутствовать скрытая предвзятость, которую ИИ может снизить (n = 18; например, «Со скрытой предвзятостью сложно справиться, а технологии могут частично устранить ее»).

Ответы судей содержат больше опасений, чем преимуществ, связанных с использованием ИИ в судебном процессе. Часто выражалась обеспокоенность по поводу целостности данных и программы (n = 73; например, «Проблема в том, что авторы программного обеспечения тоже люди; они могут создать предвзятую систему, которая будет казаться неоспоримой, что еще более опасно»). Многие выражали опасения, что их могут заменить (n = 30; например, «Какой антиутопический кошмар ожидает нас, когда судей заменит так называемый искусственный интеллект!»). Кроме того, некоторые судьи отмечали, что использование искусственного интеллекта усугубит предвзятость (n = 26; «Компьютерные программы могут устранить человеческие предрассудки, но они создадут новые»). Другие судьи упоминали, что ИИ не хватает объективного мышления (n = 12; «Я думаю, что машина не склонна будет сомневаться в каких-либо данных или их полноте»).

Обсуждение

В данном исследовании изучались мнения судейского сообщества в отношении ИИ, используемого при принятии решений об освобождении под поручительство и вынесении приговора, а также степень перспективности ИИ для устранения предвзятости. В целом наше исследование показало, что судьи имеют внутренний локус контроля, с неприятием относятся к алгоритмам, а также испытывают обеспокоенность по поводу процессуальной справедливости и легитимности, связанной с использованием ИИ для принятия решений. Чтобы разработка и внедрение технологий позволили минимизировать негативные и максимизировать позитивные аспекты ИИ, необходима поддержка судейского сообщества. Первый шаг на этом пути – понимание опасений судей. При разработке систем ИИ и обучении судей работе с ними необходимо использовать психологические теории. В частности, те юрисдикции, которые хотят разработать справедливые процедуры с использованием ИИ, должны сосредоточиться на конкретных проблемах, обозначенных в нашем исследовании. Одной из таких проблем является предоставление обвиняемым и жертвам права голоса, что составляет необходимый компонент процессуального правосудия. Судья может учитывать как рекомендации ИИ, так и мнение обвиняемого и потерпевших. Нужно информировать судей об этих гарантиях, что уменьшит их опасения и позволит им с большей готовностью поддержать разработку и использование ИИ в системе правосудия.

Исследование позволяет сделать ряд важных выводов. Во-первых, большинство судей (64 %) не считают, что ИИ устранит предвзятость при принятии решений об освобождении под поручительство и вынесении приговора, и в целом негативно относятся к ИИ. Однако часть судей (36 %) убеждены, что ИИ может уменьшить предвзятость. Доля судей, испытывающих неприятие алгоритмов, оказалась больше, чем тех, кто одобряет их применение, поскольку в целом судьи выражали обеспокоенность по поводу их недостатков; однако многие признают, что технологии в целом полезны. Значит, большинство судей еще не уверены в том, что ИИ есть место в зале суда, и защитникам ИИ придется потрудиться, чтобы убедить их.

Комментарии некоторых судей отражают высокий внутренний локус контроля: они заявили, что использование ИИ в качестве инструмента принятия решений заставляет их чувствовать недостаток контроля, неопределенность и угрозу их праву на судейское усмотрение. Это согласуется с имеющимися исследованиями локуса контроля и инструментов принятия решений, согласно которым лица, ориентированные на внутренний локус контроля, в большей степени склонны доверять своим суждениям, а не ИИ или другим людям, принимающим решения (Sharan & Romano, 2020). Таким образом, вероятно, что только судьи с низким внутренним локусом контроля будут готовы принять ИИ, даже если он будет максимизировать преимущества при использовании (например, уменьшать предвзятость).

В ответах отмечалось, что ИИ представляет угрозу для процессуального правосудия и легитимности. Некоторые судьи указали, что без человеческого фактора правосудие невозможно. Кроме того, они выразили серьезную обеспокоенность по поводу отношения к людям как к статистическим данным. Прозрачность неоднократно отмечалась как один из способов обеспечения процессуальной справедливости (Lee et al., 2019) и легитимности (de Fine Licht & de Fine Licht, 2020). Кроме того, считается, что если лица, принимающие решения, полагаются исключительно на искусственный интеллект, то их легитимность низка (Starke & Lunich, 2020). Судьи не должны полностью полагаться на ИИ при принятии судебных решений, но должны использовать его в качестве полезного инструмента. Как судьи, положительно оценивающие перспективы ИИ, так и те, кто не согласен с ними, считают, что эти инструменты принятия решений должны быть *вспомогательными*, а не *заменять* судей. Эти опасения необходимо учитывать как при разработке ИИ, так и при обучении работе с ним.

Следствия

Настоящее исследование может послужить материалом для повышения квалификации юристов. В ближайшем будущем судьи столкнутся с различными технологиями на основе ИИ – от инструментов принятия решений до способов представления доказательств в суде. Для успешной интеграции этих инструментов в систему правосудия они должны будут иметь соответствующую подготовку. Чтобы донести информацию об этих технологиях до субъектов права, ученые должны понять, как повысить доверие судей к этим технологиям и научить их взаимодействовать с ними. Данное исследование дает представление о мнениях судей, которые следует учитывать при разработке этих технологий. Так, неуверенность судей при обращении с ИИ во многом вызвана непрозрачностью его алгоритмов. В будущем предстоит создать обучающие курсы для судей, объясняющие суть работы ИИ и факторы, которые учитываются при оценке риска. Снижение неопределенности и повышение прозрачности может помочь судьям более уверенно взаимодействовать с ИИ.

Для обеспечения процессуальной справедливости и легитимности необходимо разработать стандартизированные этические инструкции для технологий, используемых в судах. С помощью ИИ можно снизить предвзятость при освобождении под поручительство и вынесении приговора, но это требует усилий в плане развития этических рекомендаций. Fjeld с соавт. (2020) проанализировали документы различных организаций, в которых изложены принципы использования ИИ. Они выделили восемь ключевых тем в этой области: неприкосновенность частной жизни, подотчетность, безопасность и защита, прозрачность и объяснимость, справедливость и отсутствие дискриминации, контроль человека над технологией, профессиональная ответственность, развитие общечеловеческих ценностей. Многие из опрошенных нами судей выразили опасения, соответствующие этим темам. Таким образом, указанные принципы могут помочь в разработке и внедрении ИИ в судах для обеспечения процессуальной справедливости и легитимности.

Ограничения и будущие направления исследований

Наше исследование имеет множество ограничений, связанных с его дизайном, выборкой и характеристиками (например, с использованием вторичных данных). Преимуществом данной работы является то, что

в ней были выявлены и получили оценку мнения респондентов, принадлежащих труднодоступной группе граждан. Остановимся более подробно на ограничениях исследования. Во-первых, опрос включал только одно утверждение с возможностью прокомментировать его. В результате судьи могли выразить свое мнение, но объем полученной информации был ограничен. В опрос не были включены демографические сведения, поэтому мы не можем сказать, различаются ли мнения судей, например, в зависимости от их возраста, пола или стажа работы в должности.

В рамках опроса судьям предлагалось высказать свое мнение как об освобождении под поручительство, так и о вынесении приговора. Судьи могли чаще сталкиваться с одним из этих действий, чем с другим, или написать свой ответ так, что он относился только к одному из них. Если судья не уточнил этого в своем ответе, трудно понять, имел ли он в виду поручительство, вынесение приговора или и то и другое. Расширенный опрос позволил бы получить более глубокие результаты. Но и имеющиеся результаты полезны для понимания общего мнения в отношении ИИ.

Второе ограничение связано с особенностями выборки. Анкета была разослана по адресам базы данных Национального юридического колледжа (*NJC*) в Рино, однако судьи, сотрудничающие с *NJC*, могут отличаться от других судей, не имевших возможности принять участие в опросе. Кроме того, судьи, решившие принять участие в опросе, могут отличаться от тех, кто решил этого не делать. Так, судьи, имеющие определенное мнение за или против ИИ, могли откликнуться с большей вероятностью, чем те, чье мнение еще не сложилось. Таким образом, наша выборка, скорее всего, не представляет мнения всех судей. Важно отметить, что многие судьи в этой выборке могут быть незнакомы с ИИ, а опрос не позволял оценить степень их информированности. Таким образом, невозможно сказать, отличаются ли мнения судей, знакомых с ИИ, от мнения тех, кто с ним незнаком. В будущих исследованиях следует проверить гипотезу воздействия, согласно которой контакт с объектом (например, с ИИ) влияет на отношение к нему (Zajonc, 2001). Более широкая выборка позволила бы исследователям оценить отношение судей с бо́льшим диапазоном мнений и опыта.

Последнее существенное ограничение связано с характером анализа вторичных данных. Хотя такой анализ является ценным методом для проведения исследования, у него также есть некоторые ограничения. Во-первых, мы не контролировали процесс сбора данных и формулировку вопросов. Кроме того, мы не контролировали факторы, которые могли повлиять на результаты. Например, *NJC* собрал эти данные в январе 2020 г., до пандемии *COVID*-19. Учитывая значительный масштаб пандемии, существует вероятность того, что отношение судей к ИИ изменилось. Например, многие судьи научились пользоваться программой *Zoom*. Такой опыт мог сделать судей более открытыми к использованию ИИ, учитывая возросшую зависимость от технологий в связи с пандемией. Более того, опыт работы с технологией часто определяет предпочтения человека в отношении ИИ (Kramer et al., 2018). Как упоминалось выше, этот феномен может быть объяснен эффектом воздействия, который предполагает, что постоянное воздействие стимула усиливает положительное отношение к нему (Zajonc, 2001).

Наша работа может служить основой для проведения будущих исследований. Во-первых, поскольку наша анкета содержала только один вопрос, считаем, следует провести более обширное экспериментальное изучение опыта судей в использовании и принятии технологий, а также других индивидуальных различий. Кроме того, учитывая возможную предвзятость отбора респондентов, в будущем исследовании следует опираться на репрезентативную выборку судей, чтобы обеспечить обобщаемость результатов. Наконец, поскольку общественное мнение может существенно влиять на формирование политических и правовых решений, в будущих исследованиях важно изучить, как широкая общественность воспринимает использование ИИ в судебной системе.

Заключение

Решения, касающиеся освобождения под поручительство и вынесения приговора, имеют огромное значение, и судьи, естественно, стремятся обеспечивать справедливость и непредвзятость таких решений. Хотя ИИ способен *минимизировать* некоторые предубеждения, присущие человеческим решениям, он также может *привнести* предвзятость в решения, если данные, на которые он опирается, являются необъективными. У ИИ есть и другие преимущества, такие как ускорение процесса принятия решений и снижение нагрузки на судей, поэтому он имеет много сторонников. Хотя ИИ может решить множество проблем, до сих пор не существует стандартных норм или инструкций, которые бы определяли использование этого инструмента судьями при

принятии решений об освобождении под поручительство и вынесении приговора. Необходимо разработать справедливые алгоритмы и стандартизированные процедуры в этой области, а также обеспечить доверие судей, которые будут их использовать. Первый шаг в данном направлении состоит в том, чтобы понять, как судьи относятся к ИИ. В целом судьи не полностью доверяют ИИ, т. е. большинство из них не готовы к внедрению ИИ в том виде, в котором они его сейчас понимают. Судьи перечислили ряд преимуществ и опасений, связанных с использованием ИИ при принятии решений об освобождении под поручительство и вынесении приговора. Эти мнения могут стать основой для разработки таких процедур ИИ, которые позволят максимизировать его преимущества и минимизировать риски. Просвещение по вопросам ИИ должно быть направлено на снятие опасений судей, которые могут неправильно понимать принципы работы или процедуры ИИ.

Учитывая, что степень неопределенности играет важную роль в повышении доверия к технологиям ИИ, понятие локуса контроля служит уникальной отправной точкой для понимания того, как судьи воспринимают использование ИИ в судебном процессе. В частности, для многих судей характерен внутренний локус контроля, в результате чего они рассматривают ИИ как угрозу потери своей автономии и снижения контроля над решениями, принимаемыми в суде. Важно понимать, как использование ИИ в судах влияет на восприятие принципов процессуального правосудия и легитимности. Некоторые суды уже внедряют эту технологию, однако исследований, изучающих ее последствия для правовой системы, недостаточно. Ученые должны активно участвовать в разработке и внедрении ИИ, что будет способствовать повышению справедливости в системе правосудия.

Список литературы / References

Andrews, D. A., & Bonta, J. (2010). *The psychology of criminal conduct*. Routledge.

Andrews, P. (2022, October 13). Designing for legitimacy. *Apolitical*. https://apolitical.co/solution- articles/en/designing-for-legitimacy

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias. In K. Martin (Ed.), *Ethics of data and analytics* (pp. 254–264). Auerbach Publications.

Audette, A. P., & Weaver, C. L. (2015). Faith in the court: Religious out-groups and the perceived legitimacy of judicial decisions. *Law & Society Review, 49*(4), 999–1022. https://doi.org/10.1111/lasr.12167

Barabas, C., Virza, M., Dinakar, K., Ito, J., & Zittrain, J. (2018, January). Interventions over predictions: Reframing the ethical debate for actuarial risk assessment. *Proceedings of Machine Learning Research, 81,* 62–76. https://proceedings.mlr.press/v81/barabas18a.html

Belur, J., Tompson, L., Thornton, A., & Simon, M. (2021). Interrater reliability in systematic review methodology: Exploring variation in coder decision-making. *Sociological Methods & Research, 50*(2), 837–865. https://doi.org/10.1177/0049124118799372

Buhlmann, M., & Kunz, R. (2011). Confidence in the judiciary: Comparing the independence and legitimacy of judicial systems. *West European Politics, 34*(2), 317–345. https://doi.org/10.1080/01402382.2011.546576

Burstein, P. (2003). The impact of public opinion on public policy: A review and an agenda. *Political Research Quarterly, 56(1),* 29–40. https://doi.org/10.1177/106591290305600103

Burstein, P. (2006). Why estimates of the impact of public opinion on public policy are too high: Empirical and theoretical implications. *Social Forces, 84*(4), 2273–2289. https://doi.org/10.1353/sof.2006.0083

Bushway, S. H., & Piehl, A. M. (2001). Judging judicial discretion: Legal factors and racial discrimination in sentencing. *Law & Society Review*, *55*(4), 733–764. https://doi.org/10.2307/3185415

Buskey, B., & Woods, A. (2018). *Making sense of pretrial risk assessments*. National Association of Defense Lawyers. https://www.nacdl.org/Article/June2018- MakingSenseofPretrialRiskAsses

Cassata, C. (2019, December 20). Facebook using artificial intelligence to help suicidal people. *Healthline*. https://www.healthline.com/health-news/facebook-artificial-intelligence-help-suicidal-people

Clawson, R. A., Kegler, E. R., & Waltenburg, E. N. (2001). The legitimacy-conferring authority of the US Supreme Court: An experimental design. *American Politics Research, 29*(6), 566–591. https://doi.org/10.1177/1532673X01029006002

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist, 34,* 571–582. http://dx.doi.org/10.1037/0003- 066x.34.7.571

de Fine Licht, K., & de Fine Licht, J. (2020). Artificial intelligence, transparency, and public decision-making: Why explanations are key when trying to produce perceived legitimacy. *AI & Society, 35*(4), 917–926. https://doi.org/10.1007/s00146-020-00960- w

Demuth, S., & Steffensmeier, D. (2004a). Ethnicity effects on sentence outcomes in large urban courts: Comparisons among White, Black, and Hispanic defendants. *Social Science Quarterly*, *85*(4), 994–1011. https://doi.org/10.1111/i.0038- 4941.2004.00255.x

Demuth, S., & Steffensmeier, D. (2004b). The impact of gender and race-ethnicity in the pretrial release process. *Social Problems, 51*(2), 222242. https://doi.org/10.1525/sp.2004.51.2.222

Diab, D. L., Pui, S. Y., Yankelevich, M., & Highhouse, S. (2011). Lay perceptions of selection decision aids in US and non-US samples. *International Journal of Selection and Assessment, 19*(2), 209–216. https://doi.org/10.1111/i.1468-2389.2011.00548.x

Dietvorst, B. J., & Bharti, S. (2020). People reject algorithms in uncertain decision domains because they have diminishing sensitivity to forecasting error. *Psychological Science*, *31*(10), 1302–1314. https://doi.org/10.1177/0956797620948841

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology. General, 144*(1), 114–126. https://doi.org/10.1037/xge0000033

Eastwood, J., Snook, B., & Luther, K. (2012). What people want from their professionals: Attitudes toward decision-making strategies. *Journal of Behavioral Decision Making*, *25*(5), 458–468. https://doi.org/10.1002/bdm.741

Farnsworth, S. J. (2003). Congress and citizen discontent: Public evaluations of the membership and one's own representative. *American Politics Research*, *31*(1), 66–80. https://doi.org/10.1177/1532673X02238580

Fine, A., Le, S., & Millera, M. K. (2023). Content Analysis of Judges' Sentiments Toward Artificial Intelligence Risk Assessment Tools, *Criminology, Criminal Justice, Law & Society, 24*(2), 31–46.

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). *Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI* (Research Publication No. 2020-1). Berkman Klein Center. https://dx.doi.org/10.2139/ssrn.3518482

Garrett, B., & Monahan, J. (2019). Assessing risk: The use of risk assessment in sentencing. *Judicature, 103(*2), 6–16. https://iudicature.duke.edu/articles/assessing-risk-the-use-of-risk-assessment-in- sentencing/

Gibson, J. L. (2006). Judicial institutions. In R. A. Rhodes, S. A. Binder, & B. A. Rockman (Eds.), *The Oxford handbook of political institutions* (pp. 514-534). Oxford University Press.

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment, 12*(1), 19–30. https://doi.org/10.1037/1040-3590.12.1.19

Gurr, T. (1974). Persistence and change in political systems, 1800-1971. *American Political Science Review, 68*(4), 1482–1504. https://doi.org/10.2307/1959937

Harris, H. M., Gross, J., & Grumbs, A. (2019). *Pretrial risk assessment in California*. Public Policy Institute of California. https://www.ppic.org/wp-content/uploads/pretrial-risk-assessment-in-california.pdf

Helm, J. M., Swiergosz, A. M., Haeberle, H. S., Karnuta, J. M., Schaffer, J. L., Krebs, V. E., Spitzer, A. I., & Ramkumar, P. N. (2020). Machine learning and artificial intelligence: Definitions, applications, and future directions. *Current Reviews in Musculoskeletal Medicine, 13,* 69–76.

Henman, P. (2020). Improving public services using artificial intelligence: Possibilities, pitfalls, governance. *Asia Pacific Journal of Public Administration, 42*(4), 209–221. https://doi.org/10.1080/23276665.2020.1816 188

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors, 57*(3), 407–434. https://doi.org/10.1177/0018720814547570

IBM Cloud Education. (2020, June 3). *What is artificial intelligence?* https://www.ibm.com/cloud/learn/what-is- artificial-intelligence

Jones, L. (2020, October 6). ProPublica's misleading machine bias. *Medium*. https://medium.com/@llewhinkes/propublic as-misleading-machine-bias-19c971549a18

Knowles, B., Richards, J. T., & Kroeger, F. (2022). *The many facets of trust in AI: Formalizing the relation between trust and fairness, accountability, and transparency.* arXiv. https://arxiv.org/abs/2208.00681

Kramer, M. F., Schaich Borg, J., Conitzer, V., & Sinnott-Armstrong, W. (2018, December). When do people want AI to make decisions? *Proceedings of the 2018 AAAI/ACM Conference On AI, Ethics, and Society*, 204–209. https://doi.org/10.1145/3278721.3278752

Lee, M. K., Jain, A., Cha, H. J., Ojha, S., & Kusbit, D. (2019). Procedural justice in algorithmic fairness: Leveraging transparency and outcome control for fair algorithmic mediation. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 1–26. https://doi.org/10.1145/3359284

Lee, N. T., & Lai, S. (2022, May 17). The U.S. can improve its AI governance strategy by addressing online biases. *Brookings*. https://www.brookings.edu/blog/techtank/20 22/05/17/the-u-s-can-improve-its-ai- governance-strategy-by-addressing-online- biases/

Lemons, M. A., & Jones, C. A. (2001). Procedural justice in promotion decisions: Using perceptions of fairness to build employee commitment. *Journal of Managerial Psychology, 16*(4), 268–281. https://doi.org/10.1108/02683940110391517

Leventhal, G. S. (1980). What should be done with equity theory? In K. J. Gergen, M. S. Greenberg, & R. H. Willis (Eds.), *Social exchange: Advances in Theory and Research* (pp. 27–55). Springer.

Lind, E. A., & Tyler, T. R. (1988). *The social psychology of procedural justice.* Springer Science & Business Media.

Lindgren, S., & Holmstrom, J. (2020). Social science perspective on artificial intelligence: Building blocks for a research agenda. *Journal of Digital Social Research*, *2*(3), 115. https://doi.org/10.33621/idsr.v2i3.65

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, *151*, 90–103. https://doi.org/10.1016/i.obhdp.2018.12.005

McKay, C. (2020). Predicting risk in criminal procedure: Actuarial tools, algorithms, AI and judicial decision-making. *Current Issues in Criminal Justice, 32*(1), 22–39. https://doi.org/10.1080/10345329.2019.1658694

Miller, M. K., & Chamberlain, J. (2015). "There ought to be a law!": Understanding community sentiment. In M. K. Miller, J. A. Blumenthal, J. Chamberlain (Eds.), *Handbook of community sentiment* (pp. 3–28). Springer. https://doi.org/10.1007/978-1-4939-1899-1

Monahan, J., & Skeem, J. L. (2016). Risk assessment in criminal sentencing. *Annual Review of Clinical Psychology, 12*(1), 489–513. https://doi.org/10.1146/annurev-clinpsy- 021815-092945

Mossman, D. (1994). Assessing predictions of violence: Being accurate about accuracy. *Journal of Consulting and Clinical Psychology*, *62*(4), 783–792. https://doi.org/10.1037/0022-006X.62.4.783

National Conference of State Legislatures. (2022, January 5). *Legislation related to artificial intelligence*. https://www.ncsl.org/research/telecommunications-and-information-technology/2020-legislation-related-to-artificial-intelligence.aspx

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man and Cybernetics. Part A, Systems and Humans, 30*(3), 286–297. https://doi.org/10.1109/3468.844354

Perry, W. L. (2013). Predictive policing: The role of crime forecasting in law enforcement operations. *Rand Corporation*.

Pinch, T. J., & Bijker, W. E. (1984). The social construction of facts and artefacts: Or how the sociology of science and the sociology of technology might benefit each other. *Social Studies of Science, 14*(3), 399–441. https://doi.org/10.1177/030631284014003004

Ramirez, M. D. (2008). Procedural perceptions and support for the U.S. Supreme Court. *Political Psychology, 29*(5), 675–698. https://doi.org/10.1111/i.1467- 9221.2008.00660.x

Rigano, C. (2019). Using artificial intelligence to address criminal justice needs. *National Institute of Justice Journal, 280,* 1–10. https://www.oip.gov/pdffiles1/nii/252038.pdf

Ritchie, K. L., Cartledge, C., Growns, B., Yan, A., Wang, Y., Guo, K., Kramer, R. S. S., Edmond, G., Martire, K. A., San Roque, M., & White, D. (2021). Public attitudes towards the use of automatic facial recognition technology in criminal justice systems around the world. *PloS One*, *16*(10), e0258241-e0258241. https://doi.org/10.1371/iournal.pone.0258241

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs: General and Applied, 80*(1), 1–28. https://doi.org/10.1037/h0092976

Rotter, J. B., Chance, J. E., & Phares, E. J. (1972). *Applications of a social learning theory of personality.* Holt, Rinehart, and Winston.

Rueda, J., Rodriguez, J. D., Jounou, I. P., Hortal- Carmona, J., Ausin, T., & Rodriguez-Arias, D. (2022). "Just" accuracy? Procedural fairness demands explainability in AI-based medical resource allocations. *AI & Society*, 1–12. https://doi.org/10.1007/s00146-022-01614-9

Schlesinger, T. (2005). Racial and ethnic disparity in pretrial criminal processing. *Justice Quarterly*, *22*(2), 170–192. https://doi.org/10.1080/07418820500088929

Scott, A. (2021). *Difference between algorithm and artificial intelligence*. Data Science Central. https://www.datasciencecentral.com/differen ce-between-algorithm-and-artificial- intelligence/

Sharan, N. N., & Romano, D. M. (2020). The effects of personality and locus of control on trust in humans versus artificial intelligence. *Heliyon*, *6*(8), e04572. https://doi.org/10.1016/i.heliyon.2020.e04572

Sherman, S. J. (1973). Internal-external control and its relationship to attitude change under different social influence techniques. *Journal of Personality and Social Psychology*, *26*(1), 23–29. https://doi.org/10.1037/h0034216

Shroff, R. (2019, September 25). Artificial intelligence explained in simple terms. *Medium*. https ://medium.com/mytake/artificial-intelligence-explained-in-simple-english- part-1-2-1b28c1f762cf

Smith, P. B., Dugan, S., & Trompenaars, F. (1997). Locus of control and affectivity by gender and occupational status: A 14 nation study. *Sex Roles, 36*(1–2), 51–77. https://doi.org/10.1007/BF02766238

Spohn, C., & Holleran, D. (2000). The imprisonment penalty paid by young, unemployed Black and Hispanic male offenders. *Criminology*, *38*(1), 281–306. https://doi.org/10.1111/L1745-9125.2000.tb00891.x

Starke, C., & Lunich, M. (2020). Artificial intelligence for political decision-making in the European Union: Effects on citizens' perceptions of input, throughput, and output legitimacy. *Data & Policy*, *2*(1). https://doi.org/10.1017/dap.2020.19

Thibaut, J. W., & Walker, L. (1975). *Procedural justice: A psychological analysis*. L. Erlbaum Associates.

Turner, K. B., & Johnson, J. B. (2005). A comparison of bail amounts for Hispanics, Whites, and African Americans: A single county analysis. *American Journal of Criminal Justice*, *30*(1), 35–53. https://doi.org/10.1007/BF02885880

Tyler, T. R. (2006). Psychological perspectives on legitimacy and legitimation. *Annual Review of Psychology, 57*(1), 375–400. https://doi.org/10.1146/annurev.psych.57.10 2904.190038

Victor, A. (2021, July 24). 10 uses of artificial intelligence in day to day life. *Daffodil*. https://insights.daffodilsw.com/blog/10-uses-of-artificial-intelligence-in-day-to-day-life

Wallston, B. S., & Wallston, K. A. (1978). Locus of control and health: A review of the literature. *Health Education Monographs*, *6*(1), 107–117. https://doi.org/10.1177/109019817800600102

Western, B. (2006). *Punishment and inequality in America*. Russell Sage Foundation.

Zadgaonkar, A. V., & Agrawal, A. J. (2021). An overview of information extraction techniques for legal document analysis and processing. *International Journal of Electrical and Computer Engineering, 11*(6), 5450–5457. https://doi.org/10.11591/iiece.v11i6.pp5450-5457

Zajonc, R. B. (2001). Mere exposure: A gateway to the subliminal. *Current Directions in Psychological Science, 10*(6), 224–228. https://doi.org/10.1111/1467-8721.00154

Zelditch, M., Jr. (2018). Legitimacy theory. In P. J. Burke (Ed.), *Contemporary social psychological theories* (pp. 340–371). Stanford University Press.

ПРИЛОЖЕНИЕ / APPENDIX

Выделенные темы, категории внутри каждой темы, их определения, количество упоминаний каждой категории в ответах, согласие или несогласие в каждой категории

Identified Themes, Categories within Each Theme, Their Definitions, the Number of Times That a Category Appeared in Their Response, and Whether They Agreed or Disagreed with the Category

Тема / Theme	Категория / Category	Рабочее определение / Operational Definition	Количество упоминаний / Number of Times Appeared
Общественное мнение / Community Sentiment	Поддержка в целом / Overall Comment Support	Отношение комментатора к ИИ в целом / Measurement of how supportive the comment was towards AI	325 Отрицательных / Negative = 229; нейтральных / Neutral = 33; положительных / Positive = 63
Неприятие технологий/ принятие технологий / Algorithm Aversion/ Algorithm Appreciation	Польза / Utility	Считают ли судьи, что ИИ может принести пользу (например, «Я думаю, ИИ может помочь мне при принятии решений») / Whether or not the judge thinks the tool will be (e.g., I think it could help me make decisions)	58 Польза / Utility Benefit = 48; отсутствие пользы / No utility benefit = 10
	Этика / Ethics	Этическое преимущество или недостаток (например, «Нельзя относиться к людям как к статистическим данным») / Ethical advantage or disadvantage (e.g., it would be wrong to treat people like statistics)	15 Этическое преимущество / Ethical advantage = 0; этический недостаток / Ethical disadvantage = 15
Локус контроля / Locus of Control	Угроза автономии/ судебному усмотрению / Threat to Autonomy/ Judicial Discretion	Ощущают ли судьи угрозу своей автономии или праву на судебное усмотрение (например, «Судьи должны самостоятельно принимать решения») / Judge's perceived threat of autonomy and discretion (e.g., Judges should make their own decisions)	61 Угроза автономии / Threat to autonomy = 61
	Высокая степень неопределенности / High Level of Uncertainty	Уровень неопределенности в представлении судей о том, как работают алгоритмы ИИ (например, «Я не совсем понимаю, как ИИ может помочь») / Judges' level of uncertainty about AI algorithms and how they work. (e.g., I am unsure of how AI can help)	40 Высокая степень неопределенности / High uncertainty = 40
	Контроль / Control	Ощущают ли судьи контроль над процессом (например, «ИИ – это черный ящик, невозможно понять, как он принимает решения») / Judge's perceived amount of control in the process (e.g., The AI is a black box and there is no way to understand the process behind its decision)	12 Отсутствие контроля / No control = 12
Процессуальное правосудие / Procedural Justice	Угроза правосудию / Threat to Justice	Ощущают ли судьи угрозу правосудию со стороны ИИ (например, «Жесткая привязка к любой системе без контроля со стороны здравого смысла – это неправильно и может привести к плохим последствиям») / Judges' perceptions of AI as a threat to justice (e.g., Further, rigid adherence to any system without a common sense review is problematic and can lead to very unjust results.)	31 Угроза правосудию / Threat to Justice = 31
	Уникальность ситуаций / Not One Size Fits All	Справедливое вынесение решения требует учета всех аспектов ситуации, а не только того, что заложено в программе (например, «Люди – сложные создания, программа не может судить их») / Sentencing and fairness demand judges to consider the whole story of the – not just whatever is programmed in the code (e.g., humans are complicated and a code cannot judge that)	39 Все ситуации уникальные / Not one size fits all = 39
	Человеческий фактор / Human Element	Мнения судей об участии человека в процессе правосудия (например, «Процесс принятия решений должен осуществляться людьми») / Judges' perceptions about having human involvement (e.g., humans need to be)	84 Да / Yes = 84

Окончание приложения / End of Appendix

Тема / Theme	Категория / Category	Рабочее определение / Operational Definition	Количество упоминаний / Number of Times Appeared
Легитимность / Legitimacy	Угроза легитимности системы / Threat to perceived legitimacy of the system	Мнения людей о системе правосудия в случае участия ИИ в процессе принятия решений (например, «Где грань между устранением предвзятости и инструкциями по вынесению приговоров? Судьи лишатся полномочий по вынесению приговоров») / How people might view the justice system if an AI is helping decisions rather than the judge making the decisions themselves (e.g., "Removing bias" sounds like "sentencing guidelines," which means sentencing decision authority is being taken from judges)	16 Угроза легитимности / Threat to legitimacy = 16
	Необходимы рекомендации / Guidelines needed	Необходимость тщательной разработки рекомендаций/инструкций/надзора/проверок/аудита/научных исследований (например, «Должны быть некие установленные нормы») / Need strict programming/guidelines/testing/oversight/checks/audits/scientific research (e.g., there needs to be some sort of set guidelines)	19 Необходимы рекомендации / Guidelines needed = 19
Выгоды / Benefit	Предвзятость судей / Judge Bias	Судьи признают наличие собственной предвзятости / Judges' admittance of their own biases	18 Предвзятость судей / Judge bias present = 18
	Желание обучиться / Willingness to Learn	Желание судей узнать об ИИ (например, «Мне необходимо изучить исследования в этой области») / Judge's willingness to learn about AI (e.g., I need to see research in this area)	10 Желание обучиться / Willing to learn = 10
Опасения / Concern	Предвзятость в данных / Bias in Data	Существует предвзятость в данных в результате ранее принятых предвзятых решений в системе уголовного правосудия (например, «Неверные данные приводят к неверным решениям») / There is bias within the data based on previously biased decisions within the criminal justice system. (e.g., garbage in garbage out)	73 Предвзятость в данных / Bias in data = 73
	Замена судей / Replace Judges	Опасения судей о полной замене их искусственным интеллектом (например, «Неужели судебный процесс могут вести роботы») / This is in reference to judges' fear about being replaced by AI (e.g., how about full robot courts)	30 Замена судей / Replace judges = 30
	Предвзятость усугубится / Make Bias Worse	Скрытая предвзятость может усугубиться (например, «По моему опыту, небелые подсудимые чаще контактируют с полицией и поэтому имеют больше криминального прошлого. Это важный элемент алгоритма, который приводит к неверным результатам») / Make implicit bias worse (e.g., My experience is that non-white defendants have longer criminal histories due to more police contact. A defendant's criminal history is a significant element of the algorithm, which produces disparate results)	26 Предвзятость усугубится / Make bias worse = 26
	Отсутствие объективности / Lack of Objective Reasoning	ИИ не учитывает объективные факторы (например, «ИИ не может учесть тонкие аспекты человеческого взаимодействия, которые говорят больше, чем любые тесты») / AI's lack of objective reasoning (e.g., AI cannot take into account the subtle aspects of human interaction, which is more telling than any paper test)	12 Отсутствие объективности / Lack of objective reasoning = 12

Примечание: Поскольку анализировались вторичные данные, одобрения Экспертного совета организации не требовалось. Эти данные были частично описаны в популярном ключе в краткой новостной заметке в бюллетене *Judicial Edge Today*, издаваемом Национальным юридическим колледжем (*NJC*) в Рино, штат Невада. Заметка доступна на сайте https://www.judges.org/news-and-info/judges-remain-skeptical-on-whether-artificial-intelligence-can-make-decisions-more-fairly-than-they-can

Об авторах

Анна Файн получила степень магистра психологии в университете штата Аризона. В настоящее время является соискателем степени PhD по программе междисциплинарной социальной психологии университета Невады в Рино. В рамках своего исследования она изучает отношение судей и широкой общественности к искусственному интеллекту и применение ИИ в обществе, в частности в судебной системе. Анна Файн благодарит Шона Марша и Монику К. Миллер за их неизменную поддержку ее исследований.

Стефани Ли обучается на первом курсе магистратуры в области социологии университета Невады в Рино. Она получила степень бакалавра в области педагогики и получает степень магистра, чтобы изучать некоммерческие и неправительственные образовательные организации на международном уровне. Среди ее научных интересов – некоммерческие и неправительственные организации, системы образования и учебные программы в различных странах, проблемы социальной справедливости и противодействия расизму. Стефани Ли выражает благодарность своим наставникам Клейтону Пиплз и Монике К. Миллер за постоянные консультации и поддержку.

Моника К. Миллер, доктор права, PhD, является профессором университета Невады в Рино. Она также сотрудничает с департаментом уголовного правосудия и программой на соискание степени PhD по междисциплинарной социальной психологии и преподает в программе по изучению процессуальной системы. Ее исследования лежат на стыке правовых и социальных дисциплин, фокусируясь на теоретическом обосновании повышения благосостояния в мире.

About the Authors

Anna Fine, M. S., graduated from Arizona State University with a Master of Science in Psychology. She is a PhD student in the Interdisciplinary Social Psychology PhD Program at the University of Nevada, Reno. Her research interests include how judges and the general public perceive artificial intelligence and its implementation within society, especially within the court system. She would like to acknowledge both Shawn Marsh and Monica K. Miller for their consistent encouragement to follow my research interests. Correspondence concerning this article should be addressed to Anna Fine, Interdisciplinary Social Psychology Ph.D. Program, Mailstop 1300 1664 N. Virginia Street, Reno, NV 89557. Email: afine@nevada.unr.edu.

Stephanie Le is a first-year Master of Sociology student at the University of Nevada, Reno. Her bachelor's degree is in Education and is obtaining her Master's degree in Sociology in order to pivot out of direct instruction into the realm of global non-profit and non-government organizations related to Education. Her research interests involve non-profit and non-government organizations, Global Education pedagogy and curriculum design, social justice, and anti-racism. She would like to acknowledge and thank her two faculty advisors Clayton Peoples and Monica K. Miller for their continued advice and support.

Monica K. Miller, J.D., PhD, is a Foundation Professor at the University of Nevada, Reno. She has appointments in the Criminal Justice Department and the Interdisciplinary Social Psychology PhD program and teaches in the Judicial Studies program. Her research is at the cross-section of law and social sciences, with focus on theory-based research that promotes well-being in the real world.

История статьи / Article history

Дата поступления / Received 04.12.2023
Дата одобрения после рецензирования / Date of approval after reviewing 10.01.2024
Дата принятия в печать / Accepted 10.01.2024