

DOI: [10.14515/monitoring.2024.5.2580](https://doi.org/10.14515/monitoring.2024.5.2580)



А. П. Казун

МОЖЕТ ЛИ ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ ПРОГНОЗИРОВАТЬ РЕШЕНИЯ СУДА? СИСТЕМАТИЧЕСКИЙ ОБЗОР МЕЖДУНАРОДНЫХ ИССЛЕДОВАНИЙ

Правильная ссылка на статью:

Казун А. П. Может ли искусственный интеллект прогнозировать решения суда? Систематический обзор международных исследований // Мониторинг общественного мнения: экономические и социальные перемены. 2024. № 5. С. 100—122. <https://doi.org/10.14515/monitoring.2024.5.2580>.

For citation:

Kazun A. P. (2024) Can Artificial Intelligence Predict Judicial Decisions? A Systematic Review of International Research. *Monitoring of Public Opinion: Economic and Social Changes*. No. 5. P. 100–122. <https://doi.org/10.14515/monitoring.2024.5.2580>. (In Russ.)

Получено: 05.03.2024. Принято к публикации: 15.08.2024.

МОЖЕТ ЛИ ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ ПРОГНОЗИРОВАТЬ РЕШЕНИЯ СУДА? СИСТЕМАТИЧЕСКИЙ ОБЗОР МЕЖДУНАРОДНЫХ ИССЛЕДОВАНИЙ

КАЗУН Антон Павлович — кандидат социологических наук, директор Института анализа предприятий и рынков, доцент департамента прикладной экономики, Национальный исследовательский университет «Высшая школа экономики», Москва, Россия
E-MAIL: akazun@hse.ru
<https://orcid.org/0000-0002-0091-5388>

Аннотация. Развитие технологий искусственного интеллекта и появление открытых баз данных судебных решений привели к стремительному совершенствованию алгоритмов, позволяющих классифицировать юридические документы и прогнозировать принимаемые судьями решения. В статье мы анализируем корпус международных исследований, посвященных вопросу о том, насколько точно ИИ может предсказывать решения судей и, как следствие, сможет ли он в перспективе заменить судью-человека. Ответ на этот вопрос складывается из анализа двух ключевых аспектов — возможности и точности прогнозирования судебных решений, а также различных ограничений, связанных с применением ИИ.

Анализ международного опыта показывает, что в последние годы точность прогнозов выросла, однако качество моделей сильно зависит от специфики задач и доступных данных. Большинство исследований анализируют решения судов высшего уровня различных стран мира, что сильно снижает их прикладной потенциал в части работы с массовыми категориями дел. Кроме того, опасения вызывает использование моделей, действующих по принципу «черного ящика», поскольку их решения трудно интерпретировать. Несмотря на стремительное

CAN ARTIFICIAL INTELLIGENCE PREDICT JUDICIAL DECISIONS? A SYSTEMATIC REVIEW OF INTERNATIONAL RESEARCH

Anton P. KAZUN¹ — *Cand. Sci. (Soc.), Director at the Institute for Industrial and Market Studies; Assistant Professor at the Department of Applied Economics*
E-MAIL: akazun@hse.ru
<https://orcid.org/0000-0002-0091-5388>

¹ HSE University, Moscow, Russia

Abstract. Advancements in artificial intelligence technologies and the emergence of open databases containing judicial decisions have led to rapid improvements in algorithms capable of classifying legal documents and forecasting decisions made by judges. This article examines a body of international research dedicated to how accurately AI can predict judges' decisions and whether it could potentially replace human judges in the future. The answer to this question is formed by analyzing two key aspects: the capability and accuracy of predicting judicial decisions and the various constraints associated with using AI.

Analysis of international experience shows that the accuracy of predictions has increased in recent years; however, the quality of the models depends greatly on the specificity of the tasks and the available data. Most studies analyze decisions from higher courts worldwide, significantly reducing their practical potential for dealing with mass categories of cases. Moreover, concerns have arisen regarding the use of models that operate on a “black box” principle, as their decisions are difficult to interpret. Despite the rapid development of AI technologies, the complete replacement of judges is unlikely because of the range of methodological limitations, including insufficient quality and volume of data, issues

развитие ИИ-технологий, полная замена судей вряд ли возможна в ближайшее время ввиду целого ряда методологических ограничений, включая недостаточное качество и объем данных, проблему интерпретируемости, сложность понимания юридического и культурного контекста, ограничения переноса на другие правовые системы. Однако ИИ-технологии возможно использовать для сокращения издержек по работе с материалами дела.

Ключевые слова: искусственный интеллект, предсказание судебных решений, машинное обучение, глубокое обучение, классификация юридических документов, точность алгоритмов

Благодарность. Исследование выполнено за счет гранта Российского научного фонда № 23-78-10073 «Разработка и апробация методики автоматизированного анализа текстов приговоров российских судов для социально-правовых исследований (на примере насильственных преступлений)» (см. подробнее: <https://rscf.ru/project/23-78-10073/>).

Введение

Уже с 1950-х годов исследователи пытались количественными методами спрогнозировать решения судей [Kort, 1957], но до недавнего времени технологии позволяли достичь лишь весьма ограниченного результата в этой области [Ashley, Brüninghaus, 2009]. В последние годы с развитием технологий искусственного интеллекта (далее ИИ) и расширением доступности открытых судебных данных качество моделей, предсказывающих решения судей, существенно выросло [Medvedeva, Wieling, Vols, 2022]. Это привело к расширению дискуссии о том, может ли ИИ в обозримом будущем заменить профессионального судью или по крайней мере упростить работу судей [Shi, Sourdin, Li, 2021; Sourdin, Cornes, 2018; Taylor, 2023; Xu, 2022]. В том числе с 2017 по 2022 г. резко возросло количество связанных с юридической тематикой докладов, представленных на профильных конференциях по компьютерным наукам [Cui, Shen, Wen, 2023]. В настоящем исследовании мы попробуем на основе систематического анализа международного опыта описать возможности и риски, которые несет в себе появление технологии ИИ в сфере судебного производства, а также ответить на вопрос о том, возмож-

with interpretability, challenges in understanding legal and cultural context, and limitations in transferring models to other legal systems. However, AI technologies can be used to reduce the costs associated with case material handling.

Keywords: artificial intelligence, prediction of judicial decisions, machine learning, deep learning, legal document classification, algorithmic accuracy

Acknowledgments. The research was funded by the Russian Science Foundation grant No. 23-78-10073 “Development and approbation of the algorithm for automated analysis of the court decisions texts for socio-legal studies (based on cases of violent crimes)” (see more: <https://rscf.ru/project/23-78-10073/>).

но ли заменить судью на ИИ. Нужно отметить, что в статье мы будем говорить исключительно о «слабом ИИ», умеющем решать конкретные задачи (в нашем случае анализировать юридические тексты и прогнозировать исходы), оставляя в стороне вопрос о создании «сильного», или общего, искусственного интеллекта, способного воспроизводить весь мыслительный процесс человека.

Для юридической сферы обсуждение вопроса о внедрении ИИ, пожалуй, не менее релевантно, чем для медицины, поскольку в обоих случаях на кону нередко стоят жизни и здоровье людей. ИИ способен анализировать юридические тексты в таком объеме и с такой скоростью, которые не доступны человеку. Однако у внедрения ИИ в судебную сферу — или же, шире, в сферу правоприменения — «цена ошибки» может быть высока. Так, в фильме 2001 г. «Особое мнение» главный герой, работавший в корпорации, предсказывавшей тяжкие преступления до их совершения, сам стал жертвой работы этого алгоритма, когда его обвинили в преступлении, которого он не совершал. Хотя до подобного будущего нам далеко, с развитием ИИ возможность предугадывать преступления в статистическом смысле становится вполне реальной [Gerber, 2014]. Помимо обозначенных этических вопросов, повсеместное внедрение ИИ поднимает и другие. Например, такая практика создает проблему защиты конфиденциальной информации, а также повышает актуальность вопроса о приватности данных обычных граждан. Однако на другой чаше весов лежит возможность существенно сократить издержки по принятию судебных решений, а возможно, и повысить качество и беспристрастность правосудия.

В российском контексте обсуждение последствий внедрения ИИ в сферу правоприменения также релевантно, поскольку развитие данной технологии является одним из национальных приоритетов. В январе 2024 г. В. В. Путин поручил Верховному суду, Генпрокуратуре, Следственному комитету, МВД и Министерству юстиции РФ продумать возможность использования ИИ в расследовании преступлений¹. В Послании Федеральному собранию от 29 февраля 2024 г. было дано поручение нарастить мощности отечественных суперкомпьютеров в десять раз к 2030 г.² Хотя первые инициативы в сфере правоприменения касаются борьбы с преступностью с применением видеокамер и технологии распознавания лиц, нет сомнений, что ИИ будет использован и в судебной практике. В частности, эксперименты по созданию «умных судов» уже проводятся в Китае [Rusakova, 2021; Shi, Sourdin, Li, 2021]. Некоторые российские судьи были скептически настроены к перспективе их замены ИИ³, однако на 2024 г. анонсирован запуск системы «Правосудие онлайн»⁴, которая сделает первый шаг к автоматизации типичных составляющих судебного процесса. Кроме того, летом 2024 г. судья Верховного суда РФ О. Зателепин высказал идею, что ИИ может внедряться в российскую систему

¹ Путин поручил улучшить использование ИИ для расследования преступлений // РБК. 2024. 17 января. URL: <https://www.rbc.ru/rbcfreenews/65a7a3359a794761a6913056> (дата обращения: 05.03.2024).

² Путин поручил нарастить мощность суперкомпьютеров в десять раз // РБК. 2024. 29 февраля. URL: <https://www.rbc.ru/rbcfreenews/65e07a3a9a79472386b14ad3> (дата обращения: 05.07.2024).

³ В Совете судей отвергли идею замены судьи на ИИ // Право.ру. 2023. 26 октября. URL: <https://pravo.ru/news/249529/> (дата обращения: 05.07.2024).

⁴ Куликов В. Суды планируют подключить искусственный интеллект к составлению решений // Российская газета. 2023. 25 мая. URL: <https://rg.ru/2023/05/25/robot-pomozhet-rassudit.html> (дата обращения: 05.07.2024).

правосудия для помощи судьям в принятии решений о наказании и для сокращения вероятности ошибок⁵, однако только в роли советника или помощника судьи.

Поскольку большинство российских судей перегружены работой⁶, внедрение ИИ-технологий для анализа материалов дел по меньшей мере могло бы способствовать высвобождению времени для содержательного знакомства с их ключевыми обстоятельствами. Для юристов и их клиентов появление подобных технологий существенно упростило бы подготовку к судебному разбирательству, а также дало бы возможность оценить шансы на успех [Jacob de Menezes-Neto, Clementino, 2022].

Статья структурирована следующим образом. Сначала мы опишем алгоритм поиска релевантной научной литературы, посвященной вопросу о предсказании судебных решений с помощью ИИ, и работу с ней. Затем последовательно ответим на несколько ключевых вопросов об использовании ИИ в классификации юридических текстов и предсказании решений, принимаемых судьями. Далее мы обсудим ограничения применимости данной технологии и рассмотрим ключевые этические вопросы, связанные с внедрением ИИ, а также сформулируем практические рекомендации.

Подбор источников и исследовательские вопросы

Отбор источников для исследования осуществлялся через Research Rabbit⁷ — ИИ-приложение, позволяющее строить сети цитирований между научными источниками. Преимущества данного подхода по сравнению с ручным поиском литературы по ключевым словам в базах научного цитирования Scopus и WoS состоит в том, что алгоритм выделяет ядро научной дискуссии, а также меньше зависит от выбора ключевых слов. В качестве отправной точки для построения сети источников использовалась выборка из систематического обзора [Medvedeva, Wieling, Vols, 2022], в котором были отобраны релевантные исследования за период с 2015 по 2021 г., посвященные вопросу о предсказании судебных решений. Использование Research Rabbit позволило актуализировать эту выборку, а также найти тексты по смежным тематикам, включая обсуждение вопросов этики. Точность отбора исследований дополнительно проверялась с помощью второго ресурса — Inciteful⁸. Он работает аналогичным образом и позволяет идентифицировать тексты, близкие по своему содержанию к имеющейся выборке. Используемый нами подход, как и поиск по ключевым словам, не исключает возможности упустить то или иное исследование, однако последнее точно не вносит высокого вклада в научную дискуссию ввиду отсутствия у него сильных взаимосвязей с основным корпусом литературы.

Итоговая сеть включала в себя 107 источников за период с 2004 по 2023 г., посвященных классификации судебных текстов, но большинство работ приходится на период с 2018 г., поскольку дискуссия по вопросам использования ИИ

⁵ Бурнов В. ВС: ИИ может помочь снизить число ошибок при назначении наказания // РАПСИ. 2024. 1 июля. URL: https://www.rapsinews.ru/judicial_mm/20240701/310059533.html (дата обращения: 05.07.2024).

⁶ Нагрузку на судей закрепят на законодательном уровне // Адвокатская газета. 2023. 06 декабря. URL: <https://www.advgazeta.ru/novosti/nagruzku-na-sudey-zakrepjat-na-zakonodatelnom-urovne/> (дата обращения: 05.07.2024).

⁷ URL: <https://www.researchrabbit.ai/> (дата обращения: 05.07.2024).

⁸ URL: <https://inciteful.xyz/> (дата обращения: 05.07.2024).

интенсифицировалась в последние годы. В целом полученная совокупность текстов охватывает все наиболее релевантные источники, посвященные вопросу о предсказании судебных решений. Для количественного анализа динамики качества предсказательных моделей было отобрано 34 источника за период с 2015 по 2023 г., сопоставимых по постановке задач и метрике оценки точности.

Проведенный анализ зарубежного опыта имеет ряд ограничений. Во-первых, мы не анализируем детально методологические вопросы, которым посвящен объемный корпус литературы в сфере компьютерных наук. В данной статье мы не ставим перед собой задачу описать существующие алгоритмы по классификации юридических текстов или же раскрыть способы их применения. Это ограничение в том числе позволило нам остановить построение сети в Research Rabbit на том моменте, когда новые смежные тексты перестали касаться предсказаний судебных решений и начали затрагивать связанные тематики в области математики и компьютерных наук. Во-вторых, мы не анализировали работы, посвященные предсказаниям в сфере правоприменения и преступности, фокусируясь только на стадии судебного дела. Это сделано для того, чтобы сохранить четкий фокус исследования и не уходить в другие сложные вопросы, такие как использование ИИ в качестве инструмента для поиска преступников, — эта тема представляет высокий научный интерес, но является предметом для отдельного анализа. В-третьих, мы не рассматриваем детально работы, моделирующие судебные решения на основе линейных регрессионных моделей, деревьев решений и других классических количественных методов, поскольку в последние годы эти исследования уступили дорогу более продвинутым методам, основанным на современных ИИ-технологиях. При этом в анализ были включены статьи, посвященные этическим аспектам применения ИИ в сфере правоприменения, которые не были напрямую связаны с прогнозированием решений.

Анализ источников литературы на русском языке показал, что, несмотря на наличие довольно большего числа исследований [Захарцев, Сальников, 2018; Коваленко и др., 2020; Кравчук, 2021; Степанов, Басаганов, 2022], они предлагают в основном постановку актуальных вопросов и анализируют юридические аспекты проблемы, но не систематизируют международный опыт использования ИИ-технологий для предсказания судебных решений. Таким образом, данная статья является первым в России систематическим анализом корпуса международных исследований, посвященных предсказанию судебных решений и перспективам использования ИИ в работе профессиональных судей.

На данный момент нет исследований, которые бы на основе одного из популярных алгоритмов глубинного обучения моделировали решения российских судей. Имеющиеся работы о факторах принятия судебных решений (например, [Volkov, 2016]) используют в анализе регрессионные модели. В проекте «Алгоритм света»⁹, а также в исследовании [Zhuchkova, Kazun, 2023] с помощью машинного обучения из текстов приговоров выделялась информация о партнерских или родственных отношениях между обвиняемым в убийстве и жертвой, что является, пожалуй, наиболее релевантным на текущий момент примером, относящимся к ИИ-техно-

⁹ URL: https://github.com/LanaLob/algorithm_sveta (дата обращения: 06.10.2024).

логиям. Однако основной анализ в последней работе также строился с использованием регрессионных моделей. Таким образом, настоящий систематический обзор может стать отправной точкой для появления первых отечественных исследований, прогнозирующих судебные решения.

Выборка источников позволит нам ответить на следующие ключевые вопросы. Во-первых, что именно ИИ может прогнозировать и на какие вопросы он способен дать ответ? Во-вторых, какой эмпирический материал (решения каких судов и в каких странах) брался за основу для обучения моделей? В-третьих, какими методами и подходами исследователи достигают своих результатов? Хотя третий вопрос не является ключевым для данного исследования, мы сделаем краткий обзор подходов, которые применяются для прогнозирования результатов судебных дел. В-четвертых, мы оценим точность предсказания судебных решений с помощью ИИ-технологий и поговорим об ограничениях, связанных с оценкой точности алгоритмов. Наконец, мы обсудим ограничения по использованию искусственного интеллекта, в том числе этические вопросы, связанные с его внедрением.

Что способен спрогнозировать ИИ?

На первом шаге следует определить терминологию. Есть значительная разница между тем, как термины «прогнозирование» и «классификация» употребляются в широком контексте, и тем, как они используются в компьютерных науках, а значит, и в большинстве научных статей о прогнозировании судебных решений с помощью ИИ. Когда социальные исследователи говорят о классификации, они, как правило, имеют в виду сортировку по типам или разделение на группы. В большинстве статей про роль ИИ под классификацией понимается более широкий класс явлений, в который входит в том числе предсказание результатов. В работе [Medvedeva, Wieling, Vols, 2022] проводится полезное разделение исследований о предсказаниях судебных решений с использованием ИИ на три группы: идентификация результата, категоризация текстов и предсказание решений. О каждом из этих типов мы скажем далее, но важно, что в научной литературе все они могут называться классификацией.

Термин «прогнозирование» (или «предсказание») может иметь два значения. Если у нас есть массив уже вынесенных судебных решений, то мы можем обучить алгоритм, скажем, на 80 % этих дел, а протестировать его на оставшихся 20 %. В этом случае алгоритм будет предсказывать решения судей с некоторой точностью. Не будет ошибки в том, чтобы называть это предсказанием, однако оно не имеет никакого отношения к исходам будущих дел или тем решениям, которые реальные судьи еще только должны вынести. То есть мы имеем дело с предсказанием в «слабом» значении этого термина. Предсказание в «сильном» значении означало бы, что наша программа способна предсказывать реальные решения судей в будущем. Большинство моделирующих судебные решения исследований [Medvedeva, Wieling, Vols, 2022] понимают под предсказанием именно «слабую» версию этого термина.

Мы также будем говорить о предсказании преимущественно в «слабом» значении, используя в качестве синонима термин «прогнозирование». Прогнозирование всегда предполагает определенную вероятность и доверительные интервалы,

но нужно помнить, что точность таких предсказаний может не в полной мере отражать способность модели предсказывать будущие решения в реальных делах. В целом любой прогноз, будь то прогноз инфляции, результатов выборов или судебного решения, строится на данных о прошлом и исходит из предпосылки, что в будущем действия людей (в том числе профильных специалистов) будут определяться тем же набором факторов, что и ранее. В реальности чаще всего так и происходит, что делает прогнозы в целом важным ориентиром при принятии решений, однако всегда остается риск появления новых факторов, которые «ломают» предыдущие закономерности и делают прогнозы, основанные на прошлых данных, менее надежными.

Как было отмечено выше, весь корпус литературы о классификации судебных решений можно разделить на три типа. Первый тип — поиск в тексте документа вердикта или резолютивной части. Эта задача не представляет никакой сложности для юриста-человека. Машинное обучение в данном случае применяется для первичной обработки большого объема данных и экономии времени. В этой задаче цель исследователей и программистов состоит в том, чтобы приблизить качество работы машины к результатам, которые может дать человек. Исследования, ставящие перед собой задачу идентификации результата, вполне могут достигать точности 99 %, что означает не предсказание решений судьи, а лишь способность выделить из документа место, где решение описано. В дальнейшем мы не будем относить исследования этого типа к предсказаниям судебных решений.

Второй тип исследований — категоризация юридических текстов. Задача состоит в выявлении релевантных обстоятельств дел или их характеристик, например пола обвиняемого или жертвы, обстоятельства совершения преступления и пр. Хотя на первый взгляд кажется, что второй тип задач близок к первому, категории могут строиться не на явных текстовых фрагментах, а на анализе всего документа, например, если оценивается идеологическая ориентация суда [Shaikh, Sahu, Anand, 2020]. В случае с выделением экстралегальных факторов ИИ уже может соперничать с человеком, поскольку человек не всегда может со стопроцентной точностью провести категоризацию. Однако к предсказанию судебных решений относится не каждая такая задача, а лишь та, что связана с категоризацией решений — вынесен приговор или нет, одобрена жалоба или нет, и т. д.

Наконец, третий тип задач является собственно предсказанием исхода — будет приговор оправдательным или обвинительным, признает суд нарушение прав человека или нет и т. д. Кроме того, предсказывать можно и размер наказания, будь то штраф или срок лишения свободы. В различении между вторым и третьим типом есть тонкая граница. Исследователи [Medvedeva, Wieling, Vols, 2022] относят к третьему типу только те немногие модели [Medvedeva et al., 2021; Sharma et al., 2015; Waltl et al., 2017], которые предсказывают решения судей на основе данных, доступных до вынесения решения (то есть такие модели близки к «сильному» определению термина предсказание). Предсказания же в «слабом» значении попадают во вторую категорию. Ввиду немногочисленности исследований третьего типа дальше мы будем относить к моделям, предсказывающим решения судей, исследования, попадающие как во вторую, так и в третью группу.

Решения каких судов уже прогнозирует ИИ?

Большинство исследований о предсказаниях судебных решений выполнены на данных Европейского суда по правам человека (далее — ЕСПЧ) [Chalkidis, Androutsopoulos, Aletras, 2019; O’Sullivan, Beel, 2019; Medvedeva, Vols, Wieling, 2020; Aletras et al., 2016; Kaur et al., 2019; Medvedeva et al., 2020; Quemy, Wrembel, 2020] и Верховного суда США [Kaufman, Kraft, Sen, 2019; Katz, Bommarito, Blackman, 2017; Sharma et al., 2015]. Базы данных решений этих судов находятся в открытом доступе, что позволяет тестировать на них новые алгоритмы. Остальные исследования также, как правило, фокусируются на решениях судов высших инстанций — конституционного суда Турции [Sert, Yildirim, Haşlak, 2022], верховного суда Филиппин [Virtucio et al., 2018], Тайваня [Kowsrihawatt, Vateekul, Boonkwan, 2018], Бразилии [Freitas et al., 2022] и Индии [Malik et al., 2021].

Существенно меньше исследований делаются на больших выборках судов общей юрисдикции или же судов по отдельным вопросам из различных стран мира (например, по налогам [Alarie, Niblett, Yoon, 2017; Waltl et al., 2017], по коммерческим спорам [Bagherian-Marandi, Ravanshadnia, Akbarzadeh-T, 2021]). В последнее время все чаще создаются новые открытые данные [Cui et al., 2023]. Самая крупная и известная из публичных баз данных — Cail2018 [Zhong et al., 2020], включающая 2,6 млн кейсов, рассмотренных Верховным судом Китая. Следует отметить, что лишь немногие исследования (например, [Jacob de Menezes-Neto, Clementino, 2022]) работают с данными судов нижних инстанций.

Открытые базы данных ([Alali et al., 2021; Xiao et al., 2018; Malik et al., 2021; Sebők, Kiss, Járay, 2023]) позволяют тестировать новые алгоритмы анализа судебных дел, что, в свою очередь, служит для исследователей ориентиром по качеству ИИ-моделей. Чтобы доказать, что новая модель лучше, достаточно провести тесты на одном из доступных массивов данных и превзойти последний лучший результат.

Как ИИ прогнозирует решения?

В задачу нашего обзора не входит детальное описание методологии работы с данными, поскольку применяемые способы анализа судебных кейсов многочисленны и разнообразны, а также требуют специальных навыков для освоения. Однако ниже мы кратко опишем основные типы подходов, которые используются в исследованиях, что позволит нам лучше понимать связанные с ними возможности и ограничения.

Снова несколько слов нужно сказать о терминологии [Sert, Yildirim, Haşlak, 2022: 8]. ИИ, машинное обучение и глубокое обучение можно соотносить, используя метафору матрешки, поскольку они «вложены» друг в друга. ИИ — это самое широкое понятие, в которое попадает множество методов анализа данных. Обработка естественного языка (Natural Language Processing или NLP), позволяющая компьютеру «читать» и «понимать» человеческий язык, является одним из ключевых способов использования ИИ, но далеко не единственным. В случае с прогнозированием судебных решений мы имеем дело именно с NLP. Следующий уровень — машинное обучение, которое представляет собой составную часть ИИ. Дальше — глубокое обучение (deep learning), один из вариантов машинного обучения. При использовании любого из методов тексты судебных решений, как правило, тре-

буют предобработки. Предобработка данных, в свою очередь, может строиться на различных методах машинного обучения и иметь разную степень точности.

Основное различие в методах предсказания судебных дел проходит между глубоким обучением и другими, как правило, более простыми методами машинного обучения [Alcántara Francia, Nunez-del-Prado, Alatrística-Salas, 2022]. Один из стандартных видов машинного обучения — обучение с учителем (supervised learning). Идея обучения с учителем состоит в том, что алгоритм тренируют на некоторой выборке дел, содержащей параметры, связанные с результатами, то есть решениями судей. Качество обучения алгоритма проверяют уже на другой, но аналогичной по структуре выборке, в которой содержатся только параметры дел — алгоритм же должен предсказать результат. Точность прогноза в данном случае равна проценту корректно предсказанных решений, что мы подробнее обсудим в следующем разделе. Такие традиционные формы машинного обучения, как логистическая регрессия или деревья решений, хорошо поддаются интерпретации, хотя некоторые более сложные модели (например, случайный лес) представляют собой «черные ящики», которые интерпретировать сложнее.

Методы глубокого обучения используют нейронные сети со множеством слоев, что позволяет обучать модели на больших данных. Такие модели могут работать с миллионами параметров (выделить которые вручную невозможно), поэтому они отличаются от других методов машинного обучения сложностью моделей, большими объемами данных и использованием существенных вычислительных ресурсов. Результаты глубокого обучения, в отличие от результатов традиционных методов машинного обучения, интерпретировать сложно или же невозможно, что создает серьезные ограничения, поскольку в юридической сфере интерпретируемость имеет решающее значение.

В настоящее время для классификации юридических текстов и предсказания судебных решений чаще всего используют предварительно обученные языковые модели [Song et al., 2022], такие как BERT [Devlin et al., 2019], относящиеся к моделям глубинного обучения. Этот метод предполагает, что модель сначала обучают на большом количестве текстов, а затем используют в различных новых задачах. Этот подход существенно отличается от более ранних методов [Ashley, Brüninghaus, 2009], в которых алгоритм создавался специально для конкретной задачи и работы с определенными массивами данных. Такие алгоритмы, как BERT, более универсальные, но одновременно и ограниченные ввиду того, что могут недоучитывать специфику юридических текстов. Однако их можно обучить дополнительно и создать специализированную версию для решения определенных задач, такую как LegalBERT, обученную на юридических текстах [Chalkidis et al., 2020]. Как показывают исследования, специализированные языковые модели могут давать на 1—5% более высокие показатели точности, чем общие [Song et al., 2022], но их обучение, конечно, связано с дополнительными издержками для исследователя. Упомянутые BERT и LegalBERT являются лишь двумя актуальными примерами предварительно обученных языковых моделей, однако есть и множество других. В задачи настоящего исследования не входит более глубокий анализ особенностей разных моделей, с которыми можно ознакомиться в работах [Alcántara Francia, Nunez-del-Prado, Alatrística-Salas, 2022; Song et al., 2022].

Отметим, что технология генеративного искусственного интеллекта ChatGPT также относится к классу предварительно обученных языковых моделей, хотя основная задача ChatGPT состоит в генерации текста, а не в классификации. Потому BERT и аналогичные модели все еще лучше отвечают задачам классификации юридических текстов и прогнозирования судебных решений, чем ChatGPT, который, в свою очередь, лучше приспособлен для составления юридических документов.

С какой точностью можно прогнозировать решение суда?

Как было отмечено выше, под классификацией результатов могут пониматься разные задачи — от точности выделения в тексте приговора резолютивной части до предсказания реальных решений, которые еще не были вынесены. При этом сами результаты могут быть различными: виновен или не виновен, жалоба истца принята или отклонена [Bagherian-Marandi et al., 2021], апелляция успешна или нет [Jacob de Menezes-Neto, Clementino, 2022; Waihl et al., 2017], права признаны нарушенными или нет [Sert, Yıldırım, Haşlak, 2022], и т. п. Как правило, точность прогноза оценивается для бинарных переменных, хотя размеры штрафов или сроки приговора, пусть с меньшей точностью, также можно предсказывать.

Исследователи выделяют несколько ключевых метрик для расчета точности [ibid.]: *Accuracy* (может быть переведена как аккуратность, но именно эта метрика чаще всего отражает точность), *Precision* (точность), *Recall* (отзыв) и *F1 Score*. *Accuracy* — это сумма правильно классифицированных позитивных и негативных решений, поделенная на их общее число. *Precision* и *Recall* означают соответственно долю верно предсказанных позитивных случаев среди всех предсказанных и долю верных позитивных предсказаний среди всех реальных позитивных случаев. *F1 Score* — это индекс, который часто используется для сравнения качества различных моделей. Он рассчитывается следующим образом:

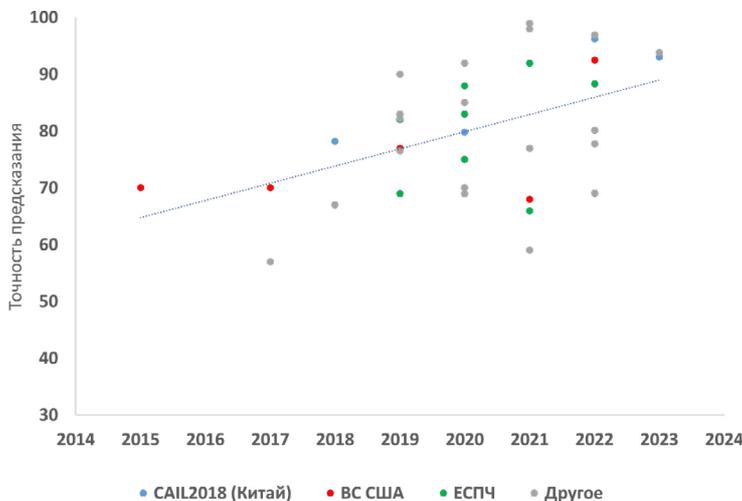
$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

При сравнении точности различных исследований нужно учитывать, какой из параметров авторы выводят в качестве точности — *Accuracy* или *F1*, поскольку это может давать различия от нескольких десятых процента до нескольких процентов. Хотя большинство работ используют эти индексы, в последнее время их критикуют за то, что они могут давать слишком оптимистичные результаты [Chicco, Jurman, 2020]. В качестве замены им предлагается *Matthews correlation coefficient* (MCC), который дает хорошие результаты лишь в том случае, если алгоритму удастся показать себя эффективно во всех четырех вариантах: верный позитивный результат, верный негативный результат, ложный позитивный результат и ложный негативный результат.

Точность алгоритмов находится обычно в диапазоне от 60% до 99%, однако более низкий показатель точности не всегда означает, что алгоритм плохой: возможно, он хороший, просто с его помощью решалась более сложная задача. По этой причине дескриптивный анализ, который мы предлагаем ниже, нужно воспринимать именно как иллюстрацию базового тренда и разброса в точности предсказания результатов судебных дел, сделанных в существующих работах.

Анализ международного опыта по отобранным нами 34 статьям позволил оценить динамику предсказательной силы моделей, классифицирующих тексты судебных решений (см. рис. 1).

Рис. 1. Динамика точности моделей, предсказывающих решения судов с 2015 по 2023 г. (для различных баз данных)¹⁰



Получившаяся картина отражает общий возрастающий тренд в предсказательной силе моделей. Однако график одновременно предлагает и весьма упрощенную картину, поскольку в разных исследованиях ставились различные цели, применялись разные алгоритмы анализа и разные оценки точности. Например, исследования [Jacob de Menezes-Neto, Clementino, 2022] на графике нет, поскольку в нем использовался альтернативный метод оценки точности, несопоставимый напрямую с другими исследованиями.

С методологической точки зрения лучше сравнивать точность решения одинаковых задач, выполненных на одних и тех же данных, таких как база решений ЕСПЧ, Верховного суда США или база данных Cail2018. Эти популярные открытые массивы данных публиковались вместе с бенчмарками (ориентирами), которые задавали для всех будущих алгоритмов базовый уровень качества. Если сравнивать прогресс внутри отдельных баз данных, то также заметен возрастающий тренд. Например, точность предсказания по китайской базе данных Cail2018 выросла с 78% [Xiao et al., 2018] в 2018 г. до 96% в 2023 г. [Cui et al., 2023]. В 2015—2017 гг. алгоритмы предсказывали решения Верховного суда США с точностью

¹⁰ Примечание: график построен автором на основе 38 оценок, выполненных в рамках 34 исследований. Из них 19 наблюдений в готовом виде взяты из систематического обзора [Medvedeva, Wieling, Vols, 2022], остальные собраны и закодированы автором. В выборку включены только оценки для решений судов, предполагающих решения с двумя исходами (1 или 0). Вертикальная ось — значения точности (Accuracy); для некоторых исследований, где отсутствует значение Accuracy, приведено значение F1; горизонтальная ось — год публикации исследования. Синяя линия на графике — линия тренда для всех наблюдений.

70% [Katz, Bommarito, Blackman, 2017; Sharma et al., 2015], а в 2022 г. появились модели, предсказывающие их на 92,5% [Alghazzawi et al., 2022]. Алгоритмы, моделирующие решения ЕСПЧ, также прогнозируют решения с точностью до 92% [Medvedeva et al., 2020]. Но, как отмечалось ранее, речь идет о предсказании в его «слабой» форме, то есть о качестве модели, обученной на уже имеющихся решениях и проверенной на них же. Эти модели будут давать адекватный прогноз о будущем только при условии, что внешние обстоятельства, такие как законодательство, состав суда, политическая обстановка и пр., не поменяются значимым образом.

Еще один интересный вопрос о точности предсказаний связан с выбором точки для сравнения. Естественным выбором кажется сравнение модели с реальными решениями. Такой подход использовался во всех исследованиях, включенных в рисунок 1. Однако есть и альтернативная точка зрения: чтобы понять, может ли ИИ заменить юриста, корректнее делать сравнение с предсказательными способностями реального человека, который является экспертом в той или иной области. Например, в работе [Jacob de Menezes-Neto, Clementino, 2022] на данных судов о более чем 765 тыс. кейсов из Бразилии сравнивалась точность предсказания исхода апелляций, которую может показать ИИ или 22 эксперта из числа судей и сотрудников судов (оценивших случайную выборку в 690 кейсов). Согласно результатам, способности ИИ превосходят способности человека почти в три раза. Впрочем, это не первое исследование, которое показывает, что ИИ может превзойти человека. В исследовании [Ruger et al., 2004] машина также превзошла экспертов-юристов в точности предсказания решений Верховного суда США, правда, не со столь высоким отрывом.

Из сказанного можно сделать несколько выводов. Во-первых, точность работы алгоритмов по классификации судебных решений сильно зависит от данных и от задач, которые ставятся перед исследователями. Во-вторых, в последние годы наблюдается явный тренд на повышение качества предсказательных моделей. При сравнении с предсказательными способностями экспертов-юристов некоторые алгоритмы демонстрируют явное превосходство над человеком. Впрочем, о стопроцентной точности, которая позволила бы заменить судью, пока говорить тоже не приходится.

Дискуссия: может ли ИИ заменить судью?

Прежде чем ответить на главный вопрос нашего исследования, суммируем основные ограничения в использовании ИИ в судебной практике.

Во-первых, результаты моделей сильно зависят от качества данных. Чтобы модель могла давать качественные предсказания, тексты судебных решений должны иметь похожую структуру, а также их должно быть достаточно много для обучения. Поэтому уникальные решения конституционных судов существующие методы анализа предсказывать не способны [Sert, Yildirim, Haşlak, 2022]. Во-вторых, отдельный технический вызов связан с работой на языках, отличающихся от английского, например, на китайском, турецком [ibid.] или португальском [Jacob de Menezes-Neto, Clementino, 2022]. Впрочем, это ограничение преодолимо — предварительно обученные языковые модели можно доучить. В-третьих, алгоритмы,

построенные на малых или специфических выборках, имеют меньше практической пользы, чем модели, построенные для судов общей юрисдикции [ibid.]. Алгоритмы научились отлично предсказывать (в «слабом» значении этого термина) решения Верховного суда США и ЕСПЧ, но это не позволяет распространить результаты на суды нижнего уровня. Вопрос о разработке моделей, предсказывающих решения судов по типичным делам, которые касаются основной массы населения, пока не был решен: имеющиеся исследования в этой области крайне немногочисленны, а универсальных решений не существует. В-четвертых, большинство моделей имеют дело с предсказаниями в «слабом» значении этого термина и не дают гарантий, что можно точно предсказать будущие решения судей. Можно описать это как проблему индукции — сколько бы много наблюдений мы ни сделали о прошлом, они не позволяют заглянуть в будущее. Таким образом, высокая точность предсказаний ИИ на рисунке 1 не должна вводить нас в заблуждение: она лишь отражает хорошую тренировку моделей на очень ограниченном количестве данных о прошлом.

Наконец, ключевая проблема моделей, основанных на глубоком обучении, состоит в том, что их сложно интерпретировать. Судья-человек может объяснить логику принятия им решений. Классические модели, такие как регрессии и деревья решений, также дают легко интерпретируемые результаты. Например, из регрессионной модели о влиянии различных факторов на решения российских судей по делам об убийствах [Zhuchkova, Kazun, 2023] несложно понять, какое влияние на приговор имеет каждый анализируемый параметр дела — пол, признание вины, повторное совершение преступления и пр. Однако методы предсказания судебных решений, основанные на глубоком обучении, представляют собой «черные ящики». Хотя модели дают весьма точные предсказания, они не позволяют исследователям понять, какие именно легальные и экстралегальные факторы стоят за каждым конкретным решением. Однако преимущество ИИ над линейными регрессионными моделями состоит в том, что последние могут недоучитывать нелинейность влияния многих факторов [Alarie et al., 2017]. Также именно модели, основанные на глубоком обучении, дают наиболее точные прогнозы.

Одна из самых высокоцитируемых работ о принятии решений на основе машинного обучения [Rudin, 2019] содержит ключевой тезис в своем названии — не следует использовать «черные ящики» для принятия решений в тех случаях, когда цена ошибки велика. В работе автор называет мифом классическую дилемму между объяснительной силой модели и ее интерпретируемостью. Интерпретируемые модели тоже могут иметь высокую предсказательную способность. Например, в исследовании [Tan et al., 2020] на 1,3 млн кейсах о преступлениях в Китае была построена поддающаяся интерпретации модель с точностью, сопоставимой или превышающей результаты моделей, действующих по принципу «черных ящиков». Ключевое преимущество моделей, позволяющих интерпретировать результаты, состоит в возможности работы над ошибкой. Если ИИ-алгоритм вынес неверный приговор (или предложил неверный курс лечения), это может оказать существенное влияние на судьбу человека. Если подобную ошибку совершает «черный ящик», мы не сможем сказать, почему она была совершена и повторится ли она в будущем. Из-за этого недостатка подобные системы мо-

гут никогда не получить высокого уровня доверия у общества. Вместе с тем это не означает, что развитие прогнозирования на основе «черного ящика» в сфере медицины или права должно остановиться: если такие алгоритмы будут обладать статистически более высокой точностью, то для некоторых задач точность и скорость могут перевесить прозрачность.

В 2016 г. Верховный суд штата Висконсин США вынес решение о том, что при принятии решений судьи могут опираться на результаты анализа, выполненного алгоритмами, даже если принцип работы алгоритма не вполне прозрачен [Beriaín, 2018]. Аргументация такого решения сводилась к двум тезисам: качество работы алгоритмов весьма высоко, а судья сам по себе достаточно компетентен, чтобы принимать взвешенное решение о том, нужно ли ориентироваться на результаты алгоритма. Фактически этот прецедент открыл дорогу использованию ИИ в качестве ассистента судьи при принятии решений в США. Поводом для этого разбирательства стала система COMPAS (Correctional Offender Management Profiling for Alternative Sanctions). По замыслу создателей COMPAS должен прогнозировать индивидуальный риск рецидивизма, что является важным обстоятельством при принятии судьями решений. Однако решение суда породило дебаты о точности работы алгоритма [Dressel, Farid, 2018], а также о том, не воспроизводит ли он расовые стереотипы, почерпнутые из обучения на решениях, принятых людьми. Как бы там ни было, без этой системы судьи все равно должны оценивать риски того, что обвиняемый совершит преступление повторно, и этот процесс не свободен от различных предубеждений [Brennan, Dieterich, 2018].

В целом идея о возможности предсказать значительную долю судебных приговоров должна поменять представление о том, что такое дискреция судьи [Tahura, Selvadurai, 2023]. Не все юридические решения имеют единственное решение. Если алгоритм ИИ будет способен предсказывать решения судей очень точно, это означает, что он должен воспроизводить не только формальный анализ легальных и экстралегальных факторов, но и то, что сами судьи назвали «психологией судебного разбирательства»¹¹. Без развития общего ИИ, о котором не шла речь в данной статье, алгоритм вряд ли будет способен понимать нюансы языка, игру слов, ценности и смыслы, воспринимаемые людьми в различных обстоятельствах дела (но даже «слабый» ИИ со временем вполне может научиться имитировать это понимание, если обучить его на достаточном объеме подобных кейсов). Отсюда же возникает важное опасение, связанное с тем, не воспроизводит ли ИИ в своем анализе стереотипы, идеологию и другие убеждения судей [Manresa-Yee, Ramis, 2022].

В этом контексте возникает еще более интересный вопрос: должен ли ИИ полностью прогнозировать решения судьи-человека или же он может встать в этом отношении на ступеньку выше? Исследования показывают [Doerner, 2015; Franklin, Fearn, 2008], что судьи в США не свободны от расовых и гендерных стереотипов. Более того, в условиях ограниченной рациональности [Albonetti, 1991], связанной с нехваткой информации и времени, именно эти стереотипы нередко становятся опорными точками [Steffensmeier, Ulmer, Kramer, 1998], помогающими

¹¹ В совете судей отвергли идею замены судьи на ИИ // Право.ru. 2023. 26 октября. URL: <https://pravo.ru/news/249529/> (дата обращения: 06.10.2024).

принимать непростые решения. Ведь ИИ мог бы следовать букве закона строже, чем человек (не говоря о более высокой скорости анализа информации). Человек не свободен от симпатий, ценностей и убеждений, психологического давления, в то время как ИИ следует этим факторам только в том случае, если его научили им следовать. Это непростой вопрос, на который пока нет ответа. Однако можно прогнозировать, что по меньшей мере население к такому повороту событий готово будет нескоро.

В исследовании [Barysè, Sarel, 2023] выдвигается важный тезис о том, что легитимность использования ИИ воспринимается населением и профессионалами по-разному в зависимости от стадии судебного процесса, на которой применяется технология. Люди в целом доверяют судьям-людям больше, чем ИИ, по этой причине они не поддерживают высокую автоматизацию работы судов на стадии принятия решений. Однако население выступает за использование технологий на стадии сбора данных, поскольку это может способствовать повышению объективности анализа, выполненного человеком.

Заключение: ИИ как ассистент судьи

Результаты исследования позволяют выдвинуть несколько практических рекомендаций о развитии ИИ-технологий для помощи юристам и судьям. Хотя ИИ пока не может заменить судью, он способен помочь судьям в их профессиональной деятельности.

Мы показали, что, несмотря на стремительное совершенствование алгоритмов, предсказывающих решения судей, большинство из них страдают от ограничений, связанных с данными. Например, алгоритмы, предназначенные для анализа решений высших судебных органов, имеют низкую внешнюю валидность. Большинство судебных процессов, касающихся граждан и бизнеса, происходят в судах нижнего уровня, прогнозировать которые ИИ пока не умеет. За исключением единственного набора данных из Китая [Xiao et al., 2018], другие публично доступные данные включают крайне ограниченное число кейсов, не превышающих несколько десятков тысяч дел. Впрочем, в России уже есть коммерческие проекты, умеющие предсказывать исходы арбитражных дел, однако их алгоритмы непрозрачны, поскольку не публиковались в рецензируемых академических изданиях.

В России в соответствии с законом от 22 декабря 2008 г. № 262-ФЗ «Об обеспечении доступа к информации о деятельности судов в Российской Федерации» публикуется огромное количество данных об уголовных, гражданских и арбитражных делах, что создает большой потенциал для обучения на них ИИ. Это означает, что, хотя на сегодняшний день не созданы ИИ модели, которые бы предсказывали решения российских судов, в этой сфере в обозримом будущем могут быть сделаны существенные прорывы. Исследования об использовании ИИ для классификации юридических текстов во всем мире все еще находятся в начале своего развития.

Другой вопрос состоит в том, нужно ли вообще внедрять ИИ-технологии в судебный процесс. Ответить на него однозначно, конечно, нельзя, и дебаты по этому поводу свидетельствуют о сложности этических вопросов. Однако сложность этических вопросов не отменяет важности развития самой технологии. Одна из пер-

спективных возможностей — это превращение ИИ в ассистента, который может помогать судьям в работе с материалами дела, включая их классификацию, поиск проблемных мест, выделение наиболее значимых обстоятельств дела. Публичная позиция одного из судей Верховного суда РФ¹² в целом поддерживает именно этот сценарий.

Важно учитывать, что большинство существующих алгоритмов предсказания судебных решений действуют по принципу «черного ящика», и это сильно ограничивает потенциал для их внедрения в юриспруденцию. В сферы с высокими ставками нельзя бездумно интегрировать алгоритмы, ход работы которых мы не понимаем. Но возможно, что дилемма между точностью решений и прозрачностью процесса их принятия является ложной [Rudin, 2019]. В таком случае в судебную практику должны внедряться именно те решения, которые позволяют не только предсказать решения, но и объяснить их. Именно последняя задача и должна стоять на повестке отечественных исследований в области использования ИИ-технологий в судебном процессе.

Список литературы (References)

1. Захарцев С. И., Сальников В. П. Судья-робот в уголовном процессе: хорошо или плохо? // Юридическая наука; история и современность. 2018. № 7. С. 176—180.
Zakharcev S. I., Salnikov V. P. (2018) The Robot Judge in Criminal Proceedings: Good or Bad? *Legal Science: History and Modernity*. No. 7. P. 176—180. (In Russ.)
2. Коваленко К. Е., Печатнова Ю. В., Стаценко Д. А., Коваленко Н. Е. Судья-робот как преодоление противоречий судебного усмотрения (юридический аспект) // Юридический вестник Дагестанского Государственного Университета. 2020. Т. 36. № 4. С. 169—173.
Kovalenko K. E., Pechatnova Yu. V., Statsenko D. A., Kovalenko N. E. (2020) The Robot Judge as a Resolution of Judicial Discretion Contradictions (Legal Aspect). *Legal Bulletin of Dagestan State University*. Vol. 36. No. 4. P. 169—173. (In Russ.)
3. Кравчук Н. В. Искусственный интеллект как судья: перспективы и опасения (Обзор) // Социальные и гуманитарные науки. Отечественная и зарубежная литература. Серия 4: Государство и право. 2021. № 1. С. 115—122.
Kravchuk N. V. (2021) Artificial Intelligence as a Judge: Prospects and Concerns (Review). *Social and Human Sciences. Domestic and Foreign Literature. Series 4: State and Law*. No. 1. P. 115—122. (In Russ.)
4. Степанов О. А., Басангов Д. А. О перспективах влияния искусственного интеллекта на судопроизводство // Вестник Томского Государственного Университета. Философия. Социология. Политология. 2022. № 475. С. 229—237.
Stepanov O. A., Basangov D. A. (2022) On the Prospects of Artificial Intelligence Impact on Judiciary. *Bulletin of Tomsk State University. Philosophy, Sociology, Political Science*. No. 475. P. 229—237. (In Russ.)

¹² Бурнов В. ВС: ИИ может помочь снизить число ошибок при назначении наказания // РАПСИ. 2024. 1 июля. URL: https://www.rapsinews.ru/judicial_mm/20240701/310059533.html (дата обращения: 05.07.2024).

5. Alali M., Syed S., Alsayed M., Patel S., Bodala H. (2021) JUSTICE: A Benchmark Dataset for Supreme Court's Judgment Prediction. *arXiv*. Art. 2112.03414. <https://doi.org/10.48550/arXiv.2112.03414>.
6. Alarie B., Niblett A., Yoon A. (2017) Using Machine Learning to Predict Outcomes in Tax Law. SSRN. <https://doi.org/10.2139/ssrn.2855977>.
7. Albonetti C. A. (1991) An Integration of Theories to Explain Judicial Discretion. *Social Problems*. Vol. 38. No. 2. P. 247—266. <https://doi.org/10.2307/800532>.
8. Alcántara Francia O. A., Nunez-del-Prado M., Alatrística-Salas H. (2022) Survey of Text Mining Techniques Applied to Judicial Decisions Prediction. *Applied Sciences*. Vol. 12. No. 20. Art. 20. <https://doi.org/10.3390/app122010200>.
9. Aletras N., Tsarapatsanis D., Preotiuc-Pietro D., Lampos V. (2016) Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective. *PeerJ Computer Science*. Vol. 2. No. 10. Art. e93. <https://doi.org/10.7717/peerj-cs.93>.
10. Alghazzawi D., Bamasag O., Albeshri A., Sana I., Ullah H., Asghar M. Z. (2022) Efficient Prediction of Court Judgments Using an LSTM+CNN Neural Network Model with an Optimal Feature Set. *Mathematics*. Vol. 10. No. 5. Art. 5. <https://doi.org/10.3390/math10050683>.
11. Ashley K. D., Brüninghaus S. (2009) Automatically Classifying Case Texts and Predicting Outcomes. *Artificial Intelligence and Law*. Vol. 17. No. 2. P. 125—165. <https://doi.org/10.1007/s10506-009-9077-9>.
12. Bagherian-Marandi N., Ravanshadnia M., Akbarzadeh-T M.-R. (2021) Two-Layered Fuzzy Logic-Based Model for Predicting Court Decisions in Construction Contract Disputes. *Artificial Intelligence and Law*. Vol. 29. No. 4. P. 453—484. <https://doi.org/10.1007/s10506-021-09281-9>.
13. Barysé D., Sarel R. (2023) Algorithms in the Court: Does It Matter Which Part of the Judicial Decision-Making is Automated? *Artificial Intelligence and Law*. Vol. 32. P. 117—146. <https://doi.org/10.1007/s10506-022-09343-6>.
14. Beriain I. D. M. (2018) Does the Use of Risk Assessments in Sentences Respect the Right to Due Process? A Critical Analysis of the *Wisconsin v. Loomis Ruling*. *Law, Probability & Risk*. Vol. 17. No. 1. P. 45—53. <https://doi.org/10.1093/lpr/mgy001>.
15. Brennan, T., Dieterich W. (2018) Correctional Offender Management Profiles for Alternative Sanctions (COMPAS). In: Singh J. P., Kroner D. G., Wormith J. S., Desmarais S. L., Hamilton Z. (eds.) *Handbook of Recidivism Risk/Needs Assessment Tools*. Hoboken: John Wiley & Sons. P. 49—75. <https://doi.org/10.1002/9781119184256.ch3>.
16. Chalkidis I., Androutsopoulos I., Aletras N. (2019) Neural Legal Judgment Prediction in English. *arXiv*. Art. 1906.02059. <https://doi.org/10.48550/arXiv.1906.02059>.

17. Chalkidis I., Fergadiotis M., Malakasiotis P., Aletras N., Androutopoulos I., Androutopoulos I. (2020) LEGAL-BERT: The Muppets Straight Out of Law School. *arXiv*. Art. 2010.02559. <https://arxiv.org/abs/2010.02559>.
18. Chicco D., Jurman G. (2020) The Advantages of the Matthews Correlation Coefficient (MCC) over F1 Score and Accuracy in Binary Classification Evaluation. *BMC Genomics*. Vol. 21. No. 1. Art. 6. <https://doi.org/10.1186/s12864-019-6413-7>.
19. Devlin J., Chang M.-W., Lee K., Toutanova K. (2019) BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. In: Burstein J., Doran C., Solorio T. (eds.) *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Vol. 1. Minneapolis, Association for Computational Linguistics. P. 4171—4186. <https://doi.org/10.18653/v1/N19-1423>.
20. Doerner J. K. (2015) The Joint Effects of Gender and Race/Ethnicity on Sentencing Outcomes in Federal Courts. *Women & Criminal Justice*. Vol. 25. No. 5. P. 313—338. <https://doi.org/10.1080/08974454.2014.989298>.
21. Franklin C. A., Fearn N. E. (2008) Gender, Race, and Formal Court Decision-Making Outcomes: Chivalry/Paternalism, Conflict Theory or Gender Conflict? *Journal of Criminal Justice*. Vol. 36. No. 3. P. 279—290. <https://doi.org/10.1016/j.jcrimjus.2008.04.009>.
22. Freitas A. L., Allende-Cid H., Santana O., Oliveira-Lage L. (2022) Predicting Brazilian Court Decisions. *PeerJ Computer Science*. Vol. 8. Art. e904. <https://doi.org/10.7717/peerj-cs.904>.
23. Gerber M. S. (2014) Predicting Crime Using Twitter and Kernel Density Estimation. *Decision Support Systems*. Vol. 61. P. 115—125. <https://doi.org/10.1016/j.dss.2014.02.003>.
24. Jacob de Menezes-Neto E., Clementino M. B. M. (2022) Using Deep Learning to Predict Outcomes of Legal Appeals Better than Human Experts: A Study with Data from Brazilian Federal Courts. *PLoS ONE*. Vol. 17. No. 7. Art. e0272287. <https://doi.org/10.1371/journal.pone.0272287>.
25. Dressel J., Farid H. (2018) The Accuracy, Fairness, and Limits of Predicting Recidivism. *Science Advances*. Vol. 4. No. 1. Art. eaao5580. <https://doi.org/10.1126/sciadv.aao5580>.
26. Cui J., Shen X., Wen S. (2023) A Survey on Legal Judgment Prediction: Datasets, Metrics, Models and Challenges. *IEEE Access*. Vol. 11. P. 102050—102071. <https://doi.org/10.1109/access.2023.3317083>.
27. Katz D. M., Bommarito M. J., Blackman J. (2017) A General Approach for Predicting the Behavior of the Supreme Court of the United States. *PLoS ONE*. Vol. 12. No. 4. Art. e0174698. <https://doi.org/10.1371/journal.pone.0174698>.

28. Kaufman A. R., Kraft P., Sen M. (2019) Improving Supreme Court Forecasting Using Boosted Decision Trees. *Political Analysis*. Vol. 27. No. 3. P. 381—387. <https://doi.org/10.1017/pan.2018.59>.
29. Kaur H., Choudhury T., Singh T. P., Shamoan Mohammad. (2019) Crime Analysis using Text Mining. In: Ming Fong A. Ch., Hong G. Y., Fong B. (eds.) *2019 International Conference on Contemporary Computing and Informatics (IC3I)*. Singapore. P. 283—288. <https://doi.org/10.1109/IC3I46837.2019.9055606>.
30. Kort F. (1957) Predicting Supreme Court Decisions Mathematically: A Quantitative Analysis of the “Right to Counsel” Cases. *American Political Science Review*. Vol. 51. No. 1. P. 1—12. <https://doi.org/10.2307/1951767>.
31. Kowsrihawatt K., Vateekul P., Boonkwan P. (2018) Predicting Judicial Decisions of Criminal Cases from Thai Supreme Court Using Bi-directional GRU with Attention Mechanism. In: Do Van T., Do Duc H., Nguyen G. (eds.) *5th Asian Conference on Defense Technology (ACDT)*. Hanoi, Vietnam. P. 50—55. <https://doi.org/10.1109/acdt.2018.8592948>.
32. Malik V., Sanjay R., Nigam S. K., Ghosh K., Guha S. K., Bhattacharya A., Bhattacharya A., Modi A. (2021) ILDC for CJPE: Indian Legal Documents Corpus for Court Judgment Prediction and Explanation. In: Zong Ch., Xia F., Li W., Navigli R. (eds.) *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Confer.* Stroudsburg: Association for Computational Linguistics. P. 4046—4062. <https://doi.org/10.18653/v1/2021.acl-long.313>.
33. Manresa-Yee C., Ramis S. (2022) Assessing Gender Bias in Predictive Algorithms using eXplainable AI *arXiv*. Art. 2203.10264. <https://doi.org/10.48550/arXiv.2203.10264>.
34. Medvedeva M., Üstun A., Xu X., Vols M., Wieling M. (2021) Automatic Judgement Forecasting for Pending Applications of the European Court of Human Rights. In: Ashley K. D., Atkinson K., Branting K., Francesconi E., Grabmair M., Walker V. R., Waltl B., Wyner A. Z. (eds.) *Proceedings of the fifth workshop on automated semantic analysis of information in legal text (ASAIL 2021), São Paulo, Brazil*. CEUR Workshop Proceedings. P. 1—12. <https://ceur-ws.org/Vol-2888/paper2.pdf>.
35. Medvedeva M., Vols M., Wieling M. (2020) Using Machine Learning to Predict Decisions of the European Court of Human Rights. *Artificial Intelligence and Law*. Vol. 28. No. 2. P. 237—266. <https://doi.org/10.1007/s10506-019-09255-y>.
36. Medvedeva M., Wieling M., Vols M. (2022) Rethinking the Field of Automatic Prediction of Court Decisions. *Artificial Intelligence and Law*. Vol. 31. P. 195—212. <https://doi.org/10.1007/s10506-021-09306-3>.
37. Medvedeva M., Xiao X., Wieling M., Vols M. (2020) JURI SAYS: An Automatic Judgement Prediction System for the European Court of Human Rights. *Frontiers in Artificial Intelligence and Applications*. Vol. 334. P. 277—280. <https://doi.org/10.3233/faia200883>.

38. O'Sullivan C., Beel J. (2019) Predicting the Outcome of Judicial Decisions made by the European Court of Human Rights. *arXiv*. Art. 1912.10819. <https://doi.org/10.48550/arXiv.1912.10819>.
39. Quemy A., Wrembel R. (2020) On Integrating and Classifying Legal Text Documents. In: Hartmann S., Küng J., Kotsis G., Tjoa A. M., Khalil I. (eds.) *Database and Expert Systems Applications*. Cham, Switzerland: Springer International Publishing. P. 385—399. https://doi.org/10.1007/978-3-030-59003-1_25.
40. Rudin C. (2019) Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. *Nature Machine Intelligence*. Vol. 1. No. 5. Art. 5. <https://doi.org/10.1038/s42256-019-0048-x>.
41. Ruger T. W., Kim P. T., Martin A., Martin A. D., Quinn K. M. (2004) The Supreme Court Forecasting Project: Legal and Political Science Approaches to Predicting Supreme Court Decisionmaking. *Columbia Law Review*. Vol. 104. No. 4. P. 1150—1209. <https://doi.org/10.2307/4099370>.
42. Rusakova E. P. (2021) Integration of “Smart” Technologies in the Civil Proceedings of the People’s Republic of China. *RUDN Journal of Law*. Vol. 25. No. 3. P. 622—633. <https://doi.org/10.22363/2313-2337-2021-25-3-622-633>.
43. Sebők M., Kiss R., Járay I. (2023) Introducing HUNCOURT: A New Open Legal Database Covering the Decisions of the Hungarian Constitutional Court for Between 1990 and 2021. *Journal of the Knowledge Economy*. Vol. 15. P. 6507—6540. <https://doi.org/10.1007/s13132-023-01395-6>.
44. Sert M. F., Yıldırım E., Haşlak İ. (2022) Using Artificial Intelligence to Predict Decisions of the Turkish Constitutional Court. *Social Science Computer Review*. Vol. 40. No. 6. P. 1416—1435. <https://doi.org/10.1177/08944393211010398>.
45. Shaikh R. A., Sahu T. P., Anand V. (2020) Predicting Outcomes of Legal Cases based on Legal Factors using Classifiers. *Procedia Computer Science*. Vol. 167. P. 2393—2402. <https://doi.org/10.1016/j.procs.2020.03.292>.
46. Sharma R. D., Mittal S., Tripathi S., Acharya S. (2015) Using Modern Neural Networks to Predict the Decisions of Supreme Court of the United States with State-of-the-Art Accuracy. In: Arik, S., Huang, T., Lai, W., Liu, Q. (eds.) *Neural Information Processing. ICONIP 2015. Lecture Notes in Computer Science*. Vol. 9490. Cham: Springer. P. 475—483. https://doi.org/10.1007/978-3-319-26535-3_54.
47. Shi C., Sourdin T., Li B. (2021) The Smart Court — A New Pathway to Justice in China? *International Journal of Court Administration*. Vol. 12. No. 1. Art. 1. <https://doi.org/10.36745/ijca.367>.
48. Song D., Gao S., He B., Schilder F. (2022) On the Effectiveness of Pre-Trained Language Models for Legal Natural Language Processing: An Empirical Study. *IEEE Access*. Vol. 10. P. 75835—75858. <https://doi.org/10.1109/ACCESS.2022.3190408>.

49. Sourdin T., Cornes R. (2018) Do Judges Need to Be Human? The Implications of Technology for Responsive Judging. In: Sourdin T., Zariski A. (eds.) *The Responsive Judge: International Perspectives*. Singapore: Springer. P. 87—119. https://doi.org/10.1007/978-981-13-1023-2_4.
50. Steffensmeier D., Ulmer J., Kramer J. (1998) The Interaction of Race, Gender, and Age in Criminal Sentencing: The Punishment Cost of Being Young, Black, and Male. *Criminology*. Vol. 36. No. 4. P. 763—798. <https://doi.org/10.1111/j.1745-9125.1998.tb01265.x>.
51. Tahura U. S., Selvadurai N. (2023) The Use of Artificial Intelligence in Judicial Decisionmaking: The Example of China. *International Journal of Law, Ethics, and Technology*. Vol. 3. P. 1—20. <https://doi.org/10.55574/pyeb5374>.
52. Tan H., Zhang B., Zhang H., Li R. (2020) The Sentencing-Element-Aware Model for Explainable Term-of-Penalty Prediction. In: Zhu X., Zhang M., Hong Y., He R. (eds.) *Natural Language Processing and Chinese Computing. NLPCC2020. Lecture Notes in Computer Science*. Vol. 12431. Cham: Springer. P. 16—27. https://doi.org/10.1007/978-3-030-60457-8_2.
53. Taylor I. (2023) Justice by Algorithm: The Limits of AI in Criminal Sentencing. *Criminal Justice Ethics*. Vol. 42. No. 3. P. 193—213. <https://doi.org/10.1080/0731129X.2023.2275967>.
54. Virtucio M. B. L., Aborot J. A., Abonita J. K. C., Avinante R. S., Copino R. J. B., Neverida M. P., Osiana V. O., Peramo E. C., Syjuco J. G., Tan G. B. A. (2018) Predicting Decisions of the Philippine Supreme Court Using Natural Language Processing and Machine Learning. In: Honiden S., Fujii R. (eds.) *2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)*. Tokyo. P. 130—135. <https://doi.org/10.1109/compsac.2018.10348>.
55. Volkov V. (2016) Legal and Extralegal Origins of Sentencing Disparities: Evidence from Russia's Criminal Courts. *Journal of Empirical Legal Studies*. Vol. 13. No. 4. P. 637—665. <https://doi.org/10.1111/jels.12128>.
56. Waltl B., Bonczek G., Scepankova E., Landthaler J., Matthes F. (2017) Predicting the Outcome of Appeal Decisions in Germany's Tax Law. Electronic Participation. In: Parycek P. et al. (eds.) *Electronic Participation. ePart 2017. Lecture Notes in Computer Science*. Vol. 10429. Cham: Springer. P. 89—99. https://doi.org/10.1007/978-3-319-64322-9_8.
57. Xiao C., Zhong H., Guo Z., Tu C., Liu Z., Sun M., Feng Y., Han X., Hu Z., Wang H., Xu J. (2018) CAIL2018: A Large-Scale Legal Dataset for Judgment Prediction. *arXiv*. Art. 1807.02478. <https://arxiv.org/abs/1807.02478>.
58. Xu Z. (2022) Human Judges in the Era of Artificial Intelligence: Challenges and Opportunities. *Applied Artificial Intelligence*. Vol. 36. No. 1. Art. 2013652. <https://doi.org/10.1080/08839514.2021.2013652>.

59. Zhong H., Xiao C., Tu C., Tianyang Zhang, Zhiyuan L., Liu Z., Liu Z., Sun M. (2020) How Does NLP Benefit Legal System: A Summary of Legal Artificial Intelligence. In: Jurafsky D., Chai J., Schluter N., Tetreault J. (eds.) *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Stroudsburg. Association for Computational Linguistics. P. 5218—5230. <https://aclanthology.org/2020.acl-main.466/>.
60. Zhuchkova S., Kazun A. (2023) Exploring Gender Bias in Homicide Sentencing: An Empirical Study of Russian Court Decisions Using Text Mining. *Homicide Studies*. <https://doi.org/10.1177/10887679231217159>.