

УДК 004+81'33

Е.А. ПАХОМОВ, Н.Р. ТИТАРЕНКО

(*pahomow.egor2014@yandex.ru, nikita.titarenko2014@yandex.ru*)

Национального государственного исследовательского Нижегородского университета им. Н.И. Лобачевского

ИСПОЛЬЗОВАНИЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ПРОИЗВОДСТВЕ ЛИНГВИСТИЧЕСКИХ ЭКСПЕРТИЗ*

Данная статья посвящена изучению проблемы использования искусственного интеллекта в производстве лингвистических экспертиз при помощи исследования экспертных систем. Производится изучение понятия искусственного интеллекта и экспертных систем, сравнение того, как их изучали отечественные и зарубежные ученые. Особое внимание уделяется таким системам, как “Galaktika-ZOOM”, “TextAnalyst 2.0”, “MonoConc / ParaConc”, «НКРЯ», и возможности их применения в экспертной деятельности, анализируются возможности этих систем и их характеристика

Ключевые слова: *искусственный интеллект, судебно-экспертная деятельность, языкознание, лингвистика, экспертные системы.*

Актуальность данной темы обусловлена тем, что на сегодняшний день искусственный интеллект уже внедрен и используется во многих сферах человеческой деятельности. В том числе и в сфере судебной экспертизы. Так, ИИ помогает анализировать большие объемы данных, содержать полезные базы данных и в конечном итоге помогать эксперту с поиском необходимой ему информации.

В сфере лингвистических экспертиз на данный момент не существует ни одной экспертной системы, которую можно было бы использовать постоянно при проведении лингвистических экспертиз, эксперты ограничены небольшим перечнем систем, которые могут помочь найти нужную информацию в тексте. Разработки в области создания единой системы для проведения конкретных лингвистических экспертиз, как, например, система «ПАПИЛОН» [16], которая используется для проведения дактилоскопических экспертиз, не ведется.

Цель данной работы заключается в исследовании ряда экспертных систем и возможности использования данных систем в производстве лингвистических экспертиз.

Идея создания экспертной системы на основе искусственного интеллекта возникла в 1970-х гг., когда исследователи занялись построением базы знаний, которая по их задумке должна была обладать элементами искусственного интеллекта.

Одной из главных характеристик XXI в. является быстрое накопление разной информации и, как следствие, её огромные объемы. Для того, чтобы накопленные массивы данных можно было использовать необходима обработка информации и её интерпретация, которая требует колоссальных ресурсов. В связи с тем, что человек уже в полной мере не может самостоятельно справляться с текущими потоками данных, требуется включение в его деятельность новых технологий.

Судебная экспертиза, как экспертный институт, в ходе своего развития постоянно использовала и адаптировала под свои нужды передовые и эффективные для своего времени методы и средства исследования, например, методы химии и физики, фотографии, а в дальнейшем и компьютерные технологий, которые развились в такие роды экспертиз, как баллистические, ботанические и многие другие [15].

Об использовании компьютерных технологий в судебной экспертизе писали такие ученые, как Р.М. Ланцман [9, с. 95], А.И. Винберг [5, с. 75–77] Л.Е. Ароцкер [1, с. 182–190]. Ученые в этих работах анализировали как компьютеры могут применяться в почерковедческой экспертизе (Р.М. Ланцман), или при даче экспертом вывода при условии неполного знания им исследуемого явления (А.И. Винберг), или при исследовании вещественных доказательств (Л.Е. Ароцкер). Д.В. Бахтеев, Л.В. Тарасова [3, с. 249–254]. проанализировали использование ИИ в арбитражных судах как дальнейшей ступенью автоматизации делопроизводства. Одной из последних работ об искусственном интеллекте в систе-

* Работа выполнена под руководством Новоселова А.Н., кандидата филологических наук, доцента кафедры судебной экспертизы Нижегородского государственного университета им. Н.И. Лобачевского.

ме расследование преступлений является статья А.Ю. Афанасьева [2, с. 34], в которой он рассуждал о возможности использования ИИ в расследовании преступлений. Исследователь пришел к выводу, что на современном этапе развития ИИ вместо того, чтобы заменять субъекты раскрытия преступлений системой ИИ, эффективнее будет наладить их взаимодействие, отдав им решение стандартных задач.

Зарубежные ученые вопрос использования искусственного интеллекта в судебной экспертизе также рассматривали. Так, Мэтью Роберт Беннет и Марсин Будка 18 августа 2021 г. [19], наглядно продемонстрировали как искусственный интеллект, может быть применен в трасологической экспертизе, и возможно ли заменить эксперта искусственным интеллектом. Они решили, что современные нейросети не обладают достаточным для эксперта-трасолога восприятием, нейросеть могла определять только новые подошвы, поношенные у нее определить не получилось, т. к. на них было слишком много разных признаков. Камшад Мошин, анализировал возможность использования ИИ в расследовании преступлений, распознавании следов, предоставлений им заключений эксперта [18].

Анализируя работы отечественных и зарубежных ученых, мы пришли к выводу, что исследование возможностей ИИ пошло разными путями. Так, отечественные ученые исследовали возможности ЭВМ в судебной экспертизе, а затем перешли к исследованию возможностей ИИ в судебном процессе. Тогда как зарубежные ученые не отходили от судебной экспертизы и продолжали анализировать ИИ в рамках производства конкретных видов экспертиз. Это обусловлено, на наш взгляд, тем, что зарубежная наука начала раньше изучать возможности ИИ, что привело к дифференциации предметных областей и создало специальные области исследований, конкретно под судебную экспертизу. В то же время отечественная наука только начала исследовать ИИ применительно к правовой системе. Однако, несмотря на различные подходы к проблеме ИИ в расследовании преступлений, все ученые согласны в том, что на данном этапе развития искусственный интеллект, может только дополнять человека, решая типовые задачи.

Так, например, в ходе производства лингвистических экспертиз активно применяется компьютерная лингвистика, где при установлении автора конкретного текста, с помощью марковской матрицы человека, которая отслеживает особенности лексикона конкретного человека, характерные орфографические или пунктуационные ошибки и затем при сопоставлении получившейся матрицы с эталонной делается вывод, является ли автором исследуемого текста подозреваемый в этом человек [4, с. 203].

С помощью математических методов и компьютерных алгоритмов было создано множество баз данных. Одной из которых является экспертная система “OLDMEN” [10], применяемая для установления возраста исполнителя подписей, выполненных от имени лиц пожилого или старческого возраста.

Таким образом, использование компьютерных технологий и внедрение математических методов обработки информации дали сделать очередной шаг в развитии форм использования специальных знаний.

Дальнейшей перспективой развития судебной экспертизы мы видим использование в производстве конкретных видов экспертиз технологий искусственного интеллекта.

Российские ученые, например, Т.А. Гаврилова, В.Ф. Хорошевский, под искусственным интеллектом понимают «одно из направлений информатики, целью которого является разработка аппаратно-программных средств, позволяющих пользователю-непрограммисту ставить и решаться свои, традиционно считающиеся интеллектуальными задачи, общаясь с ЭВМ на ограниченном подмножестве естественного языка» [6, с. 15].

Н.Н. Новичков выделил определения зарубежных ученых, под искусственным интеллектом они понимают – свойство автоматических систем брать на себя отдельные функции интеллекта человека, например, выбирать и принимать оптимальные решения на основе ранее полученного опыта и рационального анализа внешних воздействий [13, с. 9].

Сравнивая эти определения, мы пришли к выводу, что определение, данное Н.Н. Новичковым, более общее, чем определение, данное Т.А. Гавриловой, В.Ф. Хорошевским, в котором представлено больше конкретики. Так, в их определении сказано, что ИИ представляют из себя аппаратно-

программные средства, позволяющих пользователю-непрограммисту ставить и решать свои интеллектуальные задачи, общаясь с электронно-вычислительными машинами на естественном языке.

В 2001 г. П. Джексон дал следующее определение. «Экспертная система – это программа для компьютера, которая оперирует со знаниями в определенной предметной области с целью выработки рекомендаций или решения проблем, решающая следующие задачи» [7, с. 622]:

- извлечение информации из первичных данных (таких как сигналы, поступающие от гидролокатора);
- диагностика неисправностей (как в технических системах, так и в человеческом организме);
- структурный анализ сложных объектов (например, химических соединений);
- выбор конфигурации сложных многокомпонентных систем (например, распределенных компьютерных систем);
- планирование последовательности выполнения операций, приводящих к заданной цели (например, выполняемых промышленными роботами) [Там же].

В своем определении П. Джексон, анализирует многозадачность экспертной системы, указывая на широкий спектр их применения в деятельности человека.

В.К. Финн в 1986 г. дает следующее определение. Экспертные системы – компьютерные системы, принимающие решения, имитирующие решения задач квалифицированными экспертами. Фактически экспертные системы являются интеллектуальными системами первого поколения с Решателями задач, реализующими правила вывода в виде продукций – высказываний «если X, то Y». Экспертные системы применяются для решения задач медицинской диагностики, планирования, контроля, проектирования и некоторых др. задач [17, с. 350].

Сравнив определения, мы увидели следующие отличия в понятии экспертной системы отечественным и зарубежным учеными, В.К. Финн, считает, что экспертная система, принимает решения, имитируя квалифицированного эксперта, тогда как П. Джексон, деятельность экспертной системы основана не на имитации, а на использовании знаний предметной области, при этом указывая на то, что экспертная система самостоятельно решений не принимает, только дает рекомендации к решению проблем, благодаря решению перечисленных им задач.

Одной из отличительных особенностей экспертных систем искусственного интеллекта является использование эвристик, которые симулируют механизм мышления человека эксперта. В результате чего, цепочка рассуждений оказывается уникальной для каждой ситуации в ходе решения задачи.

Положительным моментом использования искусственного интеллекта в экспертных системах и экспертизе, будет сохранение большого объема времени, т. к. будет ускорен процесс поиска информации.

Однако, в реалиях отсутствия единой системы для производства лингвистических экспертиз, необходимо обучать экспертов пользованию рядом систем, функционал которых может разительно отличаться. Например, в одной из систем есть функция поиска предложений по контексту, а в другой системе эта функция ограничена поиском определенных слов, использованных в контексте. В качестве другого примера можно представить, что одна из систем является базой данных, поиск, в которой ограничивается уже доступной информацией, а другая система содержит в себе базу данных и может проводить поиск по конкретным критериям в исследуемом тексте. Исходя из приведенных примеров, мы понимаем, что обучать эксперта будет не просто, т. к. в каждой системе присутствует свой функционал и особенности ее использования.

Далее представлены примеры различных систем, которые используются экспертами в определенных ситуациях.

На наш взгляд, данные ниже системы лучше всего подходят для работы экспертов-лингвистов, т. к. эти системы являются наиболее разработанными, обновляемыми. Эти системы обладают наиболее удобным и большим функционалом, для поиска по тексту и обработке представленной информации.

Galaktika-ZOOM – Система позволяет выявлять значимые слова и словосочетания документа, проводить поиск документов по вводимым пользователем ключевым словам с учетом их синонимов, а также формировать отчеты по частоте встречаемости слов в текстах. Программа обеспечивает обработку русскоязычных текстов. Алгоритмы основаны на использовании морфологического и статистического анализа [8].

Данная система позволяет экспертам быстро анализировать текст на наличие призывов к экстремизму, самоубийствам или иным противозаконным деяниям, по средствам поиска конкретных слов и словосочетаний в тексте, а также их синонимов.

TextAnalyst 2.0 – Данная система использует построения семантической сети понятий, выделяемых из обрабатываемого текста, со ссылками на контекст. Позволяет анализировать текст путем построения иерархического дерева тем/подтем, затрагиваемых в тексте. Также имеется возможность реферирования текста. Лингвистический анализ проводится в основном на основе стемминга. Морфологический анализ реализован для сравнительно небольшого количества слов. Из лингвистического анализа исключаются не только стоп-слова, но и все глаголы. При поиске не учитывается порядок слов. Синтаксический и семантический машинный анализ тестов не реализован [15].

Данная система позволяет находить ключевые слова, используемые в одном контексте. Это полезно при наличии в тексте скрытых под определенным контекстом с призывом к противозаконным действиям, данная система способна выделить из текста предложения, словосочетания, использованные в похожем контексте, хотя точность варьируется в зависимости от размера документа или текста.

MonoConc / ParaConc – Программы построения конкордансов (в том числе мультязычных) для одного или двух сравниваемых тестов. Реализованы возможности различных режимов поиска: с помощью символов маскирования, регулярных выражений, контекстно-чувствительного поиска [13].

Различные варианты сортировки. Функции частоты встречаемости слов и словосочетаний. Различные режимы вывода конкорданса. Есть возможность генерации индексов и словариков. Для построения параллельного конкорданса (для двух сравниваемых текстов) необходимо обеспечить их соответствие по структуре (параграфам, предложениям).

Позволяет сравнить несколько текстов и быстро определить:

Регулярные в данном тексте выражения и на основе их определить, был ли написан текст одним человеком или несколькими, на основе сравнений данных выражений. Также при помощи функций данной системы возможно сравнить два текста, предположительно написанных одним человеком, и помочь эксперту определить, так ли это на самом деле.

Национальный Корпус Русского Языка (НКРЯ) – это интернет-сервис, в котором собраны художественные и нехудожественные тексты: древнерусских посланий с современными инструкциями к бытовым приборам. При помощи данной системы можно отыскать конкретные словосочетания и их вариации, контексты. В отличие от интернет-поисковиков “Google” или “Yandex”, все слова в базе имеют лингвистическую разметку: род, падеж, лицо, залог и т. д. За счет этого поиск можно проводить по множеству характеристик [12]. НКРЯ в судебно-лингвистических экспертизах используется для работы с конкретными словосочетаниями и словами, в частности, данную систему возможно использовать для исследования текстов, содержащих призывы к терроризму или иным противозаконным деяниям, путем поиска конкретных слов, словосочетаний и проверки их контекста.

Данная система используется экспертами лингвистами чаще, чем другие, т. к. она является наиболее полной и удобной. В ней собрано большое количество возможностей по поиску конкретных словосочетаний по разным признакам

Все вышеперечисленные системы были выбраны, т. к. обладают рядом особенностей. Выделяя все эти особенности, можно отметить такие из них, как: нахождение ключевых слов в определенных контекстах в системе “TextAnalyst 2.0”, возможность сравнения двух текстов в программах “MonoConc” и “ParaConc”, что не может делать ни одна другая, общедоступная система. В дополнение необходимо

отметить систему «НКРЯ», которая содержит в своей базе огромное количество словосочетаний, предложений, слов, и имеет удобный инструментарий по работе с ними.

Мы пришли к выводу, что использование искусственного интеллекта и экспертных систем, построенных на нем, существенно облегчает работу эксперта при проведении лингвистических экспертиз. На сегодняшний день, в сфере лингвистических экспертиз нет не единой экспертной системы, не единого перечня экспертных систем, которые используются экспертами. Все экспертные системы, которыми пользуются при проведении лингвистических экспертиз, разработаны для других целей, тем не менее они помогают экспертам облегчить производство экспертиз, сэкономить время, увеличить производительность.

Мы считаем, что в будущем необходимо создать официальный перечень систем, которые помогают в производстве лингвистических экспертиз, а также проводить обучение экспертов по их использованию. Также необходимо привлекать ученых и разработчиков для создания экспертных систем, которые будут направлены на помощь в производстве конкретных лингвистических экспертиз, а также разработки методики их использования. В этом случае будет возможным использование официальных систем, которые будут помогать искать нужный текст по необходимым и настроенным критериям.

В области использования ИИ в судебной экспертизе в России еще многое не изучено по сравнению с зарубежными странами. При дальнейшем исследовании искусственного интеллекта и внедрении его в судебно-экспертную деятельность, будет возможно создание экспертных систем на основе искусственного интеллекта, направленных на производство лингвистических экспертиз. Такие системы позволят эксперту находить нужную информацию, проводить автоматическое сравнение текста, использовать встроенные в систему базы данных, ориентироваться по текстам с помощью ключевых слов.

Литература

1. Ароцкер Л.Е. Организационные и процессуальные вопросы использования электронно-вычислительных машин в экспертной практике // Криминалистика и судебная экспертиза. Киев: РИО МВД УССР, 1969. Вып. 6. С. 182–190.
2. Афанасьев А.Ю. Дискуссионная трибуна // Библиотека криминалиста. 2018. № 3(38). С. 28–34.
3. Бахтеев Д.В., Тарасова Л.В. Применение искусственного интеллекта в деятельности арбитражных судов РФ: перспективные направления и проблемы // Вестник Костром. гос. ун-та. 2020. Т. 26. № 4. С. 249–254.
4. Беллман Р. Введение в теорию матриц / под ред. В.Б. Лидского; пер.: В.Я. Катковник, Р.А. Полуэктов, М.С. Эпельман. М.: Книга по требованию, 1969.
5. Винберг А.И. Вывод эксперта при неполном знании изучаемого явления // Советское государство и право. 1975. № 6. С. 75–77.
6. Гаврилова Т.А., Хорошевский В.Ф. Базы знаний интеллектуальных систем. СПб.: Питер, 2001.
7. Джексон П. Введение в экспертные системы / пер. с англ. В.Т. Тертышного. 3-е изд. М. [и др.]: Вильямс, 2001.
8. Корпорация «Галактика»: [сайт]. URL: <https://galaktika.ru>.
9. Ланцман Р.М. Кибернетика и криминалистическая экспертиза почерка. М.: Наука, 1968.
10. Липовский В.В., Косолапов И.Л. Разработка экспертной системы “OLDMEN” // Криминалистика и судебная экспертиза: межведомственный науч.-методич. сб. Киев: РИО МООН УССР, 1997. Вып. 48. С. 96–101.
11. Национальный корпус русского языка: [сайт]. URL: <https://ruscorpora.ru>.
12. Новичков Н.Н. Англо-русский словарь по нанотехнологиям. М.: Информ. Агентство АРМС-ТАСС, 2010.
13. Официальная программа для поиска в тексте “MonoConc”: [сайт]. URL: <https://monoconc.com>.
14. Официальный сайт программного обеспечения для аналитического анализа текста: [сайт]. URL: <https://analyst.ru/inde>.
15. Приказ МВД России от 29.06.2005 № 511 (ред. от 30.05.2022) «Вопросы организации производства судебных экспертиз в экспертно-криминалистических подразделениях органов внутренних дел Российской Федерации». [Электронный ресурс]. URL: https://www.consultant.ru/document/cons_doc_LAW_55315/ (дата обращения: 03.05.2023).
16. Система «ПАПИЛОН»: [сайт]. URL: <https://www.papillon.ru>.
17. Финн В.К., Поспелова Д.А., Лахути Д.Г. [и др.]. Интеллектуальные системы и общество: сб. ст. 2-е изд., испр. и доп. М.: URSS: КомКнига, 2006.
18. Kamshad M. Artificial Intelligence in Forensic Science Maharishi International University “SSRN Electronic Journal”. [Электронный ресурс]. URL: https://www.researchgate.net/publication/356263846_Artificial_Intelligence_in_Forensic_Science (дата обращения: 02.05.2023).
19. Matthew R.B. We trained AI to recognise footprints, but it won't replace forensic experts yet // The Conversation. August 18, 2021. [Электронный ресурс]. URL: <https://theconversation.com/we-trained-ai-to-recognise-footprints-but-it-wont-replace-forensic-experts-yet-161686> (дата обращения: 02.05.2023).

EGOR PAKHOMOV, NIKITA TITARENKO

National Research Lobachevsky State University of Nizhny Novgorod

THE USE OF ARTIFICIAL INTELLIGENCE IN THE PRODUCTION OF LINGUISTIC EXPERTISES

This article is devoted to the study of the problem of using the artificial intelligence in the production of linguistic expertise with the help of the study of expert systems. The concept of artificial intelligence and expert systems is studied, comparing how they were studied by the domestic and foreign scientists. The special attention is paid to such systems as “Galaktika-ZOOM”, “TextAnalyst 2.0”, “MonoConc / ParaConc”, “National Corpus of the Russian Language”, and the possibility of their application in expert activities. The possibilities of these systems and their characteristics are analyzed.

Key words: *artificial intelligence, forensic activity, study of language, linguistics, expert systems.*