УДК 543+004.8

Томуева А.Ш.
доктор философии по химии
кафедра общей химии и технологии обучения химии
Азербайджанский государственный педагогический университет
(г. Баку, Азербайджан)

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В ОПРЕДЕЛЕНИИ СТРУКТУРЫ ХИМИЧЕСКИХ СОЕДИНЕНИЙ: СОВРЕМЕННЫЕ МЕТОДЫ И ПЕРСПЕКТИВЫ

Аннотация: *в статье систематизированы современные подходы к применению искусственного интеллекта при установлении структуры химических соединений. Рассмотрен переход от традиционной спектроскопической интерпретации и квантово-химических расчётов к машинному обучению, глубоким нейронным сетям, графовым моделям и Transformer-архитектурам. Особое внимание уделено анализу данных ядерного магнитного резонанса, масс-спектрометрии и инфракрасной спектроскопии. Показано, что современные модели способны решать как прямую задачу «структура-спектр», так и обратную задачу «спектр-структура». Обсуждается интеграция искусственного интеллекта с теорией функционала плотности (DFT). По опубликованным данным, для молекул до 19 тяжёлых атомов гибридная модель CNN-Transformer включала правильную структуру в первые 15 кандидатов в 69,6% случаев и сокращала пространство структурного поиска до 11 порядков величины [6]. Сделан вывод о перспективности комплексного подхода, объединяющего эксперимент, квантово-химическое моделирование и искусственный интеллект.*

Ключевые слова: *искусственный интеллект, химическое соединение, молекулярная структура, машинное обучение, глубокое обучение, нейронная сеть, ЯМР-спектроскопия, масс-спектрометрия, инфракрасная спектроскопия, теория функционала плотности, трансформер-архитектура, молекулярный граф, структурный анализ.*

Введение.

Установление структуры химического соединения является одной из фундаментальных задач современной химии. Определение состава молекулы, последовательности химических связей, пространственной конфигурации и особенностей электронной структуры необходимо для объяснения физико-химических свойств вещества, прогнозирования реакционной способности и разработки новых материалов, катализаторов и биологически активных соединений.

Традиционный структурный анализ основан на совместном использовании ядерного магнитного резонанса (NMR), масс-спектрометрии (MS), инфракрасной и Raman-спектроскопии, рентгеноструктурного анализа и других методов. Рост сложности исследуемых систем приводит к увеличению объёма экспериментальных данных и числа возможных структурных решений, что создаёт предпосылки для автоматизации интерпретации.

В предыдущем исследовании авторов показано, что DFT позволяет оптимизировать геометрию комплексных соединений, оценивать длины связей металл–лиганд, распределение электронной и спиновой плотности и относительные энергии различных структур [1]. Там же подчёркнута необходимость сопоставления вычислительных результатов с рентгеноструктурными, спектроскопическими и магнитными экспериментальными данными [1]. Данный принцип непосредственно переносится на ИИ-ориентированный структурный анализ: интеллектуальная модель должна ускорять поиск, но окончательное заключение требует физико-химической проверки.

Цифровизация науки и промышленности создаёт общую методологическую основу для внедрения интеллектуальных методов обработки информации. В работах С.Р. Нуриевой цифровые технологии рассматриваются как важное направление современного научно-технологического развития [2], а взаимосвязь физических параметров с технологическими и экономическими показателями анализируется в контексте энергоэффективности промышленных предприятий [3]. Эти исследования формируют междисциплинарный контекст применения вычислительных методов к химическим данным. В смежных исследованиях цифровая трансформация также рассматривается как один из факторов изменения условий функционирования промышленных предприятий [10].

Цель работы заключается в системном анализе современных методов применения искусственного интеллекта для определения структуры химических соединений, оценке количественных результатов ИИ-моделей и выявлении перспектив интеграции ИИ с экспериментальными и квантово-химическими методами.

1. Эволюция методов определения структуры химических соединений.

Современная структурная химия развивается по пути последовательного усложнения способов интерпретации данных: от экспертного анализа спектров к квантово-химическому моделированию и далее к интеллектуальным системам, способным выявлять многомерные зависимости в больших массивах экспериментальной информации.

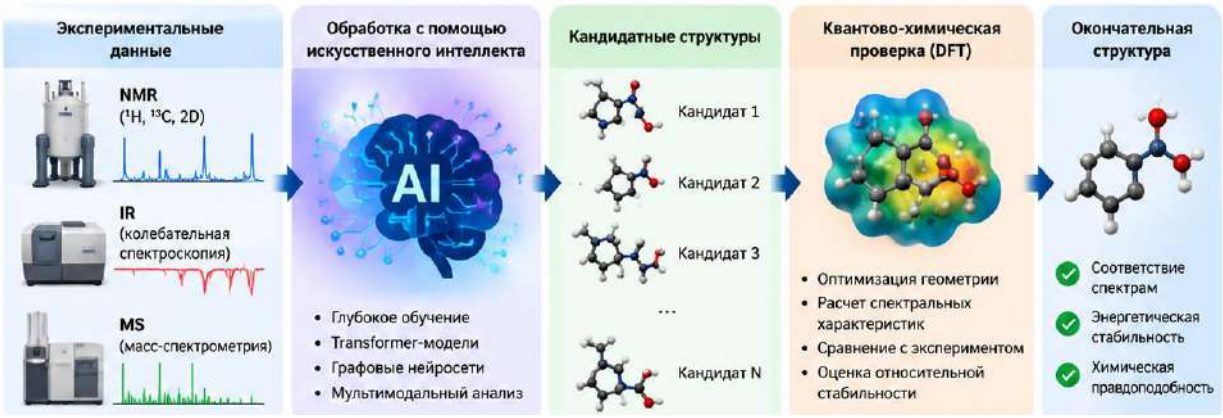


Рисунок 1. Общая схема ИИ-ориентированного определения структуры химических соединений.

В представленной схеме искусственный интеллект занимает промежуточный уровень между экспериментом и физически обоснованной проверкой. Такой подход принципиально отличается от идеи полной замены традиционных методов: ИИ сокращает число кандидатов и автоматизирует интерпретацию, тогда как DFT и эксперимент обеспечивают проверку предложенного решения.

2. Основные методы искусственного интеллекта в структурной химии.

Химические данные могут быть представлены в форме молекулярных дескрипторов, SMILES-последовательностей, молекулярных графов, пространственных координат атомов или непосредственно спектральных сигналов. Выбор архитектуры определяется видом входных данных и задачей исследования.

Таблица 1. Основные интеллектуальные методы в структурном анализе.

Метод	Представление данных	Основная задача	Преимущество	Ограничение
Machine Learning	Дескрипторы	Классификация и прогноз	Простота и скорость	Требует выбора признаков
Deep Learning	Спектры, массивы данных	Автоматическое извлечение признаков	Работа со сложными зависимостями	Требует больших выборок
CNN	NMR/MS/IR сигналы	Выделение локальных спектральных признаков	Эффективна для сигналов	Зависит от качества данных
GNN	Молекулярный граф	Анализ атомов и связей	Учитывает топологию молекулы	Вычислительная сложность
Trans-former	Последовательности и спектры	Генерация/идентификация структуры	Учитывает дальние зависимости	Высокая стоимость обучения
ИИ + DFT	Спектры + расчётные параметры	Ранжирование и проверка	Скорость + физическая обоснованность	Требует согласования моделей

3. Искусственный интеллект в NMR-структурном анализе.

NMR-спектроскопия является одним из наиболее информативных методов установления структуры органических соединений. ИИ-системы позволяют решать две взаимосвязанные задачи: «структура → спектр» и «спектр → структура». Вторая задача особенно сложна, поскольку одному набору спектральных признаков может соответствовать большое число потенциальных молекулярных графов.

В работе по автоматизированному определению структуры по одномерным данным ^{1}H- и ^{13}C-NMR использована мультитасковая архитектура CNN–Transformer. Для молекул размером до 19 тяжёлых атомов правильная структура входила в Top-15 в 69,6% случаев без предварительного задания молекулярной формулы, а пространство поиска сокращалось вплоть до 11 порядков величины [6].

При использовании информации о 957 простых субструктурах размером не более семи атомов правильная структура входила в Top-15 в 93,2% случаев. Для уверенно предсказанного присутствия субструктуры сообщалась точность 96,3%, а для её отсутствия — 99,8% [6]. Эти показатели относятся к конкретной модели и тестовым условиям и не должны интерпретироваться как универсальная точность искусственного интеллекта.

Таблица 2. Количественные характеристики ИИ–NMR модели [6].

Показатель	Значение
Максимальный размер молекул	до 19 тяжёлых атомов
Число простых субструктур	957
Субструктуры → структура, Top-15	93,2%
^{1}H/^{13}C NMR → структура, Top-15	69,6%
Уверенное присутствие субструктуры	96,3%
Уверенное отсутствие субструктуры	99,8%
Сокращение пространства поиска	до 11 порядков величины

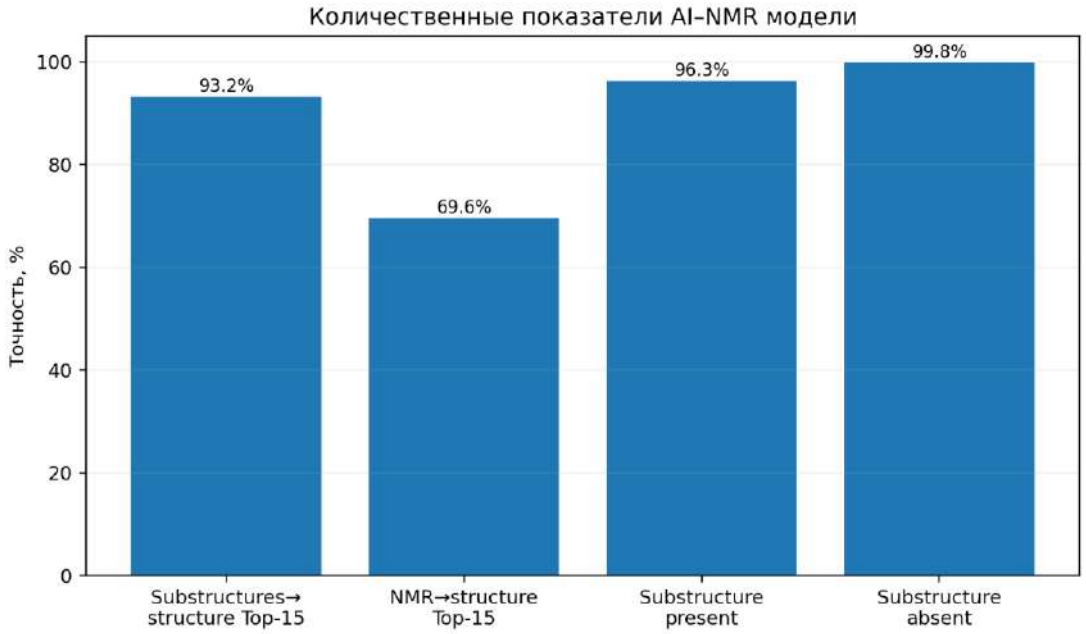


Рисунок 2. Количественные показатели ИИ–NMR модели по данным [6].

4. Интеграция искусственного интеллекта с DFT.

Одним из наиболее перспективных направлений является объединение ИИ с квантово-химическими расчётами. DFT обеспечивает физически содержательную оценку геометрии, электронной структуры и спектральных параметров, тогда как ИИ позволяет быстро ранжировать большое число кандидатов.

В предыдущей работе авторов DFT рассматривается как уровень количественного анализа, следующий за классическими моделями комплексных соединений [1]. Показано, что выбор функционала и базисного набора существенно влияет на расчёт, особенно для близких по энергии спиновых состояний, поэтому вычислительные результаты необходимо интерпретировать критически и сопоставлять с экспериментом [1].

Рациональная стратегия структурного анализа состоит в использовании ИИ для сокращения химического пространства, после чего ограниченное число наиболее вероятных структур подвергается DFT-оптимизации и спектральной проверке.



Рисунок 3. Интегрированная схема ИИ–DFT определения молекулярной структуры.

Таблица 3. Функции ИИ и DFT в интегрированном анализе.

Этап	Искусственный интеллект	DFT
Первичный анализ массива данных	Высокая эффективность	Нецелесообразно для массового перебора
Генерация кандидатов	Да	Нет
Ранжирование кандидатов	Быстро	Вычислительно затратно
Оптимизация геометрии	Модельно-зависимо	Да
Электронная структура	Приближённый ML-прогноз	Физически обоснованный расчёт
Финальная проверка	Вероятностная	Расчёт + эксперимент

5. Искусственный интеллект в масс-спектрометрии.

Масс-спектрометрия предоставляет информацию о молекулярной массе, составе и характере фрагментации. В ИИ-ориентированном анализе применяются прогнозирование MS/MS-спектров, поиск в библиотеках, классификация и ранжирование структур-кандидатов.

Графовые нейронные сети особенно перспективны, поскольку фрагментация определяется топологией молекулы и устойчивостью отдельных связей. Однако обратная задача «MS/MS → структура» остаётся неоднозначной, поэтому наибольшую ценность представляет объединение масс-спектрометрии с NMR и другими независимыми аналитическими данными.

6. Искусственный интеллект в IR-спектроскопии и мультимодальном анализе.

Инфракрасная спектроскопия традиционно используется для идентификации функциональных групп, однако глубокое обучение позволяет использовать практически весь спектральный профиль. В опубликованном исследовании Transformer-модель была предварительно обучена на 634 585 моделированных IR-спектрах и адаптирована на 3453 экспериментальных спектрах [9].

Для соединений с 6–13 тяжёлыми атомами точность определения полной структуры составила 44,4% для Top-1 и 69,8% для Top-10. При определении молекулярного каркаса показатели достигали 84,5% и 93,0% соответственно [9].

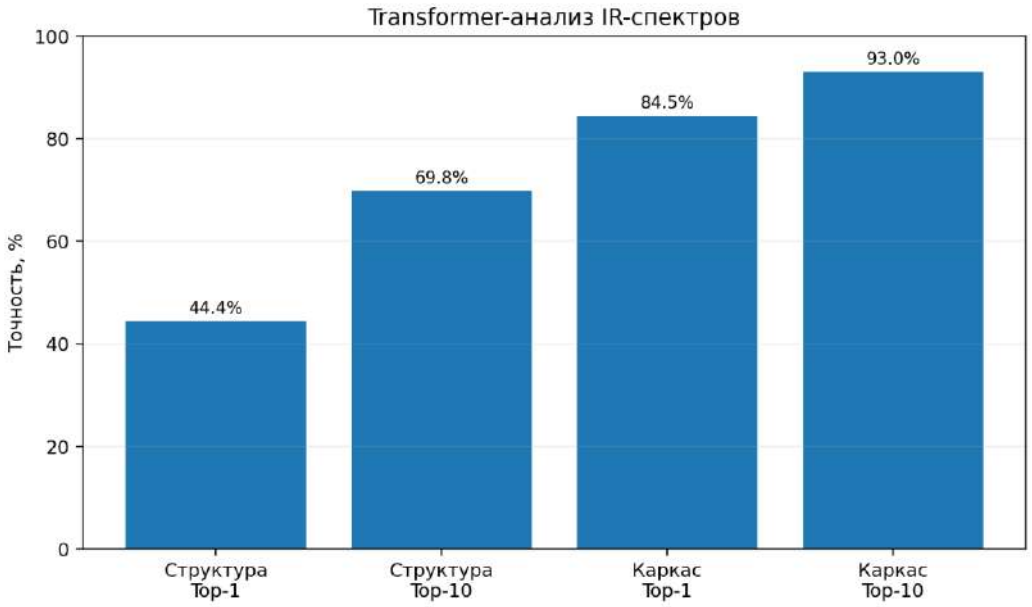


Рисунок 4. Результаты Transformer-модели при анализе IR-спектров по данным [9].

Таблица 4. Информационная роль аналитических методов в мультимодальной ИИ-системе.

Метод	Основная информация	Роль в ИИ-системе
^{1}H NMR	Окружение атомов водорода	Распознавание фрагментов
^{13}C NMR	Углеродный скелет	Проверка локального окружения
MS/MS	Масса и фрагментация	Ранжирование кандидатов
IR	Колебания и функциональные группы	Структурные ограничения
Raman	Колебательные характеристики	Дополнительная идентификация
DFT	Геометрия и электронные параметры	Физическая проверка

7. Сравнение традиционного и ИИ-ориентированного анализа.

Таблица 5. Сравнение подходов к установлению структуры.

Критерий	Традиционный анализ	ИИ-ориентированный анализ
Интерпретация	Экспертная	Автоматизированная/ полуавтоматизированная
Большие массивы данных	Трудоёмко	Высокая скорость
Генерация кандидатов	Ограниченная	Автоматическая
Зависимость от эксперта	Высокая	Снижается, но сохраняется
Интерпретируемость	Обычно высокая	Зависит от архитектуры
Новые химические классы	Экспертная адаптация	Риск снижения точности вне обучающей области
Проверка	Эксперимент	Эксперимент + DFT + экспертная оценка

ИИ обеспечивает наиболее заметное преимущество на этапе сокращения пространства поиска и ранжирования кандидатов. Однако статистическая вероятность не является самостоятельным доказательством структуры. Это соответствует методологическому принципу, сформулированному в предыдущей работе: сложный расчёт должен дополняться физической интерпретацией [1].

8. Ограничения применения искусственного интеллекта.

Первое ограничение связано с качеством и репрезентативностью обучающих данных. Ошибочные спектры, неточная разметка и недостаточное представительство отдельных классов соединений непосредственно отражаются на качестве прогноза.

Второе ограничение связано с переносимостью моделей. Высокая точность на тестовой выборке не гарантирует аналогичного результата для соединений, находящихся вне химического пространства обучения.

Третья проблема — интерпретируемость. Для химического исследования важно не только получить структурный ответ, но и установить, какие спектральные и физико-химические признаки обусловили выбор конкретного кандидата.

Четвёртое ограничение — необходимость экспериментальной верификации. ИИ должен рассматриваться как инструмент поддержки решения, а не как самостоятельная замена NMR, MS, рентгеноструктурному анализу или квантово-химическим расчётам.

Существенной методологической проблемой является различие между качеством модели на внутренней тестовой выборке и её надёжностью в реальной лабораторной практике. Спектры одного и того же вещества могут изменяться в зависимости от растворителя, концентрации, температуры, параметров прибора и способа предварительной обработки сигнала. Поэтому перенос ИИ-модели на данные, полученные в другой лаборатории или на другом спектрометре, требует дополнительной проверки. Для структурной химии особенно важно формировать стандартизованные наборы данных, содержащие не только спектр и конечную структуру, но и условия регистрации эксперимента. Такой подход позволяет уменьшить риск того, что алгоритм будет распознавать особенности конкретного набора данных вместо химически значимых закономерностей.

Отдельного внимания требует оценка неопределённости прогноза. Выдача единственной структуры без указания альтернативных кандидатов может создавать ложное ощущение достоверности. Более корректным является ранжирование нескольких возможных решений с указанием степени уверенности и последующей независимой проверкой средствами NMR, MS, IR, DFT или рентгеноструктурного анализа. Уже рассмотренные результаты показывают практическую значимость метрик Top-k: правильная структура не всегда занимает первое место, но может присутствовать среди ограниченного числа наиболее вероятных кандидатов [6, 8, 9]. Следовательно, ИИ целесообразно использовать как систему поддержки химика, сокращающую пространство поиска, а не как источник окончательного структурного заключения.

Проблема интерпретируемости непосредственно связана с доверием к автоматизированному анализу. Для исследователя важно понимать, какие химические сдвиги, корреляции, фрагментационные пики или полосы поглощения повлияли на выбор структуры. Перспективным направлением является развитие объяснимого искусственного интеллекта, способного связывать внутренние признаки модели с конкретными спектральными областями и структурными фрагментами. Такая интерпретация облегчает выявление ошибочных решений и делает результат пригодным не только для автоматической классификации, но и для полноценного химического анализа.

9. Перспективы развития.

Основным направлением дальнейшего развития является переход от одномодальных систем к мультимодальным моделям, одновременно использующим NMR, MS, IR/Raman, молекулярную формулу и расчётные параметры.

В системе NMRMind Transformer-модель была предварительно обучена приблизительно на 45 млн одномерных NMR-спектров, а последующая настройка выполнялась на 2,2 млн 1D- и 2D-NMR-спектров. При использовании полного набора входных данных сообщалась точность Top-1 92,07%, а при использовании только NMR без молекулярной формулы и заранее заданных фрагментов — 85,10% [7].

Другое современное направление связано с расширением размера анализируемых молекул. Для соединений, содержащих до 40 неводородных атомов, Transformer-модель, использующая только одномерные данные ^{1}H- и ^{13}C-NMR, достигла точности Top-15 60,4% [8]. Эти результаты показывают, что область автоматизированного структурного анализа постепенно распространяется на более сложные химические системы.

Особенно перспективна мультимодальная архитектура, в которой разные аналитические методы рассматриваются не как независимые источники информации, а как взаимодополняющие ограничения одной структурной задачи. NMR характеризует локальное химическое окружение атомов и связи между фрагментами, MS/MS отражает массу и закономерности фрагментации, IR и Raman дают информацию о колебательных характеристиках и функциональных группах, а DFT позволяет проверить геометрию, относительную устойчивость и расчётные спектральные параметры. Объединение этих данных потенциально уменьшает число структур, совместимых с экспериментом, и повышает устойчивость решения при неполных или зашумлённых входных данных.

Современные результаты NMRMind демонстрируют значение масштабного предварительного обучения: модель использовала около 45 млн одномерных спектров, а затем была настроена на 2,2 млн 1D- и 2D-NMR-спектров [7]. Важным является не только достигнутый показатель Top-1, но и сама методология обучения на разнородных модальностях. Она показывает возможность создания систем, способных работать при различной полноте исходной информации. Для практической химии это существенно, поскольку экспериментатор не всегда располагает одновременно молекулярной формулой, всеми типами NMR-корреляций и дополнительными спектроскопическими данными.

Расширение области применимости на более крупные молекулы остаётся отдельной задачей. По мере роста числа неводородных атомов резко увеличивается число потенциальных изомеров и структурных вариантов. Поэтому результат, полученный для небольших органических соединений, нельзя автоматически переносить на природные продукты, координационные соединения или другие структурно сложные системы. Данные о применении одномерного NMR к соединениям размером до 40 неводородных атомов показывают продвижение в этом направлении [8], однако дальнейший прогресс требует сочетания более информативных экспериментальных данных с химическими ограничениями и физически обоснованными вычислительными методами.

В перспективе ИИ-система структурного анализа должна включать не только генерацию ответа, но и автоматическую самопроверку. После построения кандидата могут прогнозироваться его NMR-, IR- и MS-характеристики, которые затем сопоставляются с исходным экспериментом. При существенном расхождении кандидат должен понижаться в рейтинге или исключаться. На следующем уровне DFT может использоваться для проверки ограниченного набора наиболее вероятных структур.

Такой замкнутый цикл «эксперимент — ИИ — расчёт — сравнение — уточнение» соответствует логике современной вычислительной химии и снижает риск принятия статистически вероятного, но физически несогласованного решения.

10. Результаты и обсуждение.

Проведённый анализ показывает, что искусственный интеллект формирует новый вычислительный уровень структурной химии. Его наиболее значимая функция состоит не только в ускорении обработки спектров, но и в существенном сокращении пространства возможных структур.

По данным рассмотренной модели CNN–Transformer, для молекул до 19 тяжёлых атомов правильная структура входила в Top-15 в 69,6% случаев, а использование субструктурной информации повышало соответствующий показатель до 93,2% [6]. Для NMRMind при полном наборе входных данных сообщалась точность Top-1 92,07% [7]. Для более крупных систем, содержащих до 40 неводородных атомов, показатель Top-15 составлял 60,4% [8].

Результаты IR-анализа демонстрируют, что интеллектуальное структурное прогнозирование распространяется и за пределы NMR: для молекулярного каркаса Transformer-модель достигала 84,5% Top-1 и 93,0% Top-10 [9].

При этом прямое ранжирование моделей по приведённым процентным показателям некорректно, поскольку в исследованиях используются разные наборы данных, размеры молекул, входные модальности и критерии Top-k. Поэтому количественные показатели следует рассматривать как характеристики конкретных систем, а не как универсальную меру точности ИИ.

Таблица 6. Обобщённые количественные результаты рассмотренных ИИ-систем.

Подход	Условия	Показатель
CNN–Transformer для NMR [6]	до 19 тяжёлых атомов, $^{1}H/^{13}C$	Top-15: 69,6%
Substructure–Transformer [6]	957 субструктур	Top-15: 93,2%
NMRMind [7]	полный набор входов	Top-1: 92,07%
NMRMind [7]	только NMR	Top-1: 85,10%
Transformer NMR [8]	до 40 неводородных атомов	Top-15: 60,4%
Transformer IR [9]	полная структура	Top-1: 44,4%, Top-10: 69,8%
Transformer IR [9]	молекулярный каркас	Top-1: 84,5%, Top-10: 93,0%

Заключение.

Искусственный интеллект является одним из наиболее динамично развивающихся инструментов современной структурной химии. Методы машинного и глубокого обучения позволяют автоматизировать интерпретацию спектральных данных, распознавать структурные фрагменты, генерировать и ранжировать молекулярные кандидаты.

Наиболее перспективным является интегрированный подход, объединяющий экспериментальную спектроскопию, ИИ и квантово-химическое моделирование. Искусственный интеллект обеспечивает высокоскоростной предварительный анализ и сокращение химического пространства, DFT — количественную оценку геометрии и электронной структуры, а эксперимент — окончательную проверку.

Предыдущая работа авторов показала целесообразность многоуровневого перехода от физически наглядных моделей к DFT [1]. В настоящем исследовании этот принцип расширен за счёт включения ИИ как дополнительного уровня обработки химической информации.

Такой подход позволяет сохранить физическую интерпретируемость и одновременно использовать преимущества современных цифровых технологий.

Дальнейшее развитие мультимодальных моделей, объяснимого искусственного интеллекта и физически обоснованных нейронных сетей может привести к созданию автоматизированных систем структурной идентификации, способных работать с комплексными экспериментальными данными при сохранении обязательной химической и экспериментальной верификации.

СПИСОК ЛИТЕРАТУРЫ:

1. Томуева А.Ш., Нуриева С.Р. Применение электростатической теории к комплексным соединениям: от классических моделей к квантово-химическим подходам. Рукопись, представленная авторами, 2026
2. Нуриева С.Р. Роль физики в развитии современных цифровых технологий // Вестник науки. 2026. Т. 4. № 8 (101). С. 326–339
3. Нуриева С.Р. Физические основы энергоэффективности и их влияние на экономическую эффективность промышленных предприятий // Вестник науки. 2026. Т. 1. № 7 (100). С. 57–63
4. Ahmadova M., Karimova M., Karimov F., Majidov R., Abdiyeva G., Bayramova L., Nuriyeva S. Integrated diagnostic assessment of light industry clusters development in the socio-economic system of Azerbaijan // Technology Audit and Production Reserves. 2026. Vol. 3, No. 4 (89). P. 74–80. DOI: 10.15587/2706-5448.2026.363059
5. Cramer C.J., Truhlar D.G. Density functional theory for transition metals and transition metal chemistry // Physical Chemistry Chemical Physics. 2009. Vol. 11. P. 10757–10816. DOI: 10.1039/B907148B
6. Hu F., Chen M.S., Rotskoff G.M., Kanan M.W., Markland T.E. Accurate and Efficient Structure Elucidation from Routine One-Dimensional NMR Spectra Using Multitask Machine Learning // ACS Central Science. 2024. Vol. 10. No. 11. P. 2162–2170. DOI: 10.1021/acscentsci.4c01132
7. Xue X., Sun H., Sun J., Patiny L., Liu X., Chen K., Yan J., Li L., Liu X., Xu S. et al. NMRMind: A Transformer-Based Model Enabling the Elucidation from Multidimensional NMR to Structures // Analytical Chemistry. 2025. Vol. 97. No. 41. P. 22603–22614. DOI: 10.1021/acs.analchem.5c03783

8. Hu F., Tubb J.M., Argyropoulos D., Golotvin S., Elyashberg M., Rotskoff G.M., Kanan M.W., Markland T.E. Pushing the Limits of One-Dimensional NMR Spectroscopy for Automated Structure Elucidation Using Artificial Intelligence // Journal of Chemical Information and Modeling. 2026. Vol. 66. No. 12. P. 6992–7002. DOI: 10.1021/acs.jcim.6c00905

9. Alberts M., Laino T., Vaucher A.C. Leveraging Infrared Spectroscopy for Automated Structure Elucidation // Communications Chemistry. 2024. Vol. 7. Art. 268. DOI: 10.1038/s42004-024-01341-w

10. Керимова М.Г., Керимов Ф.Дж., Аббасова Ш.Р. Экономические факторы повышения конкурентоспособности предприятий легкой промышленности // Вестник науки. 2026. Т. 2. № 8 (101). С. 22–38

Tomueva A.Sh.

PhD in Chemistry

department of general chemistry and chemistry teaching technology

Azerbaijan State Pedagogical University

(Baku, Azerbaijan)

ARTIFICIAL INTELLIGENCE IN DETERMINING STRUCTURE OF CHEMICAL COMPOUNDS: MODERN METHODS AND PERSPECTIVES

Abstract: *article systematizes modern approaches to application of artificial intelligence in determining structure of chemical compounds. Transition from traditional spectroscopic interpretation and quantum chemical calculations to machine learning, deep neural networks, graph models and Transformer architectures is considered. Special attention is paid to analysis of nuclear magnetic resonance, mass spectrometry and infrared spectroscopy data. It is shown that modern models are capable of solving both direct «structure-spectrum» task and inverse «spectrum-structure» task. Integration of artificial intelligence with density functional theory (DFT) is discussed. According to published data, for molecules up to 19 heavy atoms, hybrid CNN-Transformer model included correct structure in first 15 candidates in 69.6% of cases and reduced structural search space by up to 11 orders of magnitude [6]. Conclusion is made about prospects of complex approach combining experiment, quantum chemical modeling and artificial intelligence.*

Keywords: *artificial intelligence, chemical compound, molecular structure, machine learning, deep learning, neural network, NMR spectroscopy, mass spectrometry, infrared spectroscopy, density functional theory, transformer architecture, molecular graph, structural analysis.*