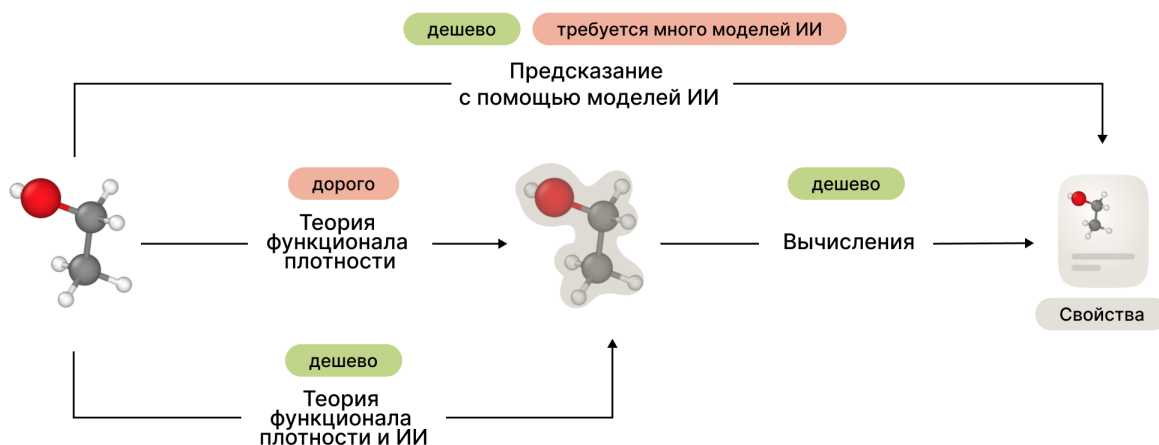


Что происходит между атомами и электронами

Ученые из России создали самую большую в мире базу данных для квантовой химии

Разработка новых лекарственных препаратов и материалов зависит от качества предсказания физических и химических свойств будущего продукта. Один из новых и популярных подходов к решению подобных задач – применение методов на стыке квантовой химии и искусственного интеллекта. Однако для обучения моделей ИИ предсказанию свойств молекул необходим доступ к соответствующей информации о каждой из их многочисленных характеристик, а качество предсказания будет зависеть от количества и разнообразия данных. Ученые из Института искусственного интеллекта AIRI создали и выложили в открытый доступ крупнейший в мире набор данных по квантовой химии, чтобы расширить возможности исследований в области поиска новых материалов и разработки лекарств. Исследование [опубликовано](#) в журнале Physical Chemistry Chemical Physics.



Прогнозирование свойств молекулы — важный этап создания нового препарата, и машинное обучение способно ускорить и упростить этот процесс. Проверка работы моделей ИИ для химии, в отличие от популярных моделей для генерации изображений или текстов, очень трудозатратна: нужно пойти в «мокрую» лабораторию, провести эксперименты в реальном мире, синтезировать структуру и затем оценить каждое её свойство. Вместо дорогостоящих экспериментов некоторые свойства можно оценить с помощью методов квантовой химии. Например, решение уравнения Шредингера помогает понять, что происходит между атомами и электронами, смоделировать поведение молекулы или материала и вычислить их теоретические свойства. Объем вычислений, необходимых для точного решения уравнения Шредингера, экспоненциально растет с увеличением числа электронов, и здесь на помощь ученым приходят нейронные сети. Именно они позволяют эффективно «перенести» дорогие эксперименты в цифровую плоскость. Вместо того, чтобы предсказывать конкретное свойство молекулярной структуры, эти методы направлены на оценку молекулярной конформации — то есть трехмерного расположения атомов в молекуле путем предсказания ее квантовых свойств.

Чтобы сделать применение методов ИИ в квантовой химии повсеместным, научному сообществу необходимо большее количество специализированных данных. По словам Артура Кадурина, руководителя научной группы «Глубокое обучение в науках о жизни» AIRI, подавляющее большинство исследований, недавно проведенных в этой области, ограничивается экспериментами на небольшом количестве структур разных веществ, что ставит под сомнение применимость моделей ИИ в реальных задачах индустрий.

С целью решения этой проблемы ученые Института AIRI при поддержке коллег из Сколтеха и ПОМИ РАН собрали 5 340 152 конформаций для 1 004 918 подобных лекарственных молекул, а также их квантовые свойства, и выложили базу данных в открытый доступ на маркетплейс артефактов машинного обучения DataHub. [Доступ к датасету](#) возможен через платформу ML Space. В дополнение к данным в набор включили 4 модели для предсказания энергии молекулярной конформации и 2 модели для работы с теорией функционала плотности.

Хотя получить модели, близкие к химической точности, по-прежнему сложно, экспериментальные данные показывают, что большие наборы данных приводят к повышению качества моделей ИИ. Научная группа планирует дальше пополнять уже собранную базу и надеется, что эта работа сделает *in silico* эксперименты более доступными, а точность предоставляемых результатов приблизится к качеству лабораторных исследований.

Вопросы: pr@airi.net

Научно-исследовательский Институт искусственного интеллекта AIRI — автономная некоммерческая организация, занимающаяся фундаментальными и прикладными исследованиями в области искусственного интеллекта. На сегодняшний день более 90 научных сотрудников AIRI задействовано в исследовательских проектах Института для работы совместно с глобальным сообществом разработчиков, академическими и индустриальными партнерами.