

User name:
hidden by privacy settings

Check ID:
1016256547

Check date:
16.05.2024 15:09:58 EEST

Check type:
Doc vs Internet

Report date:
16.05.2024 15:10:34 EEST

User ID:
100013793

File name: **Ісайович**

Page count: **57** Word count: **12143** Character count: **94623** File size: **1.20 MB** File ID: **1016043703**

2.15% Matches

Highest match: **1.49%** with Internet source (<http://catalog.liha-pres.eu/index.php/liha-pres/catalog/download/239/6070/13661-1?..>)

2.15% Internet sources 128

Page 59

No Library search was conducted

0% Quotes

Exclusion of quotes is off

Exclusion of references is off

0% Exclusions

No exclusions

Modifind

Text modifications detected. Find more details in the online report.

Replaced characters 1

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДЕРЖАВНИЙ ВИЩИЙ НАВЧАЛЬНИЙ ЗАКЛАД
«УЖГОРОДСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ»
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ХІМІЇ ТА ЕКОЛОГІЇ

Кафедра аналітичної хімії

Дипломна робота магістра

«ШТУЧНИЙ ІНТЕЛЕКТ У ХІМІЇ ТА ХІМІЧНІЙ ОСВІТІ»

Виконала: студентка II –го курсу магістр

спец. 014.06 Середня освіта. Хімія

Тегза Вікторія Антонівна

Керівник: к.х.н., доц. (Студеняк Я.І.)

Рецензент: к.х.н., доц. (Глух О.С.)

Ужгород – 2024

ЗМІСТ

ВСТУП	5
ОГЛЯД ЛІТЕРАТУРИ 1. ШТУЧНИЙ ІНТЕЛЕКТ ТА ЙОГО РОЛЬ В ОСВІТНЬОМУ ПРОЦЕСІ.....	8
1.1. Короткі відомості про ШІ та його історію розвитку.....	8
1.2. Досягнення та проблеми інтеграції ШІ у навчальному процесі.....	12
РОЗДІЛ 2. ВИКОРИСТАННЯ ШТУЧНОГО ІНТЕЛЕКТУ НА УРОКАХ ХІМІЇ ПЕДАГОГАМИ ТА ШКОЛЯРАМИ.....	15
2.1. ШІ при підготовці до домашнього завдання.....	15
2.2. Перевірка знань учнів із залученням ШІ.....	17
2.3. Проведення занять за допомогою ШІ.....	19
2.4. ШІ у позакласній роботі.....	20
РОЗДІЛ 3. ВИКОРИСТАННЯ ШТУЧНОГО ІНТЕЛЕКТУ В СИНТЕЗІ ТА ПОШУКУ НОВИХ РЕЧОВИН.....	24
3.1. Автоматизоване проектування лікарських засобів.....	24
3.2. Оптимізація каталізаторів і матеріалів.....	25
3.3. Генерація нових молекул за допомогою глибокого навчання.....	28
3.4. Пошук нових матеріалів з певними фізико-хімічними властивостями.....	38
3.5. Машинне навчання для прогнозування хімічних властивостей.....	41
3.6. Комп'ютерне моделювання молекулярних структур.....	44
3.7. Розвиток хімічних реакцій з використанням алгоритмів інтелектуальної обробки даних.....	48
РОЗДІЛ 4. ЕТИЧНІ АСПЕКТИ ВИКОРИСТАННЯ ШІ НА УРОКАХ ХІМІЇ.....	52
4.1. Розгляд можливих етичних питань та впливу на процес навчання.....	52
4.2. Переваги та недоліки використання ШІ на роках хімії.....	55
ВИСНОВОК.....	61

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ.....	63
ДОДАТОК.....	70

Перелік умовних скорочень

ШІ – штучний інтелект

ML – машинне навчання

ІКТ – інформаційно-комунікаційні технології

RPA – роботизована автоматизація процесів

ВСТУП

Середина 20 століття ознаменувалася початком інтенсивного розвитку глобального електронного інформаційного середовища. Широке поширення високошвидкісної комп'ютерної техніки та стрімкий розвиток ІКТ відкрили принципово нові широкі горизонти для інноваційного розвитку людського суспільства в цілому. У 21 столітті розширення Інтернету речей створило передумови для створення цифрових копій реального суспільства в кожному сегменті людської діяльності, від виробництва до фінансів, охорони здоров'я, освіти, транспорту, робототехніки тощо. Наявність доступу до всіх інформаційних ресурсів без обмежень за часом і місцем доступу створила умови для роботи з інформаційною прикладною моделлю в режимі реального часу.

Пандемія COVID-19 зіграла роль каталізатора масового впровадження нових ІКТ для підтримки віддаленої роботи майже в усіх сферах людської діяльності. За оцінками експертів ЮНЕСКО, пандемія створила проблеми для освітніх процесів для близько 1,6 млрд студентів у понад 190 країнах світу, що в свою чергу призвело до прискореного розвитку та застосування сучасних освітніх технологій. У свою чергу, нові освітні технології необхідно було адаптувати для використання в різних класах з різним рівнем професійної та комп'ютерної грамотності, що призвело до впровадження методів навчання на основі штучного інтелекту (ШІ). Прискорене впровадження нових ІКТ підтверджується, наприклад, такими даними. Міжнародна корпорація даних (IDC) прогнозує, що «обсяг технологічної індустрії перевищить 5,3 трильйона доларів у 2022 році». Відмінною рисою таких революційних перетворень стало інтенсивне поширення кількості проведених комунікацій, що в свою чергу призвело до розвитку систем ШІ. Прогнозується, що до 2028 року світовий ринок штучного інтелекту становитиме «641,30 мільярдів доларів із середньорічним темпом зростання 36,1%». Моделі віддалених робочих місць стають ближчими до реального світу завдяки використанню інструментів віртуального світу.

Актуальність проблеми. Сьогодні впровадження елементів ШІ є масовим і спонтанним. Як показав аналіз, багато вчених приділяють увагу питанням розширення сфери застосування ШІ та розвитку платформ навчання цифрових технологій з використанням ШІ. Їх виконання визначається емпіричними судженнями. Така ситуація пояснюється тим, що технології ШІ розвиваються настільки швидко, що використання ретроспективного аналізу є дуже проблематичним. Можна констатувати, що питанню моделювання процесів із залученням технологій ШІ особливо у хімічній освіті приділяється недостатня увага, у цій сфері існують невирішені проблеми.

Зв'язок роботи з науковими програмами, планами, темами. Робота виконувалась у рамках науково-педагогічних завдань кафедри аналітичної хімії.

Мета і задачі дослідження - з'ясувати базові можливості, оцінити переваги та недоліки використання штучного інтелекту в прикладних аспектах хімії та хімічної освіти.

Завдання:

1. Розглянути відомості про штучний інтелект та його роль в освітньому процесі.
2. Оцінити використання ШІ на уроках хімії, оглянути проведення занять з використанням ШІ.
3. Описати використання ШІ при синтезі та пошуці речовин і удосконаленні алгоритмів інтелектуальної обробки даних
4. Описати переваги та недоліки ШІ на уроках хімії

Об'єкт дослідження – інтеграція можливостей штучного інтелекту в хімію та хімічну освіту.

Предмет дослідження. Роль штучного інтелекту в освітньому процесі при вивченні хімії та при вирішенні практичних завдань хімії і хімтехнології.

Методи дослідження – пошук та критичний огляд літератури,

Наукова новизна одержаних результатів, проаналізовано сучасний стан розвитку штучного інтелекту у плані його практичного застосування на уроках

хімії та хімічній науці. З'ясовано основні переваги та недоліки використання штучного інтелекту.

Практичне значення одержаних результатів, викладений матеріал, міркування та висновки можуть бути використані в освітньому процесі та пошуковій роботі з хімії.

Особистий внесок здобувача, робота виконана самостійно, основні положення обговорювались із науковим керівником.

Апробація результатів дослідження, основні положення роботи доповідались на науковій студентській конференції ДВНЗ «Ужгородський національний університет», секція «Хімічних наук та екології» (25 травня 2023 р.)

Структура та обсяг роботи. Дана робота містить 71 ст., 4 роз. 54 джерел, 2 табл., 10 рис., 1 дод.

ОГЛЯД ЛІТЕРАТУРИ 1. ШТУЧНИЙ ІНТЕЛЕКТ ТА ЙОГО РОЛЬ В ОСВІТНЬОМУ ПРОЦЕСІ.

1.1. Короткі відомості про ШІ та його історію розвитку.

Термін штучний інтелект (ШІ) не був придуманий до середини 20 століття, але ідея побудови інтелектуальних систем захоплювала людей протягом тисячоліть. Це концепція, яка передувє промисловій революції та винаходу першого цифрового комп'ютера. Він також був частим джерелом натхнення для творців фентезі та наукової фантастики.

З моменту офіційного визначення в 1950-х роках сфера штучного інтелекту пережила багато злетів і падінь, хвилювань і розчарувань, проривів і тупиків. За останні кілька десятиліть штучний інтелект досяг значних успіхів, але він також не зміг виконати багато своїх раних обіцянок – принаймні поки що.

У 1950 році, коли штучний інтелект формується як нова наукова галузь, Айзек Азімов пише книгу «Я, робот», у якій обговорює мирне співіснування людей і автономних роботів за допомогою своїх трьох законів робототехніки.[1-3]

З винаходом цифрового комп'ютера в середині 20-го століття вчені почали запитувати, чи зможуть комп'ютери вийти за рамки простого виконання обчислень і виконання жорстко закодованих інструкцій.

Іншими словами, чи може ця машина продемонструвати інтелект? Чи могло воно думати як людина? Це питання відкриває низку філософських запитів, оскільки визначення людського інтелекту може включати складні поняття, як-от свідомість.

Алан Тюрінг, засновник сучасної інформатики, запропонував метод вимірювання машинного інтелекту. У своїй основоположній праці «Обчислювальна техніка та інтелект» Тюрінг переформулював питання «Чи може машина мислити?» запитати, чи може машина демонструвати людський інтелект своєю поведінкою та імітувати здатність мислити.

Далі він запропонував тест для машин, названий імітаційною грою, яка мала б визначити, чи є машина розумною чи ні. Це стало відомим як тест Тюрінга, і в його початковому форматі він вважав машину розумною, якщо вона була здатна вести текстову розмову у спосіб, який неможливо відрізнити від людини.

Отже, як виглядає тест Тюрінга? Уявіть на мить, що ви ведете онлайн-розмову за допомогою текстових повідомлень. Ви можете задавати запитання та оцінювати відповіді. Якщо ви взаємодієте з машиною, але не можете сказати, що це машина за відповідями, які вона дає, ми можемо сказати, що вона пройшла тест Тюрінга та є розумною.

Ідея Тюрінга отримала велику підтримку, але також і критику. Незважаючи на побоювання, що цей тест може бути недостатнім для оцінки машинного інтелекту, він все ще широко використовується сьогодні в нових формулюваннях.

Головне занепокоєння Тюрінга полягало в тому, що машини будуть обмежені пам'яттю, і їх ніколи не вистачить, щоб пройти його тест. Він передбачив, що комп'ютер зможе пройти, якби він міг зберігати 100 МБ пам'яті, що, як він вважав у 1950-х роках, стане можливим до 2000 року. Однак пам'ять у машинах багато років тому значно перевищувала 100 МБ, а продуктивність була схожа на людську. в тесті Тюрінга зустрічаються надзвичайно рідко.

Навіть тоді деякі вчені сперечаються, чи справді випадки, коли машини нібито «пройшли», можна вважати успішною імітацією людського інтелекту. Обмеження тесту Тюрінга пролили світло на труднощі вимірювання та порівняння складності людського інтелекту зі здібностями машин.

У наступні роки після тесту Тюрінга почали формуватися галузі машинного інтелекту та мислячих машин. Термін «штучний інтелект» вперше використав Джон Маккарті під час Дартмутського літнього дослідницького проекту зі штучного інтелекту в 1956 році, також відомого просто як Дартмутська майстерня. Це було зібрання деяких найвидатніших учених,

інженерів, математиків і психологів того часу. Протягом двох місяців вони об'єднали досвід і знання з різних академічних дисциплін в єдину галузь дослідження, яку Маккарті потім правильно назвав.

Семінар у Дартмуті вважається основоположною подією в історії ШІ. На цій конференції обговорювалися багато понять, які все ще використовуються сьогодні, такі як «нейронні мережі» та «обробка природної мови».

Інтелектуальне становлення штучного інтелекту як власної сфери дослідження незабаром почало давати значні успіхи.

Багато раних комп'ютерних програм, які могли грати в ігри, розв'язувати головоломки, доводити математичні теореми та виконувати штучні міркування, були розроблені в роки, що відбулися одразу після Дартмутської майстерні. Концепції машинного навчання та штучних нейронних мереж, двох основних стовпів сучасних систем штучного інтелекту, також були формалізовані приблизно в цей час.

Уміння грати в шахи завжди було тісно пов'язане з ідеєю інтелекту. У 1950 році Клод Шеннон, засновник теорії інформації[4-6], описав першу комп'ютерну програму, яка могла грати в шахи. Шеннон окреслив два підходи, до яких комп'ютер може вдатися, щоб виграти гру:

- Машина перевіряє дошку і вичерпує всі можливі рухи.
- Машина дотримується більш розумної стратегії та враховує лише встановлену кількість ключових ходів.

Перші досягнення штучного інтелекту в 1950-х і 1960-х роках призвели до хвилі надмірної самовпевненості та нереалістичних очікувань. Видатні діячі в області штучного інтелекту, зрозуміло, були схвильовані тим, що вони вже бачили, і очікували, що такі ж вражаючі прориви продовжаться з такою ж швидкістю.

Люди набувають інтелекту, дотримуючись інструкцій і виконуючи дедуктивні міркування, але вони також навчаються через досвід і повторення методом проб і помилок. Чому те саме не повинно бути вірним для розумних машин? Це питання почали задавати собі інженери ШІ, і незабаром машини

більше не обмежувалися формальними міркуваннями. Натомість вони почали оснащуватися можливостями, необхідними для навчання на прикладах.

Машинне навчання, термін, введений Артуром Семюелом кілька десятиліть тому, повернулося з великим розмахом. Ця нова парадигма, починаючи з кінця 1980-х років, використовуватиме статистичні дані та ймовірності, щоб дозволити машинам навчатися з наявних даних і адаптуватися, спираючись на попередній досвід. Невдовзі штучний інтелект став зв'язуватися з іншими зрілими та строгими дисциплінами, такими як теорія прийняття рішень, статистика, теорія управління та оптимізація.

Так само, як і в 1950-х і 1960-х роках, окремі досягнення та інновації почали набирати обертів і прогресувати. Невдовзі відновився інтерес до штучних нейронних мереж, які представляють собою групу моделей машинного навчання, створених на основі біологічних нейронів і представлених у середині 20 століття. Це сталося завдяки переосмисленню одного з ключових алгоритмів, які використовуються для навчання таких мереж – алгоритму «зворотного поширення».

Створення Всесвітньої павутини та розвиток телекомунікаційного сектору сприяли масштабній передачі та зберіганню даних протягом 2000-х років. Ці розробки дали нейронним мережам і алгоритмам глибокого навчання паливо, необхідне для досягнення значних успіхів: великі дані.

Мабуть, найкращим прикладом є комп'ютерна програма AlphaGo, розроблена компанією зі штучного інтелекту DeepMind. Го є набагато складнішою грою, ніж шахи, але ця програма змогла навчитися грі з нуля та за кілька місяців перемогла чемпіона світу Лі Седоля у 2016 році. Перехід від основного алгоритму гри в шашки Артура Семюеля до перемоги над Каспаровим у Deep Blue зайняв 40 років. Від цього моменту до перемоги AlphaGo знадобилося приблизно половина цього часу.

1.2. Досягнення та проблеми інтеграції ШІ у навчальному процесі.

Інтеграція штучного інтелекту (ШІ) у класну кімнату володіє потенціалом до кардинальних змін способів навчання учнів і викладання вчителів. Алгоритми штучного інтелекту можуть надавати учням персоналізований відгук і рекомендації, забезпечуючи більш захоплюючий і ефективний досвід навчання. Незважаючи на ці потенційні переваги, існує також кілька проблем, пов'язаних із впровадженням ШІ у класне навчання [7-9].

Штучний інтелект стає все більш важливою частиною нашого повсякденного життя, і він має потенціал змінити спосіб роботи, спілкування та навчання. В освіті штучний інтелект може надати учням більш специфічний досвід навчання та допомогти вчителям ефективніше задовольняти потреби кожного учня. Незважаючи на ці потенційні переваги, є також кілька проблем, пов'язаних із впровадженням штучного інтелекту в клас, включаючи потребу в технічних знаннях, обмежені ресурси та етичні проблеми.

Однією з ключових переваг впровадження штучного інтелекту в класі є можливість надати учням більш персоналізований досвід навчання. Алгоритми штучного інтелекту можуть аналізувати дані учнів і адаптуватися до їхніх стилів навчання, надаючи відгуки та рекомендації, адаптовані до їхніх індивідуальних потреб і здібностей. Це може допомогти підтримати зацікавленість і мотивацію студентів і може призвести до покращення академічної успішності. Ще одна перевага впровадження штучного інтелекту в клас – це можливість поглибити розуміння цієї технології, що швидко розвивається. Включивши ШІ в навчальну програму, вчителі можуть допомогти учням розвинути критичний погляд на цю технологію та підготувати їх до викликів і можливостей епохи цифрових технологій. Нарешті, впровадження штучного інтелекту в класну кімнату також може допомогти учням розвинути важливі для 21-го століття навички, такі як вирішення проблем, критичне мислення та співпраця. Ці навички необхідні для успіху в епоху цифрових технологій, і їх можна розвинути завдяки практичному досвіду роботи з інструментами та програмами ШІ[10].

Незважаючи на те, що впровадження штучного інтелекту в клас має багато переваг, є також кілька проблем, які вчителі повинні подолати. Однією з найбільших проблем є потреба в технічній експертизі. Вчителям, які не знайомі з ШІ, може бути важко інтегрувати цю технологію у свою практику викладання, і їм може знадобитися підтримка та навчання. Ще однією проблемою є вартість інструментів і програм ШІ. Багато шкіл та університетів не мають ресурсів для придбання та підтримки технології, необхідної для впровадження штучного інтелекту в клас, і їм, можливо, доведеться шукати зовнішнє фінансування або партнерство для підтримки своїх зусиль. Нарешті, існують також етичні проблеми, пов'язані з впровадженням ШІ в клас. Оскільки штучний інтелект стає все більш досконалим, виникають занепокоєння щодо його впливу на конфіденційність, безпеку та ринок праці. Вчителі повинні знати про ці проблеми та працювати над тим, щоб їхні учні були захищені під час вивчення цієї захоплюючої технології, яка швидко розвивається[11].

Найкращі практики впровадження штучного інтелекту в класі можуть допомогти вчителям ефективно інтегрувати цю технологію у свою практику викладання та надати учням більш персоналізований та цікавий досвід навчання. Нижче наведено кілька основних найкращих практик, які слід враховувати[12]:

1. Співпрацюйте з надійним постачальником штучного інтелекту

Для успішної інтеграції ШІ в класну кімнату вкрай важливо знайти надійного партнера з штучного інтелекту. Це може бути технологічна компанія, місцевий університет або некомерційна організація, яка спеціалізується на навчанні ШІ. Правильний партнер може надати підтримку, навчання та керівництво, щоб допомогти вчителям ефективно запровадити ШІ у свою практику викладання.

2. Почніть з малого

Замість того, щоб намагатися впроваджувати штучний інтелект в усю навчальну програму, вчителям рекомендується починати з малого та

просуватися вперед. Це дозволяє вчителям набути досвіду роботи з технологіями, здобути впевненість і з часом удосконалити свою практику викладання. Наприклад, вчителі можуть розпочати з включення навчальних ігор на основі штучного інтелекту у свої уроки або використання алгоритмів штучного інтелекту, щоб надати учням персоналізований відгук щодо їхніх завдань.

3. Розвивайте етичне та критичне мислення

Впровадження штучного інтелекту в класну кімнату дає студентам можливість виробити критичний погляд на цю технологію та її вплив на суспільство. Викладачі повинні заохочувати учнів критично мислити про етичні наслідки ШІ та розглядати потенційні наслідки його широкого використання. Це може допомогти учням стати відповідальними та поінформованими цифровими громадянами, які зможуть орієнтуватися в викликах і можливостях цифрової ери.

Підсумовуючи, запровадження штучного інтелекту в класі відкриває унікальну можливість як для вчителів, так і для студентів. ШІ має потенціал, щоб надати студентам персоналізований та захоплюючий досвід навчання, а також допомогти їм розвинути такі важливі навички 21-го століття, як критичне мислення та вирішення проблем. Однак така інтеграція технологій у класну кімнату також створює низку проблем, таких як конфіденційність даних і етика, потреба в постійному навчанні та підтримці, а також потенціал для нерівного доступу до технологій і цифрових навичок[13].

РОЗДІЛ 2. ВИКОРИСТАННЯ ШТУЧНОГО ІНТЕЛЕКТУ НА УРОКАХ ХІМІЇ ПЕДАГОГАМИ ТА ШКОЛЯРАМИ.

2.1. ШІ при підготовці до домашнього завдання.

У минулому шкільні домашні завдання склалися здебільшого на папері, часто були тьмяними, а іноді займали багато часу. Учням доводилось покладатись на підручники, енциклопедії та бібліотеки, щоб зібрати інформацію та виконувати завдання. Цей підхід мав свої обмеження, такі як обмежений доступ до ресурсів, потенційний плагіат і суб'єктивне оцінювання.

З розвитком технологій шкільні домашні завдання перейшли в цифрову сферу. Тепер студенти мають доступ до онлайн-ресурсів, електронних книг і освітніх платформ, які забезпечують інтерактивне навчання. Платформи цифрових домашніх завдань також дозволяють вчителям ефективніше призначати та оцінювати домашні завдання, що приносить користь як учням, так і викладачам.

Майбутнє шкільних домашніх завдань – за інтеграцією технологій ШІ. Завдяки прогресу в обробці природної мови (NLP) і машинному навчанні (ML) помічники для виконання домашніх завдань на основі штучного інтелекту можуть надавати учням персоналізовані вказівки, відгуки та ресурси. Ці помічники можуть аналізувати сильні та слабкі сторони студентів, адаптувати рекомендації та відстежувати їхній прогрес, що призводить до покращення результатів навчання та успіху в навчанні.

ШІ для домашніх завдань з хімії долає розрив між традиційними методами навчання та індивідуальними навчальними потребами, використовуючи потужність технологій ШІ, включаючи машинне навчання та обробку природної мови. Цей потужний інструмент пропонує налаштовані шляхи навчання, інтерактивне вирішення проблем і миттєвий зворотний зв'язок, адаптований до потреб окремих учнів.

Застосування ШІ в хімії виходить за рамки допомоги в домашніх завданнях. Він пронизує різні аспекти освітньої подорожі, від досліджень і експериментів до аналізу даних. Алгоритми ШІ мають здатність передбачати

хімічні реакції, моделювати молекули та моделювати експерименти. Вони можуть обробляти величезну кількість хімічних даних, допомагаючи в інтерпретації результатів і виведенні значущих висновків.

Персоналізовані інструменти перевірки на основі штучного інтелекту, такі як налаштовані тести та картки, покращують навчання та допомагають у підготовці до іспитів. ШІ для домашніх завдань з хімії ефективно долає розрив між теоретичними знаннями та практичним застосуванням, імітуючи лабораторні експерименти та візуалізуючи хімічні процеси. Це забезпечує захоплюючий досвід навчання, який перевершує традиційні методи навчання.

Переваги:

1. Персоналізоване навчання: штучний інтелект забезпечує персоналізований досвід навчання, адаптуючись до унікальних потреб, стилю навчання та темпу кожного учня. Домашні завдання можна адаптувати, щоб зосередити увагу на областях, які потрібно вдосконалити, забезпечуючи цільову підтримку та прискорюючи академічне зростання.

2. Ефективне оцінювання: системи оцінювання на основі штучного інтелекту можуть виконувати великі обсяги завдань і надавати швидший і послідовніший зворотний зв'язок. Це економить дорогоцінний час вчителів, дозволяючи їм зосередитися на навчанні, а не виставляти оцінки вручну.

3. Академічна доброчесність: ШІ може відігравати вирішальну роль у просуванні академічної доброчесності. Алгоритми виявлення плагіату можуть ідентифікувати скопійований вміст, гарантуючи, що студенти створять оригінальну роботу. Це виховує культуру чесності та етичного навчання.

4. Покращена взаємодія: платформи домашніх завдань на базі штучного інтелекту можуть включати елементи гейміфікації, роблячи навчання більш інтерактивним і захоплюючим для учнів. Використання мультимедіа, тестів та інтерактивних вправ зберігає у студентів мотивацію та інвестує їх у виконання завдань.

Інтеграція ШІ в шкільні домашні завдання має величезне значення для освіти. Він має потенціал революціонізувати традиційну практику домашніх завдань, сприяти персоналізованому навчанню та ефективно задовольняти індивідуальні потреби учнів. Використовуючи технології штучного інтелекту, школи можуть підготувати учнів до викликів цифрової епохи, розвивати навички критичного мислення та навчатися протягом усього життя.

ШІ має потенціал трансформувати різні аспекти освіти за межі шкільних домашніх завдань. Деякі програми ШІ в освіті включають:

- Інтелектуальні системи навчання
- Віртуальні класи
- Адаптивні навчальні платформи
- Персоналізовані системи рекомендацій

Для викладачів вкрай важливо відповідально ставитися до ШІ та використовувати його можливості для створення більш захоплюючого, ефективного та інклюзивного навчального середовища[14].

2.2. Перевірка знань учнів із залученням ШІ.

Оцінювання є невід'ємною частиною викладання та навчання. Це також одна зі сфер освіти, яка охоплює впровадження штучного інтелекту. Завдяки аналізу різноманітних показників відвідуваності студентів і моделей навчання ШІ допомагає викладачам максимізувати результати навчання.

Сьогодні оцінка знань на основі штучного інтелекту використовує такі технології, як роботизована автоматизація процесів (RPA), машинне навчання, зіставлення шаблонів і обробка природної мови (NLP). Поєднання цих технологій допомагає викладачам аналізувати величезні масиви даних і вносити зміни в персоналізовані плани навчання.

За допомогою діагностичного тестування ШІ визначає прогалини в навчанні учнів і пропонує зміни в індивідуальних навчальних програмах. Багато постачальників Edtech тренують свої алгоритми ШІ відповідно до Теорії простору знань. Цей підхід визначає та відстежує «бали знань» учня

математичною мовою та формує картину «стану знань» людини з конкретного предмету.

Системи оцінювання та підрахунку балів штучним інтелектом тепер генерують унікальні тестові запитання з кількома варіантами відповіді «на льоту» та використовують адаптивне підрахунок балів, автоматично підлаштовуючи рівень складності відповідно до рівня знань учасників. Системи адаптивного тестування здатні оцінювати як вербальні, так і кількісні відповіді та базувати свої наступні запитання на попередніх відповідях людини. Таким чином, система пропонує кожному учаснику тестування індивідуальний набір запитань, що охоплюють той самий предмет або область знань.

Оцінювання ШІ широко використовується в сучасних системах управління навчанням. Наприклад, ReaLMS від VARTEQ має вбудований модуль оцінки знань для проведення тестів, оцінювання та оцінювання навичок і знань студента без необхідної фізичної присутності.

Переваги ШІ в оцінюванні знань:

Інтеграція штучного інтелекту в хімію дає ряд переваг, починаючи від оптимізації щоденного робочого процесу і закінчуючи розблокуванням важливих ідей, які прискорюють сферу.

1. Точність

Моделі ШІ пропонують високу точність у прогнозуванні молекулярних властивостей, таких як розчинність, токсичність і стабільність. Ця точність зменшує експериментальні помилки та забезпечує краще прийняття рішень. Їхній досвід інтерпретації складних даних також дозволяє точніше ідентифікувати сполуки та їхні структурні особливості, залишаючи менше поля для помилок, хоча іноді спостерігаються «збої».

2. Ефективність

Часто вважають, що штучний інтелект і машинне навчання йдуть рука в руку з робототехнікою, і цей зв'язок може допомогти пояснити переваги використання ШІ для автоматизованої лабораторної роботи. Прості або

повторювані завдання можна легко виконати за допомогою моделей машинного навчання, що звільняє час для хіміків. Крім того, ШІ може аналізувати складні та великі набори даних набагато швидше, ніж люди, надаючи відповіді швидше, ніж будь-коли, і забезпечуючи ефективний робочий процес.

3. Прискорення

Ефективність сприяє швидшому шляху до синтезу і, отже, швидшому шляху до ринку, але моделі ШІ можуть прискорити дослідницький процес різними способами. Його передбачувані властивості допомагають керувати синтезом зі швидкістю, нечуваною для ручної хімії, зменшуючи потребу в експериментах методом проб і помилок, які забирають дорогоцінний лабораторний час.

У хімії ШІ діє як примножувач сили, який підвищує можливості дослідників. Це сприяє майбутньому, де науковий прогрес відзначатиметься не лише глибиною знань, а й ефективністю та точністю, з якою ці знання набуваються та застосовуються[15].

2.3. Проведення занять за допомогою ШІ.

ШІ став цінним ресурсом для викладання та вивчення хімії. ChatGPT, інструмент штучного інтелекту, використовувався для підтримки викладання та навчання шляхом створення рішень для проблем з хімії. Стратегії глибокого навчання, такі як резюме, аналогія, дедуктивне міркування та численні представлення, застосовувалися для ефективного навчання хімічним концепціям. Інститут хімічної грамотності через комп'ютерну науку (ICLCS) доповнив навчальні програми з хімії для середньої школи обчислювальними моделями та симуляціями, покращуючи знання змісту та сприяючи дослідженню в класі. ШІ може підвищити продуктивність, персоналізувати навчання та оптимізувати адміністративні завдання, даючи вчителям більше часу та свободи адаптуватися до індивідуальних потреб учнів. Хімія вважається серйозним викликом для штучного інтелекту, і дослідження в

галузі штучного інтелекту можуть отримати значну користь, якщо зосередитися на хімії як еталоні та області застосування в реальному світі.

ШІ має багато застосувань у викладанні хімії, пропонуючи інноваційні рішення традиційних завдань. Одним із таких застосувань є інтелектуальні системи навчання. Ці системи використовують алгоритми штучного інтелекту, щоб надавати учням інструкції та підтримку в реальному часі під час проходження уроків хімії. Вони можуть відповідати на запитання, надавати пояснення та пропонувати покрокову допомогу у вирішенні проблем, доповнюючи роль викладача та сприяючи самостійному навчанню.

Крім того, ШІ можна використовувати для створення інтелектуальних освітніх ресурсів, таких як віртуальні підручники з хімії та інтерактивні навчальні модулі. Ці ресурси можна адаптувати до індивідуальних потреб студентів, представляючи інформацію в персоналізованій та привабливій формі. Алгоритми штучного інтелекту можуть аналізувати відповіді студентів і відповідним чином коригувати зміст, гарантуючи, що студенти отримують найбільш відповідні та ефективні навчальні матеріали.

Підсумовуючи, роль штучного інтелекту в сучасній освіті, зокрема в галузі хімії, є надзвичайно важливою. Його потенціал для персоналізації досвіду навчання, підвищення залучення студентів і надання інноваційних рішень для навчання робить його цінним інструментом як для викладачів, так і для учнів. У наступних розділах ми дослідимо, як штучний інтелект трансформує дослідження хімічних елементів і сполук, а також переваги та проблеми, пов'язані з його впровадженням.

2.4. ШІ у позакласній роботі.

Вирішальну роль у загальному розвитку учнів відіграє позакласна робота. Ці заходи, які проводяться поза звичайною навчальною програмою, надають студентам можливість досліджувати свої інтереси, розвивати нові навички та налагоджувати соціальні зв'язки. Позакласні заходи часто розглядаються як доповнення до академічного навчання. Вони пропонують всебічну освіту, надаючи студентам досвід, який виходить за рамки підручників і аудиторій.

Участь у позакласних заходах допомагає учням розвивати важливі життєві навички, такі як лідерство, робота в команді, планування часу та спілкування. Крім того, ці заходи дозволяють студентам досліджувати свої пристрасті та інтереси, сприяючи особистому зростанню та самопізнанню, особливо у вивченні хімії.

ШІ «відчуває» свою присутність у різних позашкільних сферах, революціонізуючи спосіб проведення та досвіду діяльності. Інтеграція штучного інтелекту в позакласну діяльність відкриває світ можливостей і потенціалу для майбутнього. Переваги та можливості, які надає ШІ:

1. Розширений доступ і інклюзивність: позакласні заходи на основі ШІ можуть подолати географічні бар'єри та фінансові обмеження. Віртуальні платформи на основі ШІ дозволяють студентам з різних місць брати участь і співпрацювати, забезпечуючи рівні можливості для всіх.

2. Персоналізовані відгуки та вдосконалення: системи штучного інтелекту можуть надавати студентам миттєвий персоналізований зворотний зв'язок, допомагаючи їм визначати сфери для вдосконалення та пропонуючи цільові пропозиції щодо розвитку. Цей персоналізований зворотний зв'язок сприяє постійному навчанню та розвитку.

3. Поглиблений розвиток навичок: ШІ може забезпечувати вдосконалене моделювання та віртуальні середовища, які пропонують реалістичні та захоплюючі враження. Це дозволяє учням розвивати та вдосконалювати складні навички в контрольованій та безпечній обстановці, прискорюючи їх навчання та опанування.

Виклики та занепокоєння штучного інтелекту в позакласній діяльності:

1. Етичні міркування: оскільки штучний інтелект стає все більш інтегрованим у позакласну діяльність, можуть виникнути етичні проблеми.

Питання, пов'язані з конфіденційністю, безпекою даних, алгоритмічним упередженням і прозорістю, необхідно вирішити, щоб забезпечити відповідальне та етичне використання технологій ШІ.

2. Взаємодія з людьми та соціальні навички. Незважаючи на те, що штучний інтелект може покращити та полегшити певні аспекти позакласної діяльності, надзвичайно важливо знайти баланс між досвідом, керованим штучним інтелектом, і взаємодією людей. Збереження можливостей для особистої співпраці, командної роботи та розвитку соціальних навичок залишається важливим.

3. Надмірна залежність від штучного інтелекту: існує ризик надмірної залежності від систем штучного інтелекту, що потенційно може зменшити роль наставників, тренерів і вчителів у позакласній діяльності. Важливо знайти правильний баланс між інтеграцією штучного інтелекту та керівництвом людини, щоб забезпечити цілісний і всебічний досвід.

Потрібні освітні реформи:

1. Адаптація навчальної програми: системи освіти повинні розглянути питання про інтеграцію тем і навичок, пов'язаних зі штучним інтелектом, у навчальну програму. Це включає навчання студентів етиці ШІ, грамотності даних, алгоритмічному мисленню та відповідальному використанню технологій ШІ.

2. Підготовка вчителів і професійний розвиток: Педагоги повинні отримати відповідну підготовку та можливості професійного розвитку, щоб зрозуміти технології штучного інтелекту, їх застосування та вплив на позакласну діяльність. Це дасть змогу вчителям ефективно включати ШІ у свою практику викладання та скеровуватиме студентів у діяльності, інтегрованій з ШІ.

3. Співпраця між освітніми установами та експертами зі штучного інтелекту: спільні зусилля між навчальними закладами та експертами зі штучного інтелекту можуть сприяти розробці інструментів, платформ і ресурсів на основі штучного інтелекту, спеціально розроблених для позакласних заходів. Ця співпраця може гарантувати, що інтеграція ШІ узгоджується з освітніми цілями та підтримує навчання студентів.

Підготовка учнів до позакласних занять з інтегрованим ШІ:

1. Цифрова грамотність і обчислювальне мислення: учні повинні мати навички цифрової грамотності та розуміння обчислювального мислення. Це дозволяє їм орієнтуватися в технологіях штучного інтелекту, критично оцінювати створений штучним інтелектом контент і приймати обґрунтовані рішення.

2. Креативність і здатність до адаптації: заохочення креативності, інновацій і здатності до адаптації серед студентів готує їх до розвитку позакласних заходів, інтегрованих зі штучним інтелектом. Акцент на таких навичках, як розв'язання проблем, критичне мислення та співпраця, допоможе студентам досягти успіху в середовищах, керованих ШІ.

3. Етична обізнаність і відповідальне використання: студентів слід ознайомити з етичними аспектами штучного інтелекту, включаючи конфіденційність, упередженість і справедливість. Прищеплення почуття відповідальності та етичного усвідомлення гарантує, що учні розуміють наслідки своїх дій у позакласних заходах, інтегрованих з ШІ.

РОЗДІЛ 3. ВИКОРИСТАННЯ ШТУЧНОГО ІНТЕЛЕКТУ В СИНТЕЗІ ТА ПОШУКУ НОВИХ РЕЧОВИН

3.1. Автоматизоване проектування лікарських засобів

Автоматизоване проектування ліків передбачає використання штучного інтелекту (ШІ) і обчислювальних методів для прискорення процесу відкриття та розробки ліків. Ось як це зазвичай працює:

1. Збір даних : системи штучного інтелекту збирають величезну кількість біологічних, хімічних і фармакологічних даних із різних джерел, включаючи наукову літературу, бази даних і результати експериментів.
2. Моделі машинного навчання : розширені алгоритми машинного навчання, такі як глибоке навчання, навчаються на цих даних, щоб розпізнавати закономірності та зв'язки між молекулярними структурами та їхньою біологічною активністю. Ці моделі можуть передбачити, як молекула буде взаємодіяти з біологічними мішенями, такими як білки або ферменти.
3. Віртуальний скринінг : алгоритми штучного інтелекту виконують віртуальний скринінг для виявлення потенційних препаратів-кандидатів із великих баз даних хімічних сполук. Цей процес передбачає фільтрування сполук на основі їхніх передбачуваних властивостей і ймовірності зв'язування з цільовою молекулою.
4. Молекулярний дизайн : використовуючи генеративні моделі та методи навчання з підкріпленням, системи ШІ можуть пропонувати нові молекулярні структури з бажаними властивостями. Ці структури оптимізовано за допомогою алгоритмів, які збалансовують такі фактори, як ефективність, вибірковість і безпека.
5. Тестування *in Silico* : після виявлення потенційних препаратів-кандидатів вони проходять тестування *in silico*, щоб змодельовати їх поведінку в біологічних системах. Це допомагає визначити пріоритетність найбільш перспективних кандидатів для подальшої експериментальної перевірки.

6. Експериментальна перевірка : нарешті, відібрані кандидати синтезуються та перевіряються в лабораторних експериментах і доклінічних дослідженнях для оцінки їх ефективності, безпеки та фармакокінетичних властивостей. Після цього успішні кандидати можуть перейти до клінічних випробувань.

Автоматизоване проектування ліків прискорює процес відкриття ліків, знижує витрати та підвищує ймовірність відкриття нових методів лікування різних захворювань. Однак важливо зазначити, що відкриття ліків за допомогою штучного інтелекту все ще є відносно новою сферою, і вона стикається з такими проблемами, як якість даних, можливість інтерпретації моделі та дозвіл регуляторів.

3.2. Оптимізація каталізаторів і матеріалів.

Сьогодні гетерогенний каталіз сприяє високоенергетичним селективним молекулярним перетворенням для понад 90% хімічних виробничих процесів і становить 20% усіх промислових продуктів.[16,17] Враховуючи важливість гетерогенних каталізаторів у хімічних процесах, розробка раціональних правил дизайну гетерогенних каталізаторів є одна з найбільш фундаментальних цілей у реакційній інженерії та є ключем до розробки майбутніх стійких хімічних технологій. Проте розробка каталізатора є складним процесом, що включає численні взаємодіючі фактори, такі як активні речовини (метал, оксид тощо), промотор, тип носія, метод підготовки та умови попередньої обробки,(рис. 3.1). Будь-які зміни в цих факторах можуть викликати значні зміни активності та селективності каталізатора. Те ж саме стосується робочих умов: різні комбінації змінних реакції, таких як температура, швидкість потоку сировини та склад сировини.[18] Через багатовимірну складність теоретичний дизайн каталізатора має обмежене застосування при дослідженні матеріалів для розробки каталізатора, тоді як метод проб і помилок і повторні експерименти все ще є основними стратегіями. Іншою серйозною проблемою, пов'язаною з обчислювальними підходами, є відокремлення їх метрик оцінки ефективності від реальних

композиції каталізатора, як-от отримання 3D-структури матеріалу, який можна синтезувати, як результат, від введеної швидкості або селективності.



Рисунок 3.1. Різні фактори, залучені до розробки та оптимізації матеріалу каталізатора

З розвитком комп'ютеризованих технологій зберігання та обробки даних величезні обсяги накопичених знань про каталізатори можуть бути використані для аналізу одночасно.[18] Крім того, багатовимірність проблеми проектування каталізатора може бути представлена в машино-читаному форматі для розробки моделей прогнозування. Наприклад, алгоритми ML були навчені на даних, що відповідають показникам ефективності каталізатора, отриманих у результаті високопродуктивних експериментів або обчислень у поєднанні з комбінаторний алгоритм для прогнозування матеріалів каталізатора з найкращими характеристиками. Вони також використовувалися для побудови зворотних функцій, тобто відображення властивостей матеріалів/молекул. Однак точність прогнозів цих моделей обмежена розподілом навчального набору, зокрема каталітичними структурами, варіаціями складу та зв'язуючими молекулами. З іншого боку,

представлення на основі графів у поєднанні з нейронними мережами використовуються для більш узагальненої реалізації, наприклад прогнозування енергетики на поверхні каталізатора для подальшого мікрокінетичного моделювання.[19] Однак їхні вимоги до даних можуть бути на два порядки більшими, ніж інші простіші алгоритми регресії.[20]

Розробка узагальненої схеми для гетерогенних газозфазних реакцій, яка є обчислювально ефективною, точною та доступною (дійсною для різних реакційних систем і композицій каталізаторів), не покладаючись на методи, що потребують великої кількості даних, є однією з головних цілей досліджень каталізу. У цьому огляді обговорюємо впровадження популярних теоретичних інструментів, підходів і робочих процесів, які використовуються в дослідженні матеріалів для скринінгу та оптимізації каталізатора за перерахованих обмежень. Порівнюємо їхню застосовність для різних проблемних систем, підкреслюючи останні досягнення, зроблені для того, щоб обійти поточні обчислювальні вузькі місця в дизайні каталізатора. Починаємо з інструментів, які використовуються для обчислення енергетики реакції на поверхнях каталізаторів. Ці інструменти включають методи перших принципів, їх напівемпіричні/емпіричні адаптації та молекулярну динаміку. Порівнюємо різні методи, описані в літературі, для розробки реакційних мереж, оптимізації каталізатора та скринінгу. Нарешті, підсумовуємо сучасний стан теоретичного дизайну каталізатора та обговорюємо його потенційний майбутній ландшафт на основі наявних розробок.

Реакційна мережа та відповідна енергетика, що спостерігається на поверхні каталізатора, використовуються для побудови кінетичних моделей для оцінки продуктивності каталізатора та їх подальшої оптимізації. Для побудови кінетичної моделі існують підходи зверху вниз і знизу вгору. Низхідна кінетична модель включає підходи, засновані на виразах степеневого закону та механізмі Ленгмюра–Хіншелвуда–Хогена–Ватсона (LHHW). Припущення в цих підходах більш неявні. Наприклад, припускаючи певні визначальні швидкість елементарних реакцій, квазірівноважні елементарні

реакції та найпоширеніші поверхневі проміжні продукти. Підходи «знизу вгору» — це теоретичні моделі, які не роблять таких початкових припущень щодо природи визначальних швидкостей елементарних реакцій, але можуть бути спроектованими для забезпечення різних спрощень відповідно до вимог користувача.

3.3. Генерація нових молекул за допомогою глибокого навчання

Генерація та оцінка ідей для нових молекул на основі даних є критично важливими завданнями при розробці нових ліків. На початку відкриття ліків ідеї та оцінка були здебільшого процесами, керованими вручну та інтуїцією. Вчені протестували набір сполук, дослідили дані, отримані в результаті тестів, і використали розуміння основної науки, щоб визначити, які молекули синтезувати наступними. Розвиток кількісних зв'язків структура-активність (QSAR) у 1930-х роках створив математичну основу для зв'язку властивостей, розрахованих на основі хімічних структур, із різними фізичними чи біологічними реакціями. [21] Ці прогностичні моделі QSAR лягли в основу багатьох віртуальних стратегій скринінгу, в яких великі набори молекул-кандидатів профілюються, щоб вибрати менший набір, який буде синтезовано або придбано. Щоб розширити уяву хіміків-медиків, ряд груп створили комп'ютерні програми для виконання молекулярного дизайну заново та створення ідей для нових молекул.[22] Ці програми зазвичай працюють шляхом «вирощування» існуючих молекул шляхом додавання атомів або функціональних груп у контексті сайту зв'язування білка. Потім для оцінки цих молекул і визначення пріоритетності сполук для синтезу використовували різноманітні функції підрахунку балів. Незважаючи на те, що ці підходи дизайну мали певний успіх, вони не були широко прийняті.

За останні кілька років галузі генерації молекул і прогнозування властивостей пережили відродження, що було спричинене розвитком глибокого навчання. [23] Досягнення комп'ютерного обладнання та програмного забезпечення, які вплинули на такі галузі, як комп'ютерне

бачення та мовний переклад, тепер застосовуються для прогнозування молекулярних властивостей, включаючи біоактивність, ADME та взаємодії з мішенями токсичності, такими як hERG. Замість того, щоб покладатися на набори визначених експертами молекулярних характеристик, глибоке навчання дозволило дослідникам розробити навчені представлення, які можна налаштувати для конкретних завдань. Ці уявлення в поєднанні зі здатністю глибоких нейронних мереж фіксувати дуже складні кореляції між хімічною структурою та біоактивністю дозволили цим моделям перевершити більш традиційні підходи QSAR. Глибоке навчання також змінило сферу генерації молекул. Замість того, щоб розширювати існуючі молекули або використовувати набір правил для зв'язування існуючих молекулярних фрагментів, методи глибокого навчання генерують молекули на основі мотивів, отриманих з існуючих наборів молекул. Ці мотиви можна дізнатися з великих баз даних молекул, схожих на наркотики, або менших наборів молекул, які, як відомо, мають певну біологічну активність. Поєднуючи ці генеративні підходи з прогнозними моделями, дослідники можуть створювати молекули, які відповідають певним профілям біоактивності.

Хоча галузь просунулася вперед, вона все ще стикається з низкою серйозних проблем, які спонукають поточні дослідницькі зусилля. Хоча дослідники розробляють евристику, наразі нам бракує конкретних вказівок щодо фундаментальних аспектів, таких як кількість молекул, необхідних для розробки вивченого представлення, і обсяг хімічного простору, охопленого моделлю. Оскільки генеративні моделі стають все більш поширеними, потрібно розробити вказівки для оцінки значущості цих «відкриттів» і порівняння їх із більш традиційними методами, керованими людьми. Можливо, найбільш серйозною проблемою, з якою стикається ця сфера, є брак якісних даних, доступних для створення та тестування прогнозних моделей. Щоб справді досягти прогресу, потрібно створити можливості для співпраці між тими у фармацевтичній промисловості, які генерують дані, і тими в академічних колах, які створюють нові методології.

Перспективи віртуального скринінгу спонукали десятиліття досліджень методів машинного навчання (ML) для прогнозування молекулярних властивостей, починаючи від біоактивності до біорозподілу та фізичних властивостей. У разі успіху ці методи можуть змінити процес відкриття шляхом скорочення часу та витрат, пов'язаних з експериментальним скринінгом, одночасно розширюючи простір хімії, який можна досліджувати. У той час як поточні підходи експериментального скринінгу можуть потенційно отримати доступ до мільйонів молекул, віртуальний скринінг може дозволити оцінити мільярди молекул за короткий період часу. Алгоритми ML, які використовуються у віртуальному скринінгу, спрямовані на вивчення зв'язку між конкретними молекулярними субструктурами та цільовими властивостями, подібно до того, як хіміки-медики аналізують молекули. Припускається, що за наявності великої кількості навчальних даних у формі молекул із виміряними властивостями ці алгоритми можуть систематично фіксувати складні закономірності, які можуть бути надто тонкими для сприйняття людьми. На рисунку 3.2. представлено огляд підходів машинного навчання для прогнозування властивостей молекул. Навчальний набір молекул із пов'язаними даними використовується для створення моделі, яку можна використовувати для прогнозування набору експериментальних спостережуваних.

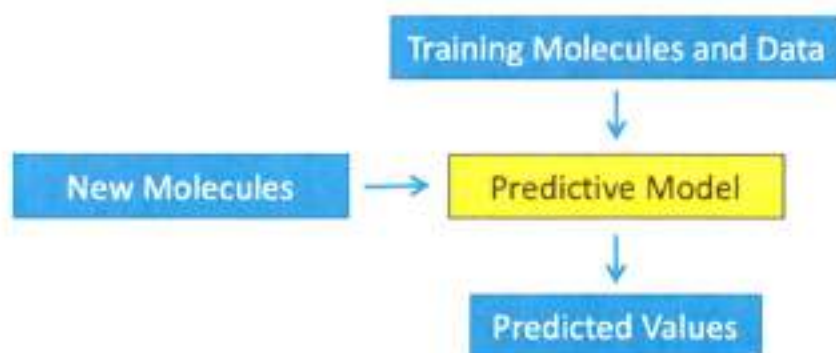


Рисунок 3.2. Схематичний огляд процесу прогнозного моделювання.

Удосконалення нейронних моделей відкрило можливість змінити цей рівень техніки. Одна з ключових відмінностей цих моделей стосується того, як

молекули представлені в алгоритмі машинного навчання. Попередні моделі представляли молекули за допомогою фіксованих експертно-скомпільованих функцій, таких як відбитки пальців і дескриптори.[24,25] Роль алгоритму ML полягала в тому, щоб навчитися зважувати ці характеристики для надання точних прогнозів. [26,27] Очевидно, що успіх цих моделей у вирішальній мірі залежить від якості вибраних характеристик, які можуть бути не оптимальними для конкретного завдання прогнозного моделювання. Навпаки, нейронні моделі вивчають власні експертні представлення ознак безпосередньо з даних, відображаючи молекулярний графік у щільний безперервний вектор. Одним із поширених механізмів такого перекладу є згортка графа,[28] техніка, яка запозичена з попередніх робіт з обробки зображень. Хоча координати цього вектора зазвичай не вловлюють характеристики, які можна інтерпретувати людиною, було показано, що вони дуже гнучкі та здатні фіксувати складні взаємозв'язки за наявності достатньої кількості даних.

Недавні публікації показали ряд випадків, коли ці нові архітектури забезпечують підвищення продуктивності в порівнянні з ненеуронними моделями. Одним з перших прикладів успіху глибокого навчання у відкритті ліків був Tox21 Challenge, де групи використовували різні підходи машинного навчання для прогнозування результатів експериментальних аналізів токсичності. Заявка групи Hochreiter, яка використовувала глибоке навчання, виграла дев'ять із 15 завдань.[29] У спільному дослідженні між Стенфордом і Merck, Feinberg et al. повідомили про середнє збільшення R² на 0,16 для моделей, побудованих на різноманітних наборах фармацевтичних даних.[30] В іншому дослідженні, проведеному MIT і консорціумом фармацевтичних партнерів, повідомлено про нейронну модель, яка дала результати, які були порівнювані або кращі за ті, що були отримані за допомогою базової моделі випадкового лісу на 11 з 19 публічних тестів і 15 з 16 власних наборів даних.

На додаток до ретроспективних тестів, нейронні моделі застосовуються перспективно. Стаття 2020 року Stokes et al. описує відкриття потужних

антибіотиків шляхом віртуального скринінгу понад 100 мільйонів молекул із бази даних ZINC. Процедура, використана для створення цієї моделі, проілюстрована на рисунку 3.3. Набір із 2335 лікарських засобів і натуральних продуктів, що продаються, перевіряли на їх здатність пригнічувати ріст *Escherichia coli*. 120 молекул, які демонструють щонайменше 80% інгібування росту, були позначені як активні (1), а решта молекул були позначені як неактивні (0). Потім була створена модель глибокого навчання та використана для вибору молекул із бази даних із 107 мільйонів комерційно доступних сполук. Подальше тестування вибраних молекул дало вісім антибіотиків, які структурно відрізнялися від молекул, використаних для навчання моделі.

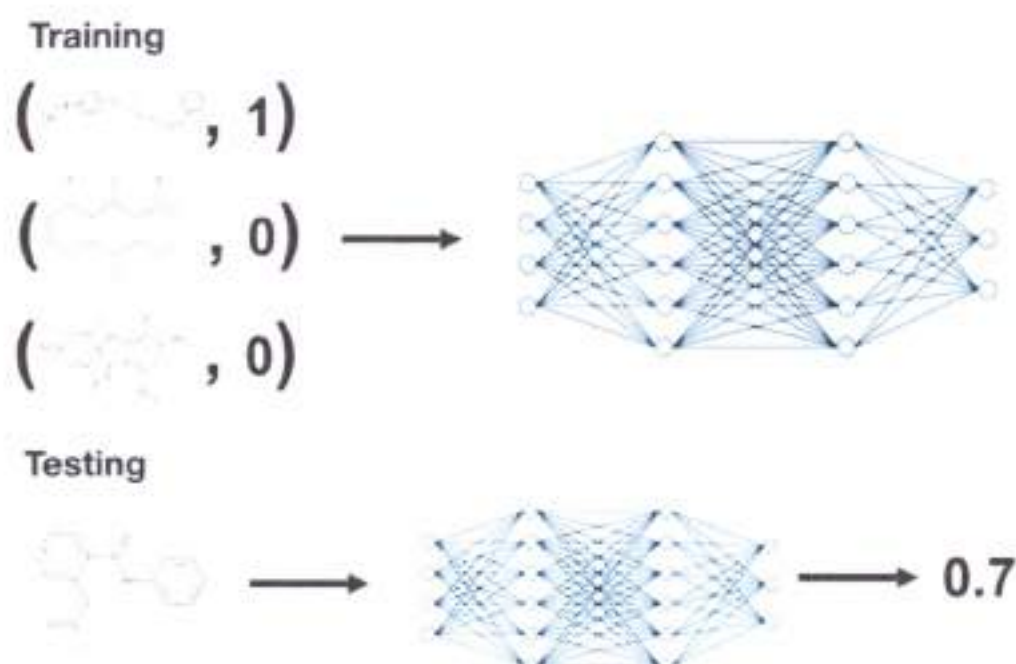


Рисунок 3.3. Процес, використаний Stokes et al. Для відкриття нового антибіотику

Протягом останніх кількох років ми спостерігали появу низки нових підходів до молекулярного дизайну *de novo*, які були натхненні методами, що виникли в таких областях, як комп'ютерне бачення та обробка природної мови. Ця тема викликала значний інтерес у спільноті машинного навчання, створивши широкий спектр нейронних архітектур, які продовжують підвищувати сучасну продуктивність. Насправді, з 2017 по 2019 рік на деяких

наборах даних абсолютний приріст сягнув понад 20% (що вимірюється успіхом у створенні молекул із бажаним профілем властивостей). [29]

Перший із цих методів, відомий як автокодер (рис. 3.4), використовує два взаємодоповнюючих завдання: навчитися безперервно кодувати молекули, що полегшує прогнозування й оптимізацію їхніх властивостей (кодування), і навчитися відображати оптимізоване безперервне представлення назад у молекулярний граф з покращеними властивостями (декодування).[31] Процедура проілюстрована на рисунку 3.5, де переходимо та декодуємо неперервне представлення для створення нових молекул. В альтернативному підході для створення нових молекул можна використовувати інший тип глибокої нейронної мережі, відомий як рекурентна нейронна мережа (RNN).[32] RNN призначений для вивчення послідовних даних, таких як текст або мова. Коли RNN навчається на корпусі відомих молекул, мережа може дізнатися, які частини молекул мають тенденцію до посднання. Після навчання мережа матиме ймовірність з'єднання атомів або функціональних груп. Ці ймовірності змінюватимуться залежно від частин молекули, які мережа вже бачила. Розглядаючи процес генерації молекул як послідовність кроків і відбираючи мережу на кожному кроці, можна генерувати молекули, які мають високу ймовірність бути дійсними та мати певну схожість з навчальними молекулами.

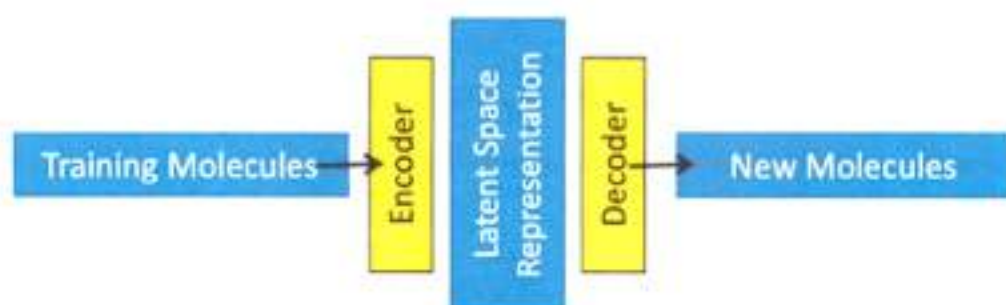


Рисунок 3.4. Схематичний огляд архітектури автокодувальника для генеративного моделювання

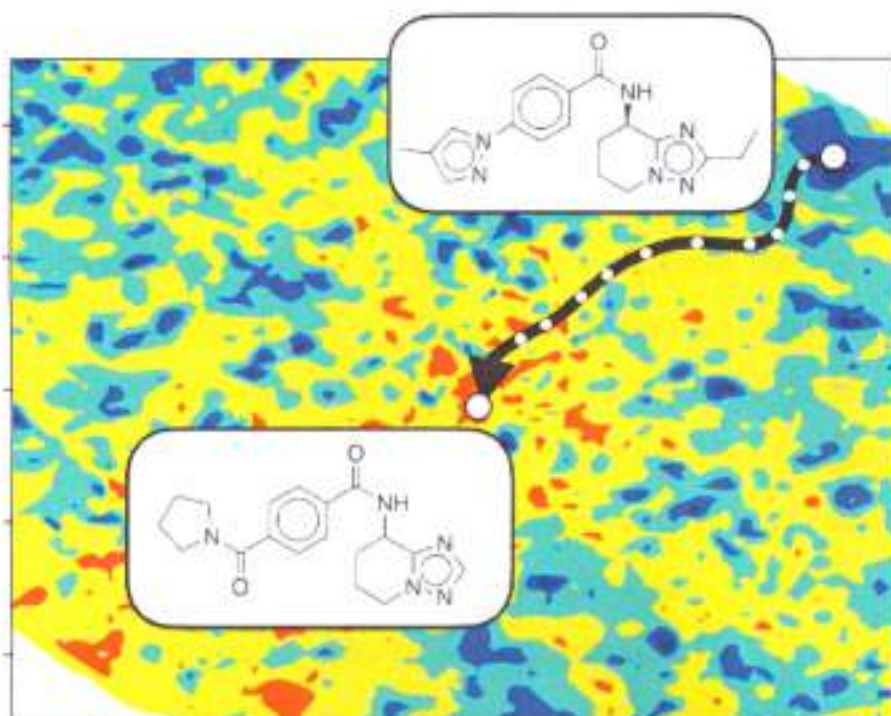


Рисунок 3.5. Використання кодера, безперервного представлення та декодера в генеративному дизайні нових молекул.

У третьому підході, відомому як навчання з підкріпленням (RL), мережа вивчає ряд кроків, необхідних для досягнення конкретної мети.[33-35] Можливо, найвідомішим прикладом RL є AlphaGo, програма, розроблена Google, яка перемогла гросмейстера у складній стратегічній грі, відомій як Go.[36]

У застосуванні RL до генерації молекул визначається серія «ходів», які дозволяють молекулі рости або змінюватися шляхом додавання та модифікації атомів і зв'язків. Потім ціль зазвичай визначається як профіль властивості на основі прогнозу моделі. Потім можна використовувати різні стратегії, щоб визначити шлях до мети.

Перші застосування генеративних моделей для молекулярного дизайну використовували рядок спрощеної молекулярної вхідної системи (SMILES), текстове представлення молекулярної структури, яке, у свою чергу, було переведено в безперервний вектор, а потім додатково оптимізовано відповідно до бажаної властивості. Потім новий оптимізований вектор був декодований у рядок SMILES, що представляє вихідну молекулу. Ключове припущення в

цьому підході полягає в тому, що оптимізація в цьому латентному молекулярному просторі легша через його гладкість, на відміну від дискретного молекулярного простору. Однак на практиці дослідження показали, що сконструйований латентний простір далеко не гладкий: молекули зі схожими структурами та властивостями не обов'язково можуть бути близькими одна до одної в сконструйованому просторі. Частково це можна пояснити впливом представлення SMILES як відправної точки — подібні молекули можуть мати дуже різні рядки SMILES. Однак існують інші сприяючі фактори, такі як обриви властивостей (де молекули з дуже подібними структурами демонструють дуже різні властивості) і обчислювальні артефакти процесу конструювання.

Останні дослідження значно вдосконалили сучасний рівень молекулярних генеративних моделей.[37-41] Виходячи за межі уявлень SMILES, дослідники розробили алгоритми кодування та декодування, які можуть ефективно обробляти молекулярні графи. Молекули фіксуються в режимі з різною роздільною здатністю, чітко фіксуючи функціональні групи, їх взаємне розташування та загальний каркас. Остання робота ще більше доповнює молекулярне представлення 3D-інформацією. Остання робота досліджувала альтернативні алгоритми пошуку оптимізованої молекули. Одним із таких підходів є навчання моделі на заданих парах вихідної молекули та її аналога з покращеними властивостями. Цей підхід можна розглядати як нейронну версію аналізу узгоджених молекулярних пар (ММРА), де модель вивчає траєкторії в прихованому просторі, що призводить до покращення властивостей. Ця формула забезпечує надійну продуктивність, навіть якщо прихований простір нерівний, а також підтримує створення різноманітних результатів.

Хоча було опубліковано багато статей, що описують генеративні моделі, дуже мало з цих публікацій повідомляють про перспективний синтез молекул на основі генеративних конструкцій. Одну з перших статей, в якій описується синтез і подальше біологічне тестування молекул, розроблених за допомогою

генеративного алгоритму, опублікували в 2018 році Мерк і його співробітники, які використовували генеративний алгоритм для успішного створення агоністів ретиноїдних рецепторів X (RXR) або пероксисомного проліфератора. активовані рецептори (PPAR). У цій роботі автори побудували кодування хімічного простору на основі 50 000 молекул, схожих на ліки, і використали літературні дані про 25 агоністів RXR і PPAR для точного налаштування генеративної моделі. Модель QSAR була використана для визначення пріоритетності синтезу п'яти молекул, запропонованих програмою. Подальші біологічні аналізи показали, що дві сполуки були подвійними інгібіторами PPAR і RXR зі значеннями EC50 від 60 нМ до 13 мкМ. Наступна стаття Zhavoronkov повідомили про генеративний дизайн інгібіторів DDR1, мішені для відкриття ліків, причетних до фіброзу. Ця стаття отримала значну увагу в популярній пресі та призвела до повідомлень про «розробку препарату за 21 день». Стаття була визнана помітною, оскільки на додаток до тестування молекул, отриманих з їх генеративних моделей у біохімічному аналізі, автори продемонстрували, що одна молекула була активною в клітинах, і продемонстрували прийнятну фармакокінетику на моделі миші. Хоча спочатку ці результати здавалися багатообіцяючими, у відповіді на цю статтю Уолтерс і Мурко було зазначено, що молекула, виділена в статті, мала разючу схожість з низкою раніше опублікованих інгібіторів DDR1, включаючи проданий на ринку препарат понатиніб (рис. 3.6). Ця суперечка підкреслює труднощі в оцінці значущості молекул, «винайдених» генеративними методами.

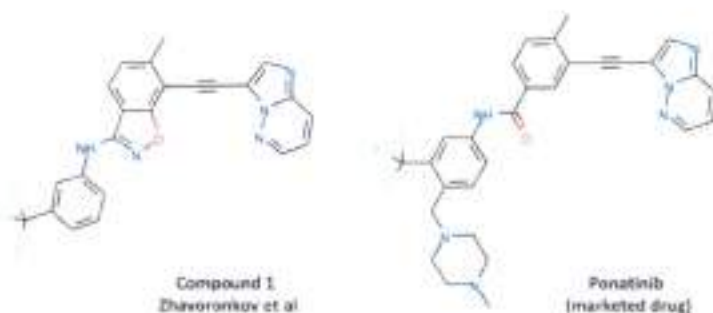


Рисунок 3.6. Порівняння Сполуки 1, створеної за допомогою генеративної моделі, і проданого препарату понатінібу.

Кілька останніх робіт запропонували методи оцінки молекул, створених генеративними моделями. Ці контрольні показники зосереджені на здатності генеративних моделей досліджувати хімічний простір і зосереджувати пошуки та використовувати відому інформацію. Тест GuacaMol, опублікований Брауном і його колегами, оцінює дослідження за допомогою серії тестів, призначених для оцінки дійсності, унікальності та різноманітності молекул, згенерованих алгоритмом. Програмне забезпечення також має цілеспрямовані методи, які оцінюють здатність генеративної моделі зосереджуватися на конкретних областях хімічного простору. Альтернативним підходом для оцінки різноманітності молекул, створених за допомогою генеративної моделі, є порівняння згенерованих молекул з молекулами, отриманими шляхом вичерпного перерахування хімічного простору. Нещодавня стаття Чжана та його співробітників порівнює функціональні можливості генерованих молекул із тими, що містяться в базі даних GDB-13, що містить вичерпний перелік органічних молекул до 13 атомів.

Інше відкрите питання для галузі полягає в тому, де застосовувати генеративні моделі в процесі відкриття ліків. Враховуючи здатність цих моделей досліджувати великі області хімічного простору, можна уявити застосування для раннього відкриття ліків. Однак навіть із досягненнями в оцінці здатності до синтезу і використання машинного навчання в плануванні синтетичних маршрутів синтезу молекул, отриманих за допомогою генеративних моделей, може виявитися складним або трудомістким. Деякі групи, можливо, віддадуть перевагу зосередити свої зусилля з виявлення хітів на комерційно доступних складених бібліотеках, які є недорогими та легко доступними. Пізніше в процесі відкриття традиційні підходи до дизайну в основному ґрунтуються на підрахунку на основі доступних реагентів. Генеративні моделі забезпечують спосіб вийти за рамки доступних синтетичних будівельних блоків. Якщо молекула, створена за допомогою

генеративної моделі, представляє переконливий набір міжмолекулярних взаємодій або потенційно долає відповідальність за ADME, команда з розробки ліків може захотіти виділити ресурси для більш складного синтезу. У міру того як генеративні моделі застосовуються все ширше, ми сподіваємося, що поле стане більш обізнаним про масштаби, обмеження та області застосування методів автоматизованого генерування молекул.

3.4. Пошук нових матеріалів з певними фізико-хімічними властивостями



Рисунок 3.7. GNoME: використання графових мереж для дослідження матеріалів

GNoME використовує два конвеєри для виявлення низькоенергетичних (стабільних) матеріалів. Структурний конвеєр створює кандидатів зі структурами, подібними до відомих кристалів, тоді як композиційний конвеєр дотримується більш рандомізованого підходу на основі хімічних формул. Результати обох конвеєрів оцінюються за допомогою встановлених розрахунків теорії функціоналу щільності, і ці результати додаються до бази даних GNoME, інформуючи наступний раунд активного навчання(рис.3.7.).

GNoME — це найсучасніша модель графової нейронної мережі (GNN). Вхідні дані для GNN мають форму графіка, який можна порівняти зі зв'язками між атомами, що робить GNN особливо придатними для відкриття нових кристалічних матеріалів.

Спочатку GNoME навчався з даними про кристалічні структури та їхню стабільність, відкрито доступними через Materials Project. Ми використовували GNoME для створення нових кристалів-кандидатів, а також для прогнозування їх стабільності. Щоб оцінити прогностичну силу нашої моделі під час поступових циклів навчання, ми неодноразово перевіряли її продуктивність за допомогою усталених обчислювальних методів, відомих як теорія функціоналу густини (DFT), які використовуються у фізиці, хімії та матеріалознавстві для розуміння структури атомів, що важливо для оцінки стабільності кристалів.

Ми використали навчальний процес під назвою «активне навчання», який значно підвищив продуктивність GNoME. GNoME генерував би прогнози для структур нових, стабільних кристалів, які потім тестувалися за допомогою DFT. Отримані високоякісні навчальні дані потім поверталися в нашу модель навчання.

Проект GNoME має на меті знизити витрати на відкриття нових матеріалів. Зовнішні дослідники незалежно створили 736 нових матеріалів GNoME у лабораторії, демонструючи, що передбачення нашої моделі щодо стабільних кристалів точно відображають реальність. Ми надали науковому співтовариству нашу базу даних нещодавно відкритих кристалів. Надаючи вченим повний каталог багатообіцяючих «рецептів» нових матеріалів-кандидатів, ми сподіваємося, що це допоможе їм перевірити та потенційно створити найкращі(рис.3.8).

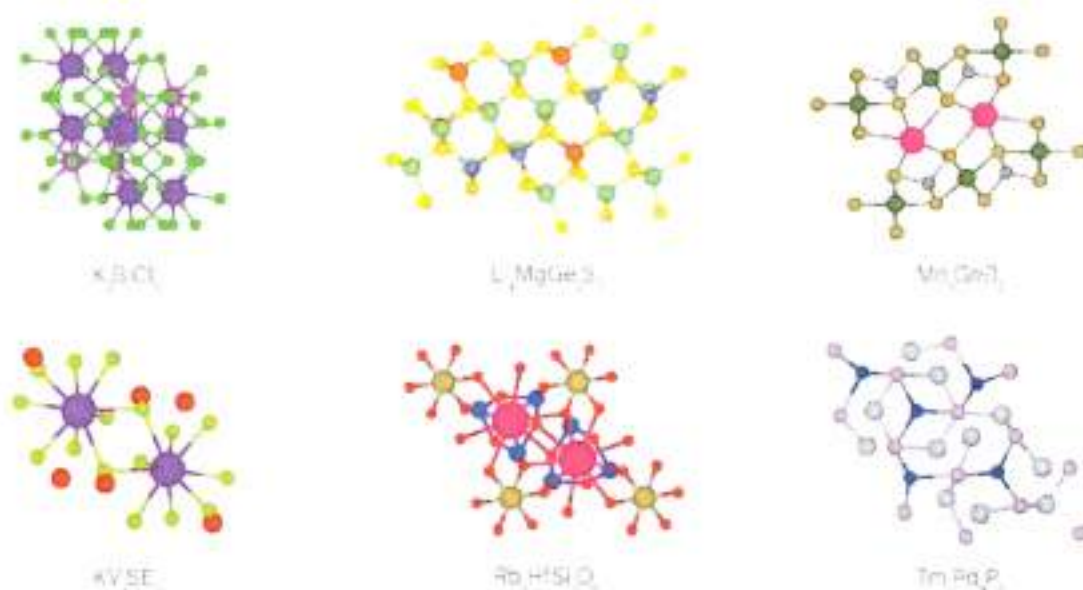


Рисунок 3.8. «Рецепти» ШІ для нових матеріалів

Після завершення наших останніх відкриттів ми провели пошук у науковій літературі та виявили, що 736 наших обчислювальних відкриттів були незалежно реалізовані сторонніми командами по всьому світу. Вище наведено шість прикладів, починаючи від першого у своєму роді лужноземельного алмазоподібного оптичного матеріалу ($\text{Li}_4\text{MgGe}_2\text{S}_7$) до потенційного надпровідника (Mo_5GeB_2).

Швидкий розвиток нових технологій на основі цих кристалів залежатиме від здатності їх виготовляти. У статті, під керівництвом наших співробітників з лабораторії Берклі, дослідники показали, що роботизована лабораторія може швидко виготовляти нові матеріали за допомогою автоматизованих методів синтезу. Використовуючи матеріали проекту «Матеріали» та інформацію про стабільність із GNoME, автономна лабораторія створила нові рецепти кристалічних структур і успішно синтезувала понад 41 новий матеріал, відкриваючи нові можливості для синтезу матеріалів за допомогою ШІ.[42]

3.5. Машинне навчання для прогнозування хімічних властивостей.

Немає штучного інтелекту (ШІ) без даних. Таким чином, доступність даних є необхідною умовою для сучасних програм машинного навчання, де мають значення як розмір, так і якість набору даних. У галузі хімії існує давня традиція збору та компіляції даних, починаючи від атомних спектрів елементів і закінчуючи макроскопічними властивостями матеріалів. Наука про дані в хімії створила тему хімічної інформатики, яка ще більше сприяє застосуванню машинного навчання в хімії. Насправді, хоча створити великий набір даних з нуля може здатися складним, багато хімічних баз даних були доступні задовго до ери машинного навчання. У таблиці 1(дод. А) перераховані вибрані популярні бази даних з хімії, багато з яких мають довгу історію збору та компіляції даних. Джерела цих даних включають відкриті патенти та дослідницькі статті, високопродуктивні експерименти щодо конкретних властивостей і розрахунки QM, як правило, на основі теорії функціоналу густини (DFT).

Бази даних хімічних реакцій мають велике значення для експериментаторів у плануванні синтетичних шляхів і особливо корисні в органічній хімії. До появи Інтернету реакції в літературі вже були проіндексовані Chemical Abstracts Service (CAS). Тепер до цих даних можна отримати доступ із SciFinder, який містить хімічну та бібліографічну інформацію з журналів, патентів, книг та інших джерел. Однак SciFinder разом із аналогічною комерційною базою даних Reaxys не можуть пакетно експортувати великі обсяги даних про хімічні сполуки та хімічні реакції, що обмежує розмір навчальних наборів даних, необхідних для глибокого машинного навчання. З цієї причини дослідники використовують методи обробки тексту, щоб отримати інформацію про реакцію з патентів Бюро патентів і товарних знаків США (USPTO) [43], які є відкритими і доступними для завантаження з Інтернету. Зовсім недавно відкрита база даних реакцій (ORD) [44] створила шаблон формату даних для зберігання хімічних реакцій,

який підтримує обмін даними загальнодоступних наборів даних хімічних реакцій. Слід зазначити, що все більше дослідників у галузі комп'ютерного синтезу тепер роблять свої бази даних загальнодоступними, наприклад, використовуючи програмне забезпечення NextMove [45], яке надає інструменти текстового аналізу з відкритим кодом для ідентифікації хімічних речовин, і діляться своїми наборами даних для завантаження та онлайн-запитів.

Для твердих матеріалів Кембріджська база структурних даних (CSD) [46] є найбільш визнаною; він збирає інформацію про органічну кристалічну структуру з літератури, включаючи дані дифракції рентгенівських променів або нейтронів, умови кристалізації та записи експериментів щодо визначення конформації. База даних неорганічної кристалічної структури (ICSD) [47] містить понад 272 000 кристалічних структур разом із молекулярною формулою, атомними координатами, параметрами комірки, просторовими групами та іншою інформацією, здебільшого визначеною експериментально. База даних Powder Diffraction File (PDF) [48] містить дані дифракції та кристалографії 1 143 236 матеріалів (випуск 2023). Спочатку PDF був колекцією однофазних рентгенограм порошкової дифракції; проте в останні роки він також частково включав записи атомних координат із CSD, ICSD, NIST тощо. База даних MatWeb охоплює широкий спектр інженерних матеріалів, таких як термопластичні та термореактивні полімери, металеві матеріали та керамічні матеріали, записуючи фізичні властивості (наприклад, водопоглинання, питому вагу), механічні властивості (наприклад, модуль пружності), термодинамічні властивості (наприклад, температура плавлення) та електричні властивості (наприклад, дипольний момент, електричний опір). Інші більш специфічні бази даних включають набори даних про старіння літій-іонних акумуляторів [49] для матеріалів літій-іонних акумуляторів від Національного управління з авіації та дослідження космічного простору (НАСА) Центру Еймса Прогностики та набір даних високопродуктивних експериментальних матеріалів (НТЕМ) [50] для неорганічних тонкоплівкових матеріалів. Перший збирає робочі профілі, такі як зарядка, розрядка та

спектроскопія електрохімічного опору матеріалу батареї, тоді як другий містить інформацію про умови синтезу, хімічний склад, кристалічну структуру та характеристики тонкоплівкових матеріалів.

Дерево рішень можна візуалізувати як карту можливих наслідків серії пов'язаних виборів, як показано на рис. 3.9(a), з наслідками, показаними у вигляді кінцевих вузлів (класи A, B і C на рис. 3.9 (a)) і варіанти як вузли в гілках (атрибути; наприклад, $x[2]$ на рис. 3.9 (a)). Щоб навчити дерево рішень, набір даних рекурсивно розбивається за вибраним атрибутом, щоб максимально класифікувати підгрупи з однаковими наслідками [51]. Цей алгоритм широко використовується для класифікації та прогнозування через його переваги, які включають пояснюваність, наявність невеликої кількості гіперпараметрів, низьку вартість обчислення та придатність для відносно невеликих наборів даних (наприклад, 200 вибірок). Однак прогноз може суттєво відрізнятись з незначною зміною даних.

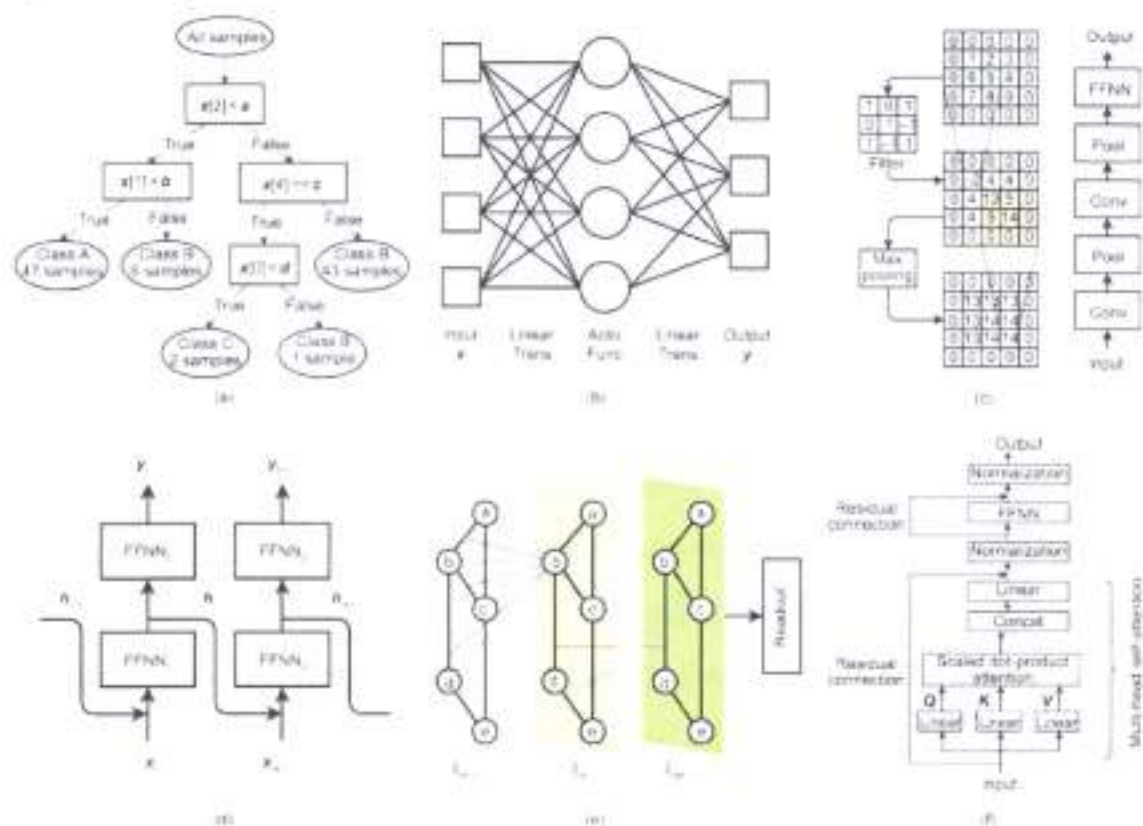


Рисунок 3.9. Шість популярних моделей машинного навчання. (a) Дерево рішень; (b) нейронна мережа прямого зв'язку (Trans: перетворення; Activ Func: функції активації); (c) згортка нейронної мережі (Conv: згортка; Pool: згортка); (d) паралельна нейронна мережа; (e) рекурсивна нейронна мережа; (f) механізм уваги.

об'єднання); (d) рекурентна нейронна мережа; (e) графова нейронна мережа; (f) трансформаторна нейронна мережа.

Кожне дерево навчається на різних піднаборах даних, випадково відібраних із вихідних даних, відомих як початкове агрегування або пакетування. Завдяки сукупності дерев рішень радіочастотна модель досягає підвищеної надійності та, отже, кращої передбачуваності. Такі моделі більше підходять для прогнозування дискретних цільових значень; таким чином, типовим застосуванням є оптимізація експериментальних змінних шляхом кореляції синтетичних умов із селективністю бажаних продуктів.

3.6. Комп'ютерне моделювання молекулярних структур

Молекулярне моделювання — це комп'ютерний засіб представлення, візуалізації та дослідження тривимірних структур і відповідних властивостей молекул. Сучасні біохімічні тексти рясніють дивовижними комп'ютерними зображеннями, що представляють біологічні молекули, тоді як більшість журналів з біологічних наук примудряються мати комп'ютерну графіку на обкладинках.

За допомогою комп'ютерів можна симулювати науково значущі картини молекул, вивчати їхні фізичні та хімічні властивості, наприклад, форму, розмір і заряд; моделювати динамічну поведінку атомів і молекул, наприклад їх коливальні, крутильні та обертальні рухи; досліджувати їх взаємодію з іншими молекулами; раціонально конструювати молекули, що представляють біологічний і клінічний інтерес; і, можливо, найважливіше, значно покращити наукову комунікацію та викладання всіх аспектів біомолекулярних наук.

Загальноновизнано, що людський мозок призначений для отримання візуальної інформації [52], отже, молекулярне моделювання має багато переваг перед традиційними підходами до вивчення структур біомолекул. Реальним фізичним моделям молекул часто бракує візуальної привабливості, вони громіздкі й крихкі, особливо коли модель велика. Для порівняння, згенеровані комп'ютером молекули, як правило, легко побудувати (хоча, звичайно, це певною мірою залежить від розглянутої молекули, доступного

програмного забезпечення та того, що відомо про структуру молекули), і вони привабливі та міцні. Молекули можна «побудувати» за допомогою введення атомних координат, імпортованих із баз даних, або за допомогою хімічних шаблонів, що зберігаються в програмі моделювання, або шляхом створення ескізу двовимірного зображення безпосередньо за допомогою VDU, або, справді, комбінації всіх трьох. Крім того, після створення зображення на дисплеї не мають тенденції до «денатурації» в руках! З іншого боку, програми можуть працювати повільно, особливо на старих типах персональних комп'ютерів.

Молекулярне моделювання можна умовно розділити на молекулярну графіку та обчислювальну хімію. Перше — це візуалізація хімічних структур і молекулярних властивостей, хоча це визначення часто розширюється, щоб включати прості маніпуляції, такі як зміна торсійних кутів хімічних зв'язків і основні геометричні розрахунки, наприклад, оцінка міжатомних відстаней. Обчислювальна хімія передбачає спроби обчислення числових властивостей молекул, найпоширенішими в молекулярних/біохімічних науках є молекулярні енергії. Основними результатами таких розрахунків є великі обсяги числових даних. Зазвичай їх аналізують за допомогою програм молекулярної графіки, тому визначення між молекулярною графікою та обчислювальною хімією стає все більш розмитим.

Молекулярна графіка пропонує кілька способів перегляду молекул, які можна використати з користю під час їх вивчення або дослідження. Глибина затінення та плавний рух структур у реальному часі дають реалістичні тривимірні зображення. Моделі можна переміщати обертанням або трансляцією. Також можна збільшувати/зменшувати певні ділянки конструкції та вирізати секції моделі, щоб отримати чіткий огляд внутрішніх особливостей. Вони також можуть бути представлені в одній із кількох форм, напр. палиця, м'яч і палиця, точкова поверхня або заповнений простір, рівномірно або в комбінаціях для виділення певних особливостей. Можливе моделювання «реалістичних» тривимірних структур і стереозображень (рис.

3.10). Можна повернути увагу до окремих частин моделі, потовщивши лінії, змінивши щільність або колір точок або провівши «стрічку» через цікаву функцію (рис. 3.10). Молекули можна порівнювати структурно, накладаючи моделі за допомогою аналізу найменших квадратів атомних координат; тоді як обертання зв'язків дозволяє досліджувати конформаційний простір, доступний молекулі. Молекулярна графіка може бути використана для відображення великої кількості числових даних, доступних з обчислювальних хімічних вправ, таких як дослідження, або симуляції, такі як ті, що пов'язані з молекулярною механікою, молекулярною динамікою та «стрибками Больцмана» (моделювання Монте-Карло), і дозволяють швидкий аналіз з використанням відповідних графічних зображень.

Це аксіома біології, що функція слідує за структурою. Це так само вірно для біомолекул, як і для будь-якої іншої біологічної сутності, тому великий інтерес до структур біологічних молекул. Біохіміки, які визначають структуру великих біомолекул, як правило, витрачають багато часу та енергії на рентгенівську кристалографію або спектроскопію ядерного магнітного резонансу, і, отже, такі методи, як обчислювальна хімія, які обходять ці зусилля, очевидно, дуже привабливі.

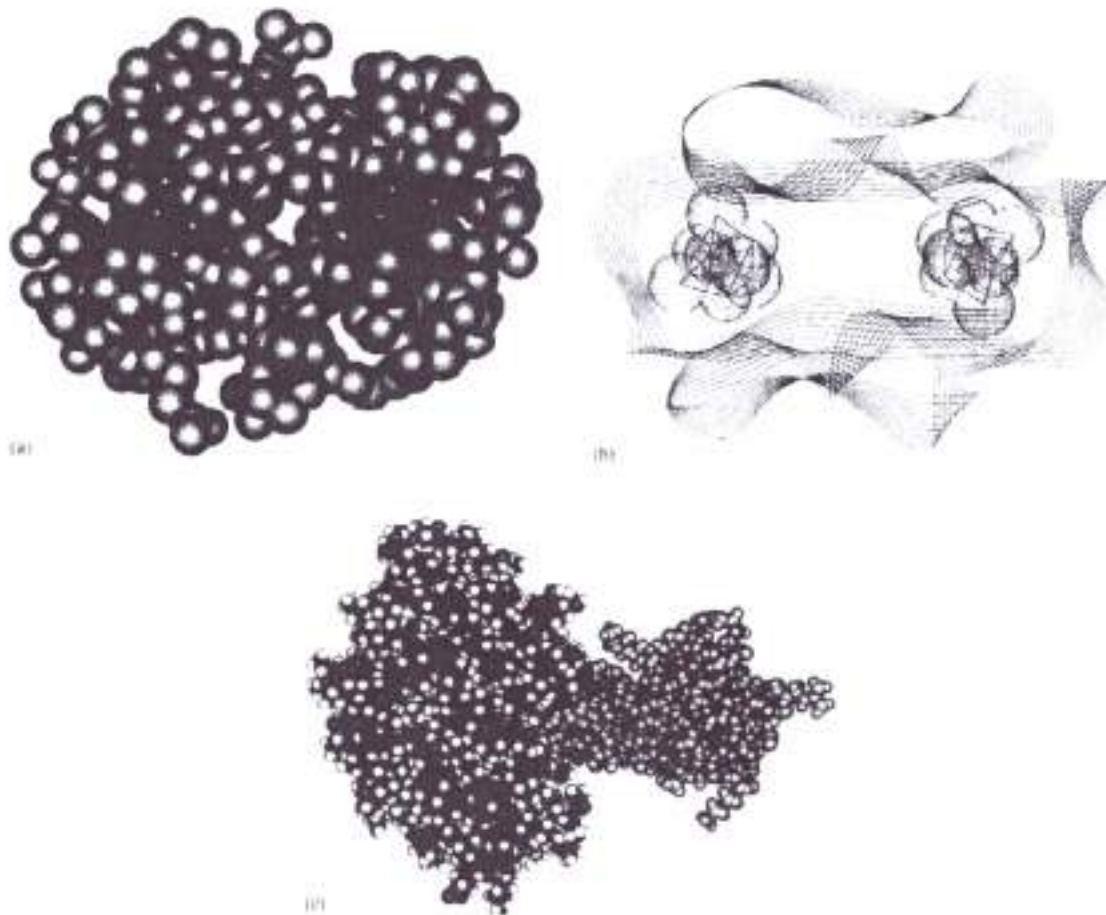


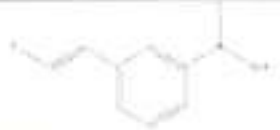
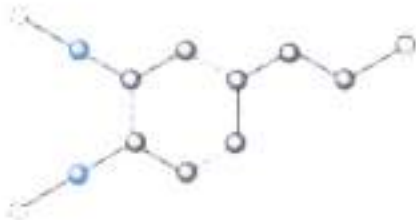


Рисунок 3.10. Показує різні способи представлення молекулярних моделей.

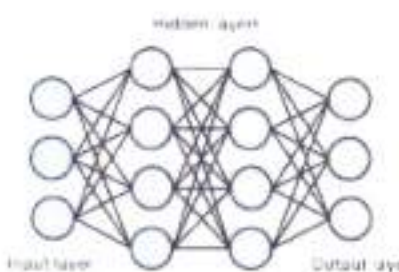

(a) Модель заповнення простору молекули ферредоксину. Атоми водню опущені, а центри переносу електронів залізо-сірка показані чорним кольором для ясності. (b) У цьому зображенні молекули ферредоксину поліпептидний скелет показаний у вигляді стрічки, щоб підкреслити його вторинну структуру, а залізо-сірчані центри — як палички, що підкреслюють їх регулярну структуру, тоді як ван-дер-ваальсові радіуси окремих заліза та сірки атоми показані як пунктирні сфери, щоб вказати загальну форму центрів. (c) Молекулярна модель Р-трипсину (багатотіньовий) у комплексі з інгібітором трипсину підшлункової залози (видобуток). Щоб повністю оцінити ці моделі, потрібен повний колір (недоступний у Biochemical Education). Усі моделі були побудовані з використанням координат, отриманих з бази даних Brookhaven Protein Database.


3.7. Розвиток хімічних реакцій з використанням алгоритмів інтелектуальної обробки даних.

Представлення органічних молекул у машинному навчанні є вирішальним кроком із далекосяжними наслідками. Це безпосередньо впливає на прогноз точність, ефективність обчислень та можливість інтерпретації моделі. Правильно вибране представлення покращує узагальнення, полегшує використання даних і пропонує цінну хімічну інформацію. Загалом цей крок значно впливає на якість і застосовність моделей машинного навчання в хімії, матеріалознавстві та відкритті ліків. Існують різні способи представлення молекул у контексті машинного навчання та обчислювальної хімії. Деякі поширені методи наведено в таблиці 2.

Таблиця 2. Молекулярне представлення в машинному навчанні

Представництво	опис	приклад
SMILES and InChI	Запроваджені в 1980-х роках спрощена система рядкового введення молекулярного введення (SMILES) і міжнародний хімічний ідентифікатор (InChI) були розроблені як стислі та зрозумілі людині способи представлення молекулярних структур. Вони представляють молекули за допомогою компактного рядка символів, які кодують інформацію про атом і зв'язок	 <chem>CC1=CC=C(C=C1)C=C</chem> <chem>CC1=CC=C(C=C1)C(=O)O</chem>
Молекулярні графіки	Молекулярний графік — це графічне зображення молекули, де кожен атом представлений як вузол, а кожен зв'язок між атомами — як ребро, що з'єднує відповідні вузли. Графік відображає зв'язок і структуру молекули, надаючи інформацію про розташування атомів і зв'язки. Це позання широко використовується в хіміоінформатичі та обчислювальній хімії для опису молекулярних структур	 Nodes:  Edges: 

<p>Графічні нейронні мережі</p>	<p>Графові нейронні мережі (GNN) — це клас моделей машинного навчання, розроблених для роботи з графоструктурованими даними, такими як молекулярні графи. GNN — це архітектури нейронних мереж, які працюють безпосередньо з графіками та вчаться обробляти та отримувати інформацію з графоструктурованих даних. GNN складаються з шарів, які оновлюють представлення вузлів шляхом агрегування інформації з сусідніх вузлів, ефективно фіксуючи локальні та глобальні структурні особливості в межах графа. GNN набули популярності в задачах, пов'язаних із графічними даними, включаючи прогнозування властивостей молекул, відкриття ліків і оптимізацію молекулярних властивостей</p>	
<p>Молекулярні відбитки пальців</p>	<p>Молекулярні відбитки представляють молекулу як двійковий або цілочисельний вектор, де кожен елемент у векторі відповідає певній субструктурі або молекулярній властивості. На наявність або відсутність певної підструктури чи властивості вказує значення відповідного елемента у векторі. Молекулярні відбитки фіксують інформацію про наявність певних функціональних груп, тип атомів та інші структурні особливості. Вони корисні для вимірювання подібності між молекулами та для швидкої ідентифікації молекул із подібними субструктурами.</p>	

	Молекулярні відбитки ширші та можуть включати різні типи субструктур і молекулярні властивості, забезпечуючи глобальне уявлення про структуру молекули																			
Кругові відбитки пальців	Кругові відбитки пальців, також відомі як кільцеві відбитки субструктури або кільцеві відбитки Моргана, є особливим типом молекулярних відбитків пальців. Вони отримуються з молекули шляхом розгляду набору атомоцентричних субструктур певного радіуса (зазвичай 2 або 3 зв'язки від центрального атома). Ці підструктури хешуються для створення двійкового або цілого вектора фіксованої довжини, де кожен елемент відповідає хешованій підструктурі. Кругові відбитки пальців фіксують наявність або відсутність специфічних субструктур, зосереджених навколо кожного атома в молекулі, надаючи спосіб представити локальне оточення кожного атома. Кругові відбитки, з іншого боку, зосереджені на локальних субструктурах навколо кожного атома, забезпечуючи більш детальне уявлення про локальне оточення молекули.	 <table border="1" data-bbox="844 759 1242 1031"> <thead> <tr> <th>0</th> <th>1</th> <th>2</th> </tr> </thead> <tbody> <tr> <td>C, ar (Sp²)</td> <td>C, ar (Sp²)</td> <td>C, ar (Sp²)</td> </tr> <tr> <td></td> <td>C, ar (Sp²)</td> <td>C, ar (Sp²)</td> </tr> <tr> <td></td> <td>C, (Sp³)</td> <td>N (Sp³)</td> </tr> <tr> <td></td> <td></td> <td>O (Sp²)</td> </tr> <tr> <td></td> <td></td> <td>O (Sp²)</td> </tr> </tbody> </table>	0	1	2	C, ar (Sp ²)	C, ar (Sp ²)	C, ar (Sp ²)		C, ar (Sp ²)	C, ar (Sp ²)		C, (Sp ³)	N (Sp ³)			O (Sp ²)			O (Sp ²)
0	1	2																		
C, ar (Sp ²)	C, ar (Sp ²)	C, ar (Sp ²)																		
	C, ar (Sp ²)	C, ar (Sp ²)																		
	C, (Sp ³)	N (Sp ³)																		
		O (Sp ²)																		
		O (Sp ²)																		

Крім дескрипторів, наведених у таблиці 2, є деякі конкретні дескриптори, які використовуються для прогнозування реакції та досліджень оптимізації, включаючи дескриптори дракона (геометричні дескриптори, дескриптори на основі 3D матриці, 3D автокореляції, дескриптори RDF, дескриптори 3D MoRSE, дескриптори WHIM, дескриптори GETAWAY,

Randic молекулярні профілі, 3D-пари атомів, дескриптори CATS 3D), дескриптори молекулярного робочого середовища (MOE), дескриптори середньої стеричної зайнятості (ASO), дескриптори розширених відбитків зв'язку (ECFP6) і дескриптори Mol2vec [53].

Відбір дескрипторів у контексті передачі фізико-хімічної інформації у вхідних даних визнано важливим, але складним завданням. Однак процес вибору дескрипторів вимагає глибокого розуміння конкретних установок проблеми, які часто проводяться фахівцями, які добре знаються на даній області. Однак необхідно зазначити, що обчислювальні витрати, пов'язані з обчисленням численних дескрипторів, можуть бути значними, потенційно впливаючи на час, необхідний для висновку. Крім того, валідність дескрипторів може бути обмежена певними регіонами у величезному хімічному просторі, що створює проблему для досягнення універсально застосовного та точного представлення [48]. Наприклад, було запропоновано багато моделей ML для прогнозування розчинності органічних молекул у воді, навчених на основі експериментальних даних. Але через різницю в хімічному просторі охоплених сполук, методів вимірювання, експериментальних умов і нестандартних представлень молекул в існуючих наборах даних розчинності існує велика різниця в прогнозуванні різних моделей. Моделі на основі представлень рядків, 3D-дескрипторів або 2D-дескрипторів показали різний ступінь точності [54].

РОЗДІЛ 4. ЕТИЧНІ АСПЕКТИ ВИКОРИСТАННЯ ШІ НА УРОКАХ ХІМІЇ

4.1. Розгляд можливих етичних питань та впливу на процес навчання.

Інтеграція штучного інтелекту (ШІ) в освітні технології відкриває двері для новаторських досягнень у навчанні та розвитку. Ця технологічна еволюція, хоч і наповнена потенціалом, створює безліч етичних проблем, які вимагають ретельного розгляду. Перш ніж поспішно впроваджувати штучний інтелект в інструменти, які підтримують навчання, дуже важливо зрозуміти не лише їхній вплив на результати навчання, але й етичні рамки, які керують їх роботою. У цій статті досліджуються ці етичні аспекти, зосереджуючись на їхньому впливі на навчання впродовж життя, професійний розвиток і ширший освітній ландшафт. Крім того, це дослідження спонукає нас поставити під сумнів фундаментальні мотиви інтеграції штучного інтелекту в освітні інструменти, гарантуючи, що його впровадження є не просто прагненням до моди, а продуманим кроком до покращення освітньої подорожі.

Прозорість штучного інтелекту має вирішальне значення в освітньому контексті. Учні та викладачі повинні чітко розуміти, як керовані ШІ інструменти приймають рішення та забезпечують освітній контент. Ця прозорість має важливе значення для того, щоб функції, інтегровані в штучний інтелект (загалом на платформах, які підтримують надання цих функцій), відповідали за їхній зміст, таким чином зберігаючи цілісність освітніх програм. Наприклад, коли функція чи навчальний інструмент, у який інтегровано штучний інтелект, рекомендує шлях навчання, і викладач, і учень мають розуміти основу цих рекомендацій. Таке розуміння зміцнює довіру та гарантує ефективне та відповідальне використання цих інструментів.

Системи штучного інтелекту сприйнятливі до упереджень, присутніх у їхніх навчальних даних. У контексті освіти дорослих, де учні мають різне походження, надзвичайно важливо розробити інклюзивні та справедливі інструменти ШІ. Щоб запобігти ненавмисній дискримінації будь-якої групи

чнів, необхідно регулярно переглядати та оновлювати алгоритми ШІ. Наприклад, систему штучного інтелекту, яка використовується для скринінгу ступу, слід ретельно перевірити, щоб переконатися, що вона не надає перевагу певним демографічним показникам над іншими, ненавмисно акріплюючи історичні упередження.

Зростаюче використання ШІ в освітніх технологіях передбачає збір і аналіз величезних обсягів персональних даних. Захист цих даних є важливою етичною проблемою. Надійні заходи конфіденційності даних і безпеки є важливими для підтримки довіри та впевненості всіх учнів. Це включає впровадження надійних методів шифрування, забезпечення анонімності даних, де це можливо, і прозорість щодо користувачів щодо використання та обробки їхніх даних. Крім того, важливим моментом є надання користувачам можливості відмовитися від обміну даними (за можливості). Це надає можливість учням контролювати свою особисту інформацію, зміцнюючи їх автономію та додатково захищаючи їх конфіденційність.

Роль викладачів в освітньому середовищі, керованому штучним інтелектом, також є етичним міркуванням. ШІ може підтримувати викладачів, автоматизуючи адміністративні завдання та надаючи інформацію про успішність студентів. Однак важливо переконатися, що штучний інтелект не замінює людський елемент у викладанні, особливо в освіті дорослих, де досвід, наставництво та особиста взаємодія мають вирішальне значення. Педагоги повинні використовувати штучний інтелект як інструмент для вдосконалення своїх методів навчання, а не як заміну професійного судження та навичок міжособистісного спілкування.

Впровадження штучного інтелекту в інструменти, які підтримують наші освітні процеси, може зробити навчання впродовж життя більш доступним для дорослих. Проте забезпечення того, щоб ці інструменти були доступні для всіх, незалежно від соціально-економічного походження, є серйозною етичною проблемою. Освітні платформи, керовані штучним інтелектом, мають бути зручними та доступними для різноманітних учнів, у тому числі

для людей з обмеженими можливостями. Це означає включення таких функцій, як розпізнавання голосу для користувачів із вадами зору, або спрощення інтерфейсів для тих, хто не розбирається в техніці, гарантуючи, що переваги ШІ в освіті доступні кожному.

Поспішаючи прийняти новітні технологічні тенденції, важливо зупинитися й розглянути фундаментальне запитання: навіщо ми впроваджуємо штучний інтелект у наші інструменти для підтримки освіти? Привабливість штучного інтелекту справді стала значним галасом, що часто призводило до його прийняття більше як модної функції, а не інструменту, орієнтованого на вирішення. Важливо оцінити конкретні проблеми, які штучний інтелект має вирішити в освітньому контексті.

Інтеграція штучного інтелекту не повинна зумовлюватися лише його новизною чи бажанням залишатися в курсі технологічних проблем. Натомість його прийняття має ґрунтуватися на чіткому розумінні потреб і проблем в освітньому секторі. Чи використовуємо ми штучний інтелект, щоб покращити персоналізоване навчання, оптимізувати адміністративні завдання чи забезпечити глибше розуміння успішності студентів? Ефективність ШІ в освіті залежить від його узгодженості з освітніми цілями та реальними потребами учнів і викладачів.

Крім того, важливо оцінити, чи впровадження ШІ приносить відчутні покращення в процес навчання чи просто додає рівень складності без суттєвих переваг. Рішення про інтеграцію штучного інтелекту має бути підкріплено науково-обґрунтованими дослідженнями та ретельним аналізом його потенційного впливу на результати навчання. Цей підхід гарантує, що штучний інтелект є не просто швидкоплинною тенденцією, а значущим доповненням, яке справді збагачує навчальний процес.

4.2. Переваги та недоліки використання ШІ на роках хімії.

ШІ мають багато з нас у виробництві хімікатів. Ці рішення допомагають автоматизувати вилучення даних, планування ланцюга поставок і тестування якості.

Оскільки обчислювальна потужність комп'ютера постійно зростає, вчені використовують фреймворки машинного навчання з відкритим кодом і алгоритми штучного інтелекту, щоб революціонізувати спосіб вирішення завдань і роботи з лікуванням, хімікатами та ліками.

Давайте зосередимося на основних застосуваннях ШІ в хімії:

1. Виявлення молекулярних властивостей:

Алгоритми штучного інтелекту можуть аналізувати хімічні дані для прогнозування та класифікації різних молекулярних властивостей, таких як токсичність, розчинність і реакційна здатність. Це робить процес виявлення швидшим і менш схильним до помилок, на відміну від ручного виявлення. Крім того, ШІ дозволяє вченим оцінити потенціал гіпотетичної молекули.

2. Розробка молекул:

Штучний інтелект може допомогти в розробці нових молекул із бажаними властивостями, створюючи віртуальні бібліотеки сполук і оптимізуючи молекулярні структури за допомогою ітераційних алгоритмів. Використання штучного інтелекту в цій галузі дозволяє робити революційні відкриття в хімічному синтезі.

3. Виявлення препаратів:

Платформи пошуку ліків на основі ШІ можуть аналізувати величезні бази даних хімічних сполук, прогнозувати їхню потенційну активність проти конкретних цілей і визначати пріоритетність кандидатів для подальшого дослідження. Зараз вчені використовують ШІ для розробки нових ефективних ліків для лікування смертельних хвороб.

4. Реакція ретросинтезу:

Хіміки використовують штучний інтелект для планування найбільш ефективних і рентабельних шляхів синтезу, генеруючи шляхи ретросинтезу та

пропонуючи оптимальні етапи реакції. Раніше цей процес проводився вручну, що займало багато часу, сил і грошей.

Зараз лише 4 з 10 хімічних компаній широко впроваджують ШІ у свою діяльність.

Причина, чому цей процес такий повільний, спричинена труднощами впровадження штучного інтелекту, включаючи його вартість, брак навичок штучного інтелекту серед співробітників тощо. Таким чином, перш ніж почати використовувати штучний інтелект у своїй лабораторії або CRO, важливо дізнатися про його проблеми:

1. Нерозвинені технології:

Незважаючи на те, що штучний інтелект широко використовується в таких сферах, як обробка природної мови або розпізнавання зображень, він лише розвивається в галузі хімії та науково-дослідних процесів. Таким чином, його технології все ще недостатньо розроблені, і адаптувати алгоритми штучного інтелекту відповідно до складних лабораторних вимог може бути складно.

2. Дефіцит навичок ШІ у робочій силі:

Щоб використовувати штучний інтелект у хімічній промисловості, потрібні експерти з технологій ШІ. На жаль, дуже складно знайти експертів як у галузі штучного інтелекту, так і в областях, необхідних для CRO, хімічних компаній і відділів досліджень і розробок.

3. Відсутність якісних даних:

Усі алгоритми ШІ покладаються на дані, щоб робити прогнози. У сучасних лабораторіях і CRO дані часто можуть бути неповними або розкиданими з кількох джерел. Крім того, часто бракує стандартизації форматів даних і протоколів, що робить процес використання ШІ ще більш складним.

4. Питання довіри та прозорості:

Хімічні та фармацевтичні галузі суворо регулюються, вимагаючи найвищого рівня прозорості та відстеження. Таким чином, відсутність

інтерпретації моделей ШІ може вплинути на процеси. Важливо працювати над озбудою довіри між регуляторами, науковцями та зацікавленими сторонами.

Незважаючи на труднощі та обмеження, штучний інтелект може поратися з чимало в хімічній промисловості. Можна трансформувати все, від відкриття ліків до матеріалознавства тощо.

Виявлення ліків може бути однією з найважливіших сфер застосування ШІ в хімії. Відкриття ліків завжди було тривалим і складним процесом, який вимагав багато ручних зусиль, грошей і часу. Це може коштувати мільярди доларів, і ви ніколи не можете бути впевнені, що все піде правильно через імовірну людську помилку.

З розвитком технологій штучного інтелекту вчені тепер можуть прогнозувати властивості нових препаратів-кандидатів, визначати потенційні цілі для розробки ліків і оптимізувати дизайн молекул ліків.

Також можна навчити моделі ШІ і ML прогнозувати активність і токсичність нових препаратів-кандидатів. ШІ також використовується для визначення мішеней для розробки ліків, таких як білки, причетні до певного захворювання.

Технології ШІ можна використовувати для розробки нових матеріалів із певними властивостями, такими як провідність, міцність і гнучкість.

Дослідники також можуть використовувати моделі ШІ і ML для прогнозування властивостей нових матеріалів і визначення майбутніх кандидатів для досліджень.

Також можна використовувати моделі глибокого навчання для прогнозування властивостей певних матеріалів, наприклад прогнозування точки плавлення на основі кристалічної структури.

Іншим застосуванням штучного інтелекту є синтез хімічних речовин, головним чином створення нових молекул або матеріалів у лабораторії. Вчені можуть використовувати ШІ для аналізу результатів попередніх

експериментів і прогнозування результатів майбутніх. Крім того, ШІ допомагає оптимізувати умови реакції для зменшення відходів.

Незважаючи на зростаюче використання штучного інтелекту в хімічній галузі та його численні застосування, це все ще відносно нова концепція, яка має низку недоліків і проблем, які можуть вплинути на процес, якщо його недостатньо досліджено та виконано неналежним чином.

1. Незрозумілість:

Алгоритми штучного інтелекту часто працюють як «чорні ящики», тому важко зрозуміти, як вони отримують свої прогнози чи рекомендації. У хімічній промисловості вчені та дослідники повинні чітко розуміти процеси та обґрунтування результатів, створених штучним інтелектом, щоб це могло спричинити серйозні проблеми.

2. Зміщення даних:

Моделі ШІ значною мірою покладаються на якість і репрезентативність навчальних даних. Якщо навчальні дані містять упередження або обмеження, модель ШІ може зберегти ці упередження, що призведе до упереджених прогнозів. Це може призвести до неточних оцінок хімічних властивостей, екологічних проблем і ризиків безпеки.

3. Обмежене узагальнення:

Моделі штучного інтелекту навчаються на конкретних наборах даних, і їм може бути важко добре узагальнити нові або невідомі дані. У хімічній галузі постійно відбуваються сучасні сполуки та реакції, тому важливо мати можливість узагальнювати дані за межами навчання.

4. Високі обчислювальні вимоги:

Навчання та розгортання моделей штучного інтелекту в хімічній галузі часто потребують значних обчислювальних ресурсів, включаючи потужне апаратне забезпечення та велике сховище даних.

5. Етичні проблеми:

Використання штучного інтелекту в хімії викликає етичні проблеми, такі як конфіденційність даних, права інтелектуальної власності та можливість

неправомірного використання отриманих штучним інтелектом знань або сполук.

6. Проблеми інтеграції:

Інтеграція штучного інтелекту в існуючі лабораторні робочі процеси та інфраструктуру може вимагати значних змін і адаптацій.

7. Людський досвід і судження:

Людське судження та критичне мислення все ще важливі для інтерпретації отриманих ШІ результатів і прийняття обґрунтованих рішень у складних сценаріях хімічних досліджень.

Технології штучного інтелекту та машинного навчання в хімічній промисловості лише на початку свого розвитку, і їх чекає процвітаюче майбутнє.

Оскільки все більше хімічних компаній використовують штучний інтелект у своїх процесах, ці технології продовжуватимуть розвиватися, щоб відповідати зростаючим потребам і вимогам цих організацій. Давайте подивимося на майбутні перспективи ШІ в хімії.

1. Прискорене відкриття ліків:

У майбутньому ШІ зробить ще більшу революцію в процесі відкриття ліків, прискоривши ідентифікацію препаратів-кандидатів.

Аналізуючи дані минулого, технологія штучного інтелекту допоможе вченим передбачити взаємодію молекул і вибрати найбільш прийнятних кандидатів для майбутніх тестів. Це може допомогти вченим скоротити час і гроші на ранніх стадіях розробки ліків, а також глибше досліджувати хімічний простір.

2. Прецизійна медицина:

Основна мета прецизійної медицини – адаптувати медичне лікування для пацієнтів відповідно до їхніх унікальних потреб і характеристик. Штучний інтелект аналізуватиме дані пацієнтів, медичні записи людей і їхній спосіб життя, щоб визначити закономірності та конкретні ризики певних захворювань і результати лікування.

3. Зелена хімія:

Зараз існує тенденція до більш стійких і екологічних хімічних методів, і ШІ може допомогти досягти цього швидше.

ШІ може аналізувати хімічні реакції та вплив на навколишнє середовище та створювати більш ефективні, екологічно чисті сполуки. Завдяки використанню ШІ можна буде мінімізувати утворення відходів і сприяти більш екологічним виробничим практикам.

4. Дизайн матеріалів і відкриття:

ШІ може допомогти вченим розробляти та відкривати нові матеріали з бажаними властивостями. За допомогою ШІ та МЛ хіміки зможуть прогнозувати нові матеріали з такими характеристиками, як провідність, міцність і каталітична активність.

5. Автоматизація та робототехніка:

Поєднання ШІ з робототехнікою може допомогти автоматизувати лабораторні процеси та підвищити продуктивність. Роботизовані системи, які використовують ШІ, можуть швидше й ефективніше виконувати рутинні повторювані завдання, такі як аналіз даних і підготовка зразків. Це допомагає дослідникам зосередитися на більш важливих завданнях і зменшити ризик людської помилки.

6. Інтеграція великих даних:

Хімічна промисловість — це також і дані. ШІ дозволяє генерувати інформацію з моделювання, експериментів, досліджень та інших джерел. Важливо правильно збирати, зберігати та аналізувати ці дані, і ШІ може допомогти в цьому. У майбутньому штучний інтелект зможе аналізувати ці дані, щоб виявити тенденції та приховані зв'язки, а також створити нові гіпотези.

ВИСНОВКИ

Стало очевидним, що впровадження штучного інтелекту як інструменту, який підтримує освіту, настільки ж складне, як і багатообіцяюче. Інтеграція штучного інтелекту вимагає сумлінного підходу, де прозорість, справедливість, конфіденційність і персоналізація є не просто ідеалами, а важливими стовпами. Він закликає педагогів адаптуватися, але не бути затьмареними технологіями, і вимагає нових інструментів штучного інтелекту, які є настільки ж інклюзивними, наскільки вони інноваційними. Найважливіше те, що це нагадує нам, що суть будь-якої інтеграції штучного інтелекту завжди має узгоджуватися зі справжніми потребами учня та тими, які сприяють успіху учня. Спираючись на впровадження штучного інтелекту на цілеспрямовані наміри та етичні міркування, ми можемо гарантувати, що цей технологічний прогрес стане справжнім союзником у пошуках збагаченого навчання та справедливого доступу до освіти.

Використання штучного інтелекту в хімічній промисловості, особливо в лабораторіях і CRO, є одним із основних способів революціонізувати та оптимізувати процеси. Це допомагає цим організаціям аналізувати дані, робити прогнози та придумувати нові рішення, які були б недоступні без ШІ.

Незважаючи на те, що застосування штучного інтелекту в хімії ще не повністю розкрило свій потенціал, дуже ймовірно, що воно змінить галузь у найближчі роки. Таким чином, настав час приєднатися до цієї тенденції та використовувати штучний інтелект якнайкраще.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. These *Three Laws of Robotics* were introduced by writer Asimov in 1950. They are intended to protect human from robots. First law, for instance, is “*A robot may not injure a human being or, through inaction, allow a human being to come to harm*”. If you’re interested, you can read the two other laws on the dedicated Wikipedia page.
2. *J.A.R.V.I.S* is a sci-fi version of a general AI, one which can work at eye level with humans. We will introduce the concepts of *narrow* and *general AI* in more detail in AI - two letters many meanings.
3. Several examples are described and illustrated in this [blog](https://aljaribook.com/en/list-of-ingenuous-mechanical-devices/)(<https://aljaribook.com/en/list-of-ingenuous-mechanical-devices/>).
4. Want to see how it looked like? Have a look at a copper engraving explaining how the Turk looked like [here](https://en.wikipedia.org/wiki/The_Turk#/media/File:Tuerkischer_schachspieler_windisch4.jpg) (https://en.wikipedia.org/wiki/The_Turk#/media/File:Tuerkischer_schachspieler_windisch4.jpg).
5. You may find a picture of a part of the Analytical Engine [here](https://en.wikipedia.org/wiki/Analytical_Engine#/media/File:AnalyticalMachine_Babbage_London.jpg)(https://en.wikipedia.org/wiki/Analytical_Engine#/media/File:AnalyticalMachine_Babbage_London.jpg).
6. Information theory is the scientific study of the quantification, storage, and communication of information. Read more about it [here on Wikipedia](https://en.wikipedia.org/wiki/Information_theory). A famous application of information theory is data compression: formats like ZIP, MP3 and JPEG were introduced to store as much information as possible while requiring the least memory possible.
(https://en.wikipedia.org/wiki/Information_theory)
7. Yann Le Cun is a researcher in AI, well known for developing convolutional neural networks. He is the head of AI at [Facebook](http://facebook.com/)(<http://facebook.com/>)
8. The Association for Computing Machinery (ACM) named Yoshua Bengio, Geoffrey Hinton, and Yann Le Cun recipients of the 2018 ACM A.M. Turing

Award (recognized as a Nobel prize equivalent for computer science). Read more about it [here](https://awards.acm.org/about/2018-turing). (<https://awards.acm.org/about/2018-turing>).

9. Chen, W., Y. Liang, and D. Liang. 2020. "Artificial intelligence in education: A review of the literature." *Educational Technology Research and Development* 68 (1): 65-83.
10. Ferrell, C. L. 2017. *The ethics of artificial intelligence*. Cambridge University Press.
11. Partridge, H., and G. Piccoli. 2018. "Artificial intelligence in education: Opportunities and challenges." *Journal of Educational Technology Development and Exchange* 1 (1): 1-15.
12. Wu, Q., and Y. Liang. 2019. "A review of artificial intelligence in education technology." *Journal of Educational Technology Development and Exchange* 2 (1): 1-16.
13. Yang, Y., and W. Chen. 2020. "The integration of artificial intelligence in education: A systematic review." *Journal of Educational Technology Development and Exchange* 3 (1): 1-17.
14. <https://scholarly.so/blog/improving-school-homework-with-ai-enhancing-learning-and-efficiency>
15. <https://varteq.com/how-ai-transforms-knowledge-assessment-practices/>
16. J. C. Védrine Heterogeneous Catalysis on Metal Oxides, *Catalysts*, 2017, 7 , 341.
17. A. F. Lee , J. A. Bennett , J. C. Manayil and K. Wilson , Heterogeneous catalysis for sustainable biodiesel production via esterification and transesterification, *Chem. Soc. Rev.*, 2014, 43 , 7887 —7916.
18. M. E. Günay and R. Yildirim , Neural network Analysis of Selective CO Oxidation over Copper-Based Catalysts for Knowledge Extraction from Published Data in the Literature, *Ind. Eng. Chem. Res.*, 2011, 50 , 12488 — 12500.

19. W. Xu and B. Yang, Microkinetic modeling with machine learning predicted binding energies of reaction intermediates of ethanol steam reforming: The limitations, *Mol. Catal.*, 2023, 537, 112940.
20. S. Back et al., Convolutional Neural Network of Atomic Surface Structures To Predict Binding Energies for High-Throughput Screening of Catalysts, *J. Phys. Chem. Lett.*, 2019, 10, 4401—4408.
21. Cherkasov, A.; Muratov, E. N.; Fourches, D.; Varnek, A.; Baskin, I. I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y. C.; Todeschini, R.; Consonni, V.; Kuz'min, V. E.; Cramer, R.; Benigni, R.; Yang, C.; Rathman, J.; Terfloth, L.; Gasteiger, J.; Richard, A.; Tropsha, A. QSAR Modeling: Where Have You Been? Where Are You Going To?, *J. Med. Chem.* 2014, 57, 4977–5010, DOI: 10.1021/jm4004285.
22. Schneider, G.; Fechner, U. Computer-Based de Novo Design of Drug-like Molecules. *Nat. Rev. Drug Discovery* 2005, 4, 649–663, DOI: 10.1038/nrd1799.
23. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* 2015, 521 (7553), 436–444, DOI: 10.1038/nature14539.
24. Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* 2010, 50, 742–754, DOI: 10.1021/ci100050t.
25. Mauri, A.; Consonni, V.; Pavan, M.; Todeschini, R. Dragon Software: An Easy Approach to Molecular Descriptor Calculations. *Match* 2006, 56, 237–248.
26. Lee, A. A.; Yang, Q.; Bassyouni, A.; Butler, C. R.; Hou, X.; Jenkinson, S.; Price, D. A. Ligand Biological Activity Predicted by Cleaning Positive and Negative Chemical Correlations. *Proc. Natl. Acad. Sci. U. S. A.* 2019, 116, 3373–3378, DOI: 10.1073/pnas.1810847116.
27. Mayr, A.; Klambauer, G.; Unterthiner, T.; Steijaert, M.; Wegner, J. K.; Ceulemans, H.; Clevert, D.-A.; Hochreiter, S. Large-Scale Comparison of

- Machine Learning Methods for Drug Target Prediction on ChEMBL. *Chem. Sci.* 2018, 9, 5441– 5451, DOI: 10.1039/C8SC00148K.
28. Kearnes, S.; McCloskey, K.; Berndl, M.; Pande, V.; Riley, P. Molecular Graph Convolutions: Moving beyond Fingerprints. *J. Comput.-Aided Mol. Des.* 2016, 30, 595– 608, DOI: 10.1007/s10822-016-9938-8.
29. Mayr, A.; Klambauer, G.; Unterthiner, T.; Hochreiter, S. Deep Tox: Toxicity Prediction Using Deep Learning. *Front. Environ. Sci.* 2016, 3, 80, DOI: 10.3389/fenvs.2015.00080.
30. Feinberg, E. N.; Sur, D.; Wu, Z.; Husic, B. E.; Mai, H.; Li, Y.; Sun, S.; Yang, J.; Ramsundar, B.; Pande, V. S. Potential Net for Molecular Property Prediction. *ACS Cent. Sci.* 2018, 4, 1520– 1530, DOI: 10.1021/acscentsci.8b00507.
31. Gómez-Bombarelli, R.; Wei, J. N.; Duvenaud, D.; Hernández-Lobato, J. M.; Sánchez-Lengeling, B.; Sheberla, D.; Aguilera-Iparraguirre, J.; Hirzel, T. D.; Adams, R. P.; Aspuru-Guzik, A. Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. *ACS Cent. Sci.* 2018, 4, 268– 276, DOI: 10.1021/acscentsci.7b00572.
32. Bjerrum, E. J.; Threlfall, R. Molecular Generation with Recurrent Neural Networks (RNNs). *arXiv (Computer Science.Machine Learning)*, May 12, 2017, 1705.04612, ver. 1. <https://arxiv.org/abs/1705.04612> (accessed 2020-10-28).
33. Popova, M.; Isayev, O.; Tropsha, A. Deep Reinforcement Learning for de Novo Drug Design. *Sci. Adv.* 2018, 4, eaap7885 DOI: 10.1126/sciadv.aap7885.
34. Zhou, Z.; Kearnes, S.; Li, L.; Zare, R. N.; Riley, P. Optimization of Molecules via Deep Reinforcement Learning. *Sci. Rep.* 2019, 9, 10752, DOI: 10.1038/s41598-019-47148-x.

35. Olivecrona, M.; Blaschke, T.; Engkvist, O.; Chen, H. Molecular de-Novo Design through Deep Reinforcement Learning. *J. Cheminf.* 2017, 9, 48, DOI: 10.1186/s13321-017-0235-x.
36. Silver, D.; Huang, A.; Maddison, C. J.; Guez, A.; Sifre, L.; van den Driessche, G.; Schrittwieser, J.; Antonoglou, I.; Panneershelvam, V.; Lanctot, M.; Dieleman, S.; Grewe, D.; Nham, J.; Kalchbrenner, N.; Sutskever, I.; Lillicrap, T.; Leach, M.; Kavukcuoglu, K.; Graepel, T.; Hassabis, D. Mastering the Game of Go with Deep Neural Networks and Tree Search. *Nature* 2016, 529, 484–489, DOI: 10.1038/nature16961.
37. De Cao, N.; Kipf, T. MolGAN: An Implicit Generative Model for Small Molecular Graphs. *arXiv (Statistics.Machine Learning)*, May 30, 2018, 1805.11973, ver. 1. <https://arxiv.org/abs/1805.11973> (accessed 2020-10-28).
38. Ma, T.; Chen, J.; Xiao, C. Constrained Generation of Semantically Valid Graphs via Regularizing Variational Autoencoders. In *Advances in Neural Information Processing Systems 32*; Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., Garnett, R., Eds.; Curran Associates, 2018; pp 7113–7124.
39. Li, Y.; Vinyals, O.; Dyer, C.; Pascanu, R.; Battaglia, P. Learning Deep Generative Models of Graphs. *arXiv (Computer Science.Machine Learning)*, March 8, 2018, 1803.03324, ver. 1. <https://arxiv.org/abs/1803.03324> (accessed 2020-10-28).
40. You, J.; Ying, R.; Ren, X.; Hamilton, W. L.; Leskovec, J. GraphRNN: Generating Realistic Graphs with Deep Auto-regressive Models. *arXiv (Computer Science.Machine Learning)*, February 24, 2018, 1802.08773, ver. 1. <https://arxiv.org/abs/1802.08773> (accessed 2020-10-28).
41. Li, Y.; Zhang, L.; Liu, Z. Multi-objective De Novo Drug Design with Conditional Graph Generative Model. *J. Cheminf.* 2018, 10, 33, DOI: 10.1186/s13321-018-0287-6.

42. <https://deepmind.google/discover/blog/millions-of-new-materials-discovered-with-deep-learning/>
43. D.M. Lowe Extraction of chemical structures and reactions from the literature [dissertation]
44. S.M. Kearnes, M.R. Maser, M. Wleklinski, A. Kast, A.G. Doyle, S.D. Dreher, et al. The open reaction database *J Am Chem Soc*, 143 (45) (2021), pp. 18820-18826.
45. S.A. Akhondi, A.G. Klenner, C. Tyrchan, A.K. Manchala, K. Boppana, D. Lowe, et al. Annotated chemical patent corpus: a gold standard for text mining *PLoS One*, 9 (9) (2014), Article e107477.
46. C.R. Groom, I.J. Bruno, M.P. Lightfoot, S.C. Ward The Cambridge Structural Database *Acta Cryst B*, 72 (Pt 2) (2016), pp. 171-179.
47. D. Zagorac, H. Müller, S. Ruehl, J. Zagorac, S. Rehme Recent developments in the Inorganic Crystal Structure Database: theoretical crystal structure data and related features *J Appl Cryst*, 52 (Pt 5) (2019), pp. 918-925.
48. S. Gates-Rector, T. Blanton The Powder Diffraction File: a quality materials characterization database *Powder Diffr*, 34 (4) (2019), pp. 352-360.
49. M. Lucu, E. Martinez-Laserna, I. Gandiaga, H. Camblong A critical review on self-adaptive Li-ion Battery Ageing Models *J Power Sources*, 401 (2018), pp. 85-101.
50. A. Zakutayev, N. Wunder, M. Schwarting, J.D. Perkins, R. White, K. Munch, et al. An open experimental database for exploring inorganic materials *Sci Data*, 5 (1) (2018), Article 180053.
51. J.R. Quinlan Induction of decision trees *Mach Learn*, 1 (1) (1986), pp. 81-106.
52. <https://iubmb.onlinelibrary.wiley.com/doi/pdf/10.1016/S0307-4412%2897%2900155-6>

- 53.R. Asahara, T. Miyao Extended connectivity fingerprints as a chemical reaction representation for enantioselective organophosphorus-catalyzed asymmetric reaction prediction ACS Omega, 7 (2022), pp. 26952-26964, 10.1021/acsomega.2c03812
- 54.J.S. Delaney ESOL: estimating aqueous solubility directly from molecular structure J. Chem. Inf. Comput. Sci., 44 (2004), pp. 1000-1005, 10.1021/ci034243x.

Таблиця 1. Список популярних хімічних баз даних, які зазвичай використовуються в ML.

Класифікація	Ім'я	Зміст	URL
Бази даних хімічних реакцій	SciFinder	Інформація про хімічні сполуки, бібліографічні дані та хімічні реакції (комерційна база даних)	https://scifinder.cas.org/
	Reaxys	Хімічна реакція та бібліографічна інформація (комерційна база даних)	https://www.reaxys.com/
	USPTO	Хімічна будова і реакція	https://www.repository.cam.ac.uk/handle/1810/244727
	ORD	Дані органічних хімічних реакцій	https://github.com/open-reaction-database

Бази даних хімічних властивостей	PubChem	Хімічні та фізичні властивості, біологічна активність і токсичність речовин	https://pubchem.ncbi.nlm.nih.gov/
	NIST	Стандартні фізико-хімічні властивості сполук	https://webbook.nist.gov/chemistry/
	ChemSpider	Будова і властивості сполук	https://www.chemspider.com
	ChemBL	Лікарські властивості біоактивних молекул	https://www.ebi.ac.uk/chembl/
Матеріальні бази даних	CSD	Органічні та металоорганічні кристалічні структури	https://www.ccdc.cam.ac.uk/
	ICSD	Неорганічні та металоорганічні кристалічні структури	https://icsd.products.fiz-karlsruhe.de/
	PDF	Дані дифракції неорганічних	https://www.icdd.com/pdfsearch/

		і органічних сполук	
--	--	------------------------	--