APPLICATIONS OF ARTIFICIAL INTELLIGENCE IN CHEMISTRY

GAJARLAWAR VEER VILAS

ROLL NUMBER: 21CY10013

IIT KHARAGPUR

VEERGAJARLAWAR737@KGPIAN.IITKGP.AC.IN

LETS START WITH THE BASICS!

- What is Artificial Intelligence?
- Artificial intelligence (AI) refers to the ability of machines to act in seemingly intelligent ways, making decisions in response to new inputs without being explicitly programmed to do so.
- What is Machine Learning(ML)?
- A subset of artificial intelligence (AI) that focuses on developing algorithms and statistical models that enable computers to learn from and make predictions or decisions based on data.

- What is Deep Learning(DL)?
- A subset of machine learning that focuses on using neural networks with many layers (hence "deep") to model and understand complex patterns in large amounts of data. It is particularly effective for tasks where data is abundant and the relationships within the data are intricate and nonlinear.

LETS START WITH THE BASICS!

- What is Chemistry/Chemical Sciences?
- It is the scientific study of the properties, composition, and behavior of matter. It explores the substances that make up our world and the changes they undergo during chemical reactions.

- What is meant by the applications of AI in Chemistry?
- Application of AI in Chemistry refers to the use of artificial intelligence techniques and tools to solve complex problems, enhance research, and streamline processes within the field of chemistry.

IMPORTANCE

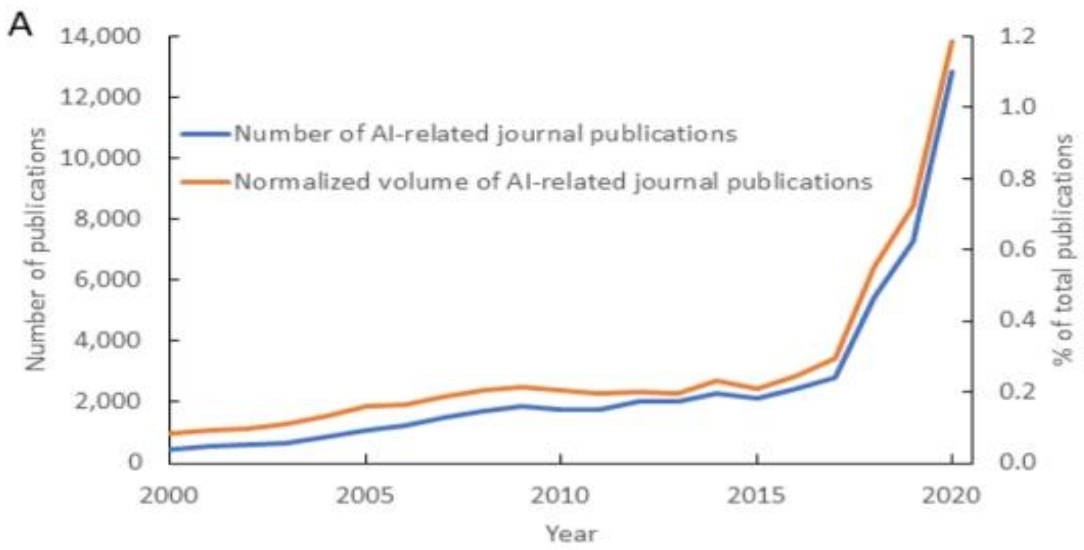
- As of 2020, over 50% of the documents on AI in chemistry were published during the past 4 years. But why?

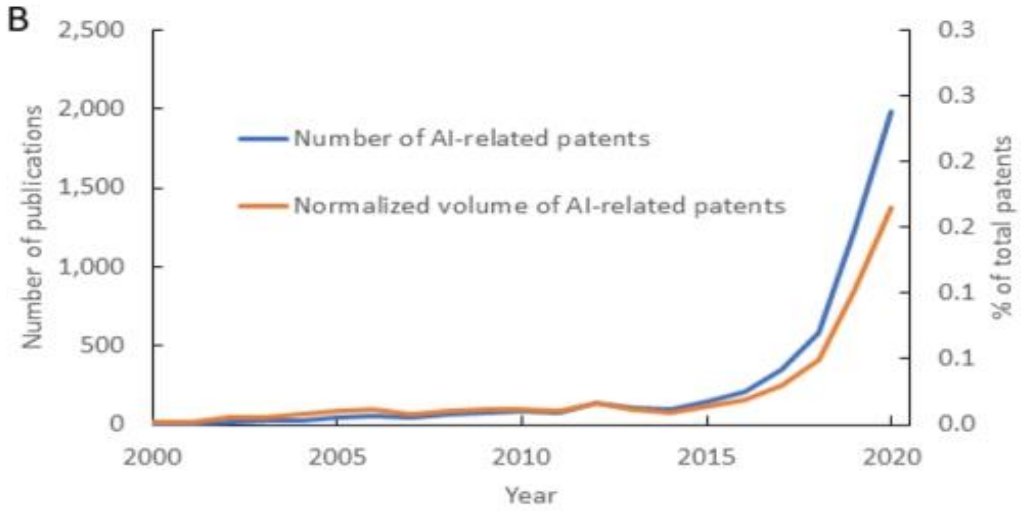
- Advances in the field of Artificial Intelligence:
- High-profile successes of deep learning projects in public data challenges starting around 2012, such as the Merck Molecular Activity Challenge and the ImageNet competition, increasingly drew research interest from the scientific community. Additionally, the introduction of open source machine learning frameworks, such as TensorFlow(2015) and PyTorch (2016), and the availability of increasingly powerful computing hardware sparked a global explosion in AI research, enabling further applications of AI to chemistry.

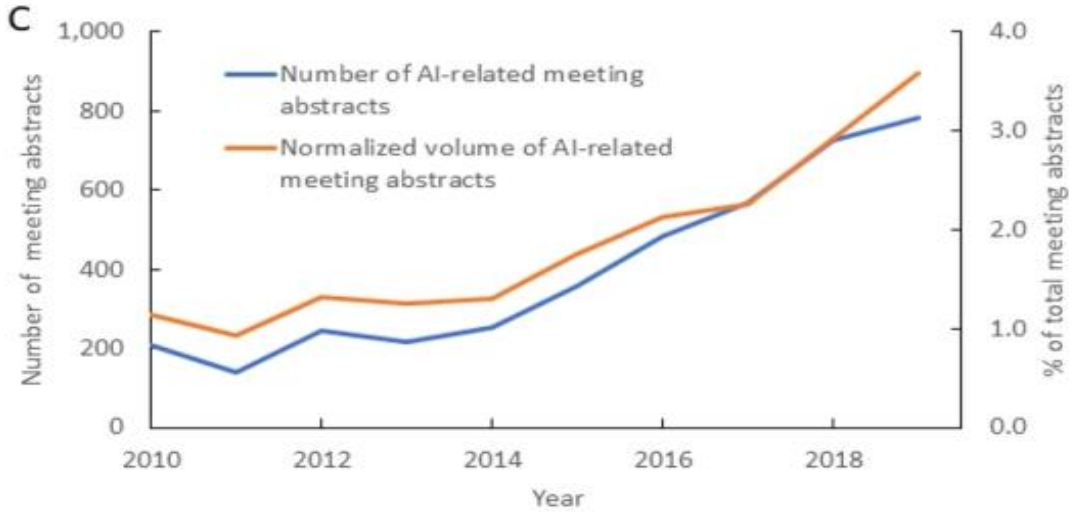
- Availability of scientific data:
- Thanks to databases like Materials Project, OQMD, CSD, PubChem, etc. there is more data available of chemical simulations than ever before. Similarly, AiiDA and NoMaD record comprehensive data provenance for ‘static’ materials simulations.

IMPORTANCE

Journal of Chemical Information and Modeling pubs.acs.org/jcim Review







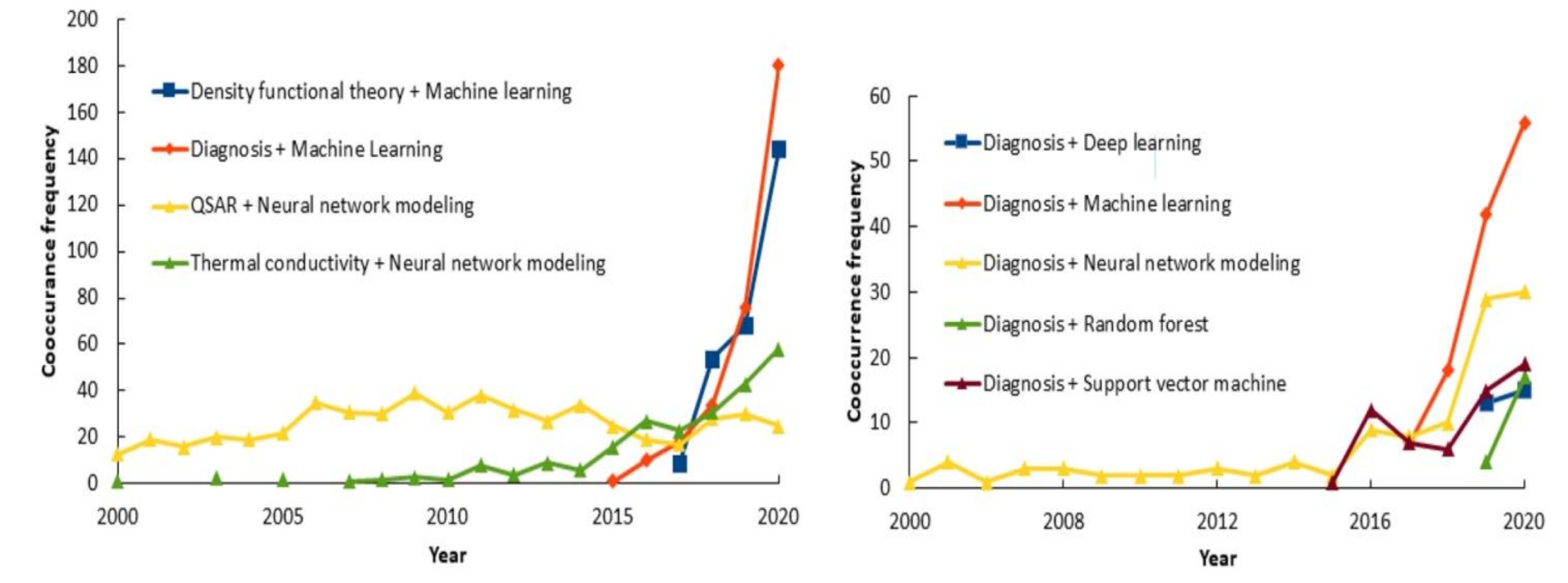
SOME OF THE MAJOR APPLICATIONS

- Detection of molecular properties: AI algorithms can analyze chemical data to predict and classify various molecular properties, such as toxicity, solubility, and reactivity. This makes the process of detection faster and less prone to errors, unlike manual detection.
-
- Drug discovery: AI-driven drug discovery platforms can analyze vast databases of chemical compounds, predict their potential activity against specific targets, and prioritize candidates for further investigation.

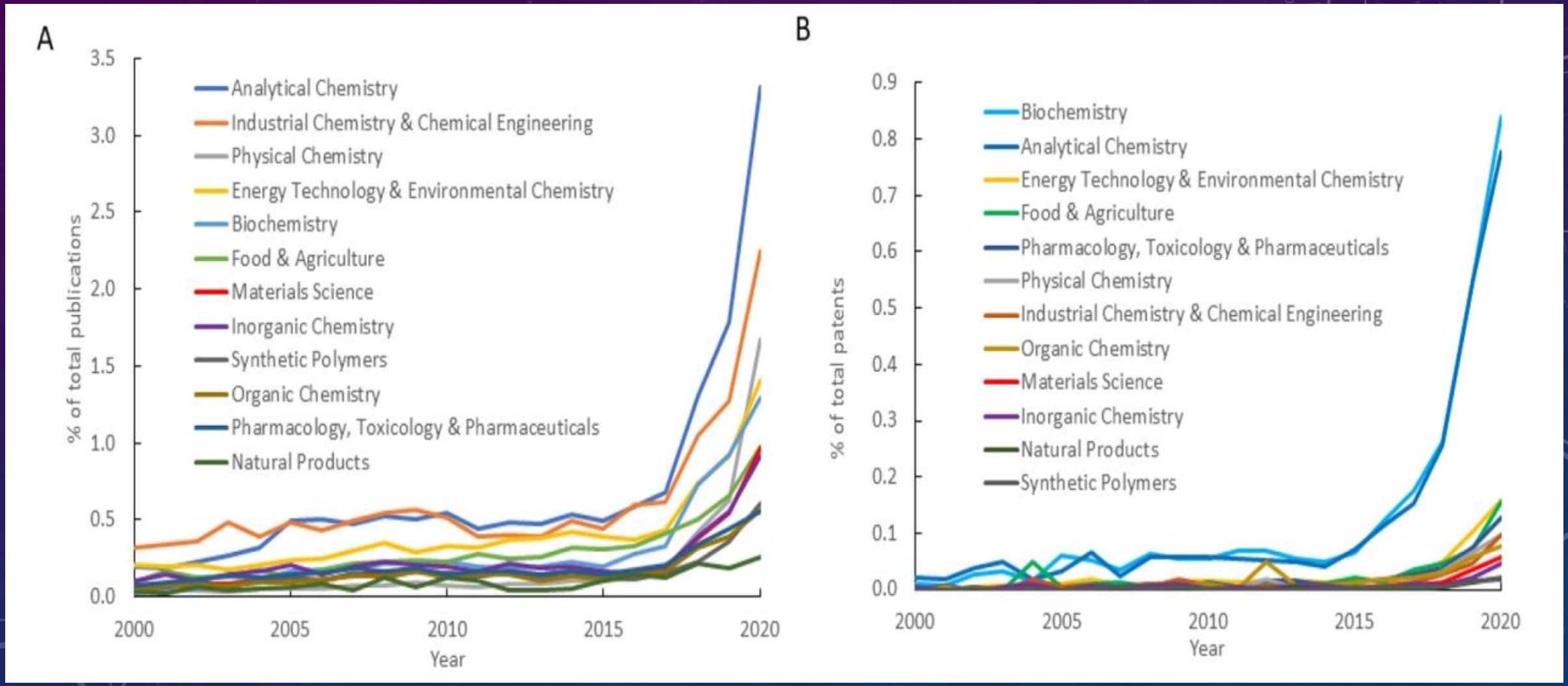
- Retrosynthesis Reactions: Chemists use AI for planning the most efficient and cost-effective synthesis routes by generating retrosynthesis pathways and suggesting optimal reaction steps.

- Material Science: Identification of properties from crystal structure.

CO-OCCURRENCE OF RESEARCH TOPICS AND AI ALGORITHMS



MAJOR AREAS OF APPLICATION



AI IN ANALYTICAL CHEMISTRY

- AI has been fundamental in managing heterogeneous data and in advanced analysis of complex spectra in areas such as spectroscopy and chromatography.
- In spectroscopy, AI has been employed to deconvolute and interpret complex spectra, a significant advancement in compound identification.
- In chromatography, algorithms are employed to process and analyze large chromatographic datasets. They can identify patterns and correlations in complex data, facilitating the identification of compounds and quantification of their components in samples.
- The application of AI in analytical chemistry is often paired with use of chemometrics which is the science of extracting information from chemical data using mathematical and statistical methods. This is done in order to cross-validate the results obtained by ML and DL models as the decision making process in them is not transparent. This is often referred to as the "Black Box" problem.

USE IN SPECTROSCOPY

- Spectroscopy is a scientific technique used to study and analyze the interaction between matter and electromagnetic radiation. It involves measuring the spectrum of light absorbed, emitted, or scattered by materials to understand their composition, structure, and properties.
- In the analysis of spectroscopic data, AI has been used to select the optimal wavelengths for analysis, a task traditionally performed using chemometric techniques.

Applications and benefits and challenges of AI in spectroscopy techniques.

Spectroscopic Technique	AI Methods	Objectives	Advantages	Examples
UV–Vis Spectroscopy	Machine Learning, Neural Networks, Deep Learning	• Accuracy in identification • Quantitative analyzes • Low concentration detection	Faster analyzes; High sensitivity; Reliable data	• Quality control of formaldehyde in industrial products [27]; • Ammonia monitoring in environmental water samples [28]; • Ethanol level determination in alcoholic beverages [29]; • Selective extraction of veterinary drugs [30]; • Glutathione quantification for oxidative stress assessment [31];
Infrared (IR) Spectroscopy	Machine Learning, Neural Networks, Deep Learning	• Detailed spectral analysis • Accurate compound detection • Quick identification	Improved interpretation; Efficient analytics; Identification of unknown compounds	• Remote petroleum trace detection in soil [32]; • Differentiation of raspberry powders [33]; • Analysis of biomass-derived bio-carbons [34]; • Cereal quality assessment [35]; • Analysis of protein secondary structures [36].
Nuclear Magnetic Resonance (NMR) Spectroscopy	Machine Learning, Neural Networks, Deep Learning	• Accurate resonance assignment; • Detailed structural analysis; • Interpretation of complex spectra	Faster analyzes; Greater precision; More detailed data	• Automated Protein Analysis [37]; • Quantification of Human Brain Metabolites [38]; • Vegetable quality monitoring [39]; • Production of New Organic Molecules [40]; • Porosity in carbonate rocks [41].
Mass Spectrometry	Machine Learning, Neural Networks, Deep Learning	• Accurate identification of masses • Precise quantitative analysis • Sensitive compound detection	Rapid sample analysis; High precision; Identification of unknown compounds	• Detection of drugs in blood and solvent samples [42]; • Rapid and objective identification and classification of câncer [43]; • Pharmacological studies of chemical substances [44]; • Improvement in accuracy and efficiency in in vitro diagnosis [45]; • Development of bioinformatics for clinical analysis [46].

USE IN CHROMATOGRAPHY

- Chromatography is a laboratory technique used to separate, identify, and quantify the components of a mixture. It involves passing the mixture through a medium where different components move at different rates, allowing them to be separated based on their physical or chemical properties.
- One of the primary uses of AI in chromatography is in data interpretation and analysis. Machine learning algorithms, for example, are employed to process and analyze large chromatographic datasets. They can identify patterns and correlations in complex data, facilitating the identification of compounds and quantification of their components in samples. ML was applied to predict chromatographic retention times, a challenge commonly addressed with chemometrics.

Integration of AI in chromatographic techniques: Objectives and advantages.

Chromatographic Technique	AI Methods	Objectives	Advantages	Examples
Gas Chromatography (GC)	Machine Learning, Neural Networks, Deep Learning	• Precise compound separation • Enhanced data analysis • Quicker process optimization	Accurate peak identification Reduced analysis time Improved method efficiency	• Classification of pyrolysis componentes [70]; • Predict and classify gasoline quality [71]; • Determination of grain quality [72]; • Predict beer aromas and quality and sensory indicators [73]; • analyze a large selection of samples [74].
Liquid Chromatography (HPLC)	Machine Learning, Neural Networks, Deep Learning	• Improved resolution • Robust data handling • Faster method development	Enhanced selectivity Streamlined workflowsGreater analytical throughput	• Quantitative structure-retention relationships for analysis of retention behavior [75]; • Optimize complex separations and improve efficiency [76]; • Predicting retention times based on mobile phase composition [77]; • Combining solute data and retention data from different columns [78]; • Simultaneous analysis of chemical substances [79].
Ionic Chromatography	Machine Learning, Neural Networks, Deep Learning	• Selective ion analysis • Automated data processing • Enhanced detection limits	Higher sensitivity Rapid result interpretation Consistent performance	• Investigation of the effects of various parameters on pesticide retention [80]; • Predicting anion retention times [81]; • Development of cation retention behavior [82]; • Predicting oligonucleotide retention times with high accuracy [83];

AI IN HIGH THROUGHPUT DATA ANALYSIS

Comparison of AI algorithms in analytical chemistry: Complexity and applicability.

Algorithm	Complexity	Data Requirements	Effective Problem Types
Neural Networks	High	High volume & diversity	Pattern recognition, predictions
Decision Trees	Medium	Moderate	Classification, regression
Support Vector Machines	High	High quality & labeled	Classification, regression
Random Forest	Medium	High volume	Classification, large datasets
Gradient Boosting	High	High volume & quality	Predictive modeling
Deep Learning	Very High	Massive volume & diversity	Complex patterns, feature extraction

DATA DATA DATA!!!

Data for both spectroscopy as well as chromatography can be obtained through a lot of simulations online.

For spectroscopy: Molecular Dynamics(MD) simulations, Quantum chemistry software.

For chromatography: Molecular Docking and Dynamics, computational chromatography software.

EXPLAINABILITY TECHNIQUES TO SOLVE THE “BLACK BOX PROBELMS

- GNNExplainer is a technique aimed at explaining Graph Neural Network (GNN) models by identifying subgraphs and node features that are most important for the model’s decisions and, for example, analyzing NMR spectra, where molecular structure affects the observed spectrum, GNNExplainer can identify substructures or bonding patterns that are most correlated with specific spectral features, helping chemists understand how certain functional groups influence the spectral properties of molecules.
- PGExplainer, on the other hand, is an approach that generates probabilistic explanations for GNN decisions. Unlike GNNExplainer, which seeks a deterministic explanation, PGExplainer works with the probability of edges in a graph contributing to the model’s decision, offering a more comprehensive and uncertain view that may be more representative of the complex reality of analytical data.
- SubgraphX is another method that focuses on identifying important subgraphs in GNNs to explain the model’s predictions.
- SME (Substructure Mask Explanation) is a technique that highlights substructures in molecules that are critical for the predictions of molecular properties made by GNNs.

INTEGRATING AI FOR DRUG DISCOVERY

- AI techniques used in drug discovery, including ML (to predict drug properties, identify potential drug candidates, and optimize chemical structures), DL (to analyze large-scale biological data, predict drug properties, and identify potential drug candidates), NLP (to analyze the scientific literature for potential drug candidates and to generate drug summaries), GM (to generate new molecules that could potentially be drug candidates), and network-based approaches (to identify potential targets for drug development).
- AI has significantly impacted the field of drug discovery, particularly in the areas of target identification and validation. This process involves identifying potential biological targets and elucidating their roles in diseases, followed by validating these targets to ensure they are directly involved in a disease mechanism and that the modulation of the target is likely to have a therapeutic effect and plays a crucial role in identifying potential drug targets by analyzing the genomic, proteomic, and metabolomic data.
- ML-based approaches, such as Kronecker regularized least squares (KronRLS), evaluate the similarities between drugs and protein molecules to determine DTBA. Correspondingly, SimBoost utilizes regression trees to predict DTBA, and considers both feature-based and similarity-based interactions.
- AI algorithms can analyze molecular structures, predict binding affinities, and prioritize compounds for further experimental testing . Also, AI techniques, such as Bayesian docking approximations and RL, can be used in molecular docking simulations . Molecular docking involves predicting the preferred orientation of a small molecule (drug candidate) when it is bound to a target protein. AI algorithms can explore the conformational space and predict the binding affinity between the drug and the target protein .

Table 1. AI methods that are frequently utilized in drug discovery.

Class	Algorithms	
Supervised learning	Regression analysis	MLR
		DT
		LR
	Classification	SVM
		CNN
		RNN
		GAN
Unsupervised learning	Clustering	k-means
		Hierarchical
	Dimensionality reduction	PCA
		t-SNE

PRODUCING A NEW DRUG

- GANs are a class of ML algorithms that consist of two neural networks, the generator and the discriminator, which are trained in an adversarial manner to generate realistic data samples.
- GANs can be employed to optimize lead compounds by generating modifications or analogs of existing drug candidates. By using the discriminator's feedback to guide the generation of new chemical structures, GANs can propose modifications that are more likely to be biologically active and have improved pharmacological properties.
- A number of studies have also shown that GANs can be used for multi-objective drug designing, where the goal is to optimize multiple drug properties simultaneously.

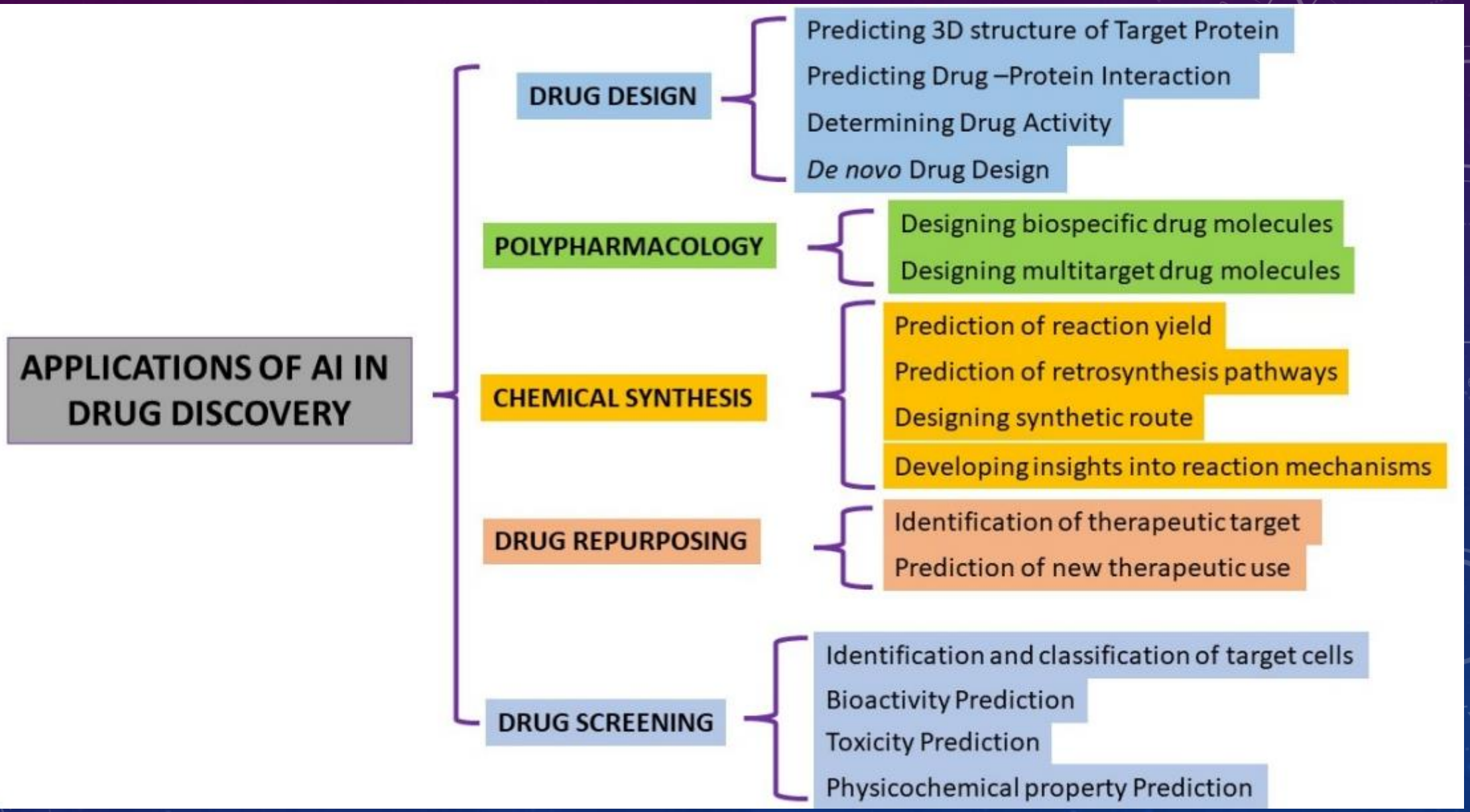

APPLICATIONS OF AI IN DRUG DISCOVERY
DRUG DESIGN
Predicting 3D structure of Target Protein
Predicting Drug –Protein Interaction
Determining Drug Activity
De novo Drug Design
POLYPHARMACOLOGY
Designing biospecific drug molecules
Designing multitarget drug molecules
CHEMICAL SYNTHESIS
Prediction of reaction yield
Prediction of retrosynthesis pathways
Designing synthetic route
Developing insights into reaction mechanisms
DRUG REPURPOSING
Identification of therapeutic target
Prediction of new therapeutic use
DRUG SCREENING
Identification and classification of target cells
Bioactivity Prediction
Toxicity Prediction
Physicochemical property Prediction

Table 3. AI tools/platforms already used in drug discovery.

Program/Platform	Description	Primary Use	Accession
DeepChem	Python-based AI system using MLP model	Candidate selection in drug discovery	https://github.com/deepchem/deepchem (accessed on 1 February 2024)
DeepNeuralNetQSAR	Python-based AI system	Can aid the detection of the molecular activity of compounds	https://github.com/Merck/DeepNeuralNet-QSAR (accessed on 1 February 2024)
Chemputer	Combination of Monte Carlo tree search and symbolic AI, including DNNs	Synthesize organic molecules	https://zenodo.org/record/1481731 (accessed on 1 February 2024)
DeepTox	AI system using DL	Chemical toxicity prediction	www.bioinf.jku.at/research/DeepTox (accessed on 1 February 2024)
AlphaFold	AI system using DL	Predicts the 3D structures of proteins	https://alphafold.ebi.ac.uk/ (accessed on 1 February 2024)
ORGANIC	Generative ML approaches and DNNs	Novel molecular materials	https://github.com/aspuru-guzik-group/ORGANIC (accessed on 1 February 2024)
PotentialNet	Neural networks, deep attention mechanisms and descriptor embeddings	The binding affinity of ligands in protein–ligand complexes.	https://www.genesistherapeutics.ai/platform.html (accessed on 1 February 2024)
Hit Dexter	ML technique, CNNs and ANNs	For predicting molecules that might respond to biochemical assays	http://hitdexter2.zbh.uni-hamburg.de (accessed on 1 February 2024)
DeltaVina	ML algorithms, including XGBoost and random forest	Scoring protein–ligand binding affinity	https://github.com/chengwang88/deltavina (accessed on 1 February 2024)
Neural graph fingerprint	CNNs	Predict properties of novel molecules	https://github.com/HIPS/neural-fingerprint (accessed on 1 February 2024)
GastroPlus	AI and predictive modeling	For pharmaceutical products (dosage form) in many animal models	https://www.simulations-plus.com/software/gastroplus/# (accessed on 1 February 2024)

A MAJOR CHALLENGE

- Despite the significant emphasis on developing theoretical strategies for efficiently constructing databases with high- fidelity data, there is a need for additional efforts to ensure that these databases are also user-friendly for interdisciplinary research, i.e., permit even non-domain-expert AI practitioners to interact with the data with minimal intervention. This accessibility is essential for facilitating the test of new algorithmic developments.
- This can be extremely useful for potential collaboration opportunities for experts from the field on Chemical Sciences and Artificial Intelligence.

REFERENCES

The following Journals and articles were used to create these slides:

1. Artificial Intelligence in Chemistry: Current Trends and Future Directions

https://pubs.acs.org/doi/abs/10.1021/acs.jcim.1c00619

2. Accelerated chemical science with AI

https://pubs.rsc.org/en/content/articlehtml/2024/dd/d3dd00213f

3. AI in analytical chemistry: Advancements, challenges, and future directions

https://www.sciencedirect.com/science/article/pii/S003991402400328X?casa_token=hLzvczXhwt4AAAAA:OpxL3wRHb8Qg1KM6NKAsmiq2_spl1pK6u6zp306eblhOVgibYs2gTonCPfYW7Pb3QgnmB-hu

4. Integrating Artificial Intelligence for Drug Discovery in the Context of Revolutionizing Drug Delivery

https://www.mdpi.com/2075-1729/14/2/233

5. The use of AI in chemical research

https://www.softformance.com/blog/the-use-of-ai-in-chemical-research/