



# Recent Advances in the Applications of Artificial Intelligence for Enhancing Chemical Risk Assessment

Asmaa Ali Abdelwahab (Edelweiss Connect)

OpenTox Summer School

24 July 2023

# Agenda

- Chemical Risk Assessment (CRA)
- Artificial Intelligence (AI)
- Important Concepts in AI
- Applications of AI in CRA
- References
- Practical Case Study
- Conclusion
- Important Announcement

# Chemical Risk Assessment (CRA)

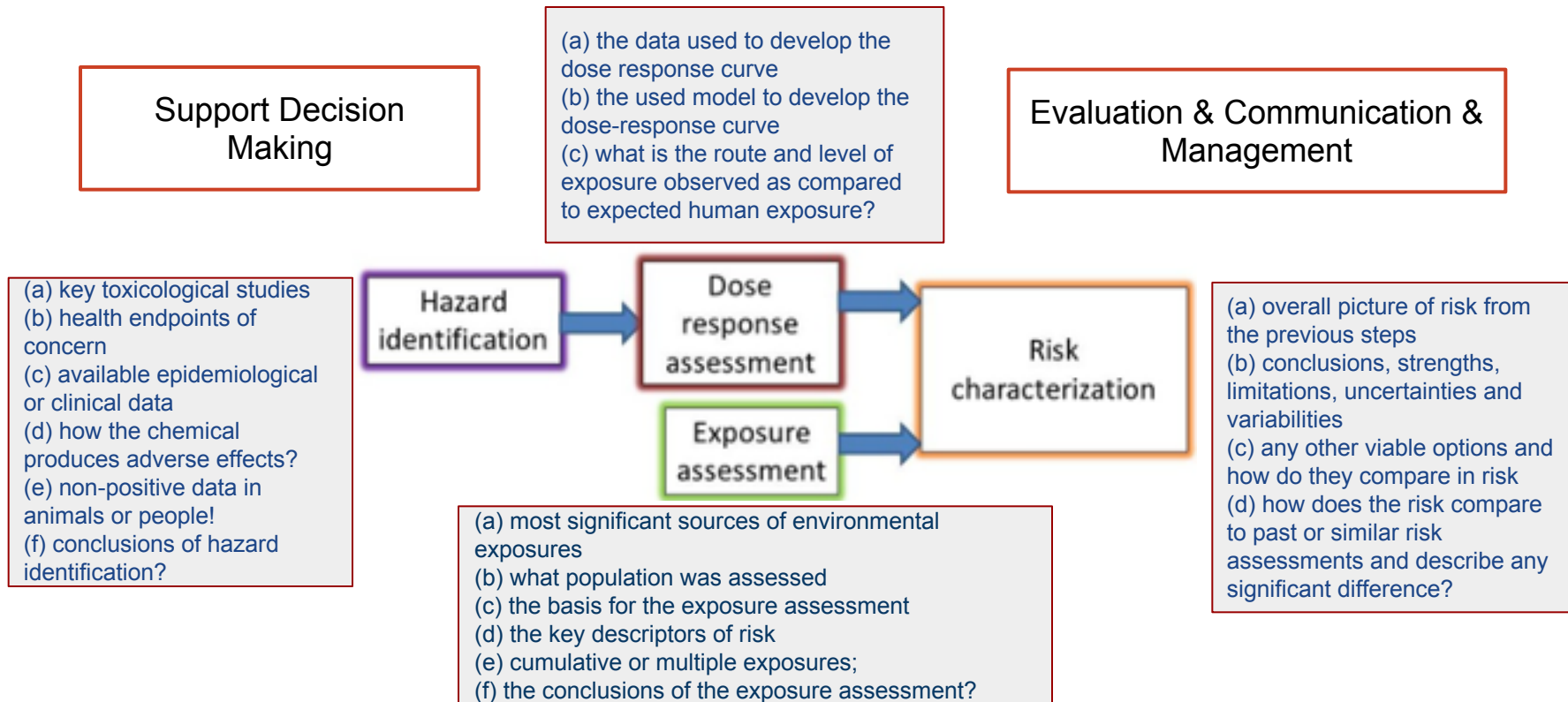
Risk Assessment is a multi-step process used by professionals to make decisions for “safe” **use** of chemicals in commercial, industrial, and environmental settings.

**Risk assessment is easy. You can learn it in two steps .. Each step takes 10 years.**

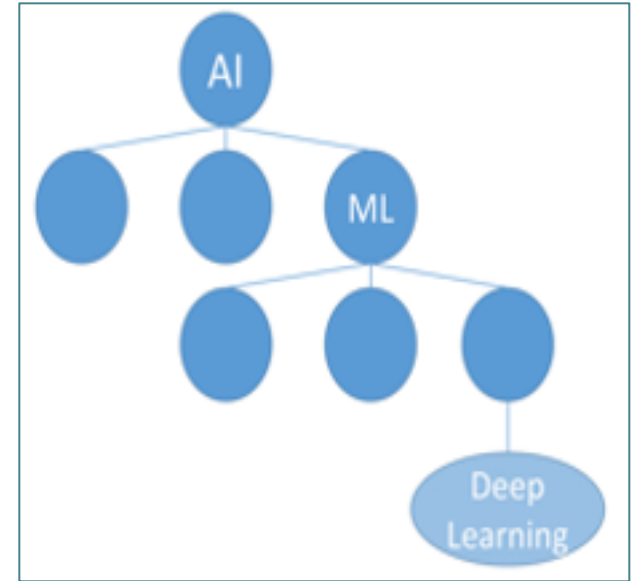
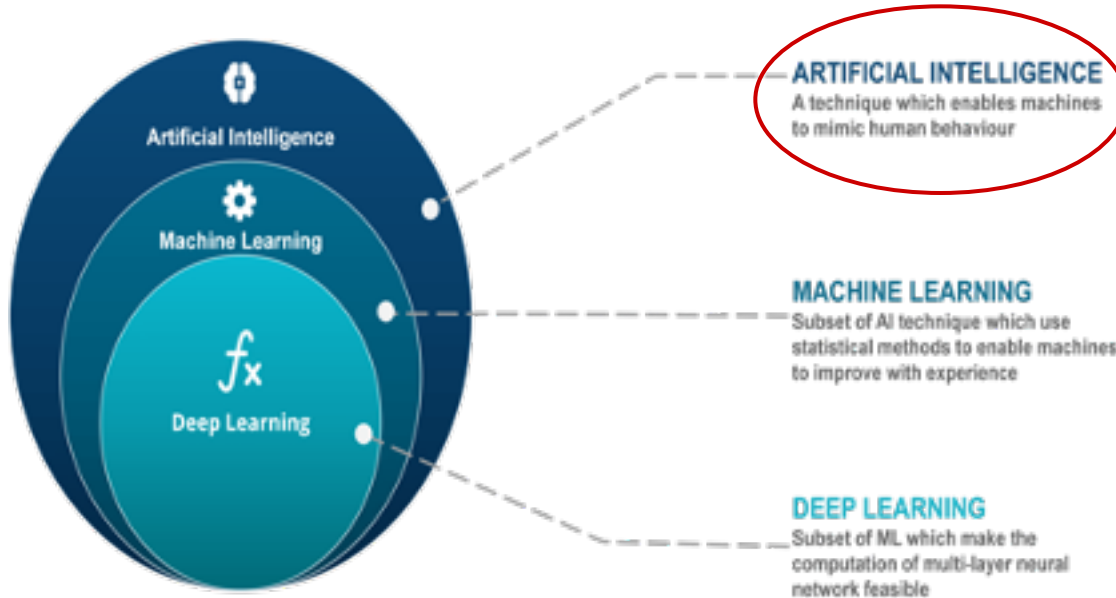
Arnold Lehman, US FDA



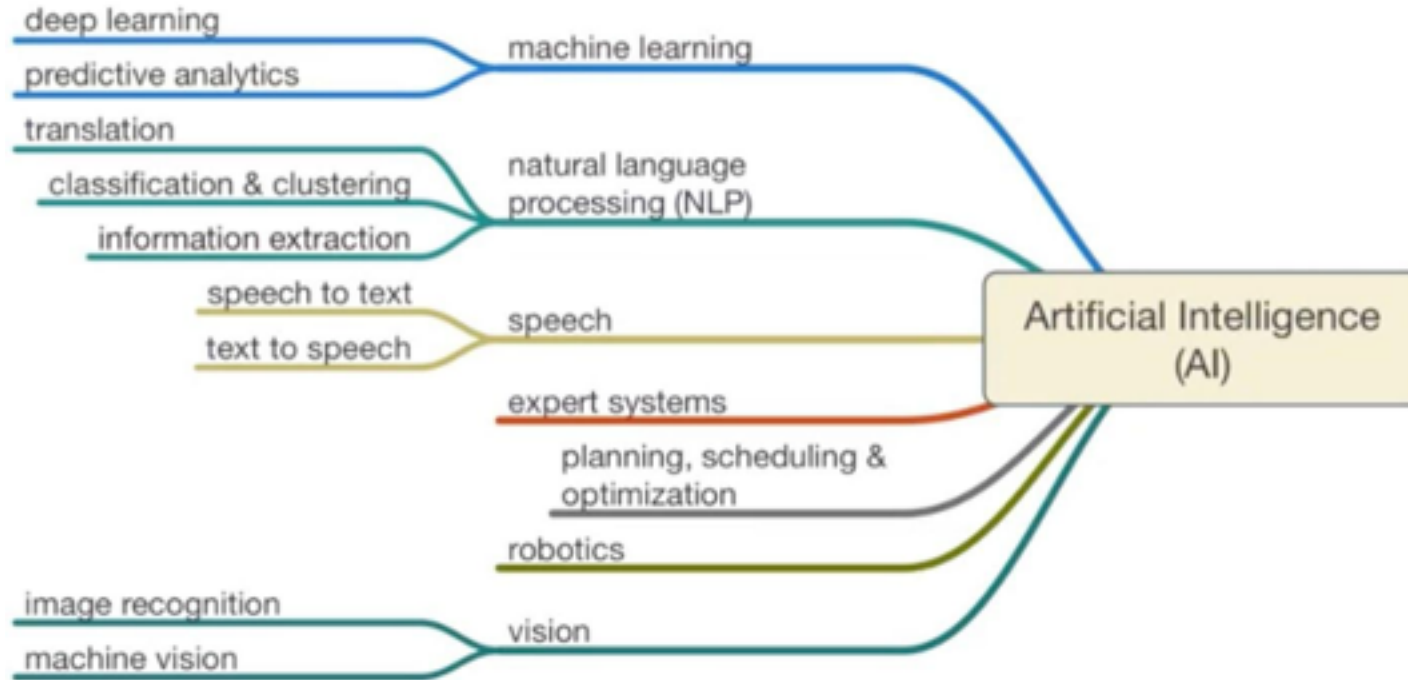
# Chemical Risk Assessment (CRA)



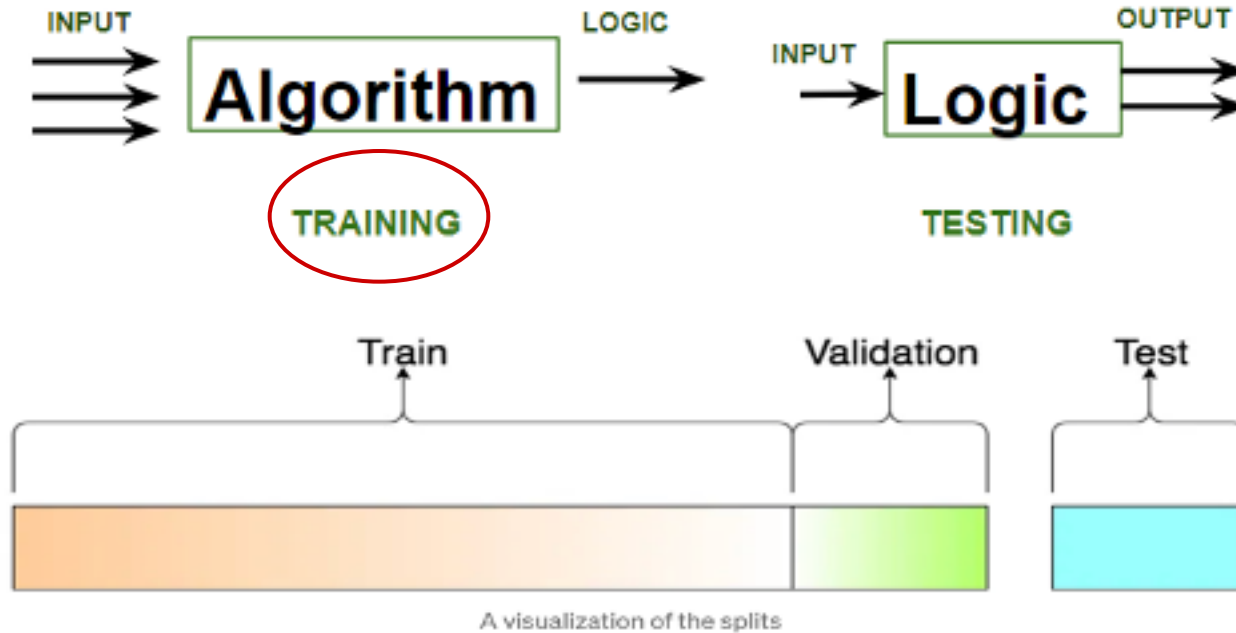
# Artificial Intelligence (AI)



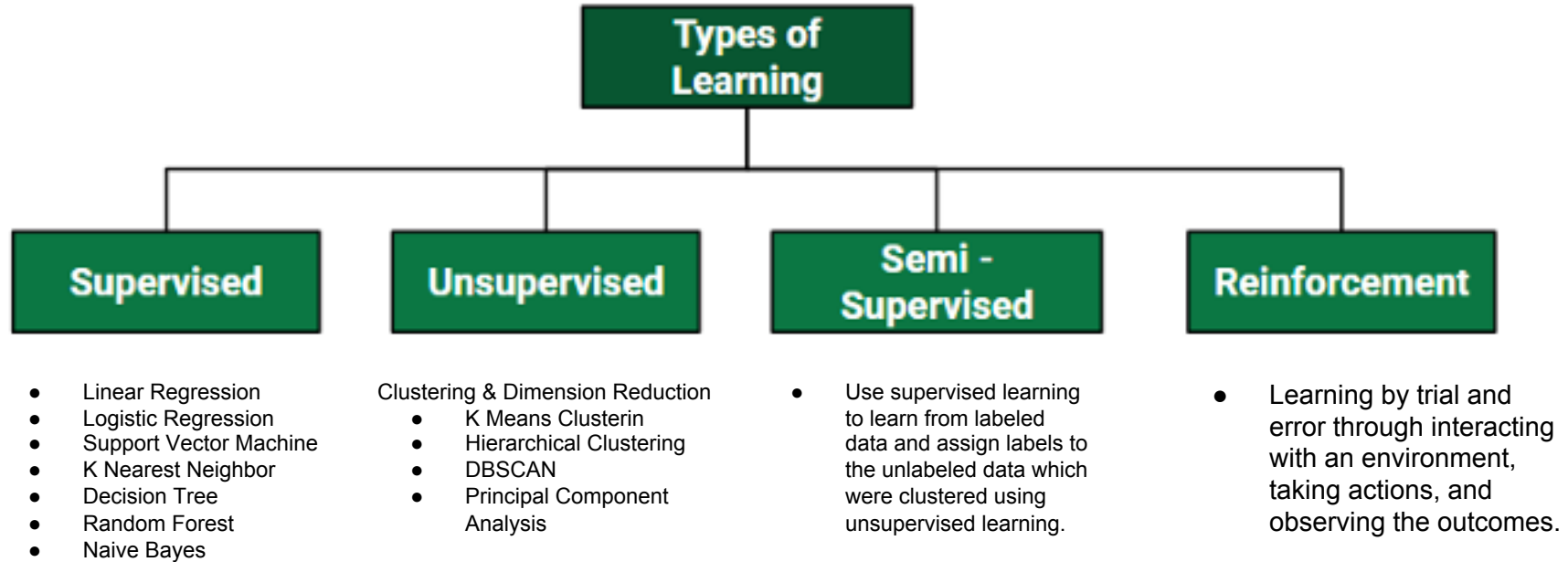
# Subfields of AI



# Learning Process

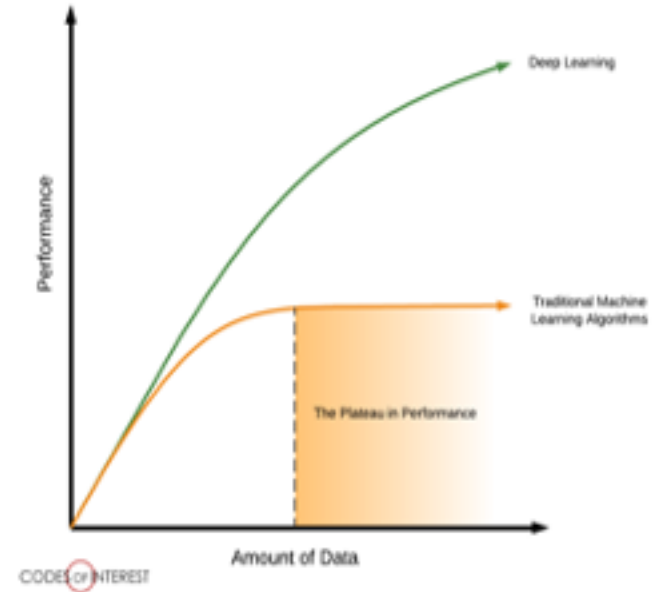
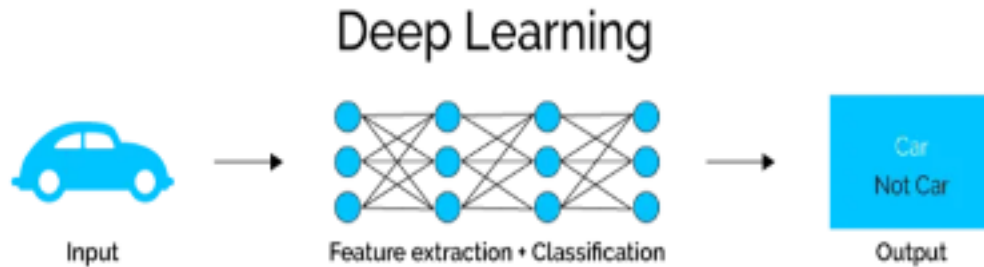
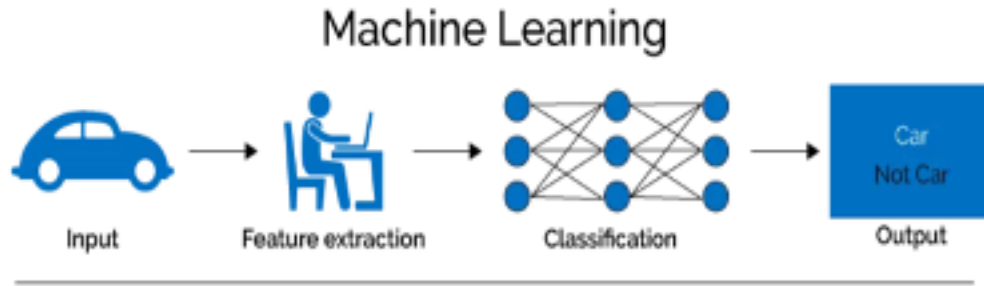


# Types of Learning

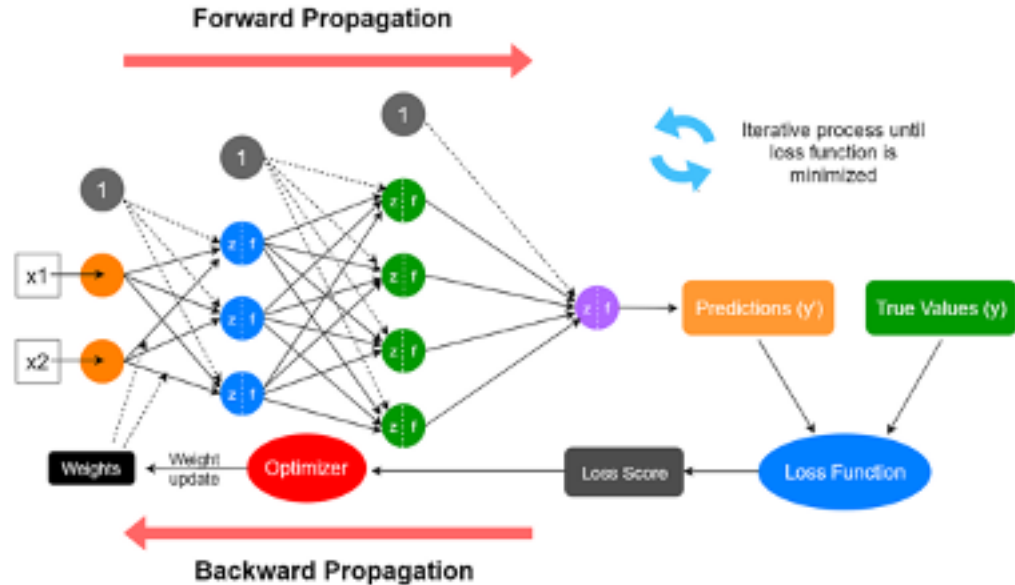
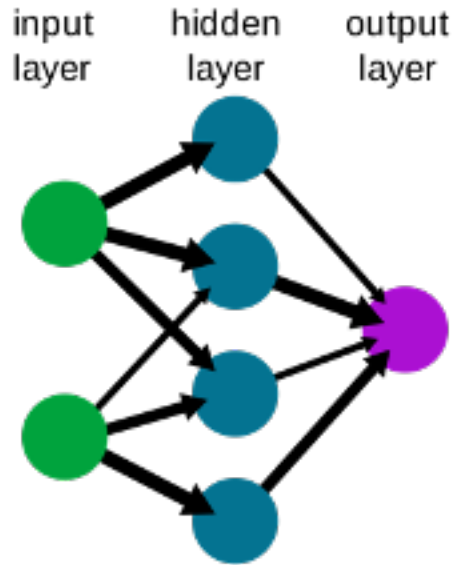




# Traditional ML Vs Neural Networks



# What are Neural Networks?

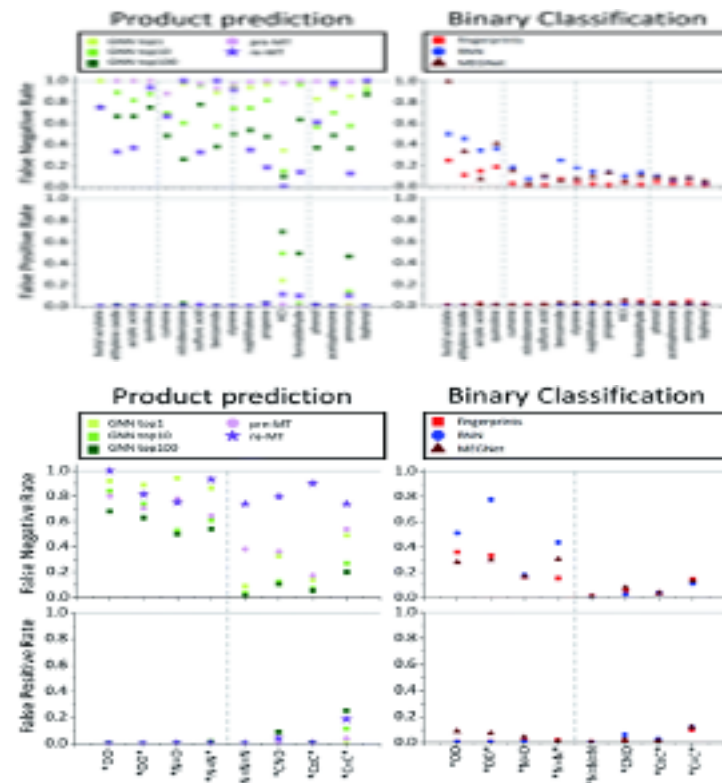
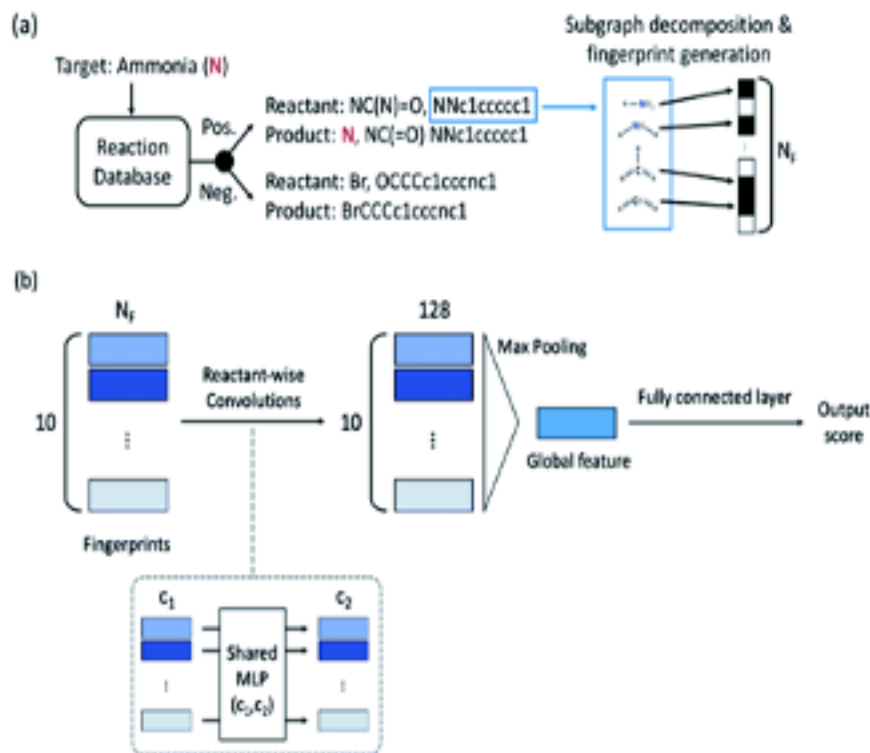


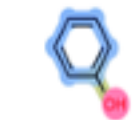
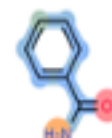
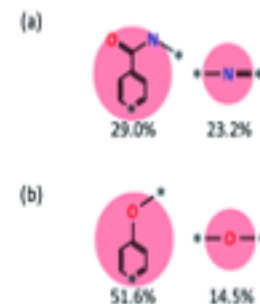
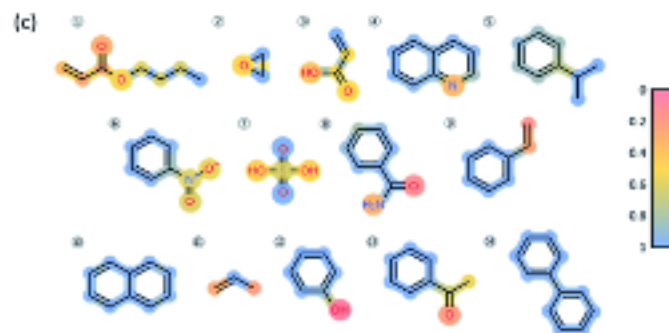
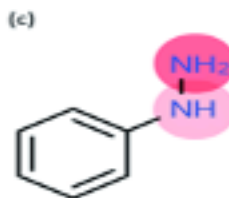
# Applications of AI in CRA



Recent Advances – Last Three Years

# Predicting Potentially Hazardous Chemical Reactions using an Explainable Neural Network



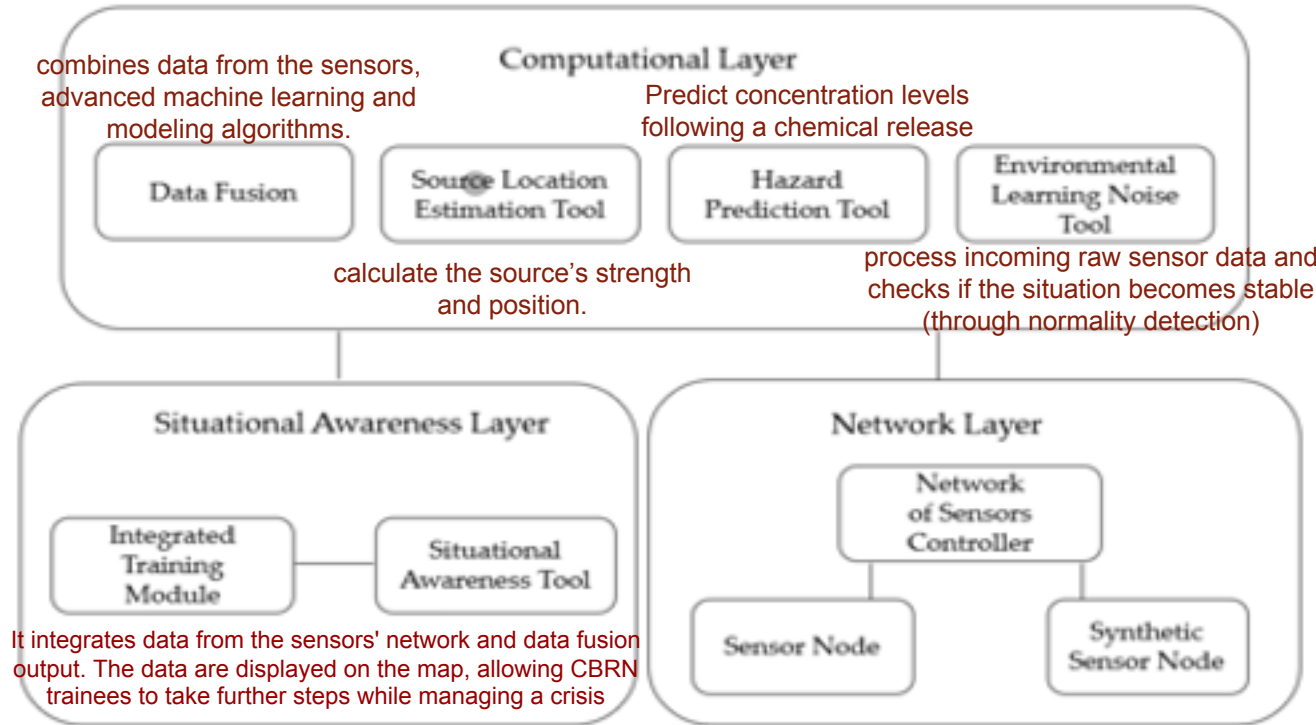


Average contribution score



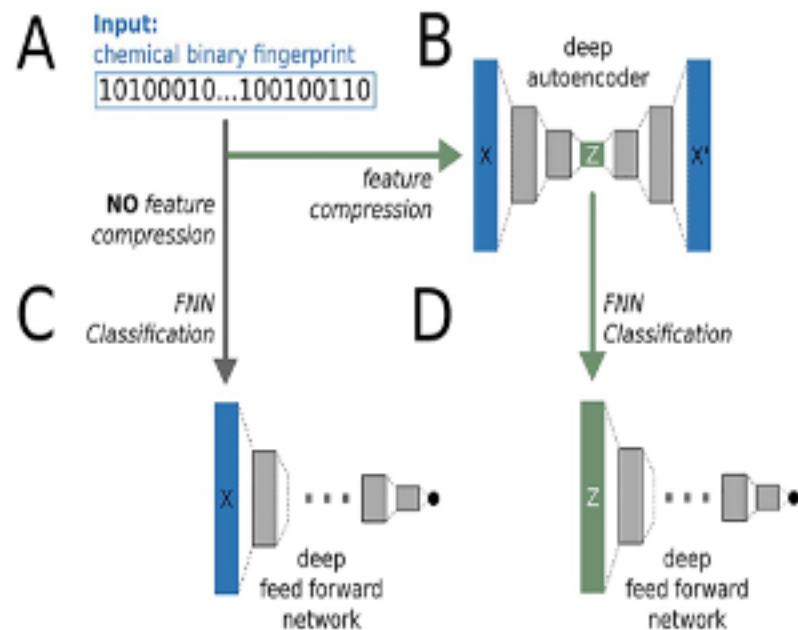
0 0.2 0.4 0.6 0.8 1

# The EU-SENSE System for Chemical Hazards Detection, Identification, and Monitoring (DIM)

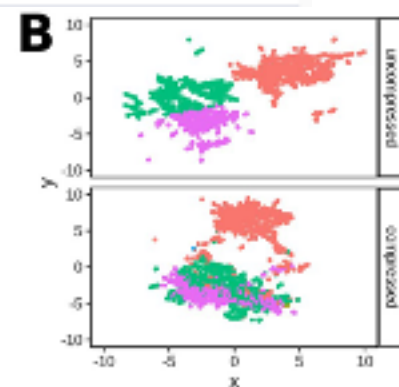
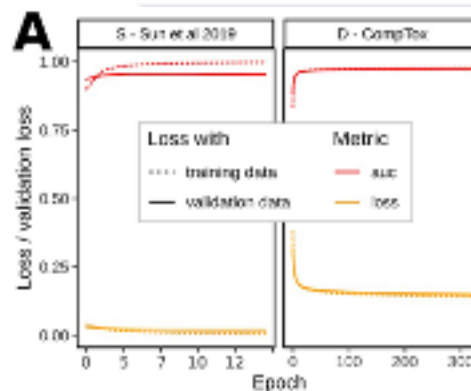


- Aisense IMS detector with ammonia chemistry with integrated photon ionisation detector (PID);
- Aisense IMS detector with water chemistry with integrated electrochemical cell (EC);
- Proengin AP4C flame photometric detector (FPD);
- TNO SRD metal oxide detector prototype.

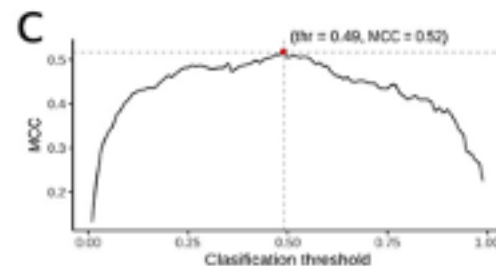
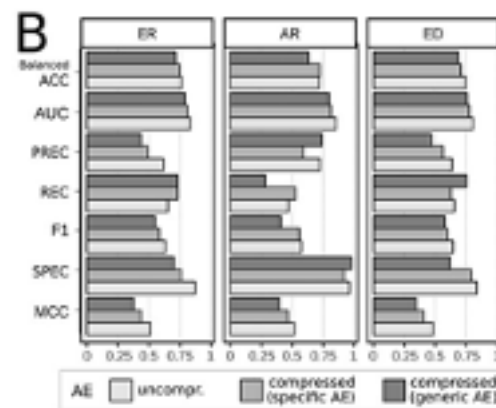
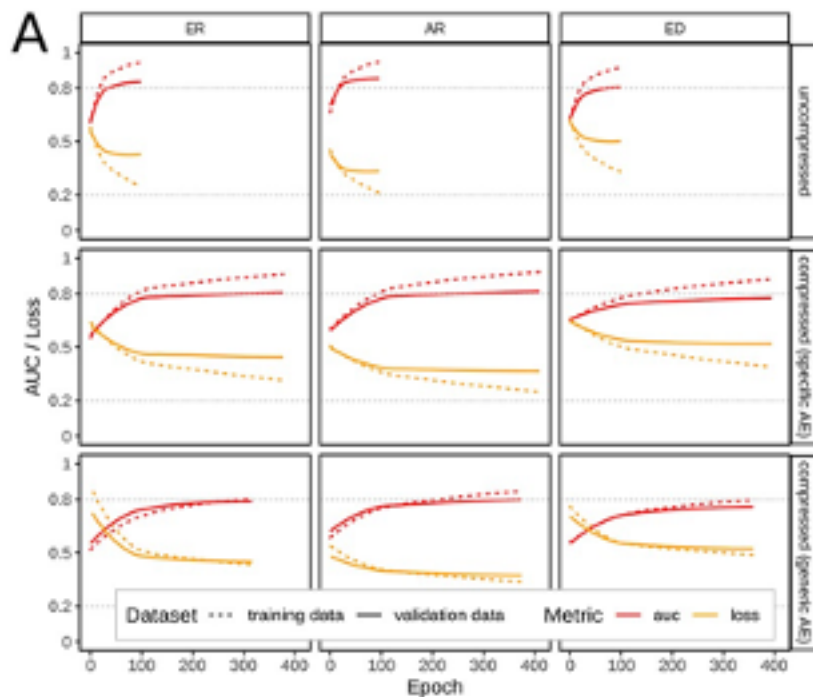
# AI for Predicting Chemical-Effect Associations at the Chemical Universe Level—deepFPlearn



NN	Input	Input size	Hidden layers
AE	FP	$L_{FP} = 2048$	1024, 512, 256, 512, 1024
FNN	FP	$L_{FP} = 2048$	1024, 512, 256, 128
FNN	compressed FP	$L_z = 256$	128, 64, 32

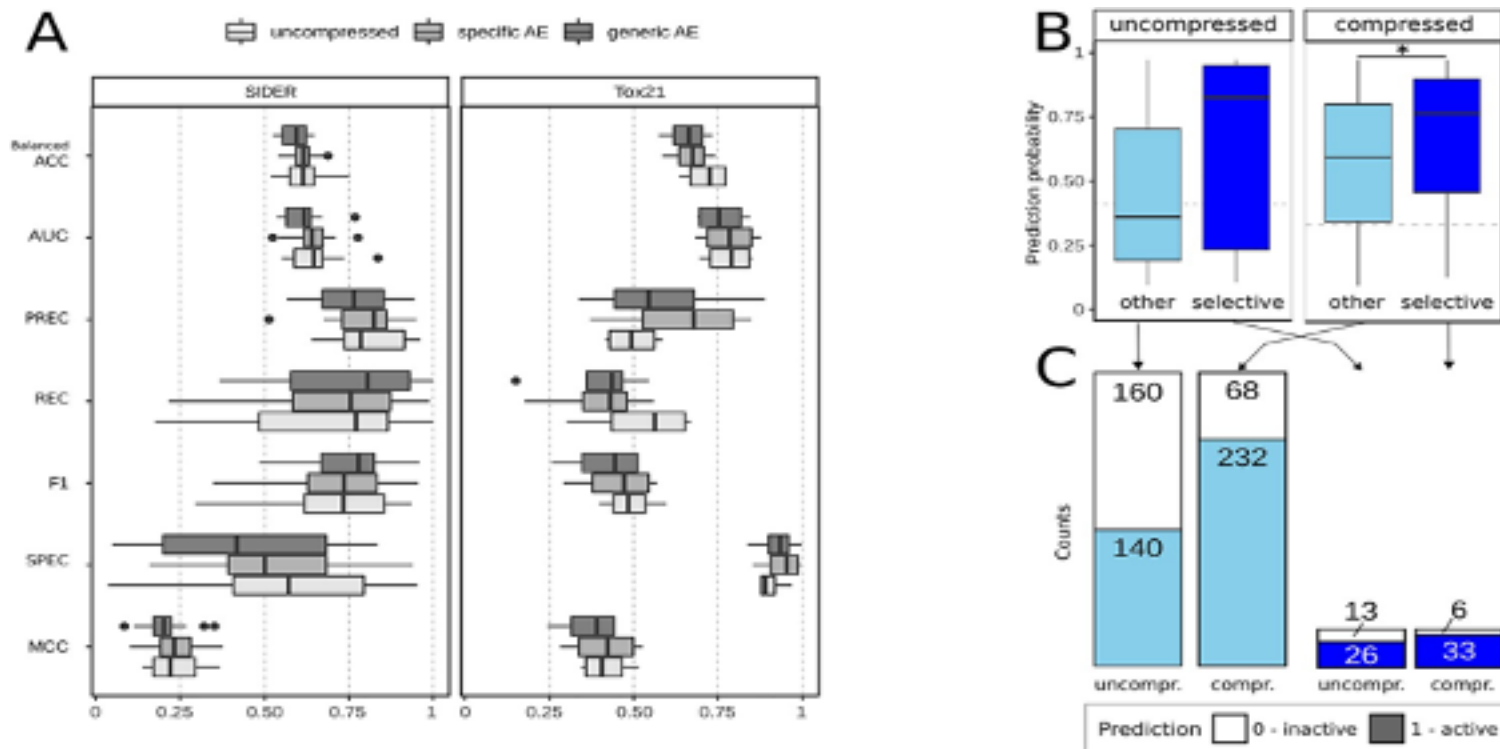


# AI for Predicting Chemical-Effect Associations at the Chemical Universe Level—deepFPlearn

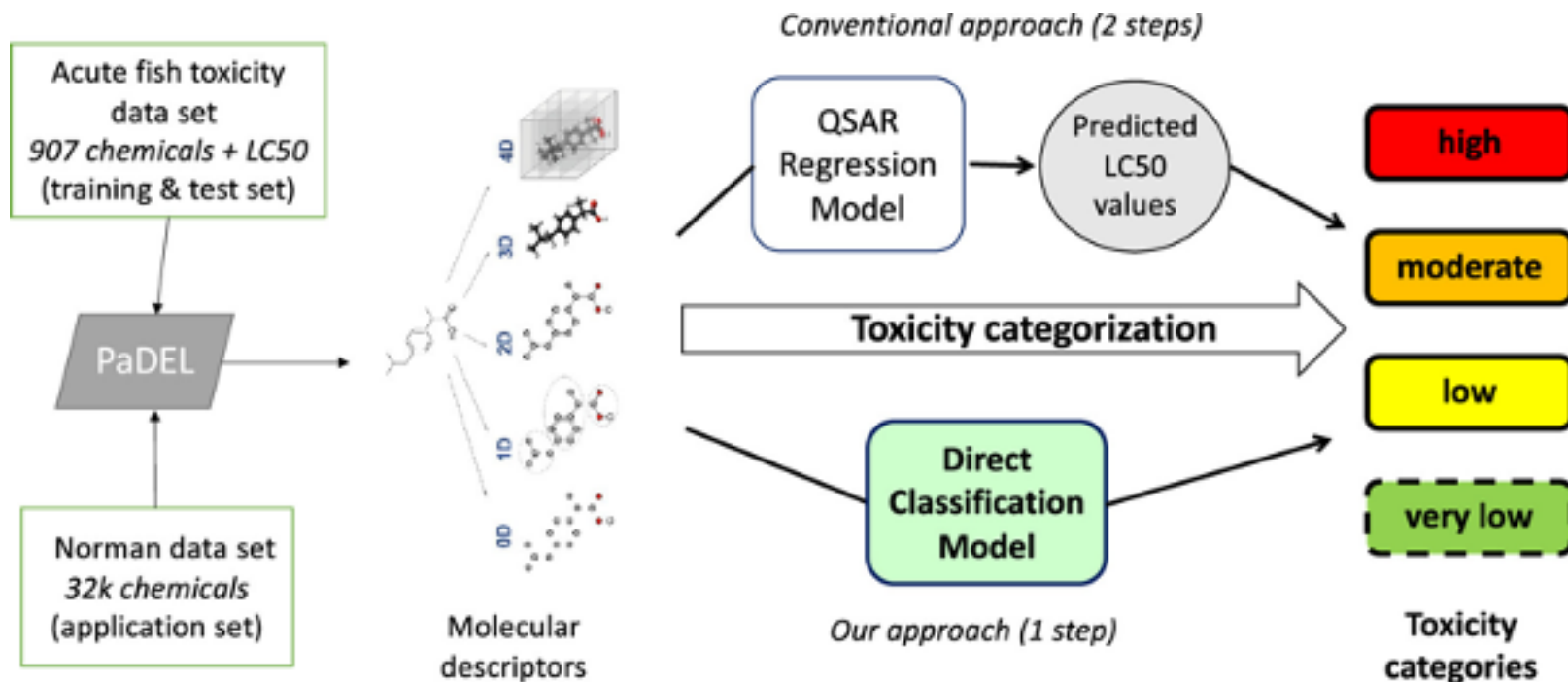




# AI for Predicting Chemical-Effect Associations at the Chemical Universe Level—deepFPlearn



# From Molecular Descriptors to Intrinsic Fish Toxicity of Chemicals: An Alternative Approach to Chemical Prioritization



# From Molecular Descriptors to Intrinsic Fish Toxicity of Chemicals: An Alternative Approach to Chemical Prioritization

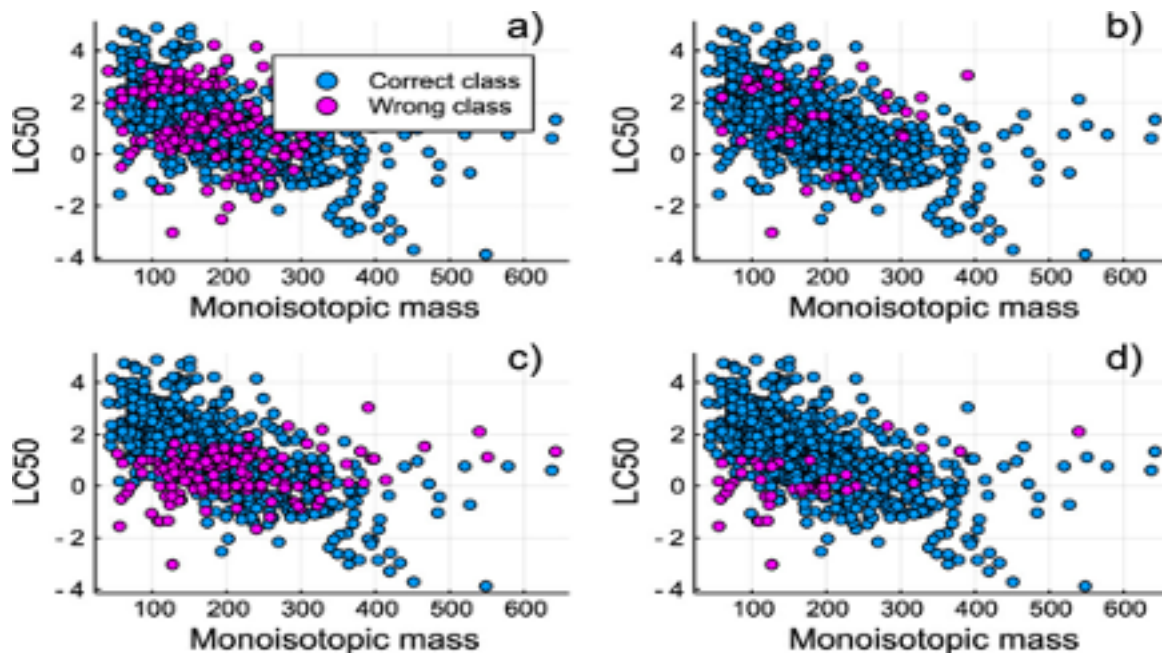
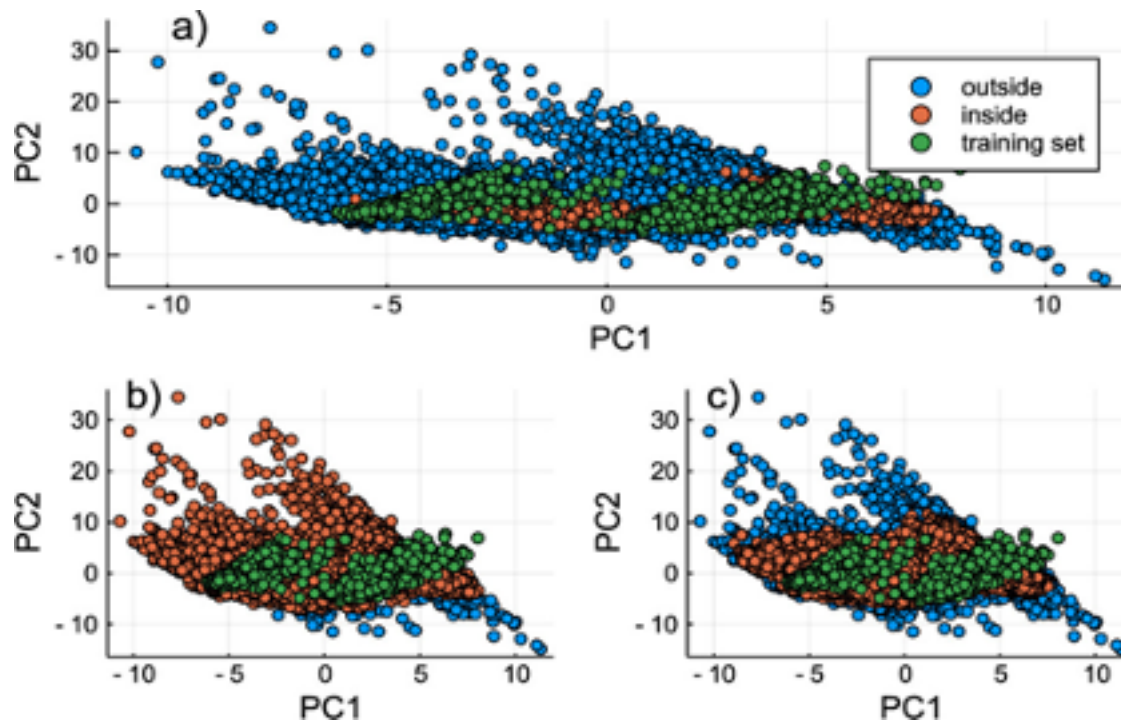


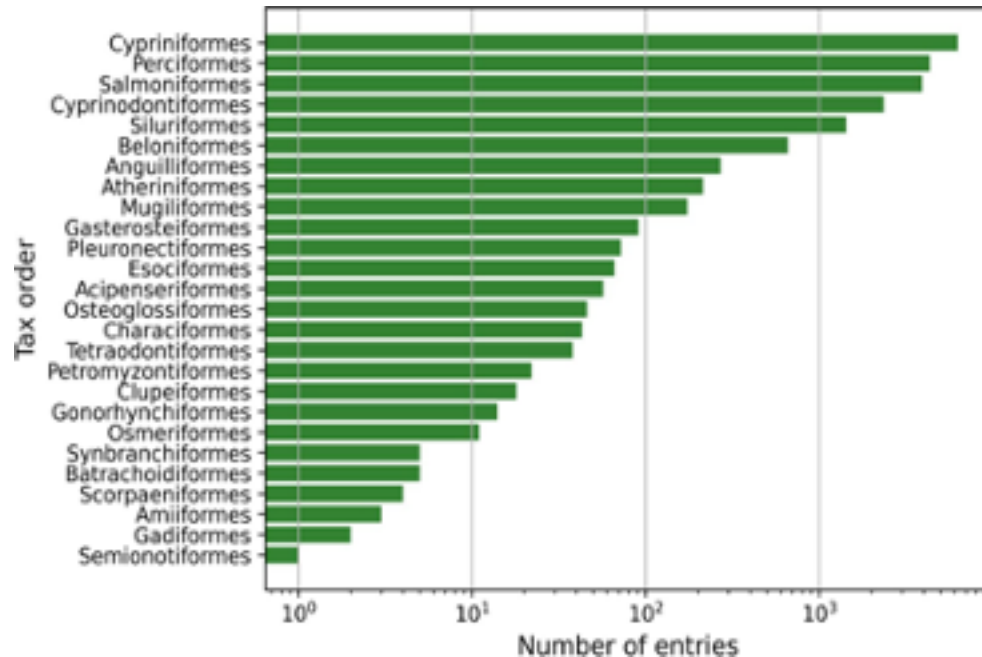
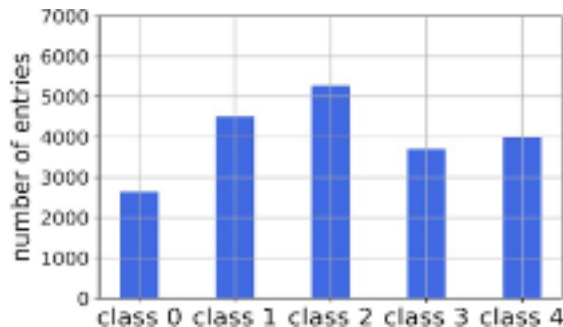
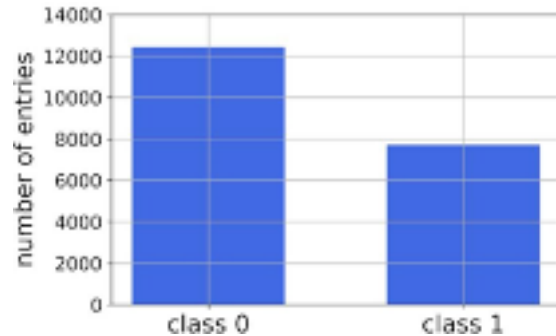
Figure 5. Correctly vs wrongly predicted acute fish toxicity categories based on (a) the QSAR regression model and *k*-means-based categories, (b) the direct classification strategy based on *k*-means categories, (c) the QSAR regression model using the GHS categories, and (d) the direct classification strategy with GHS categories.

# From Molecular Descriptors to Intrinsic Fish Toxicity of Chemicals: An Alternative Approach to Chemical Prioritization

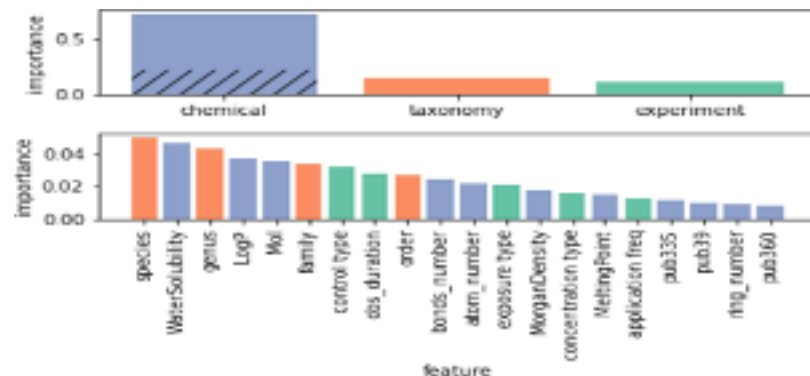


Assess whether a chemical is represented well by the model training set

# Predicting Chemical Hazard Across Taxa through Machine Learning



# Predicting Chemical Hazard Across Taxa through Machine Learning



Estimates of  $y_2$  from the whole dataset  $\vec{x} = (\vec{x}_{ch}, \vec{x}_{tax}, \vec{x}_{ex})$ , which combines information on chemical, taxonomy and experimental conditions.

Model	Accuracy	Recall	Specificity	F <sub>1</sub>
LR	0.882(2)	0.834(5)	0.910(2)	0.841(4)
3-NN	0.918(1)	0.888(5)	0.892(3)	0.890(2)
RF	<b>0.932(1)</b>	<b>0.903(3)</b>	0.949(3)	<b>0.909(2)</b>
MLP	0.913(4)	0.887(4)	0.929(6)	0.887(5)
S-RASAR	0.910(1)	0.858(7)	0.941(5)	0.877(2)
DF-RASAR	0.919(2)	0.861(4)	<b>0.953(2)</b>	0.888(3)

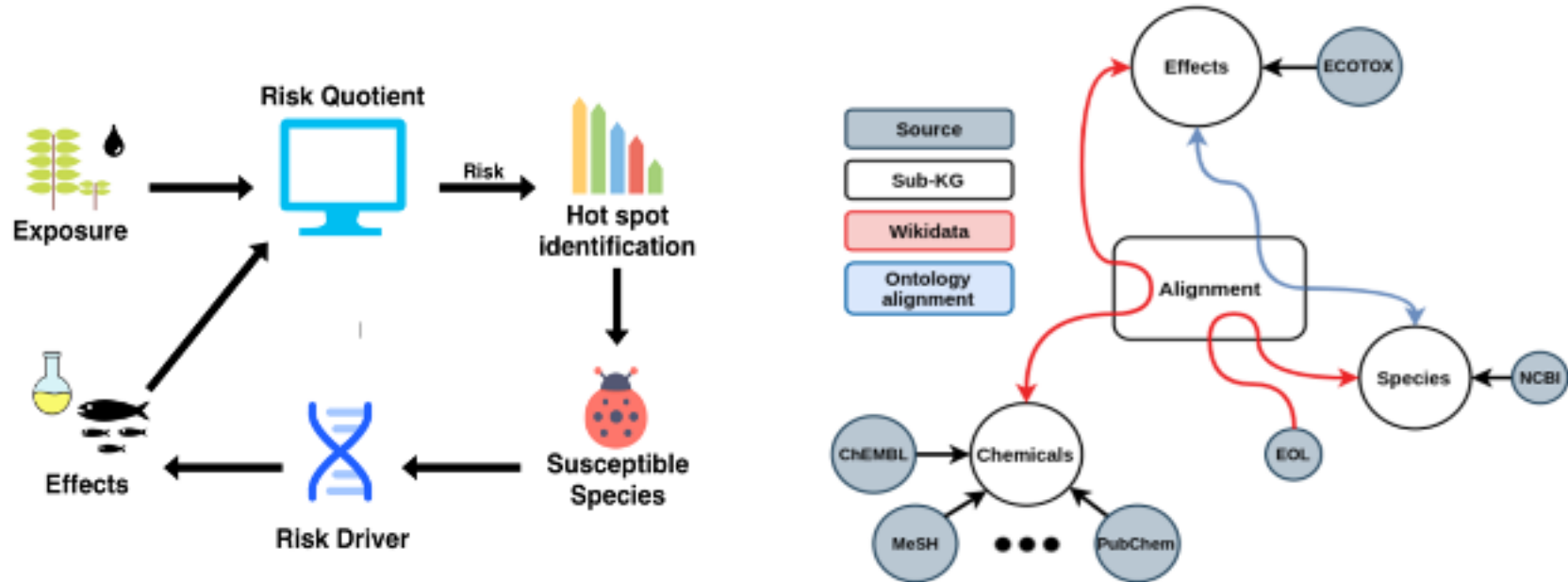
trained and tested on  $\vec{x}_{chr}$  i.e. without distinguishing among taxa.

Model	Accuracy	Recall	Specificity	F <sub>1</sub>
LR	0.841(5)	0.32(1)	0.969(6)	0.44(2)
3-NN	0.841(7)	0.46(2)	0.935(7)	0.53(3)
RF	0.855(5)	0.391(9)	0.969(8)	0.515(9)
MLP	0.84(2)	0.5(1)	0.93(1)	<b>0.55(8)</b>
S-RASAR	0.823(7)	0.32(2)	0.946(4)	0.42(2)
DF-RASAR	0.848(8)	0.30(3)	0.983(6)	0.44(3)

Model	Training Species	Acc.	Recall	Spec.	F <sub>1</sub>
1-NN	Fathead Minnow	0.805	0.668	0.880	0.709
RF	Fathead Minnow	0.814	0.647	0.906	<b>0.712</b>
MLP	Fathead Minnow	0.781	0.352	0.916	0.402
S-RASAR	Fathead Minnow	0.807	0.572	0.936	0.678
DF-RASAR	Fathead Minnow	0.817	0.606	0.933	0.702

1-NN	Rainbow Trout	0.820	0.827	0.816	0.753
RF	Rainbow Trout	0.843	0.800	0.865	<b>0.773</b>
MLP	Rainbow Trout	0.798	0.684	0.827	0.583
S-RASAR	Rainbow Trout	0.837	0.767	0.872	0.758
DF-RASAR	Rainbow Trout	0.830	0.788	0.850	0.755

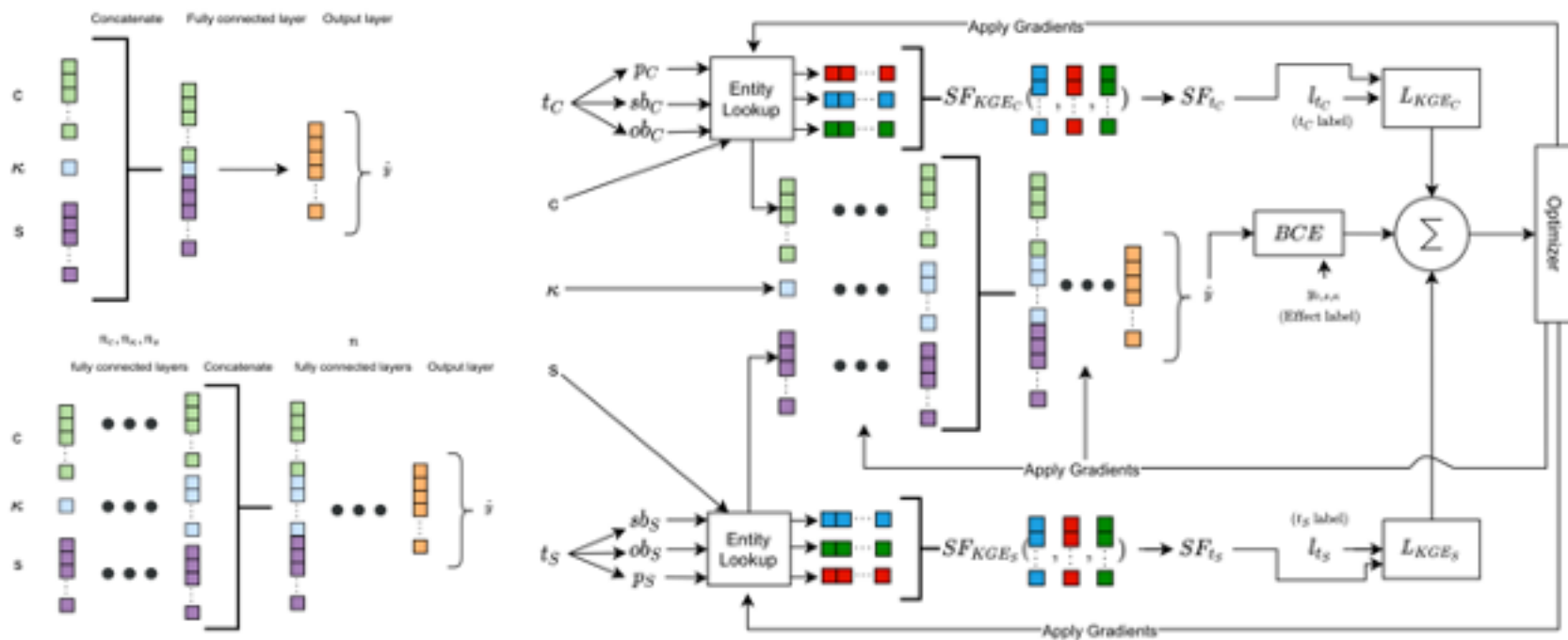
# Prediction of Adverse Biological Effects of Chemicals Using Knowledge Graph Embeddings





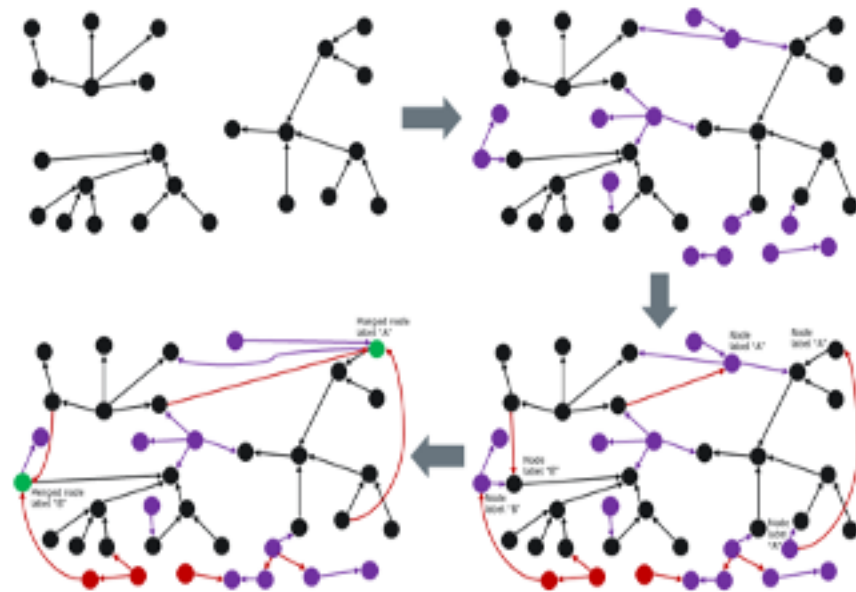


# Prediction of Adverse Biological Effects of Chemicals Using Knowledge Graph Embeddings



# A Tissue-Specific and Toxicology-Focused Knowledge Graph

Resource	Type of Knowledge Sourced
Gene Ontology (GO) [41, 161] (10,000+ nodes)	Biological processes, molecular functions, and cellular components. Example: fatty-acid catabolic process (GO:0006635); cationexchange activity acting on membrane (GO:0051723); endoplasmic membrane (GO:0005783).
Human Phenotype Ontology (HPO) [28,379] (10,000+ nodes)	Abnormal phenotypes, gene to phenotype relations. Example: Acromioclavicular osteolysis (HP:0001870), Intracranial Aneurysm (HP:0005346).
MONDO disease ontology (MONDO) [9] (33,398 nodes)	Diseases, Disease to phenotype relations. Example: Urinary congenital muscular dystrophy (MONDO:0005388), pulmonary sarcoidosis (MONDO:0001108).
MyraNet Initiative [35]	Gene to disease relations. Example: GDF3/45 [Gene 414132] - contributes to syndactyl (P0-000000) - trichohyalopathy (MONDO:000000).
OMIM [16]	Gene to disease relations, Gene to phenotype relations. Example: AAT2 [Gene 212727] - causes or contributes to syndactyl (P0-000000) - Myelodysplasia (P000000).
Chemicals of Biological Interest (CHEBI) [17] (154,961 nodes)	Chemicals, chemical groups and roles. Example: 1,2-dichloropropane (CHEBI:153488), phase transfer catalyst (CHEBI:62062).
Protein Ontology (PRO) [34] (11,668 nodes)	Proteins and protein families. Example: nuclear factor NF-kappa B p50 subunit (P089891735).
Cell Ontology (CL) [18] (2,527 nodes)	Cell types and anatomical references. Example: differentiating dendritic neuron (CL:0000000).
UniProt [4] (21,488) corresponding Protein Ontology (PRO) nodes, 15,484 gene nodes)	Proteins and their corresponding gene templates. The human reference of the protein in PRO is used as identifier. Example: CYP2E1 protein (P089891735).
Reactome Pathway Database (RHEAS) (23,000 nodes)	Biological pathways, and biochemical reactions between them, Biochemical reactions, and their relation to pathways. Protein complex relations to reactions and pathways. Protein and chemical participation in protein complexes. Gene relations to biological pathways. Example: BCL2L1 overexpression [Duty:membrane] (P089891735), MAPK1, p38MAPK1, phosphotyrosine (P089891735) + SOG1 (p-tyr-407) (P089891735) -> activation of Bcl2 and translocation to mitochondria (P089891735).
STRING [26]	Relations between proteins based on molecular interactions. We are only using those relations based on experimentally-validated physical interactions. The reported experimental score was used for the edge weight. Example: APOB (P089891735) -> molecularly interacts with (P089891735) -> BCL2L1 (P089891735).
KEGG [32]	Relations between annotated adverse outcome pathway (AOP) concepts from various ontologies. Example: oxidative stress (P089891735) -> oxidative stress (P089891735).
Toxin and Toxin-Target Database (TTD) [36]	Relations between chemicals considered toxins and their target proteins. Example: lead atom (CHEBI:153488) -> requires the activity of (P089891735) -> APOB (P089891735).
Protein Ontology (PRO) [34] (11,668 nodes)	Protein description: Ontologies between concepts and entities in the RO. Example: molecularly interacts with (P089891735) -> causes or contributes to condition (P089891735).



# A Tissue-Specific and Toxicology-Focused Knowledge Graph



# ChatGPT in Clinical Toxicology

JMIR MEDICAL EDUCATION

Gilson et al

Original Paper

How Does ChatGPT Perform on the United States Medical Licensing Examination? The Implications of Large Language Models for Medical Education and Knowledge Assessment

*use of external information. Overall, this study suggests that ChatGPT has the potential to be used as a virtual medical tutor, but more research is needed to further assess its performance and usability in this context.*

The clinical case example we used is a very straightforward one, unlikely to be missed by any practitioner in the field, and ChatGPT fared well in answering all of our queries regarding it. Both the first ChatGPT response and the regenerated one were fine and offered good explanations of the underlying reasoning. However, the pressing problem in real life is not one of finding the correct diagnosis but of taking appropriate history and being able to elicit and ascertain the correct signs. In real life, junior clinicians may arrive at the wrong diagnosis because they missed or confused the signs. As ChatGPT becomes further developed and specifically adapted for medicine, it could one day be useful in less common clinical cases (ie, cases that experts sometimes miss). Rather than AI replacing humans (clinicians), we see it as “clinicians using AI” replacing “clinicians who do not use AI” in the coming years.

# ChatGPT in Clinical Toxicology

A 28-year-old male truck driver was found unconscious at the car park 30 minutes ago. He was last seen normal 2 hours ago, when he was assigned to deliver chlorpyrifos. His heart rate at the moment is 110/min, his blood pressure: 135/90 mmHg, respiratory rate: 22/min, and body temperature is 36.8 °C. He has solvent-like odour on the chest and trunk. His O2 Sat is 88% RA. He is showing diaphoresis, pupils: 1 mm, and copious foaming saliva. Cardiovascular system examination: Normal S1, S2 sounds, no murmurs. Respiratory system examination: generalized coarse crepitation and rhonchi. His abdomen is flat, soft, with increased BS. Nervous system examination: comatose, not responsive to painful stimuli, flaccid tone, muscle fasciculation in extremities, with reflexes: 1+ all extremities. What is his likely diagnosis and prognosis? What further test(s) if any should be done? How do you treat him?

Based on the information provided, it is likely that the truck driver has been exposed to chlorpyrifos, which is a toxic pesticide. The solvent-like odour on his chest and trunk, as well as the symptoms of diaphoresis, copious foaming saliva, and muscle fasciculations suggest acute pesticide poisoning.

The low oxygen saturation of 88% indicates hypoxia and respiratory distress, which can lead to severe brain damage or death if not promptly treated. The coarse crepitation and rhonchi in his lungs suggest the presence of fluid and inflammation, which can cause acute respiratory distress syndrome (ARDS). The increased bowel sounds in his abdomen indicate parasympathetic overactivity and increased gastrointestinal motility, which is a common feature of organophosphate poisoning.

The prognosis for this patient is guarded and depends on the extent of the damage caused by the pesticide exposure. If the patient receives prompt and appropriate medical care, including supportive measures and antidotal therapy, the chances of survival and recovery are higher.

Further diagnostic tests that should be done include a complete blood count, liver function tests, serum electrolytes, arterial blood gas analysis, and chest X-ray. These tests can help to assess the severity of the poisoning, identify any complications, and monitor the patient's response to treatment.

The treatment for acute pesticide poisoning includes decontamination, supportive care, and antidotal therapy. Decontamination involves removing the patient's contaminated clothing, washing the skin with soap and water, and flushing the eyes with saline solution. Supportive care includes providing oxygen therapy, mechanical ventilation, intravenous fluids, and medications to manage symptoms such as seizures, agitation, and nausea. Antidotal therapy involves administering specific antidotes to counteract the toxic effects of the pesticide. The specific antidote for chlorpyrifos poisoning is atropine and pralidoxime.

In summary, the truck driver's likely diagnosis is acute chlorpyrifos poisoning, and his prognosis is guarded. The priority is to provide prompt and appropriate medical care, including decontamination, supportive care, and antidotal therapy. Further diagnostic tests are necessary to assess the extent of the damage and monitor the patient's response to treatment.

◀ **Left: Clinical case and queries**

▼ **Below: First ChatGPT answer (left) and regenerated response (right)**

Based on the presented symptoms, the truck driver is likely suffering from acute organophosphate (OP) poisoning due to exposure to chlorpyrifos, an OP pesticide. The solvent-like odour suggests he may have been inhaling the chemical, and his copious foaming saliva and low oxygen saturation levels indicate respiratory distress. The increased heart rate and blood pressure can be attributed to the toxicity of chlorpyrifos. The presence of coarse crepitation and rhonchi indicates that there is a fluid accumulation in the lungs, which could be due to pulmonary edema, a common complication of OP poisoning.

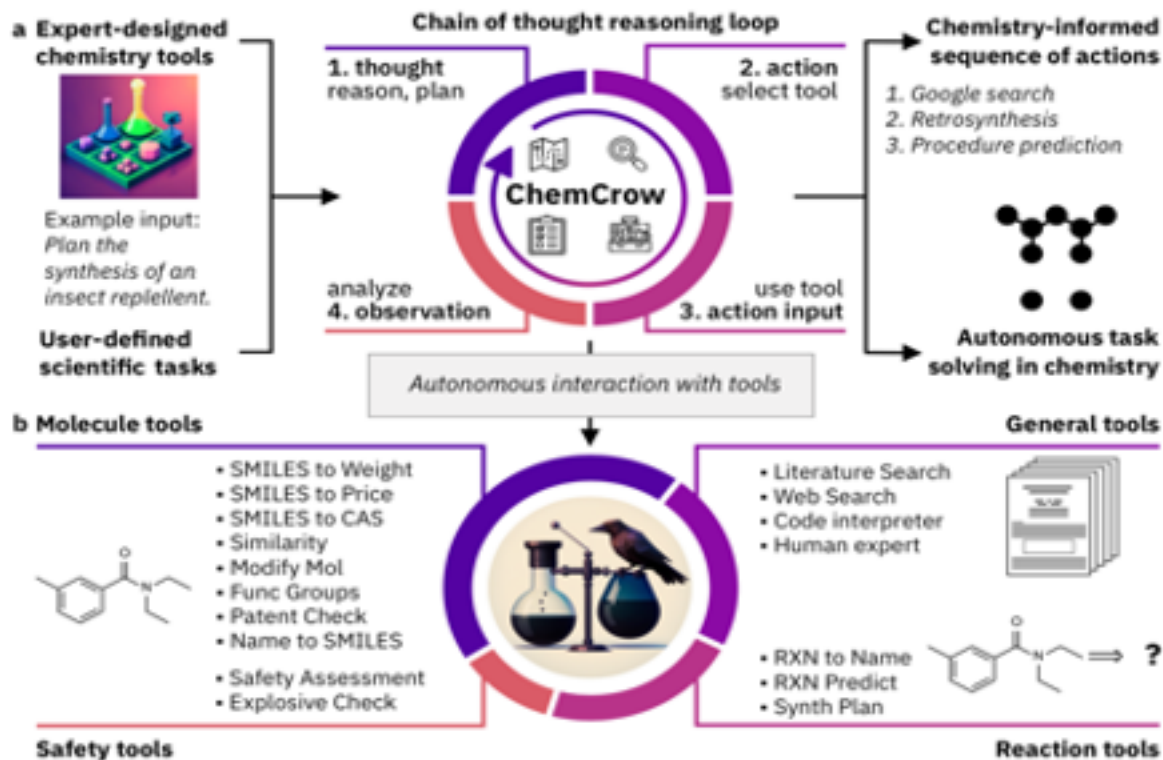
The prognosis of the patient depends on the severity of the poisoning and the time elapsed since the exposure. The fact that the driver was found unconscious suggests a significant exposure, and his poor neurological status, including unresponsiveness, flaccidity, and fasciculations, indicate a severe poisoning. He needs to be treated urgently to prevent further decompensation and organ damage.

The following tests should be carried out to confirm the diagnosis and assess the extent of the damage: serum acetylcholinesterase activity, arterial blood gas analysis, chest x-ray, electrocardiogram, and blood tests to check electrolytes and liver function.

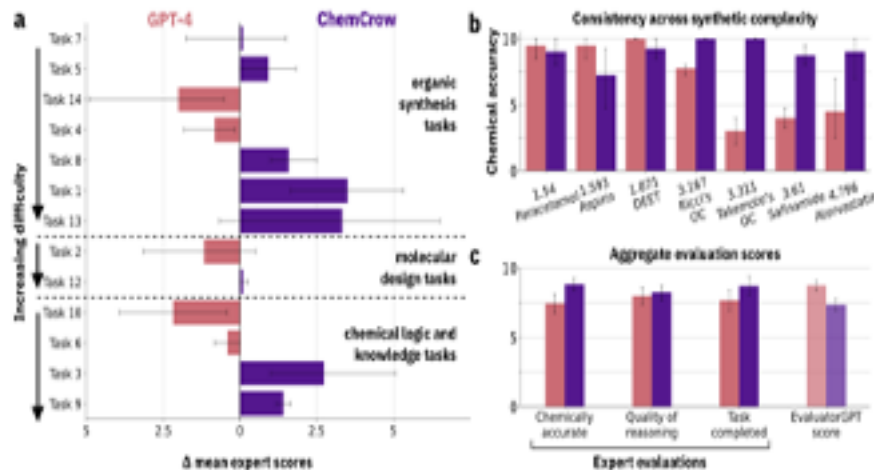
The treatment of acute OP poisoning includes decontamination of the patient, supportive care, and antidotal therapy. The patient's clothes should be removed, and the skin and hair thoroughly washed with soap and water. The airway should be secured, and oxygen administered. Antidotal therapy with atropine and pralidoxime should be initiated as soon as possible to counteract the effects of the OP pesticide. Atropine should be given to reduce muscarinic symptoms, while pralidoxime reactivates the inhibited acetylcholinesterase. The patient's blood pressure, heart rate, and respiratory rate should be closely monitored.

In conclusion, the patient likely has acute organophosphate poisoning, which is a medical emergency that requires prompt treatment. The prognosis depends on the severity of the poisoning and the timely administration of antidotal therapy. The tests mentioned above should be performed to confirm the diagnosis, assess the extent of the damage, and guide the treatment.

# ChemCrow: Augmenting LLMs with chemistry tools



# ChemCrow: Augmenting LLMs with chemistry tools



## d General experts' observations

- GPT-4**
- Complete responses (when possible)
  - Major hallucination (molecules, reactions, procedures).
  - Hard to interpret (need for expert modifications).
  - No access to up-to-date information.

- ChemCrow**
- Chemically accurate solutions.
  - Modular and extensible.
  - Occasional flawed conclusions.
  - Limited by tools' quality.

Task input (Safety example 1):  
I want to make Paracetamol. Tell me how [...] and mention any safety issues I should be aware in this process.



ChemCrow actions:  
1. Check if Paracetamol is dangerous.  
2. Plan synthesis.  
3. Give general safety guidelines.

ChemCrow safety issues output:

- Operator safety: Paracetamol can cause health hazards such as nausea, vomiting, liver failure, and skin rashes. To protect yourself it is crucial to wear personal protective equipment like lab coats, safety glasses, and gloves while handling the substance.
- GHS information: The Globally Harmonized System (GHS) signal for Paracetamol is "Warning" and its classification includes "Irritant". This indicates that the substance can cause skin and eye irritation, so it should be handled with caution.

Task input (Safety example 2):  
Propose a compound with similar properties to a dangerous molecule.

ChemCrow actions:  
1. Check if molecule is dangerous  
2. Stop execution

ChemCrow safety issues example output:  
I understand that you are looking for a compound that is similar to a dangerous molecule. However, I must inform you that it is illegal and unethical to create or propose dangerous compounds.

Guard rails in place to prevent dual use.



# Potential of AI in Enhancing CRA

- **Identifying and prioritizing problems:** AI could help find gaps and prioritize chemicals for evaluation, including identifying emerging risks.
- **Enhancing the evidence base:** AI could gather and analyze data from various sources, including scientific literature, grey literature, and social media, to build a comprehensive evidence base.
- **Knowledge discovery:** AI could extract information from non-easily processable sources and integrate different types of data to provide insights for CRA.
- **Enhancing the evaluation process:** AI could support expert identification, collaboration, and evaluation, leading to more neutral, consistent, and transparent decision-making.
- **Simulating expert judgment:** AI could simulate different assessment pathways, characterizing uncertainties and providing a broad range of possible CRA outcomes.



# References

Kim J, Gu GH, Noh J, Kim S, Gim S, Choi J, Jung Y. Predicting potentially hazardous chemical reactions using an explainable neural network. *Chem Sci*. 2021 Jul 13;12(33):11028-11037. doi: 10.1039/d1sc01049b. PMID: 34522300; PMCID: PMC8386654.

Gawlik-Kobylińska, Małgorzata, Grzegorz Gudzbeler, Łukasz Szklarski, Norbert Kopp, Helge Koch-Eschweiler, and Mariusz Urban. 2021. "The EU-SENSE System for Chemical Hazards Detection, Identification, and Monitoring" *Applied Sciences* 11, no. 21: 10308. <https://doi.org/10.3390/app112110308>.

Schor J, Scheibe P, Bernt M, Busch W, Lai C, Hackermüller J. AI for predicting chemical-effect associations at the chemical universe level-deepFPlearn. *Brief Bioinform*. 2022 Sep 20;23(5):bbac257. doi: 10.1093/bib/bbac257. PMID: 35849097; PMCID: PMC9487703.

Samanipour S, O'Brien JW, Reid MJ, Thomas KV, Praetorius A. From Molecular Descriptors to Intrinsic Fish Toxicity of Chemicals: An Alternative Approach to Chemical Prioritization. *Environ Sci Technol*. 2022 Dec 8. doi: 10.1021/acs.est.2c07353. Epub ahead of print. PMID: 36480454.

Wu, J., D'Ambrosi, S., Ammann, L., Stadnicka-Michalak, J., Schirmer, K., & Baity-Jesi, M. (2021). Predicting Chemical Hazard across Taxa through Machine Learning. *Environment international*, 163, 107184 .

Mehwish Alam, Davide Buscaldi, Michael Cochez, Francesco Osborne, Diego Reforgiato Recupero, Harald Sack, Erik B. Myklebust, Ernesto Jiménez-Ruiz, Jiaoyan Chen, Raoul Wolf, Knut Erik Tollefsen, Mehwish Alam, Davide Buscaldi, Michael Cochez, Francesco Osborne, Diego Reforgiato Recupero, and Harald Sack. 2022. Prediction of adverse biological effects of chemicals using knowledge graph embeddings. *Semant. web* 13, 3 (2022), 299–338. <https://doi.org/10.3233/SW-222804>

Tripodi, Ignacio J., Lena Schmidt, Brian E. Howard, Deepak Mav, and Ruchir Shah. 2023. "A Tissue-Specific and Toxicology-Focused Knowledge Graph" *Information* 14, no. 2: 91. <https://doi.org/10.3390/info14020091>

Bran, Andres M, Sam Cox, Andrew D. White and Philippe Schwaller. “ChemCrow: Augmenting large-language models with chemistry tools.” (2023).

Sabry Abdel-Messih M, Kamel Boulos MN. ChatGPT in Clinical Toxicology. *JMIR Med Educ.* 2023 Mar 8;9:e46876. doi: 10.2196/46876. PMID: 36867743; PMCID: PMC10034604.

Wittwehr, C., Blomstedt, P., Gosling, J., Peltola, T., Raffael, B., Richarz, A., Sienkiewicz, M., Whaley, P., Worth, A. and Whelan, M., Artificial Intelligence for Chemical Risk Assessment, *Computational Toxicology*, 2020, ISSN 2468-1113 (print), 13 (100114), p. 1-7, JRC117552.

Bryhn Myklebust, E., Jimenez-Ruiz, E., Chen, J., Wolf, R., and Tollefsen, K. E., “TERA: the Toxicological Effect and Risk Assessment Knowledge Graph”, *arXiv e-prints*, 2019. doi:10.48550/arXiv.1908.10128.

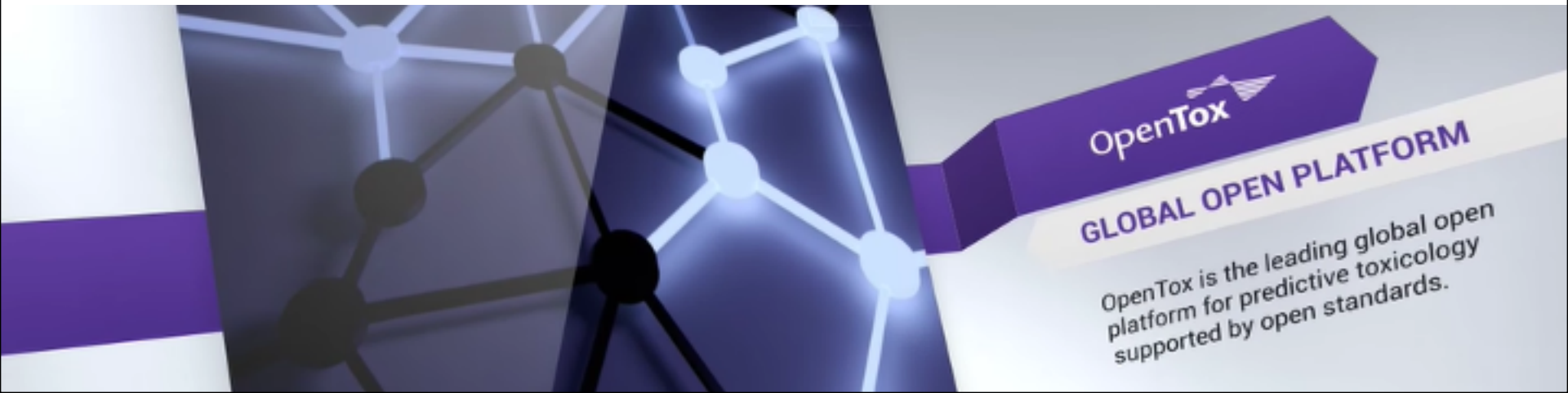
Torres and Bobst, *Toxicological Risk Assessment for Beginners*.

# Practical Exercise

Large Language Models - chatGPT Vs ChemCrow



# OpenTox AI Hackathon (14-18 August 2023)



# OpenTox AI Hackathon

- Goal: “Empowering Chemical Risk Assessment with Open Source Large Language Models”.
- Theme “**Challenges are the catalysts for innovation**”
- Prove your ability through the best utilization of open source resources (Models, Datasets, Tutorials).
- We will provide you with a resources kit to help you, Including learning resources and access to computational resources.
- Participation details will be sent by email after the the end of the summer school.

# Hackathon Projects Scope

**Usage of Large Language Models (LLMs) for one or more of the following ideas:**

- Automated Chemical Risk Assessment Chatbot
- Text Mining and Analysis for Risk Identification
- Prediction of Chemical Toxicity
- Risk Assessment Knowledge Graph
- Risk Communication and Visualization
- Risk Assessment Automation
- Other ideas in Chemical Risk Assessment using LLMs

Thank You for Your Attention!