

Navigating Potentials and Pitfalls in using Artificial Intelligence for Expedited Decision Making in Pharmaceutical Problems

Rob Smith, Ph.D.
Founder, CEO



Do what you want. Enjoy what you do.™

My Background

- CS undergrad, MS in CS (AI focus), PhD in CS (AI focus ~ applications in computational mass spectrometry)
- 8 years teaching AI, conducting academic research in AI for computational mass spectrometry.
- Many papers and grants evaluating existing algorithms, creating data sets to facilitate evaluations, creating new ones.
- Prime Labs: 4 years. We build software that leverages AI for computational mass spectrometry.



AI Crash Course



AI is not a complicated idea

- You might need deep knowledge to advance it, but not to use or understand it.
- In a nutshell:
 - A program is a recipe to solve a problem.
 - AI is a recipe that creates a recipe to solve a problem.

Obfuscation makes it harder

- There is a lot of jargon.
 - AI
 - Machine learning
 - Deep learning
 - Big data
 - Knowledge discovery
- Mostly marketing / job security.



In a nutshell

- You are building *a model*.
- A putative relationship between caus(es) and effect(s).
- Can be used to predict the future or understand the past.
- Prediction ex: Test the toxicological properties of an a new molecule.
- Understand ex: Devise plausible explanations for observed toxicological response of a tested molecule.

AI: The good and bad

- The good news: It is straightforward to apply AI to well-understood problems.
- The bad news: It is extraordinarily difficult to do something new.
- Why? Surprisingly, not because of lack of tools.
- Instead, the problem lies in *data* and *representation*.

Representation



Representation

- Standard approaches to AI funnel problems into a specific representation. Example: spreadsheet for off-the-shelf machine learning techniques.
- But nature is not found in spreadsheets....

Representation

- Machines do well with defined, repetitive tasks.
- People do well with creativity and intuition.

Representation

- Everything is simple if you know the representation that makes it so.
- Discovery isn't so much about solving problems as it is finding the representation that makes the solution obvious.
- Discovery isn't so much about doing what others can't do, but about seeing what others haven't seen. More often than not, this requires *not* seeing what others see.

Representation

- Take away: In order to succeed in applying AI to a novel task, you will need at least one *very* clever human.

Data



Data

- Things AI does well:
 - Problems for which the answer is known.
 - Problems for which there exist large quantities of data.
- Why is this a problem?

The AI Data Paradox

- The more complex the problem, the more useful AI is, BUT ALSO:
 - The more expensive data is.
 - The less accessible data is.
 - The less certain you are of the content, completeness, and meaning of your observations.

Obstacles to Data

- Must be digital.
- Must be scalable.
- Must have sufficient *breadth* (data sets) and *depth* (observed variables) to capture causative components.

Data

- Take aways:
 - You need AI because it is hard to generate large volumes of correct data, and you need large volumes of correct data to create AI.
 - In order to create large volumes of correct data, you will need *quite a few* clever humans.

The Way Forward



Practical Approaches

- Digitize the manual process.
- Create innovative tools to expedite human decisions.
- Automate or semi-automate where possible.
- Facilitate fast human feedback.

Prometheus by Prime Labs

- Extract information from data.
- Apply state of the art algorithms.
- Provide fast/easy ways for humans to find and correct discrepancies.

Prometheus by Prime Labs

- Ever-increasing volumes of training data, supplied by doing what needs doing anyway.
- Feedback loop for better solutions over time.
- Unlimited data storage and processing.

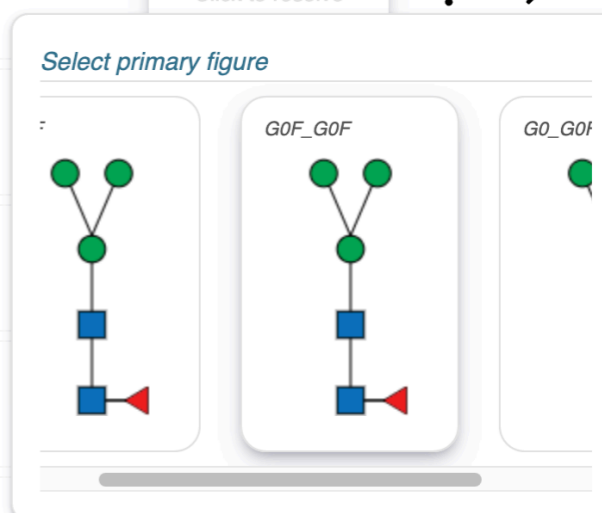
Example: Glycan Structure Elucidation

PROMETHEUS STUDIES IMPORT PROCESS REPORT DEV GLYCOPROTEOMICS TUTORIAL PRIME LABS, I...

ID	Min. Retention Time	Max. Retention Time	Max. Intensity	Figure
14	2105.995000000002	2105.995000000002	5059166	Click to resolve
13	2099.476000000002	2099.476000000002	4956196	Click to resolve
10	2027.934000000000002	2031.19399999999802	6594753	Click to resolve
9	2008.37599999999802	2018.155000000002	5461126	Click to resolve
8	1992.07900000000197	1992.07900000000197	5843492	Click to resolve
5	1893.34800000000002	1926.973000000002	13162266	Click to resolve
4	1890.340000000002	1890.340000000002	11048172	Click to resolve
1	1742.9699999999998	1762.525999999998	6031933	Click to resolve
15	2122.294000000002	2131.981999999998	8216842	Click to resolve

Select primary figure

GOF_GOF



The image shows a screenshot of a software interface for glycan structure elucidation. At the top, there is a navigation bar with tabs for 'PROMETHEUS', 'STUDIES', 'IMPORT', 'PROCESS', 'REPORT', 'DEV', 'GLYCOPROTEOMICS', 'TUTORIAL', and 'PRIME LABS, I...'. Below the navigation bar is a table with columns for 'ID', 'Min. Retention Time', 'Max. Retention Time', 'Max. Intensity', and 'Figure'. The table contains 10 rows of data. A modal window titled 'Select primary figure' is open over the table, showing three glycan structure diagrams. The first diagram is a branched structure with two green circles at the top, a green circle in the middle, a blue square below it, and a red triangle at the bottom. The second diagram is labeled 'GOF_GOF' and shows a similar structure. The third diagram is labeled 'GO_GOF' and shows a single green circle. At the bottom right of the table, there is a small glycan structure diagram with yellow, blue, and green nodes.

In Summary

- AI *is* amazing, and it will continue to drive innovation.
- But:
 - You can't get something for nothing.
 - Good AI isn't cheap or easy.
 - The next advancement in AI will probably not come from institutional sources.

Questions?

Rob Smith

rob.smith@primelabs.ms



Do what you want. Enjoy what you do.™

