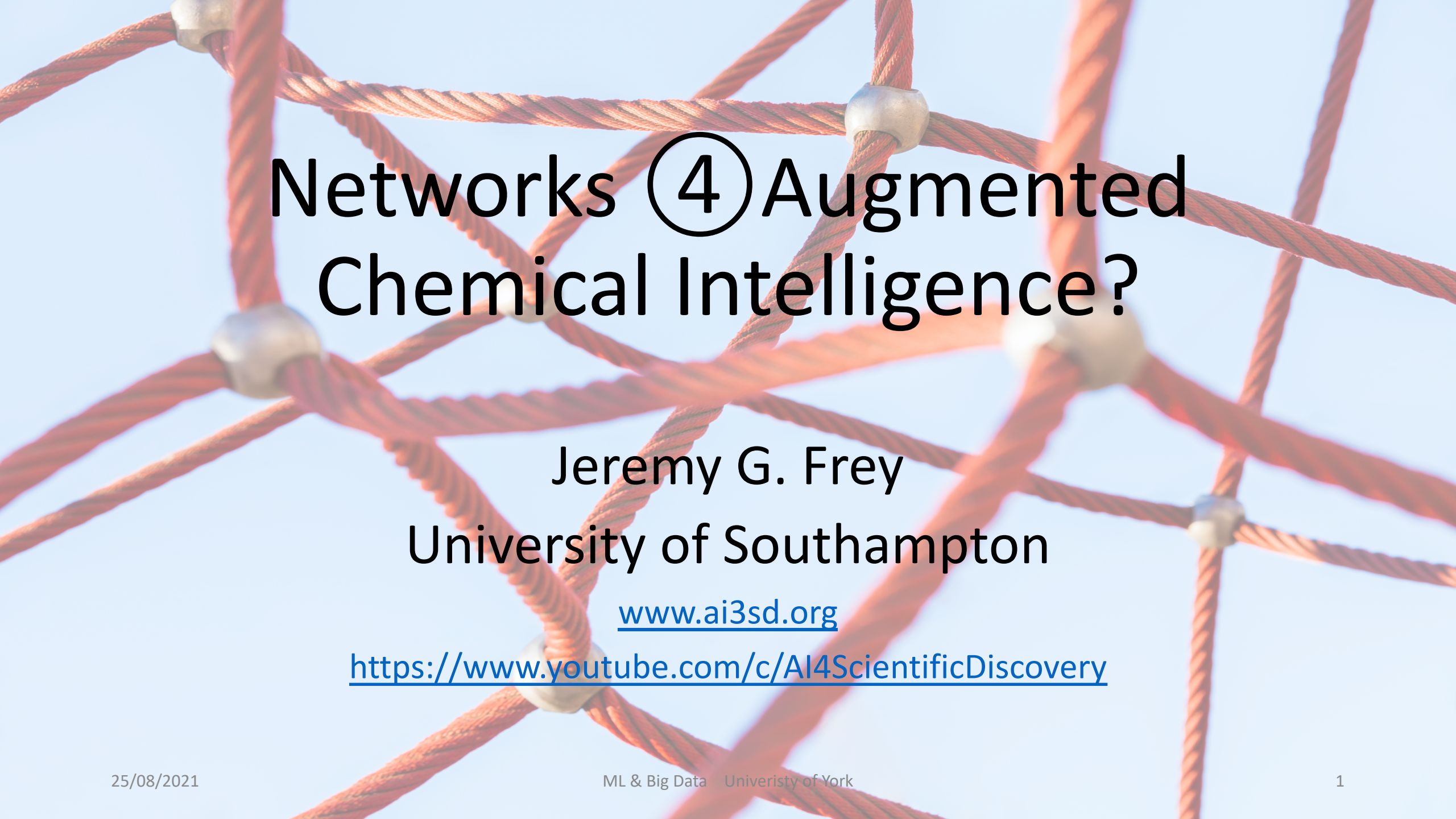


Networks ④ Augmented Chemical Intelligence?

Jeremy G. Frey

University of Southampton

www.ai3sd.org

https://www.youtube.com/c/AI4ScientificDiscovery

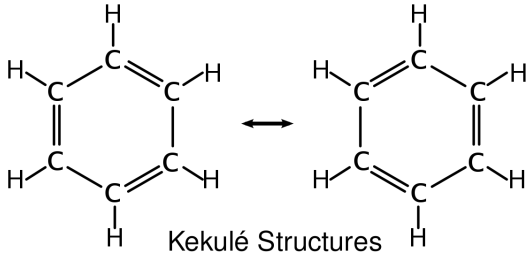
Networks ④ Augmented Chemical Intelligence?

- Surely Chemists have been building models for ages
- Representing molecules – Chemical Informatics
- Group additivity – Graphs – Networks
- What is AI/ML
- Barriers
- Limits

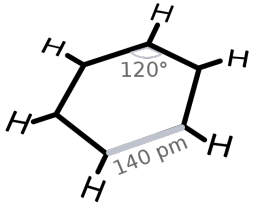
Aromaticity – why is chemistry hard for computers…..

C_6H_6

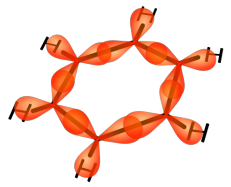
Benzene
Molecular formula



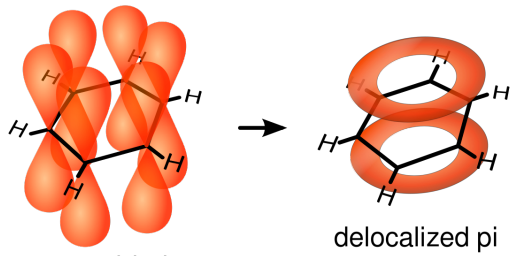
Kekulé Structures
(Resonance Forms)



Planar Hexagon
Bond Length 140 pm

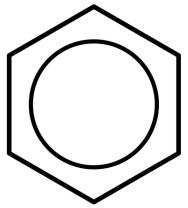


Sigma Bonds
sp^2 Hybridized orbitals



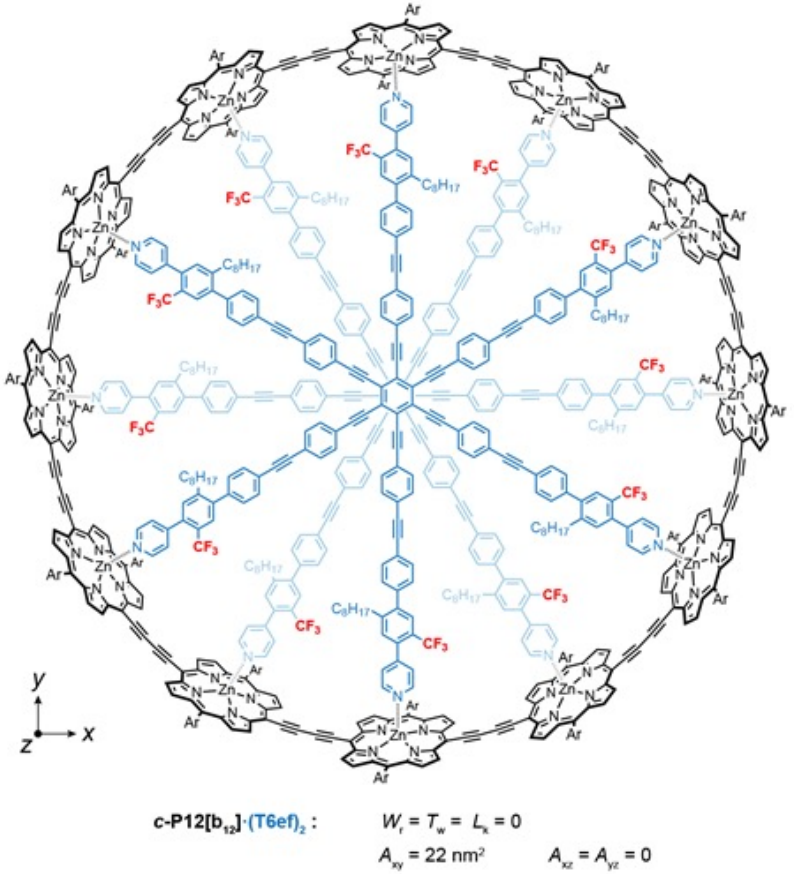
6 p_z orbitals

delocalized pi system



Benzene ring
Simplified depiction

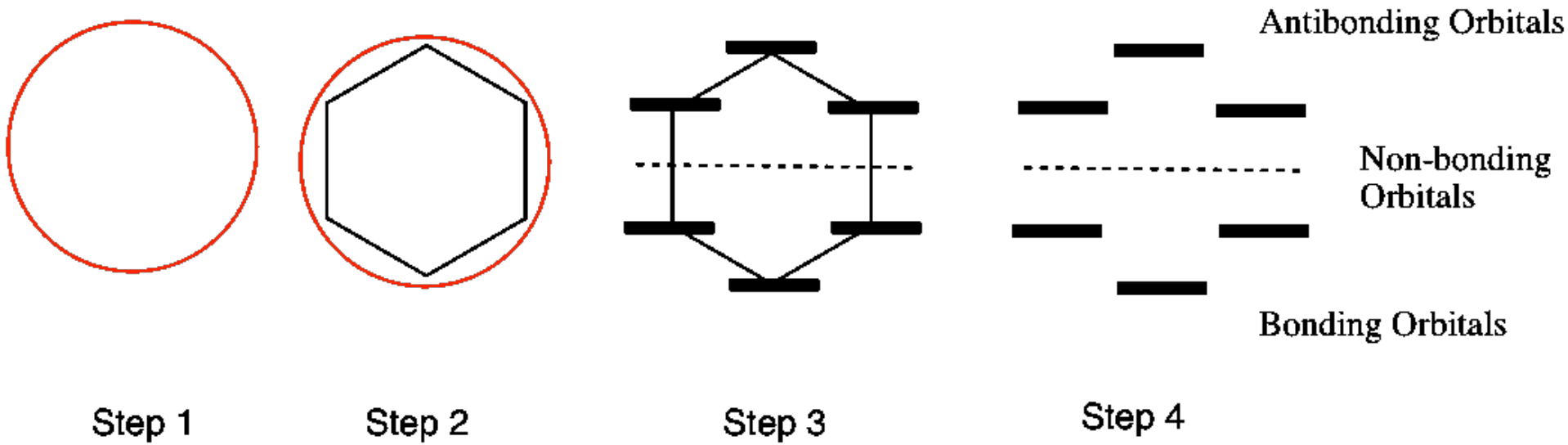
https://en.wikipedia.org/wiki/Benzene#/media/File:Benzene_Representations.svg



https://www.chemistryworld.com/features/the-search-for-the-grand-unification-of-aromaticity/4013915.article

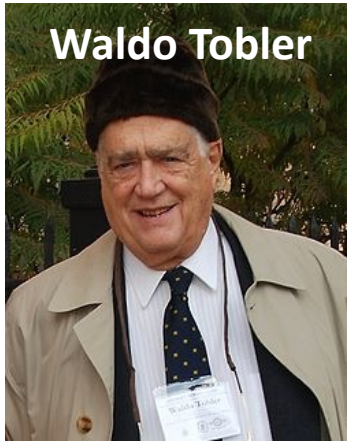
Representations of Molecular Structure

- Hückel Theory & Graph Theory



Chemical Informatics

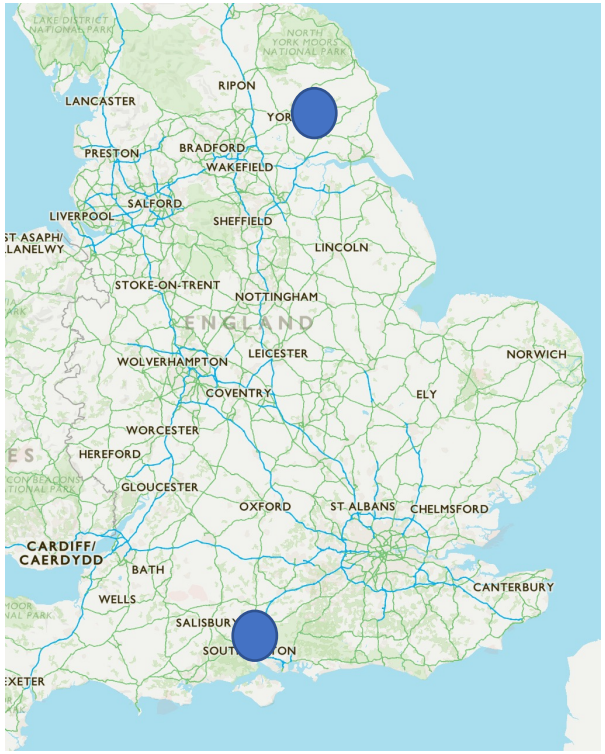
- These graphs are key to getting chemical information in and out of computers
- Allows us to process chemistry in software
- But simple input using SMILES makes life much easier
  - Benzene is c1ccccc1
- Packages like RDKit can then deal with conversion to structures in 2D and 3D and some properties
- *But what we really need is a map….*



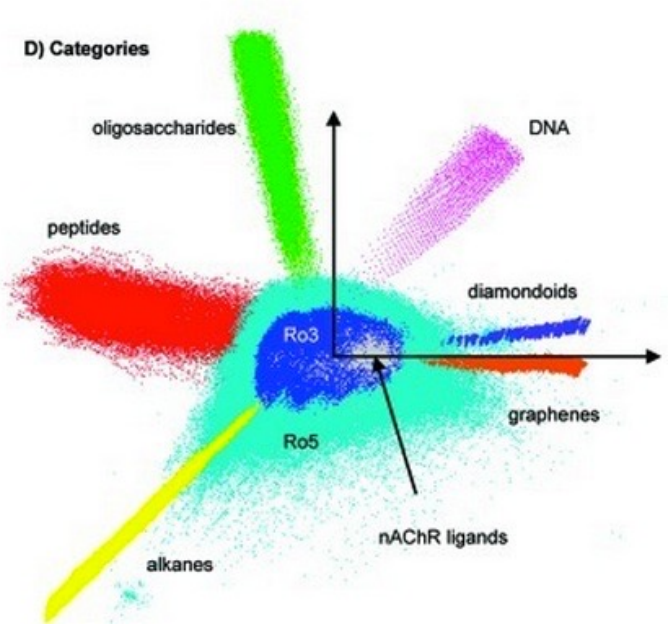
Laws of Geography

- **The First Law**
  ”Everything is related to everything else, but near things are more related than distant things”.
- **The Second law**
  "The phenomenon external to an area of interest affects what goes on inside".

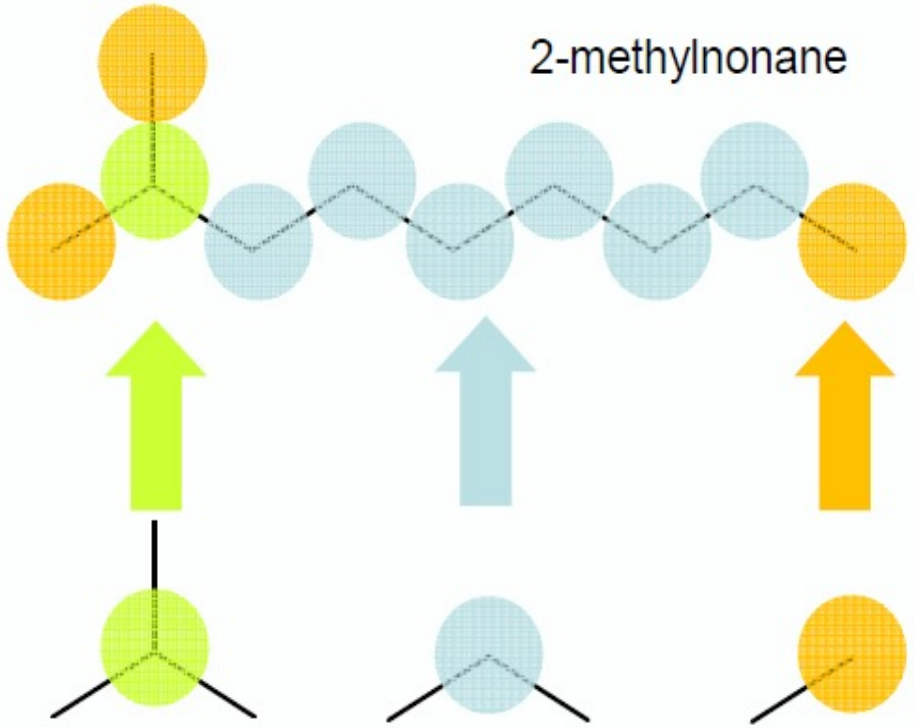
maps



Chemical Space



Possible route to a map – Group additivity

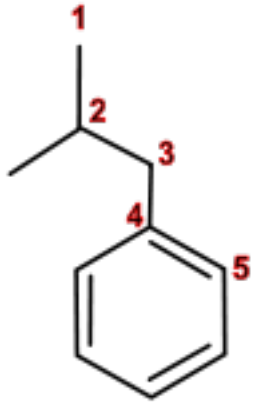


Atoms in large molecules

Group additivity

Atoms in small molecules with similar surroundings

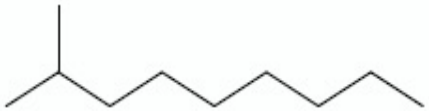
Benson's Rules



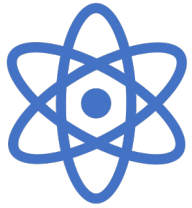
1) C -(C)(H)$_3$ . . . . . 2(-10.20)
2) C -(C)$_3$(H) . . . . . -1.90
3) C -(C$_B$)(C)(H)$_2$ . . . . . -4.86
4) C$_B$ -(C) 5.51
5) C$_B$ -(H) . . . . . 5(3.30)

-5.15 kcal/mole
(-21.6 kJ/mole)

Group definition based on surroundings (ligands)

Additivity:  = 1 + 6 + 3

QSAR



Molecules

Data

Structure

Molecular Graph

Chemical Knowledge

Use Descriptors

Chemical Space

Regression Models

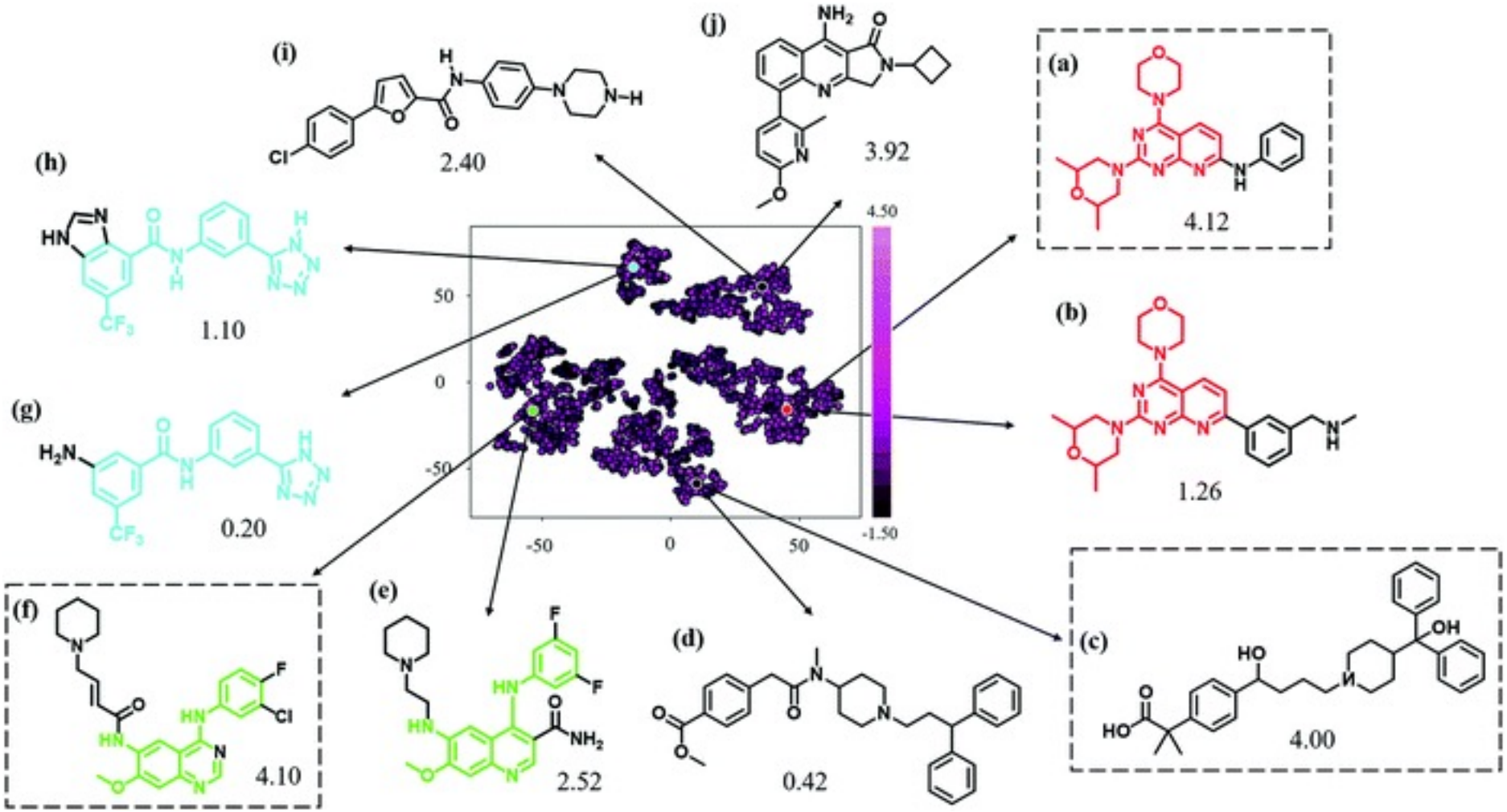
Random Forests

Statistical Models

QSAR

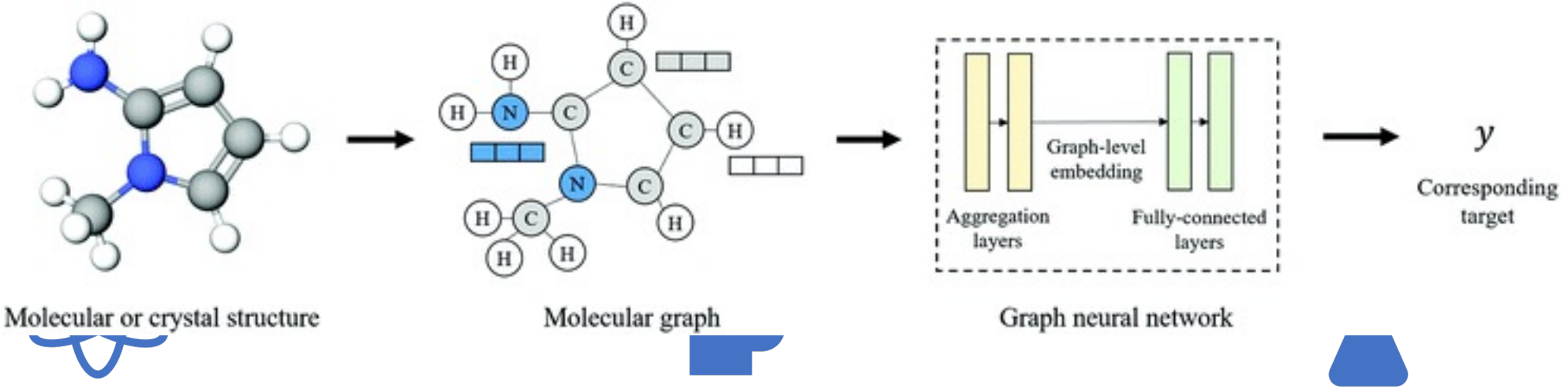
Clustering

Distribution of the molecules in the Lipophilicity dataset.



https://pubs.rsc.org/en/content/articlehtml/2020/cp/d0cp02709j

Machine Learning



Molecules

Structure

Molecular Graph

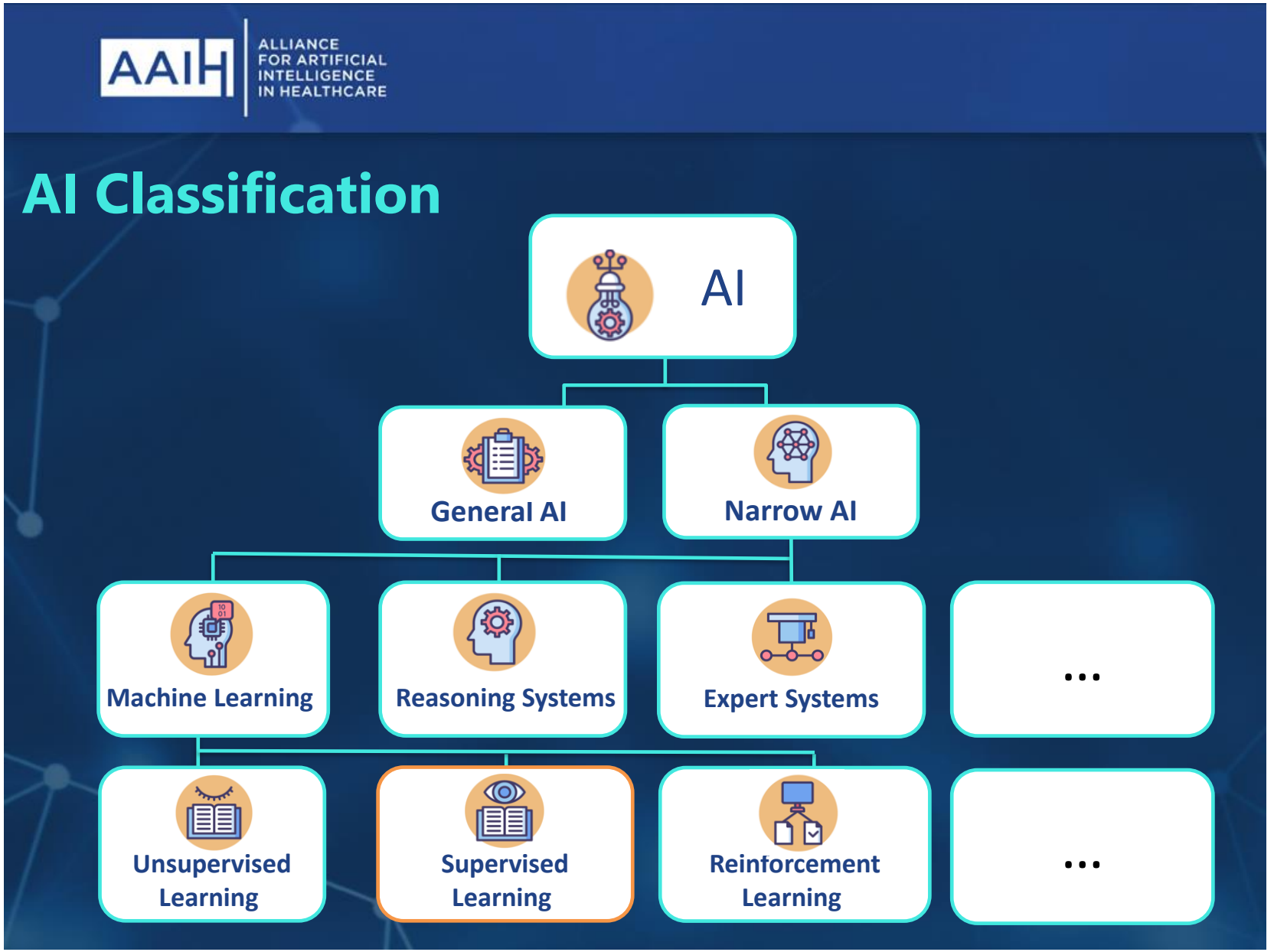
Machine Learning

Learn Relevant Descriptors

Chemical Space

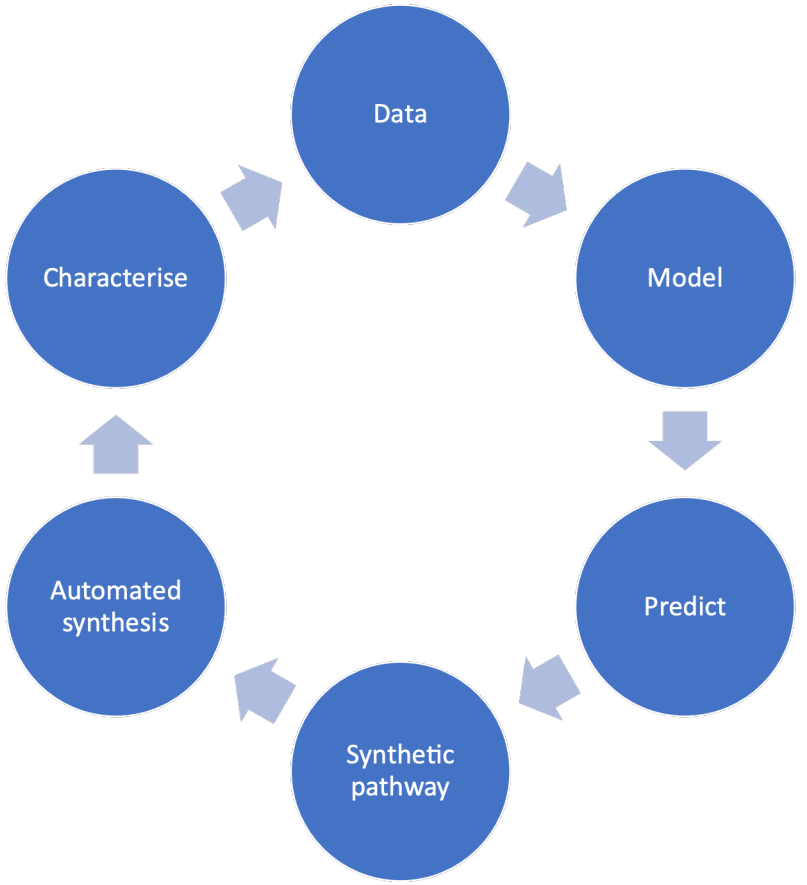
Latent Space

Can we get the learning back out of the ML Neural Network?

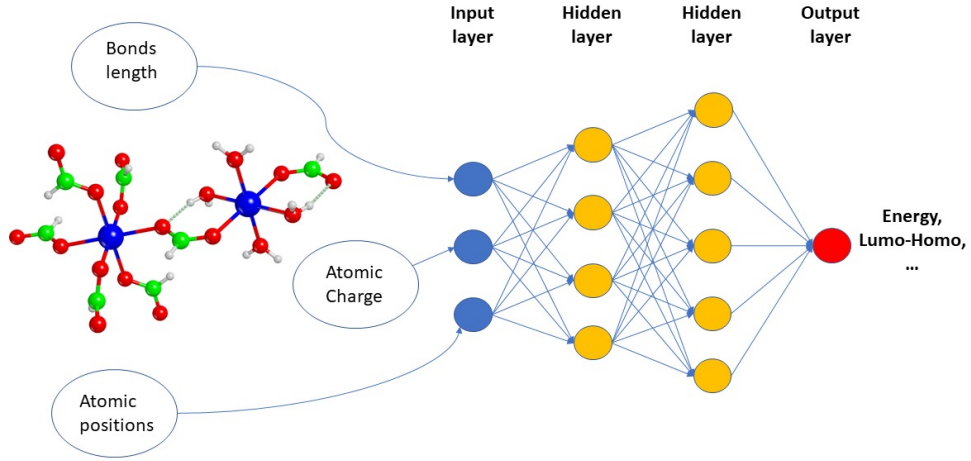

AAIH
ALLIANCE FOR ARTIFICIAL INTELLIGENCE IN HEALTHCARE
AI Classification
AI
General AI
Narrow AI
Machine Learning
Reasoning Systems
Expert Systems
...
Unsupervised Learning
Supervised Learning
Reinforcement Learning
...

Generative Models

- ML network learns the essential features for molecules with selected properties (e.g., a suitable drug candidate)
- Can be used to select suitable molecules from a library
- But a generative network ‘creates’ possible molecules
- Then check if there is a synthetic pathway to make them!



Quantum Chemistry and Machine Learning



Provides molecular and material descriptors

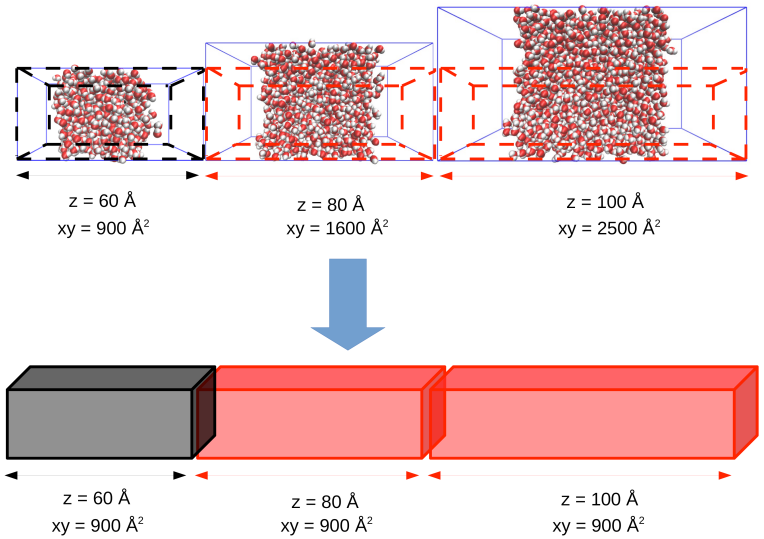
Quantum Chemistry

Machine Learning

Enhance Quantum Calculations

Molecular Alchemy & Continuous chemical space

Machine Learning and Statistical Mechanics



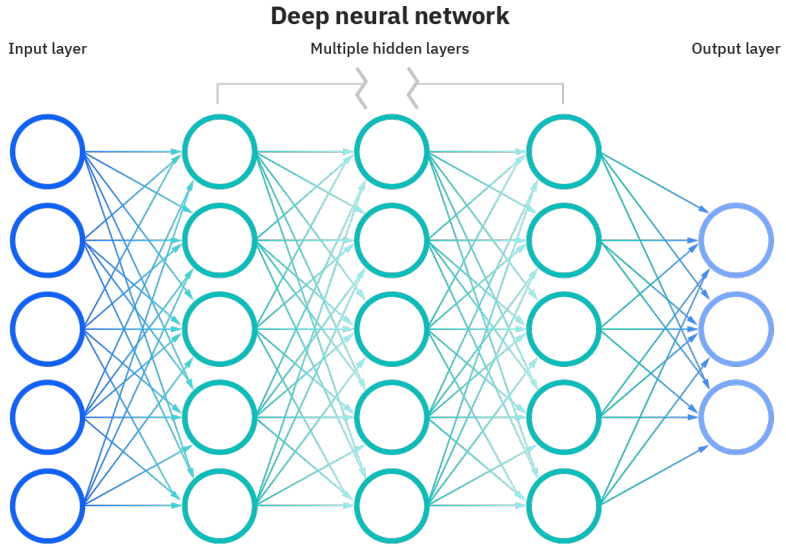
Molecular Dynamics

Develop Force Fields

Faster Integrations

Identify patterns

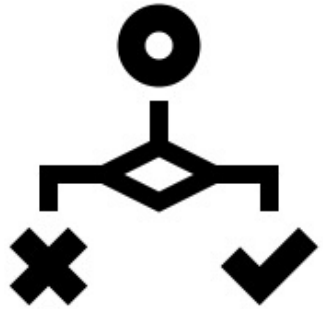
Enhanced sampling



Machine Learning

The Logic of Scientific Discovery

Types of Logical Inference
(from Ross King)

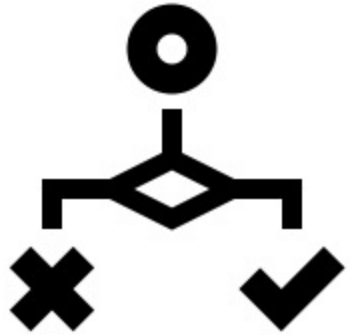


- **Deduction**
  - Rule: All swans are white
  - Fact: Daffy is a swan
  - Therefore, Daffy is white

- **Induction**
  - Rule: Daffy is a swan and white
  - Fact: Tweety is a swan and white
  - All swans are white

X Birdy is a black swan

Types of Logical Inference



- **Abduction**

- Rule: All swans are white
- Fact: Daffy is white
- Daffy is a swan

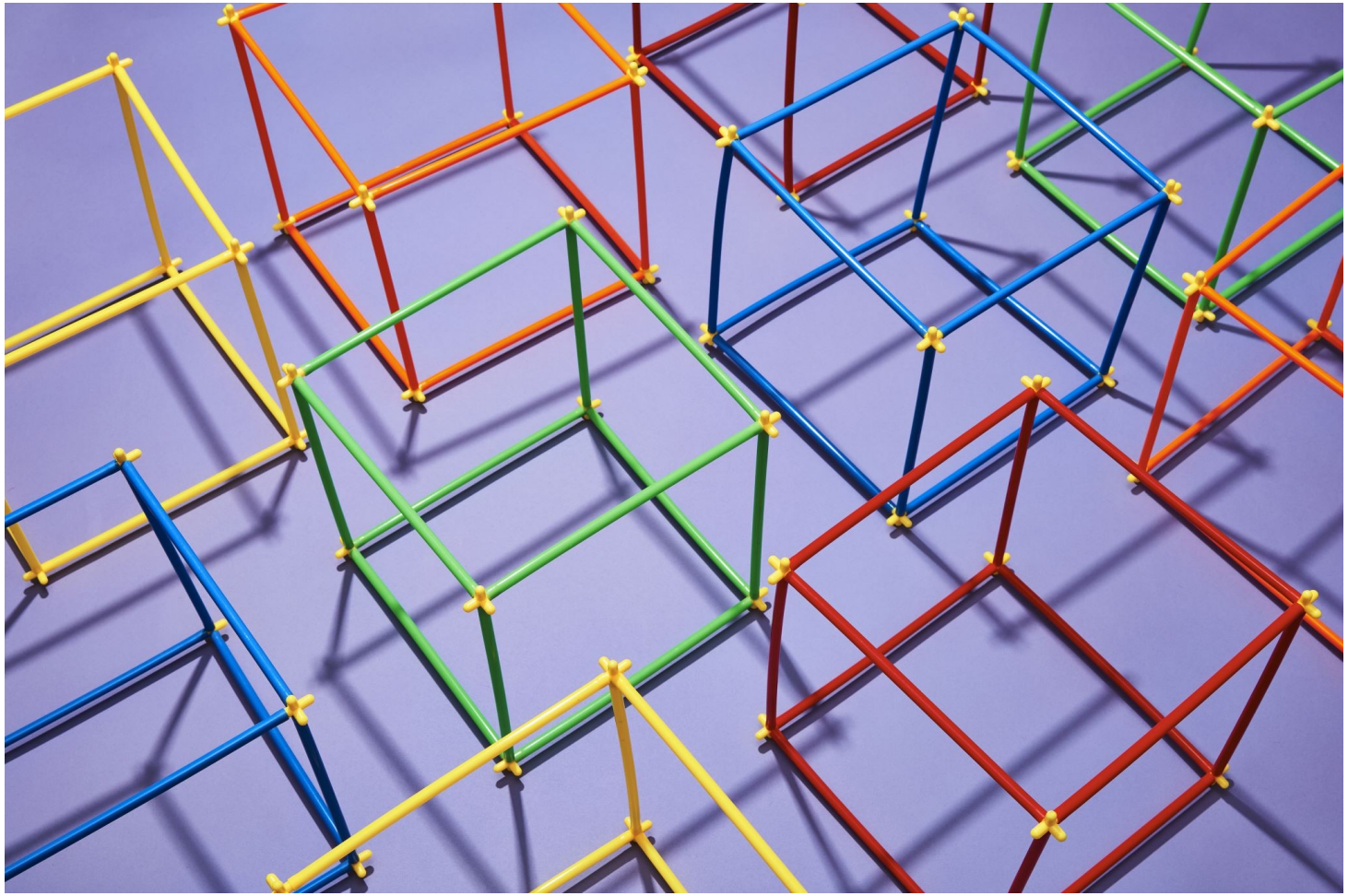
X Daffy is a Duck

Abduction from Molecules to Chemical Space

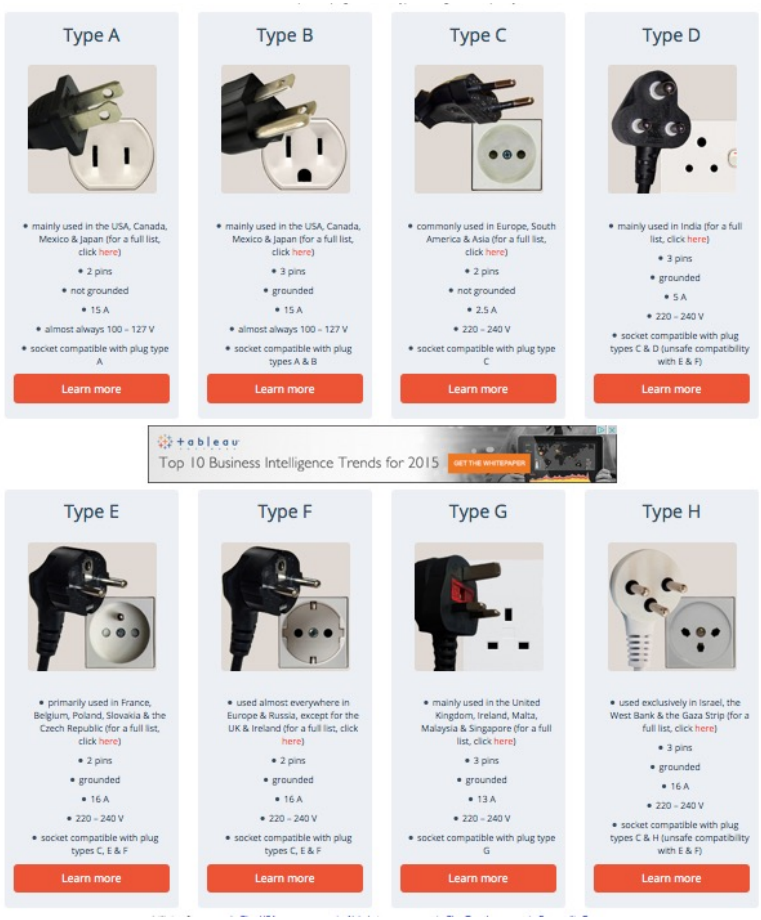
- *How to go from the study of a few molecules to ideas about the entirety of chemical space?*

Barriers to Machine Learning - DATA

- DATA
- Not enough quality data
- In machine readable form



Standards – as much time converting as extracting….



Ethical AI – The Problems of Bias

Explainable
Accountable
Reproducible

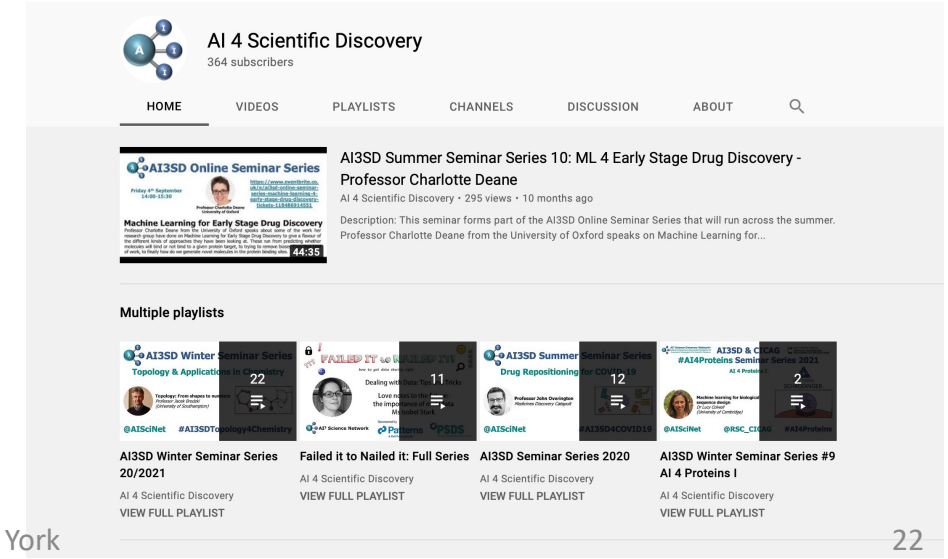
All needed for scientific discovery

The AI3SD Network

- Meetings & Workshops
- Feasibility funding
- Summer Internships
- Online Talks & YouTube Chanel
- Reports, Interviews Newsletters
- Conference

- **www.ai3sd.org**

AI3SD YouTube Cha



SKills4Scientists

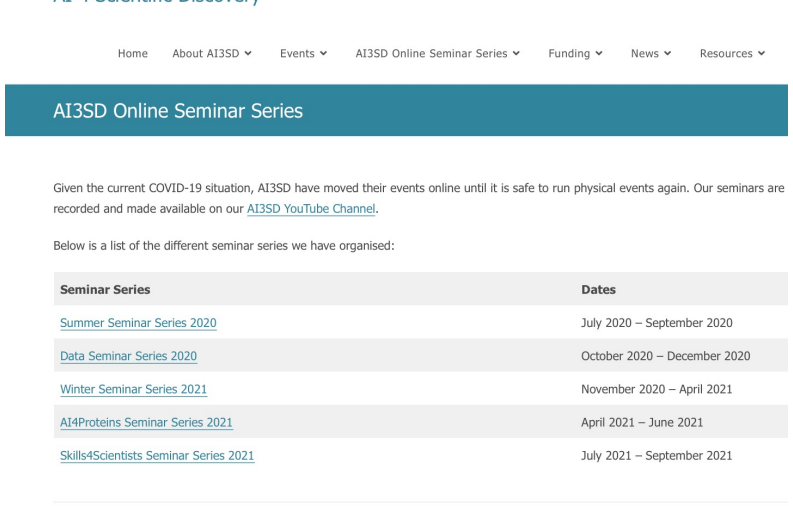
AI 4 Scientific Discovery

Home | About AI3SD | Events | AI3SD Online Seminar Series | Funding | News | Resources | Documents | Reports | Contact Us

Home

Welcome to the AI3SD Network+ (Artificial Intelligence and Augmented Intelligence for Automated Investigations for Scientific Discovery). The N
University of Southampton and aims to bring together researchers looking to show how cutting edge artificial and augmented intelligence technol
scientific discovery.

We have chosen Design and Synthesis of Chemicals and Materials (including property prediction, synthesis and manufacture) as the inter-related
which to focus the landscape for the Network+'s activities. These areas are critical to meeting the majority of the UN sustainability goals, and are
complexity of the relationships between chemical/molecular structure, physical properties and material performance renders many of these proble
computation. AI techniques hold great promise in revolutionising research in these areas.


AI 4 Scientific Discovery

Home About AI3SD Events AI3SD Online Seminar Series Funding News Resources

AI3SD Online Seminar Series

Given the current COVID-19 situation, AI3SD have moved their events online until it is safe to run physical events again. Our seminars are recorded and made available on our AI3SD YouTube Channel.

Below is a list of the different seminar series we have organised:

Seminar Series	Dates
Summer Seminar Series 2020	July 2020 – September 2020
Data Seminar Series 2020	October 2020 – December 2020
Winter Seminar Series 2021	November 2020 – April 2021
AI4Proteins Seminar Series 2021	April 2021 – June 2021
Skills4Scientists Seminar Series 2021	July 2021 – September 2021

Seminar Series

Date	Title
07/07/2021 14:00-17:00	Research Data Management
08/07/2021 14:00-17:00	Intro to Python 1
21/07/2021 14:00-17:00	Version Control & LaTeX
22/07/2021 14:00-17:00	Intro to Python 2
04/08/2021 14:00-17:00	Posters, Presentations & Reports
05/08/2021 14:00-17:00	Careers 1
18/08/2021 14:00-17:00	Ethical Research
19/08/2021 14:00-17:00	Careers 2
1-2/09/2021 10:00-17:00	Poster Symposia

AI 4 Proteins

AI 4 Scientific Discovery

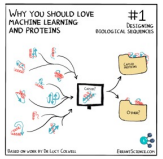
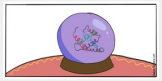
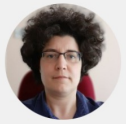
Home | About AI3SD | Events | AI3SD Online Seminar Series | Funding | News | Resources | Documents | Reports | Contact Us

AI 4 Proteins Seminar Series 2021

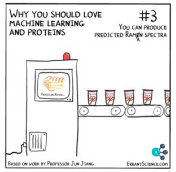
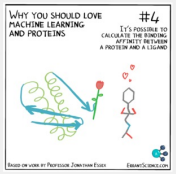
AI3SD collaborated with RSC-CICAG (The Royal Society of Chemistry – Chemical Information and Computer Applications Group) to run an **#AI4Proteins** Seminar Series in 2021. This series started on Wednesday 14th April 2021, and was made up of a set of sessions of 1-2 talks, ending with an all day virtual conference on Thursday 17th June 2021.

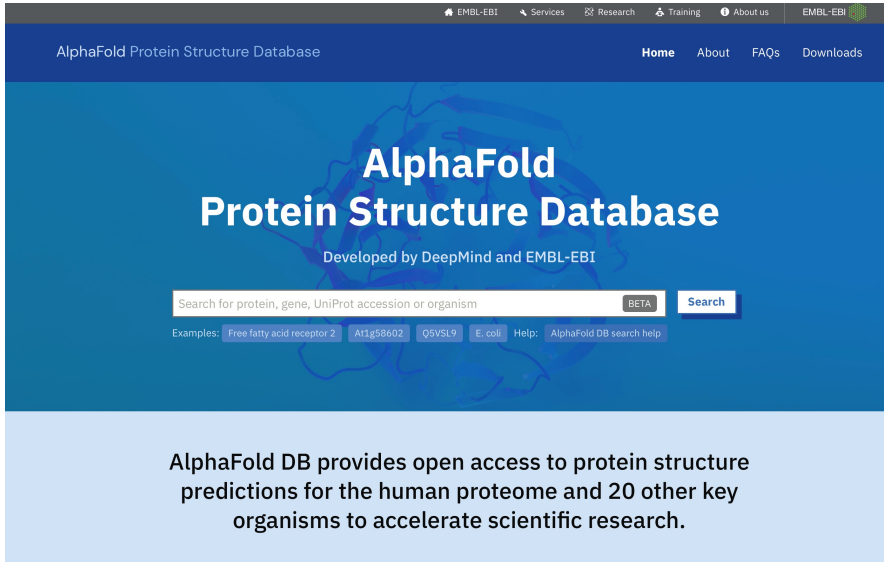
Below is the timetable for the series, with links to individual Event and Eventbrite pages. This timetable is still in a draft form and further speakers will be added as details are confirmed. A majority of the talks in these events will be recorded and made available on the AI3SD YouTube Channel.

AI4Proteins: I

Date	Talks	Speaker	Video Link	DOI Link
14/04/2021		**Machine learning for biological sequence design:** Dr Lucy Colwell (University of Cambridge)		Video Link	DOI Link
14/04/2021		**Machine learning applications for macro-molecular X-ray crystallography at Diamond:** Dr Melanie Vollmar (Diamond)		Video Link	DOI Link

AI4Proteins: II

Date	Talks	Speaker	Video Link	DOI Link
05/05/2021		**The Application of Machine Learning in Molecular Spectroscopy Study:** Professor Jun Jiang (University of Science and Technology of China)		Not available	DOI Link
05/05/2021		**Molecular Dynamics Simulations of Proteins:** Professor Jonathan Essex (University of Southampton)		Not available	Not available

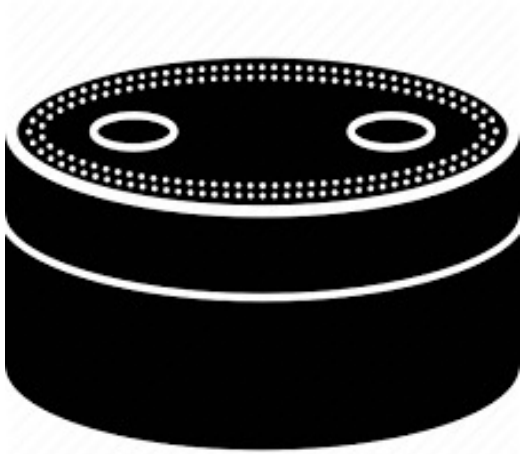


AlphaFold DB provides open access to protein structure predictions for the human proteome and 20 other key organisms to accelerate scientific research.

The Future (is ours to see)

When chemistry becomes a discipline, mathematical chemists will design new materials, predict their properties, and tell engineers how to make them — without ever entering a laboratory. We've got a long way to go on that one!

Robert A. Heinlein, “Where to?” 1950



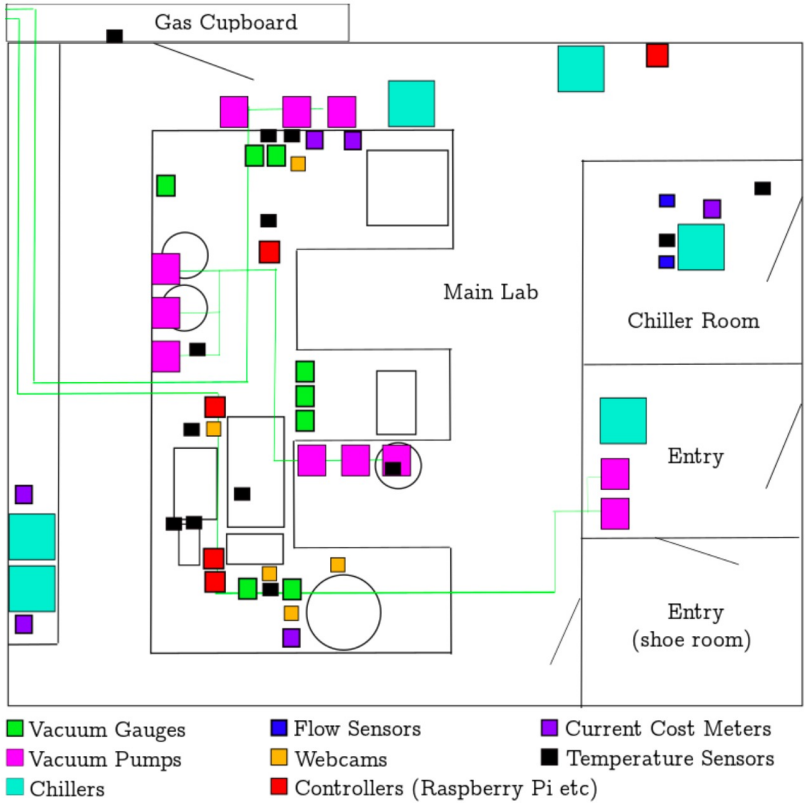
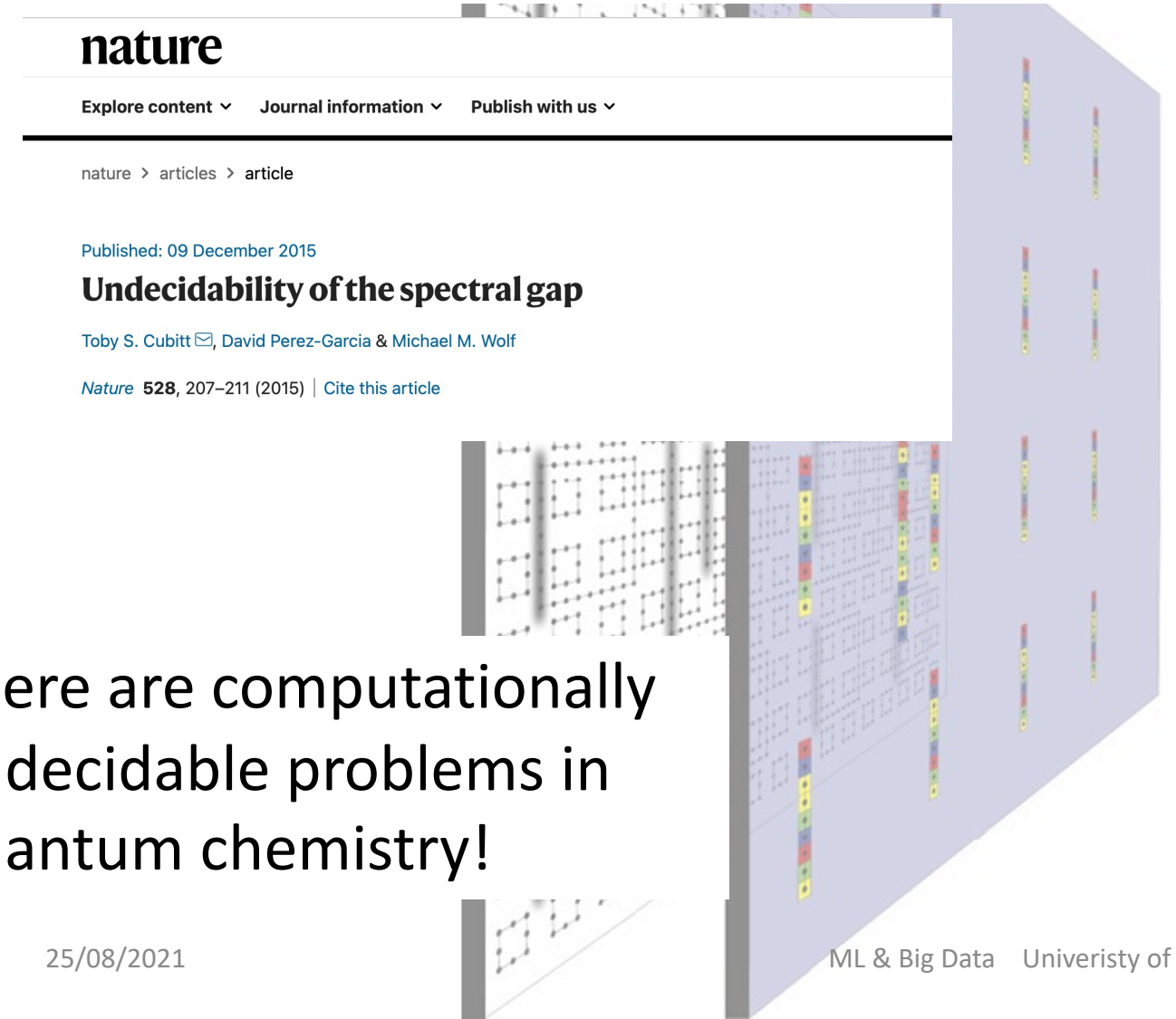


Fig. 1. Laser Lab Layout showing the different rooms and types of sensors

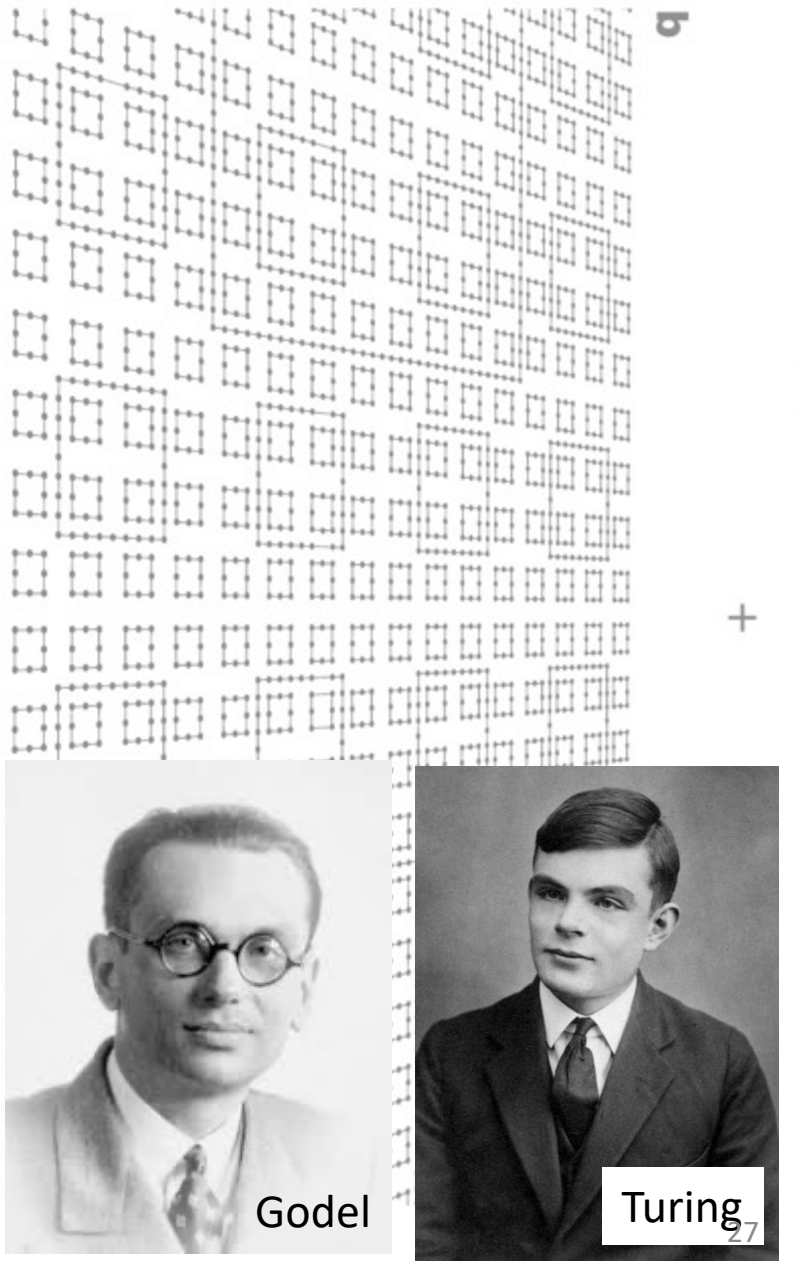
Talk2Lab

AI in the Lab

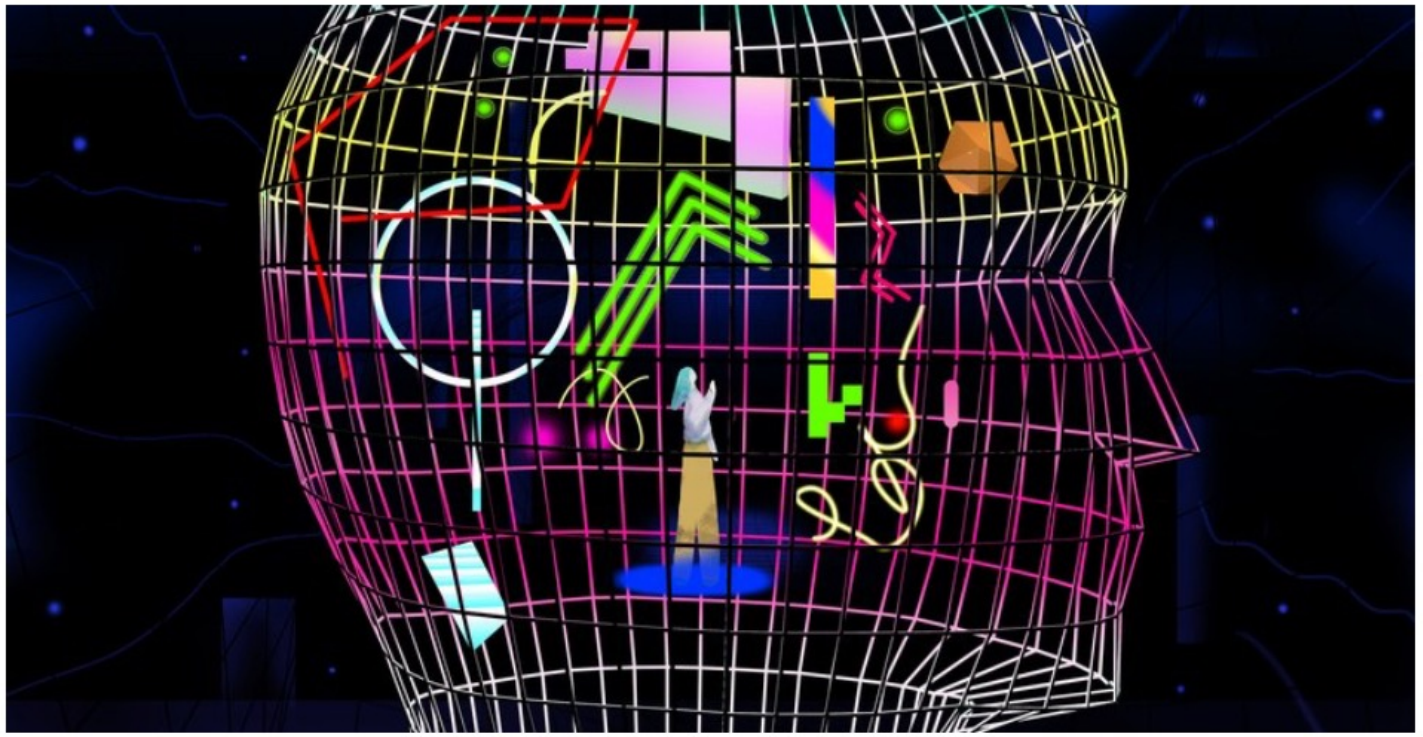
Limits of Chemical Computation



There are computationally undecidable problems in quantum chemistry!



A Key worry



Overreliance on artificial intelligence may put us in intellectual debt. Illustration by Jon Han

Overreliance on AI may put us in intellectual debt

All I am saying is that now is the time to develop the technology to deflect an asteroid

Acknowledgements

- UKRI – EPSRC - EP/S000356/1
- Samantha Kanza (Network coordinator)
- Mahesan Niranjan (co-I)
- Advisory Board (Chair John Overington)
- Nicola Knight (PSDS)
- My research group (Sam Munday, Mikey Blakey, Jack Doyle, Steve Brewer & all former members)
- Colleagues in UK and International

Thank you for listening

Trust me Mort - no electronic communications superhighway, no matter how vast and sophisticated, will ever replace the art of the schmooze