AI in science.

Maciej Marchwiany, PhD

https://www.linkedin.com/pulse/ai-powered-physics-how-ai-revolutionizing scientific-discovery-cain-ddbwc

Plan

AI concept

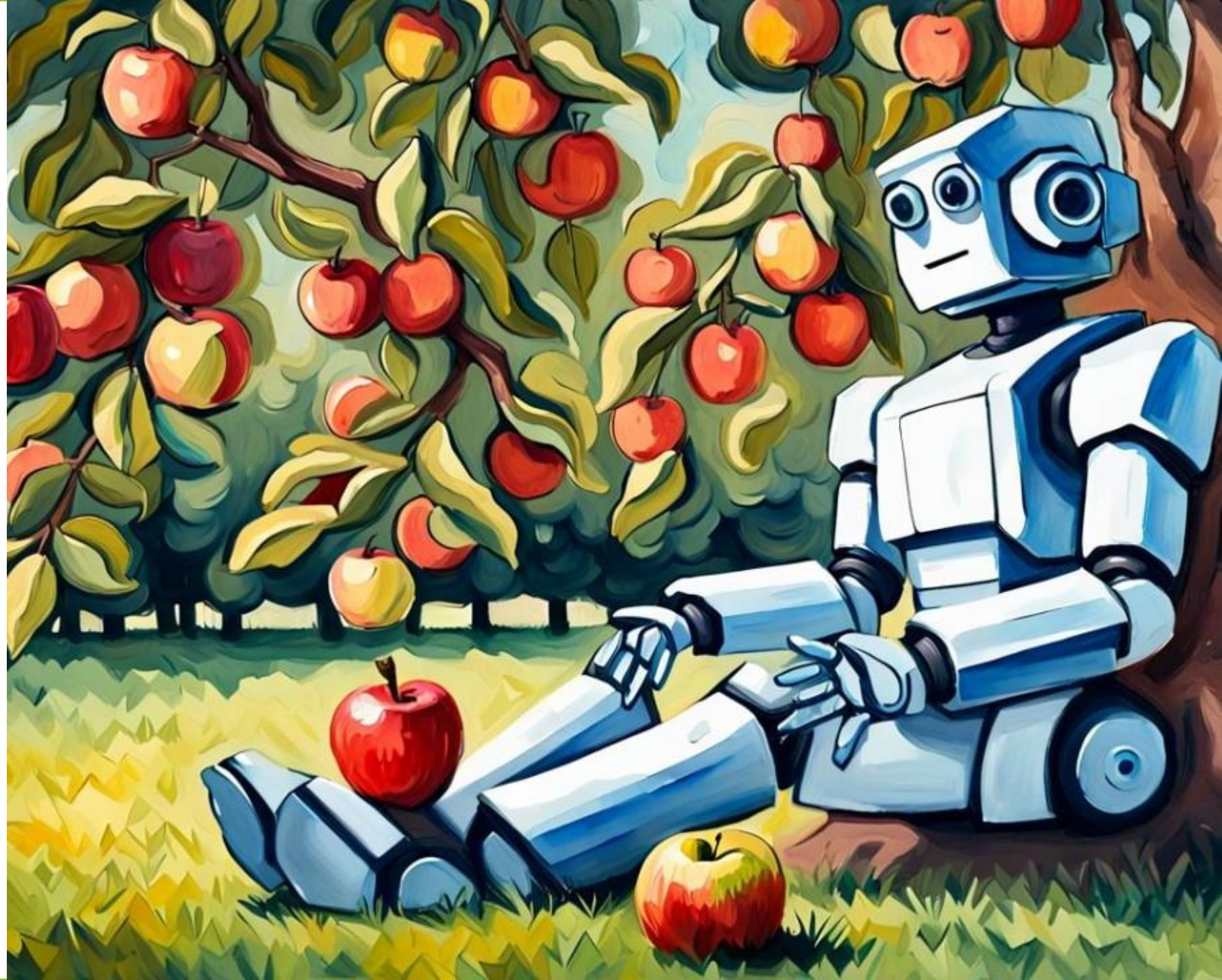
AI in science

AI as scientist

Why is AI difficult?

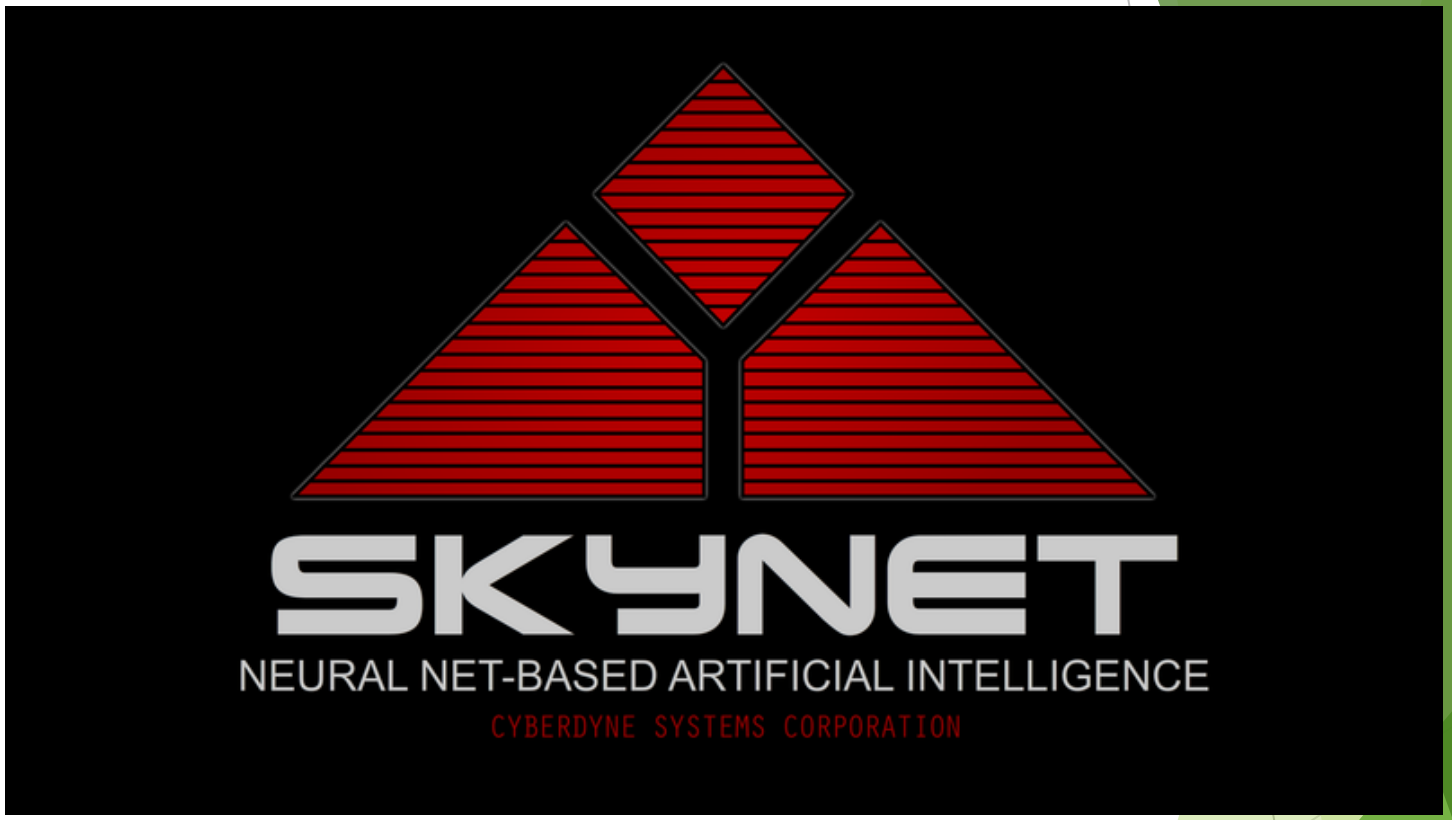
1

What is AI?



https://chat.openai.com/

What is artificial intelligence?



https://chat.openai.com/

What is artificial intelligence?

Artificial intelligence (AI) is a field of **computer science** that develops computer programs capable of thinking and making decisions, **imitating human abilities**. It uses methods such as:

- machine learning,
- natural language processing,
- image recognition,

to analyze data and solve problems. AI has applications in various fields, from medicine to industry.

Types of AI

Pattern recognition

https://chat.openai.com/
https://openai.com/dall-e-3/

NLP

Natural Language Processing (NLP) is a branch of artificial intelligence that focuses on creating computer systems that can understand, interpret and generate **human natural language**.

Various aspects are examined within NLP, such as:

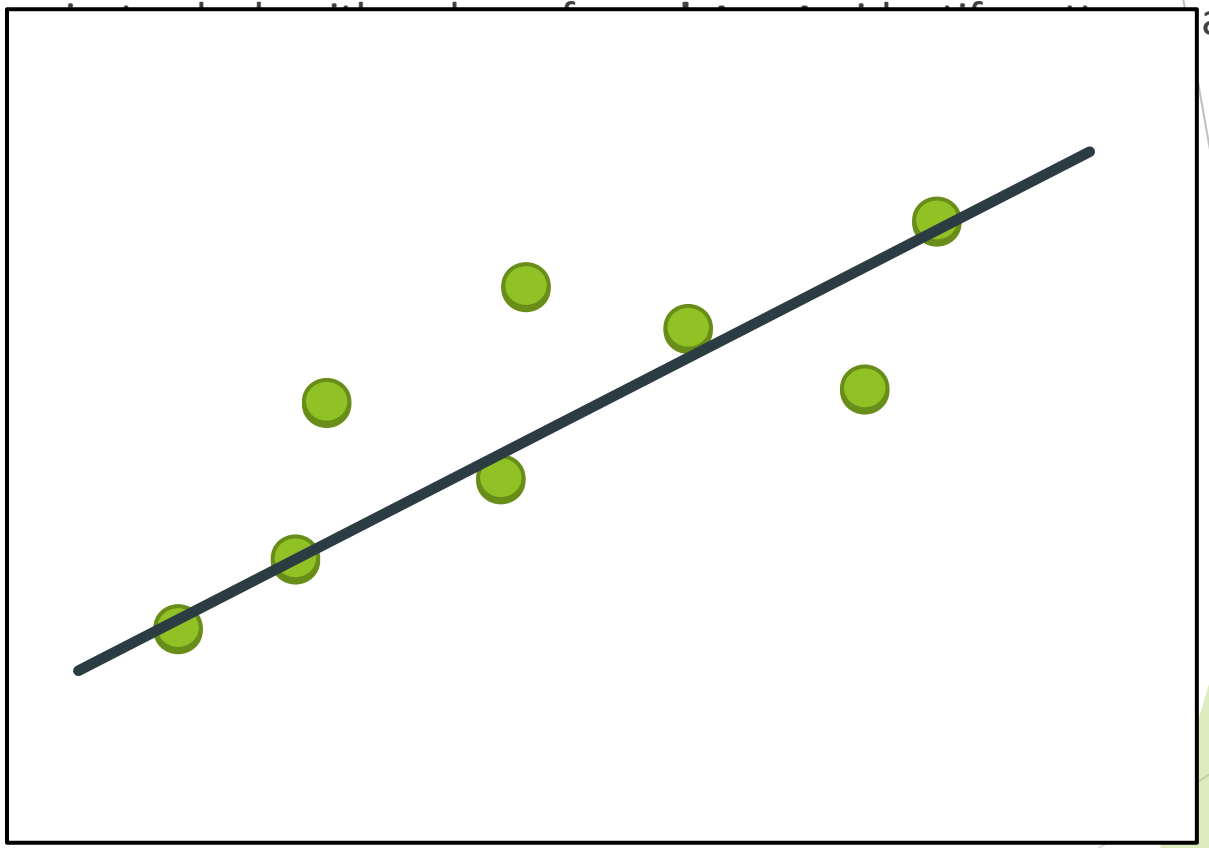
- language processing,
- syntactic analysis,
- speech recognition,
- machine translation,
- understanding the context and intention of the statement.

NLP is used in: **chatbots**, text processing systems, automatic **translations** and processing **large text sets**.

https://chat.openai.com/

Machine learning

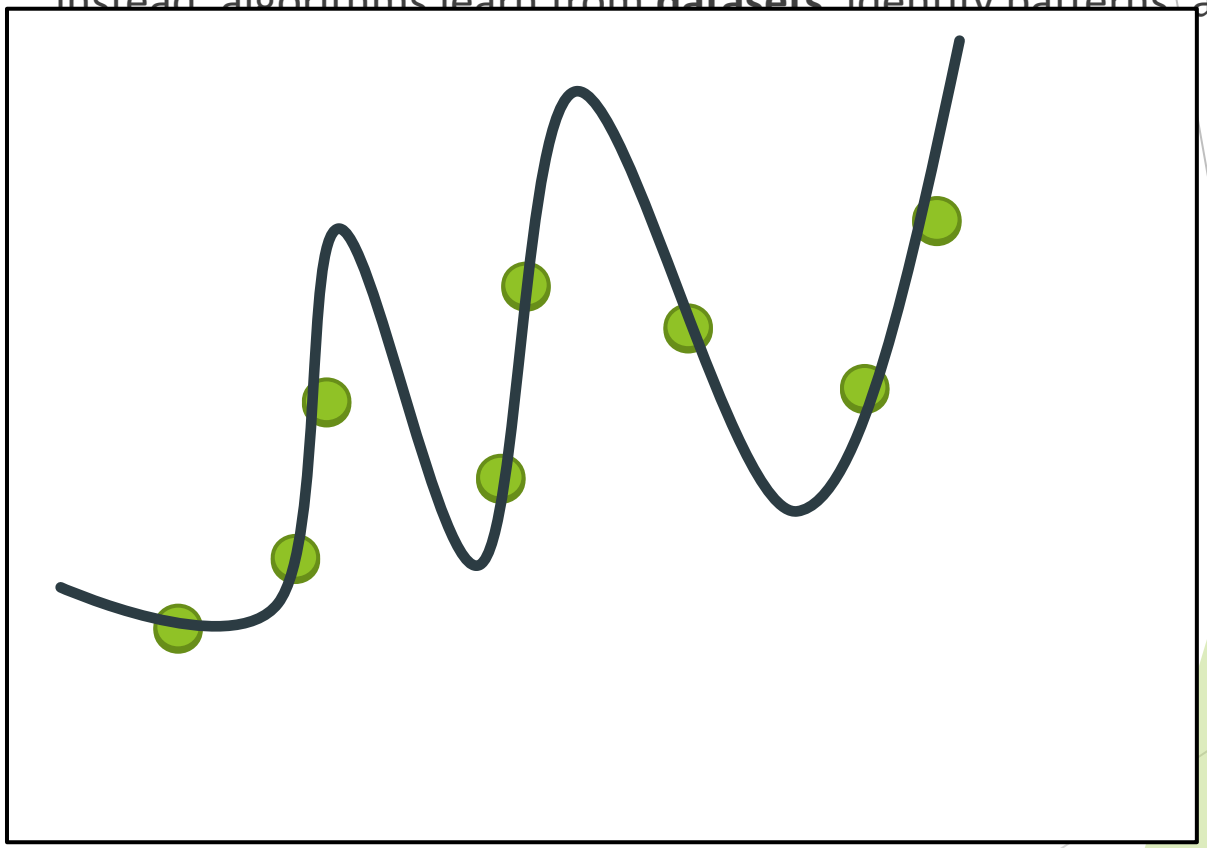
Machine learning (ML) is a field of artificial intelligence that focuses on **developing algorithms** and computer models capable of **learning independently** and **improving** their abilities **based on data**. In machine learning, the computer is **not** directly **programmed** to perform a specific task; and make decisions



https://chat.openai.com/

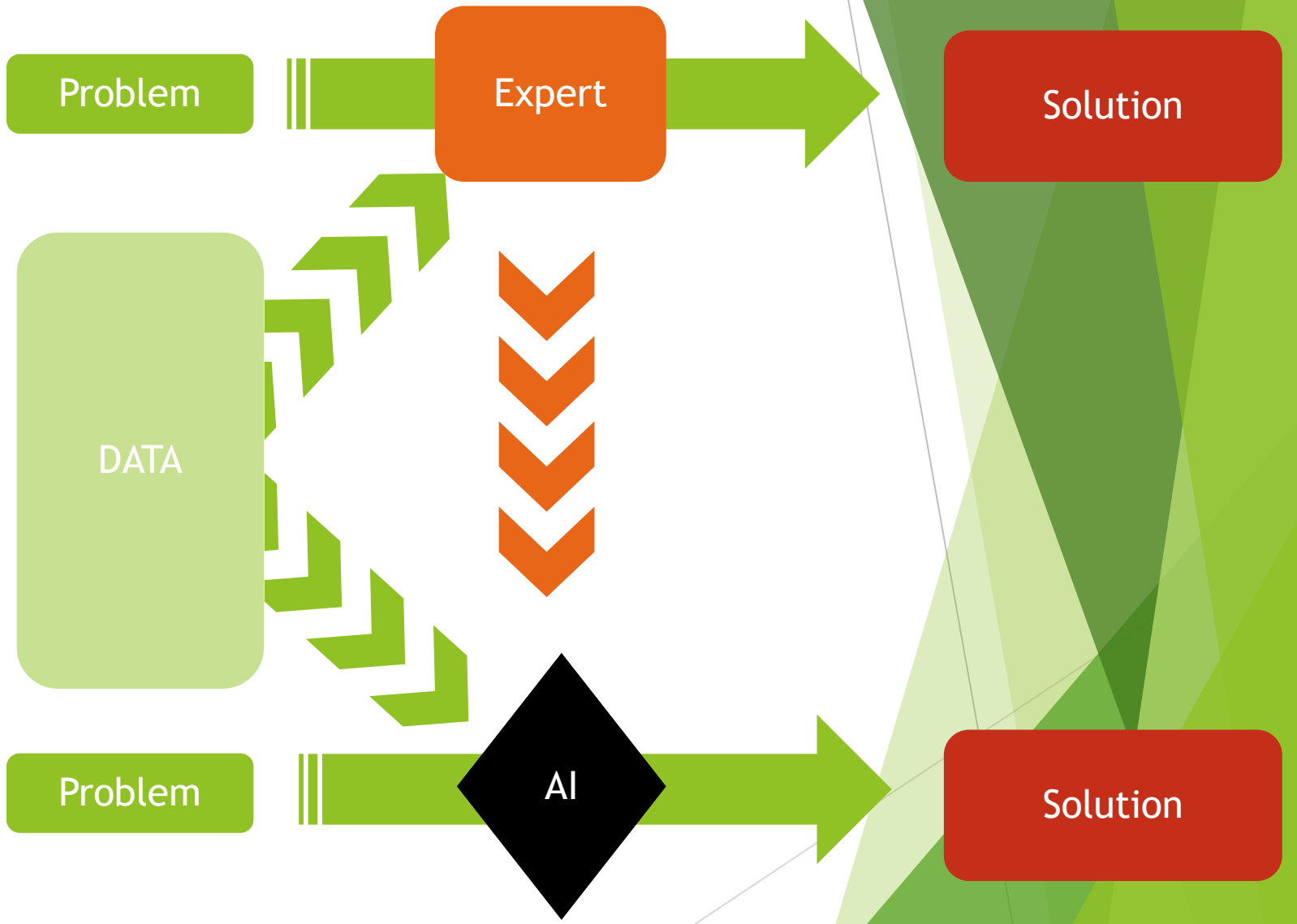
Machine learning

Machine learning (ML) is a field of artificial intelligence that focuses on **developing algorithms** and computer models capable of **learning independently** and **improving** their abilities **based on data**. In machine learning, the computer is **not** directly **programmed** to perform a specific task; instead, algorithms learn from **datasets**, identify patterns, and make decisions

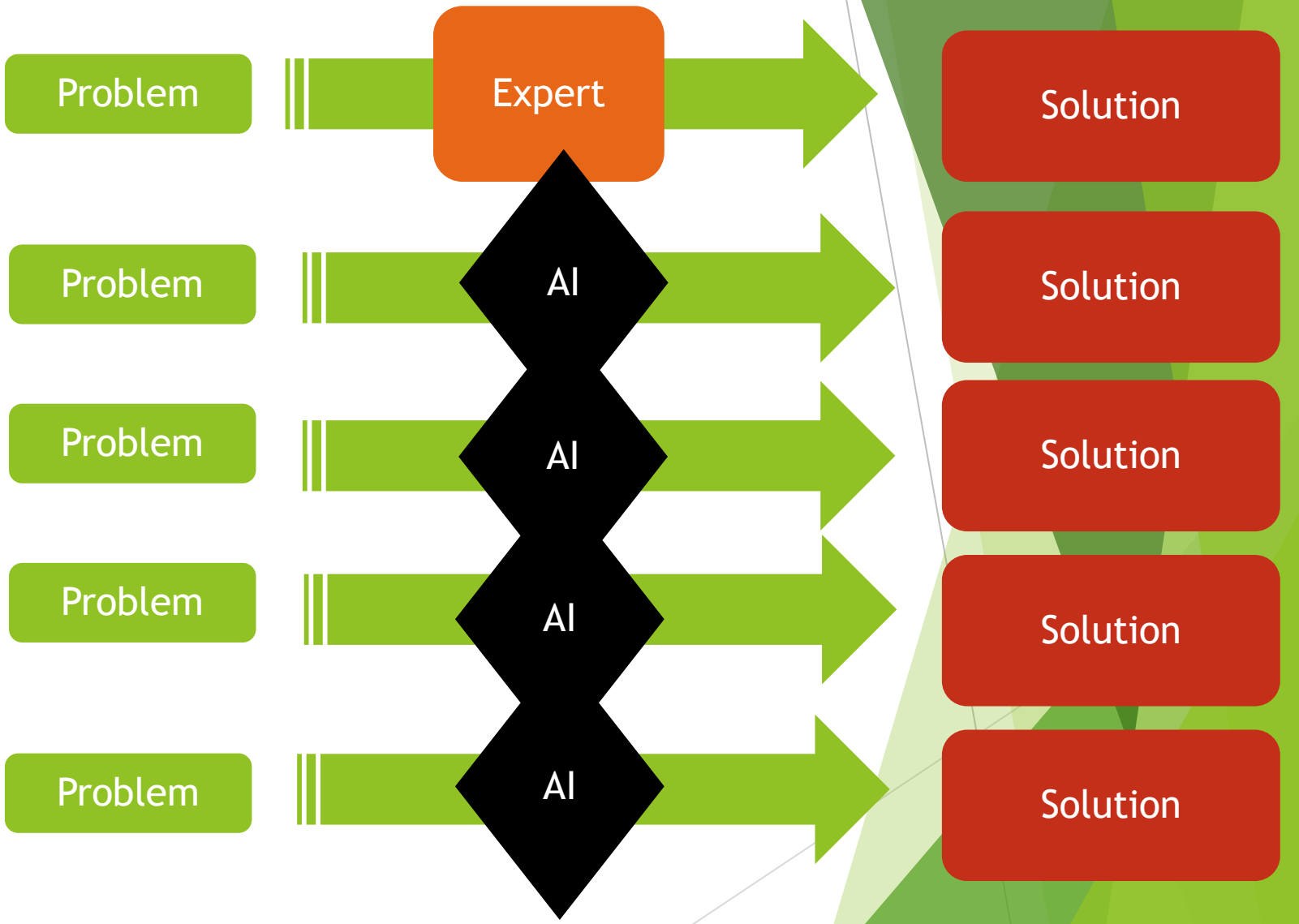


Types of AI

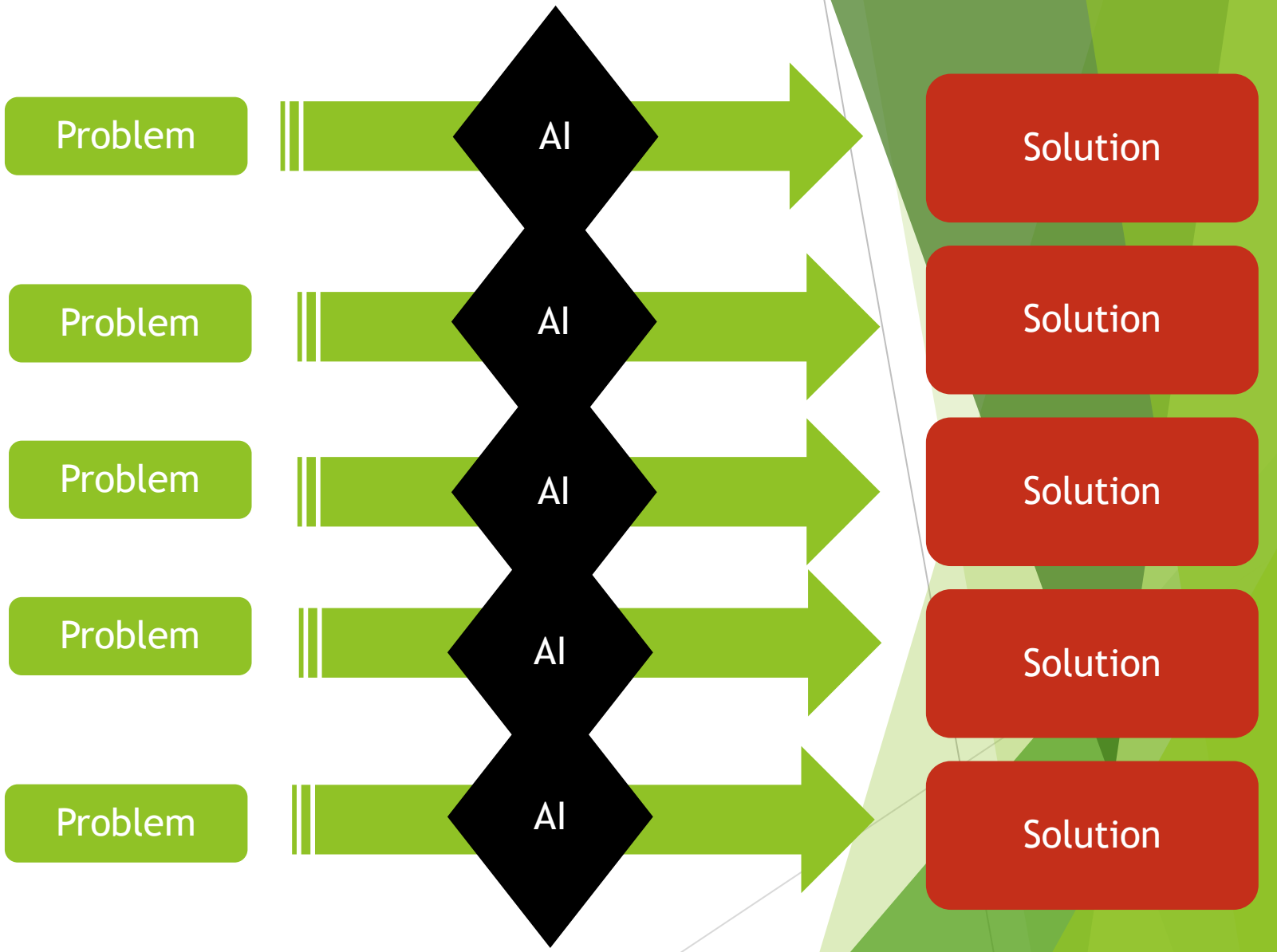
The role of artificial intelligence



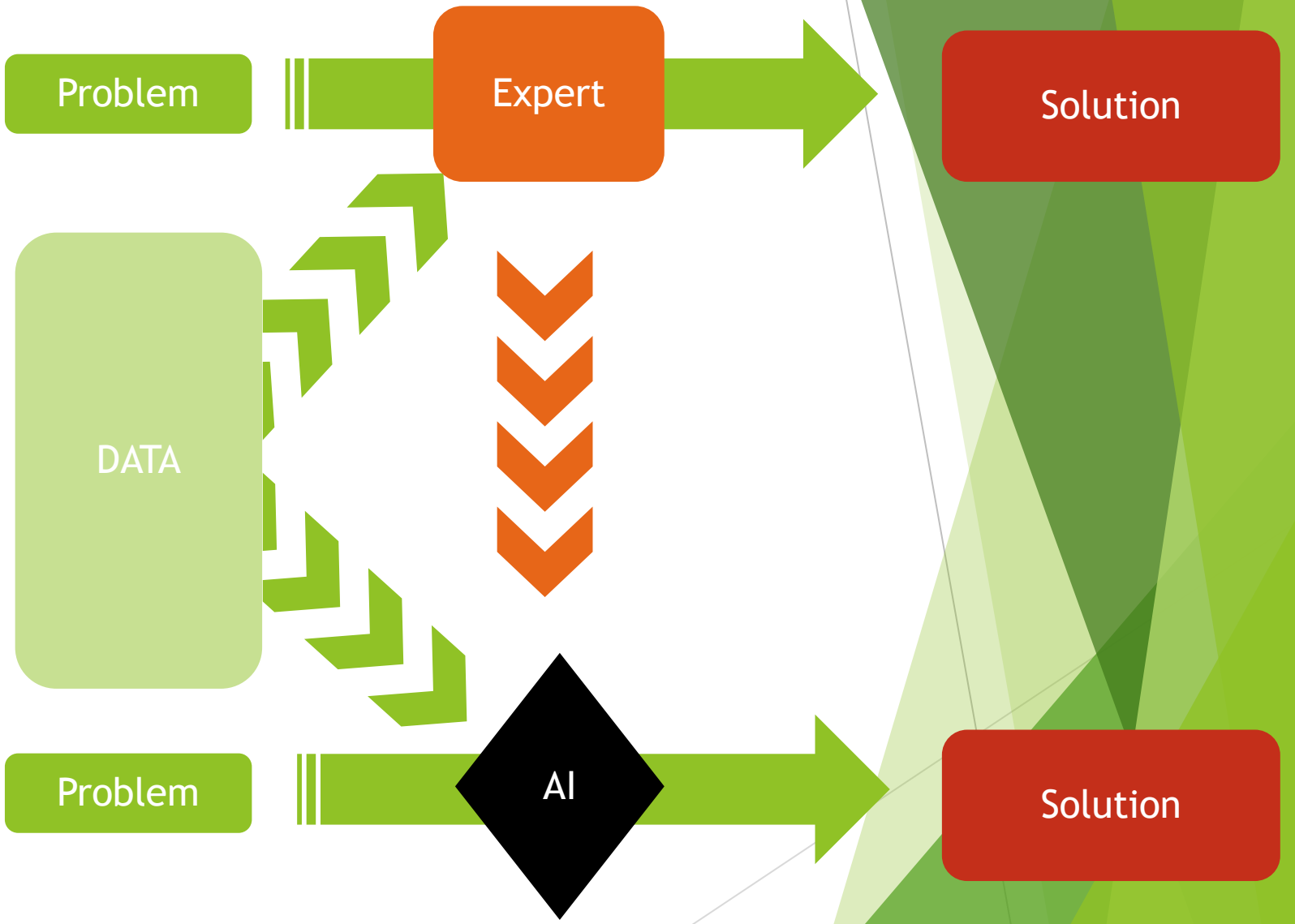
The role of artificial intelligence



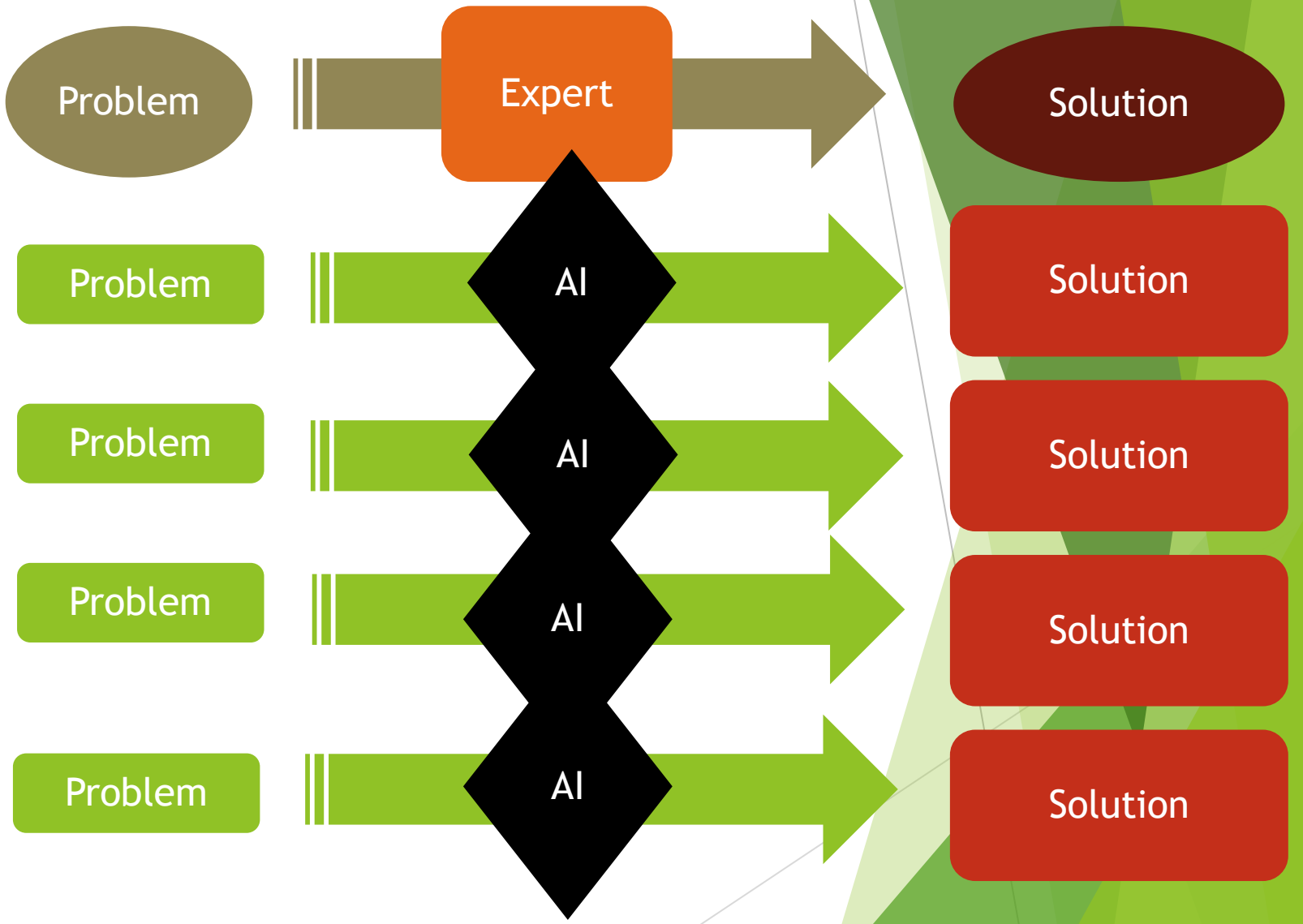
The role of artificial intelligence



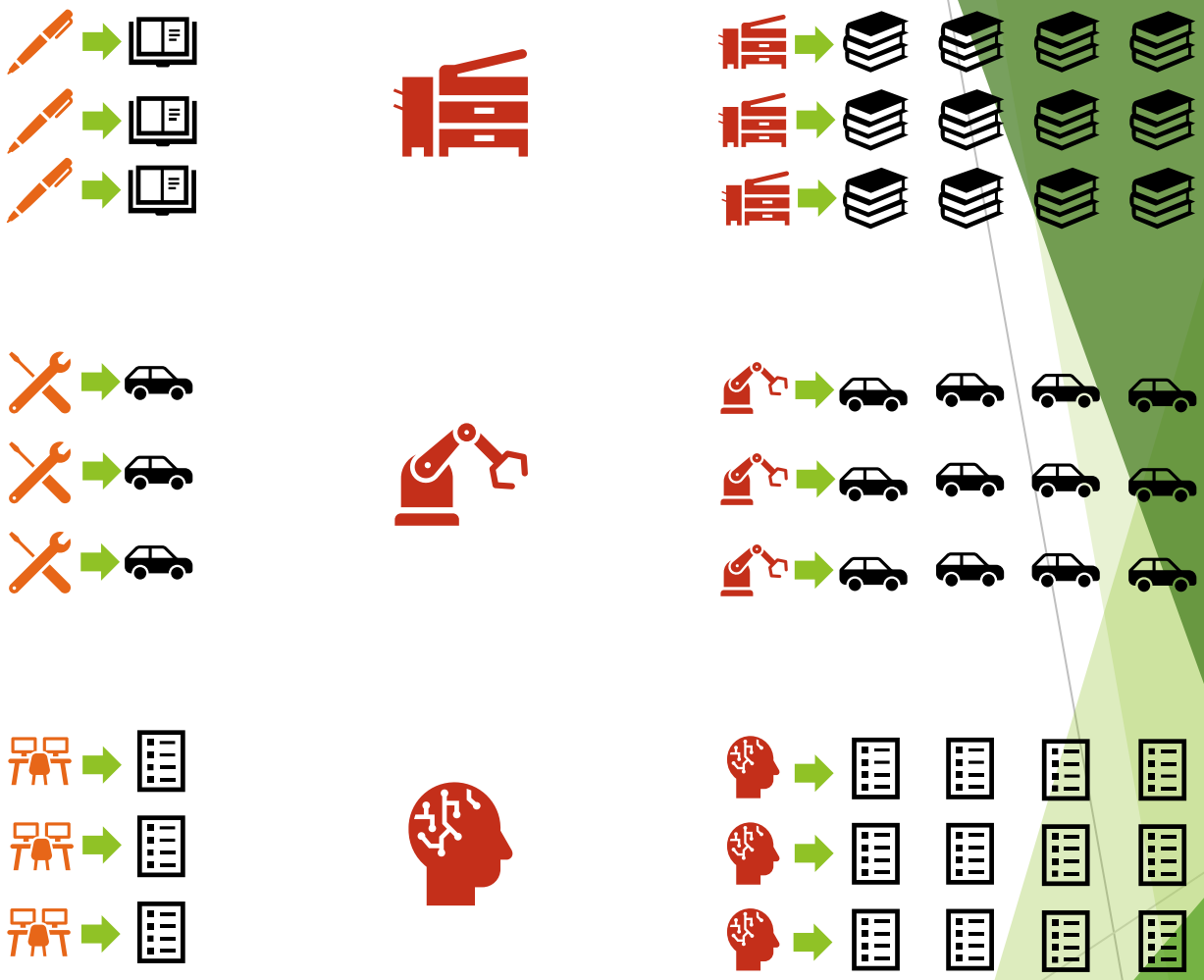
The role of artificial intelligence



The role of artificial intelligence



Revolution



Artificial intelligence in life

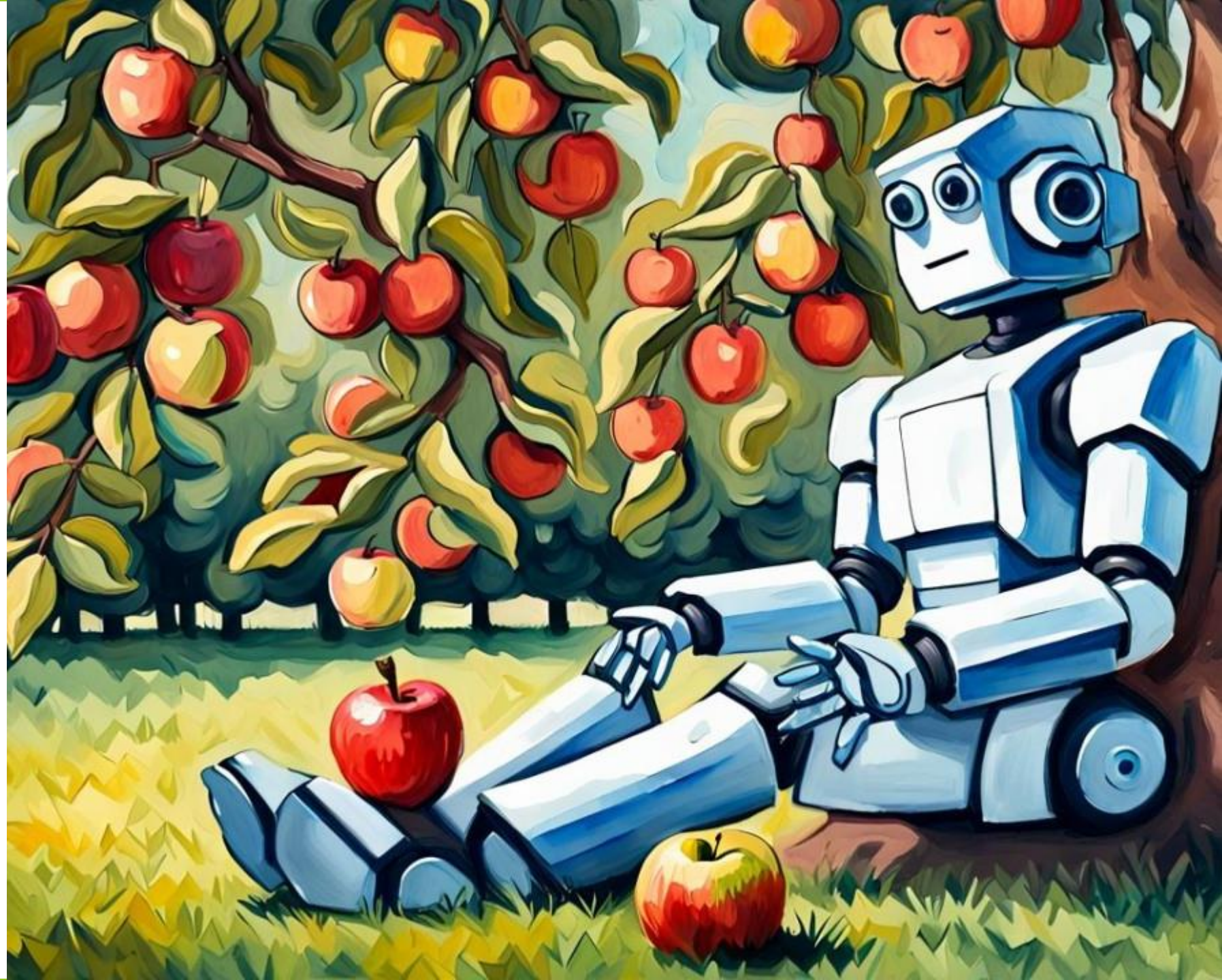
- Digital **assistants**, voice recognition systems
- Internet **search engines** and autocomplete
- Machine **translations** and grammar improvement
- Advertising personalization and **recommendation systems**
- **Automatic vehicles** and object recognition
- Spam filters and crime detection
- **Deepfake**
- Facial recognition and **profiling**
- **This presentation**

Artificial intelligence in science

- Statistical data **analysis** and image generation
- **Speeding up calculations** and approximating potentials
- Predicting chemical **reactions**
- Predicting the **outcome** of simulation and experiment
- **Optimization** of experiment **parameters** and equipment utilization
- Information retrieval and automatic **summaries**

2

AI in physics



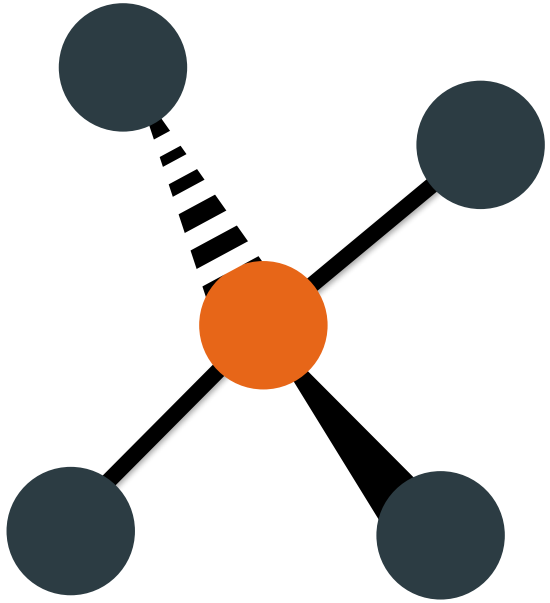
Predicting the properties of molecules

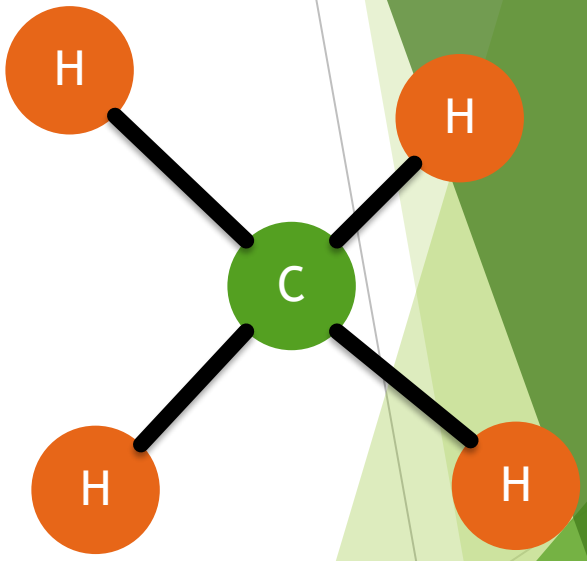
Marchwiany, M., Birowska, M., Popielski, M., Majewski, J. Jastrzębska, A. Surface-Related Features Responsible for Cytotoxic Behavior of MXenes Layered Materials Predicted with Machine Learning Approach. Materials 13 (2020).

R2 metric for predicting the total energy and cohesion energy of boron nanotubes calculated by cross-testing

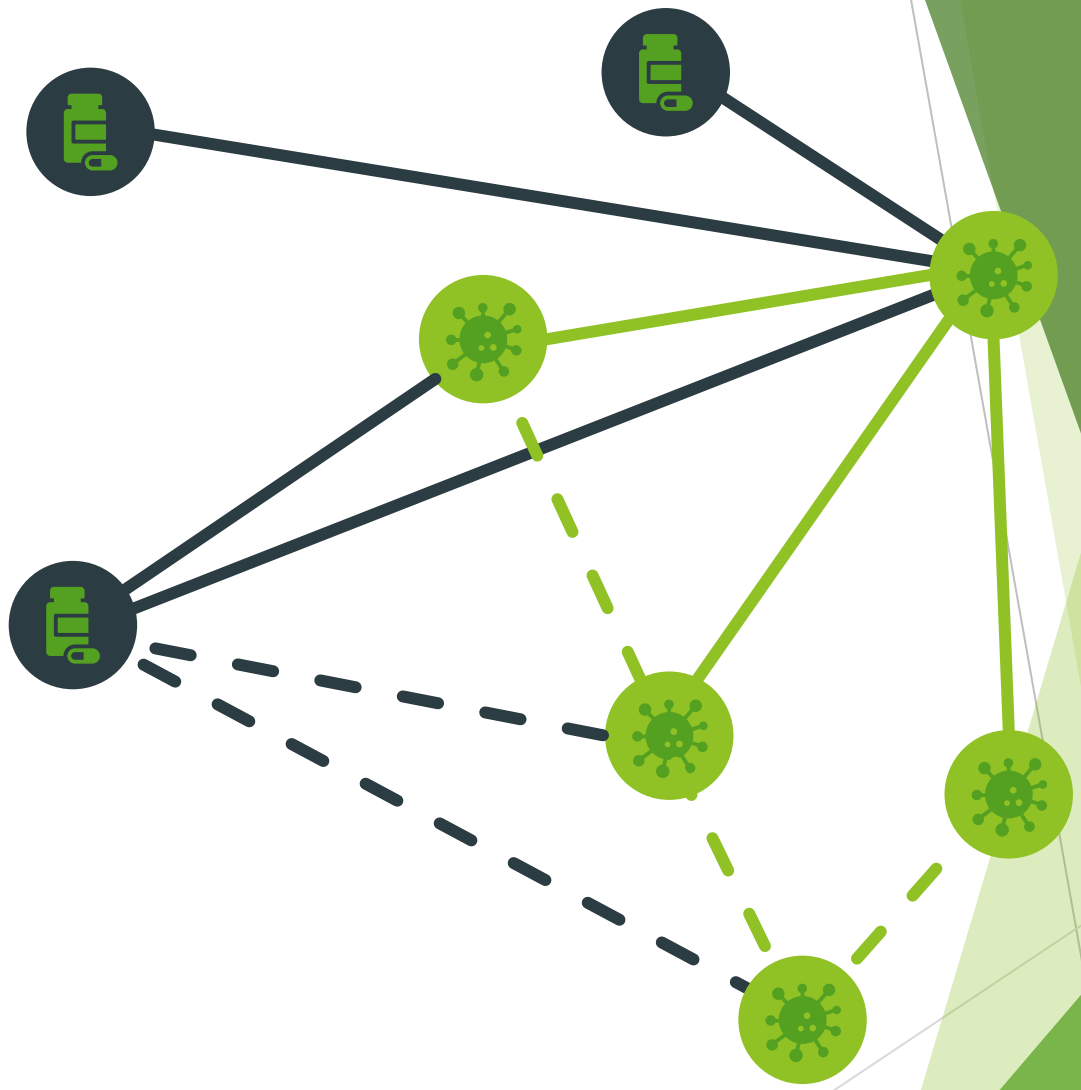
	Total energy			Cohesion energy		
	ESM	CM	SM	ESM	CM	SM
Random forest	**0.92**	**0.95**	**0.80**	**0.69**	**0.89**	**0.55**
XRT	**0.92**	**0.94**	**0.78**	**0.66**	**0.87**	**0.63**
Linear Regression	0.54	**0.95**	0.66	X	0.58	0.83
SVM – linear kernel	X	**0.95**	0.66	X	0.44	0.74
SVM – rbf kernel	X	X	X	X	X	X
SVM – sig kernrl	X	X	X	X	X	X
Ada boost – linear regression	0.49	0.90	0.50	X	0.34	0.63
Ada boost – random forest	0.92	0.95	0.77	0.51	0.90	0.52

GNN

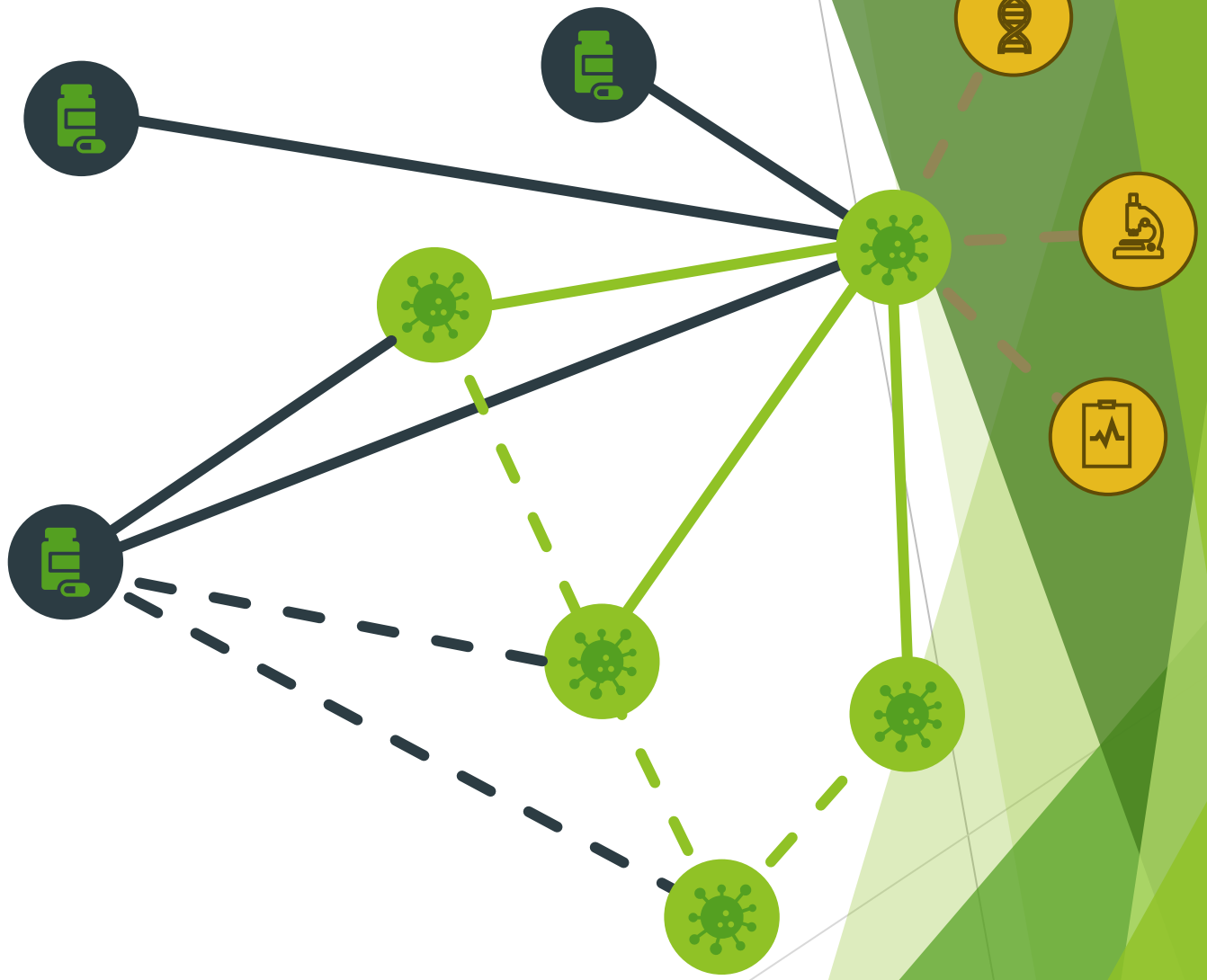




GNN



GNN



Predicting the properties of molecules

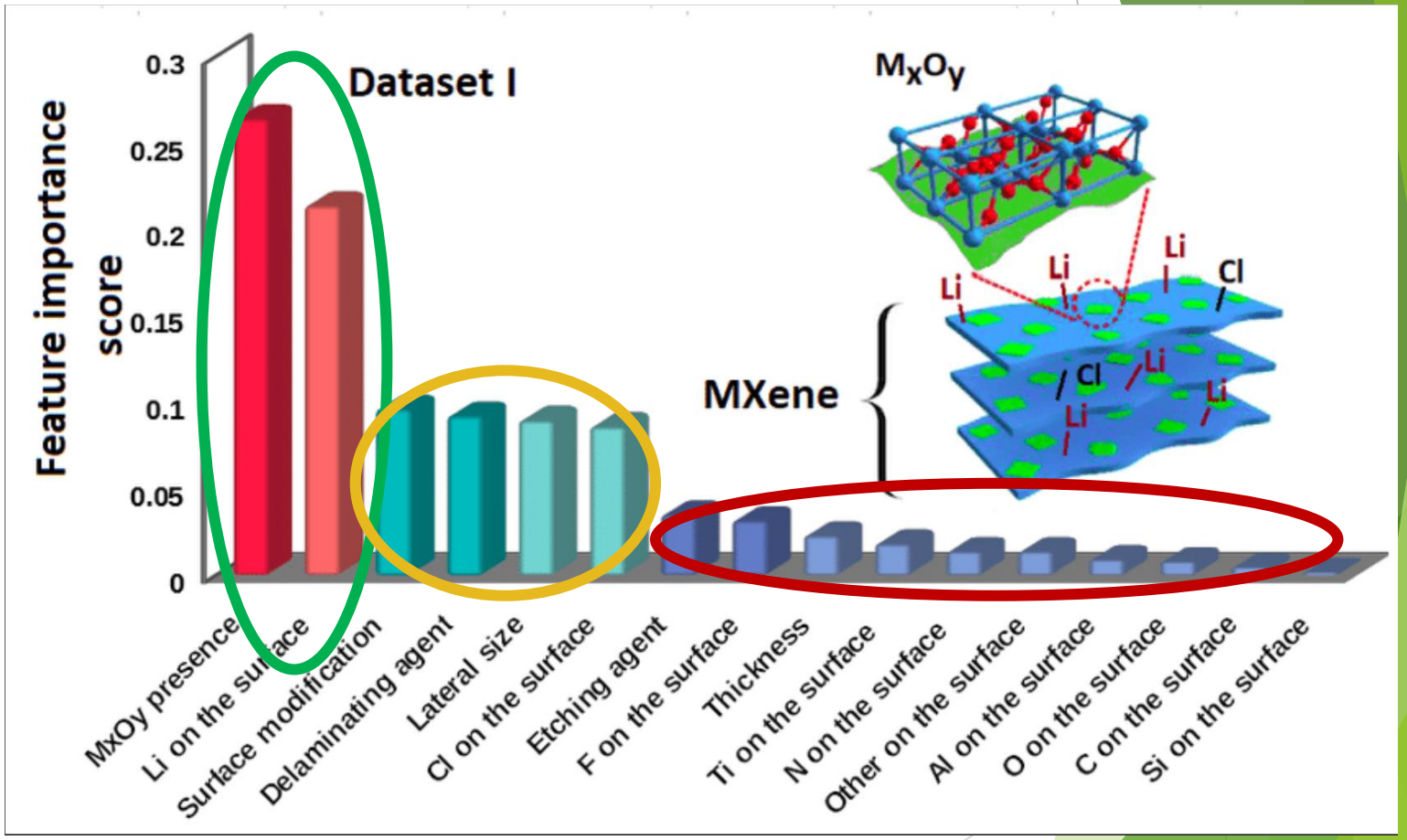
Marchwiany, M., Birowska, M., Popielski, M., Majewski, J. & Jastrzębska, A. Surface-Related Features Responsible for Cytotoxic Behavior of MXenes Layered Materials Predicted with Machine Learning Approach. Materials 13 (2020).

R2 metric for predicting the total energy and cohesion energy of boron nanotubes calculated by cross-testing

	Total energy			Cohesion energy		
	ESM	CM	SM	ESM	CM	SM
Random forest	**0.92**	**0.95**	**0.80**	**0.69**	**0.89**	**0.55**
XRT	**0.92**	**0.94**	**0.78**	**0.66**	**0.87**	**0.63**
Linear Regression	0.54	**0.95**	0.66	X	0.58	0.83
SVM – linear kernel	X	**0.95**	0.66	X	0.44	0.74
SVM – rbf kernel	X	X	X	X	X	X
SVM – sig kernrl	X	X	X	X	X	X
Ada boost – linear regression	0.49	0.90	0.50	X	0.34	0.63
Ada boost – random forest	0.92	0.95	0.77	0.51	0.90	0.52

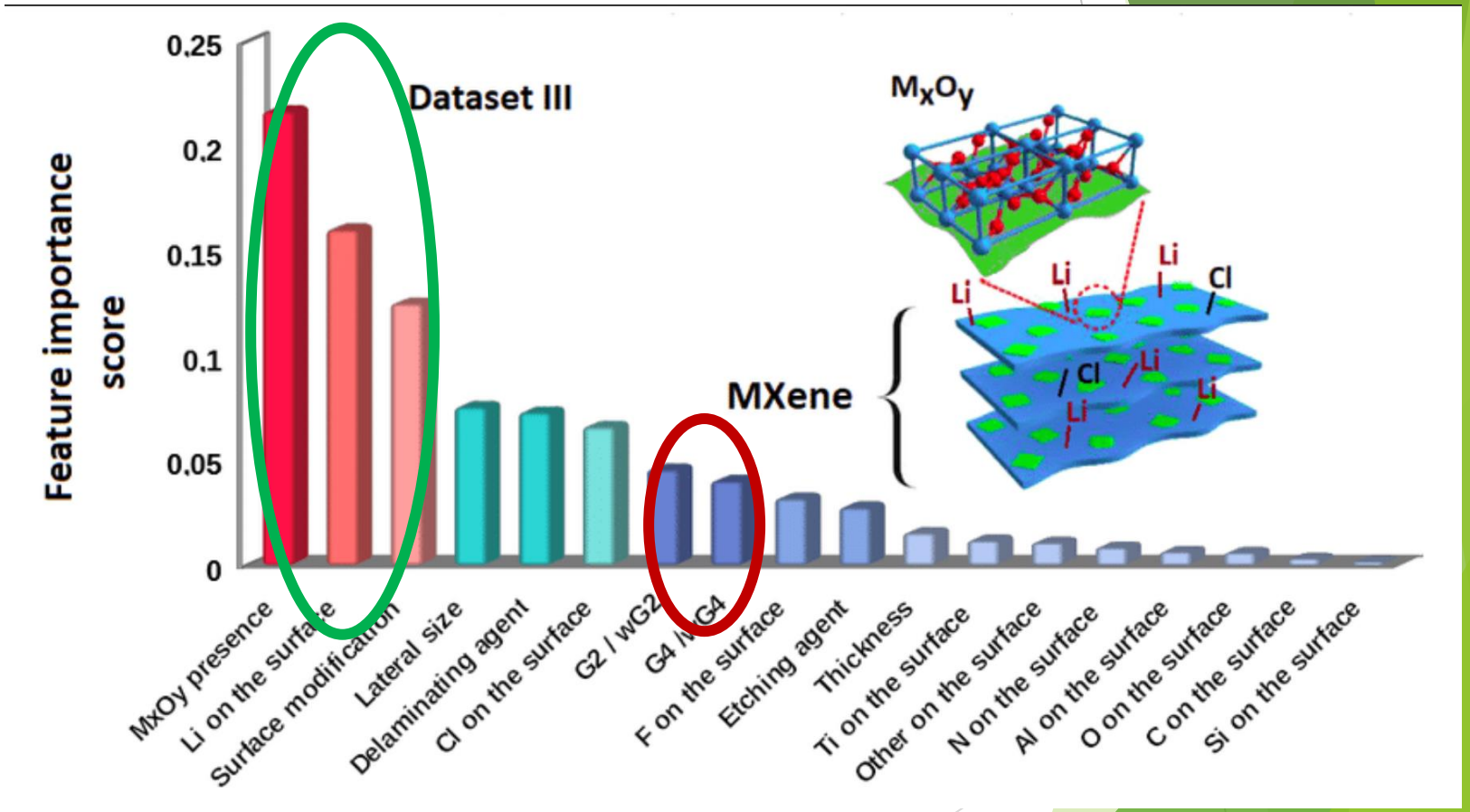
Predicting the properties of molecules

Marchwiany, M., Birowska, M., Popielski, M., Majewski, J. & Jastrzębska, A. Surface-Related Features Responsible for Cytotoxic Behavior of MXenes Layered Materials Predicted with Machine Learning Approach. Materials 13 (2020).

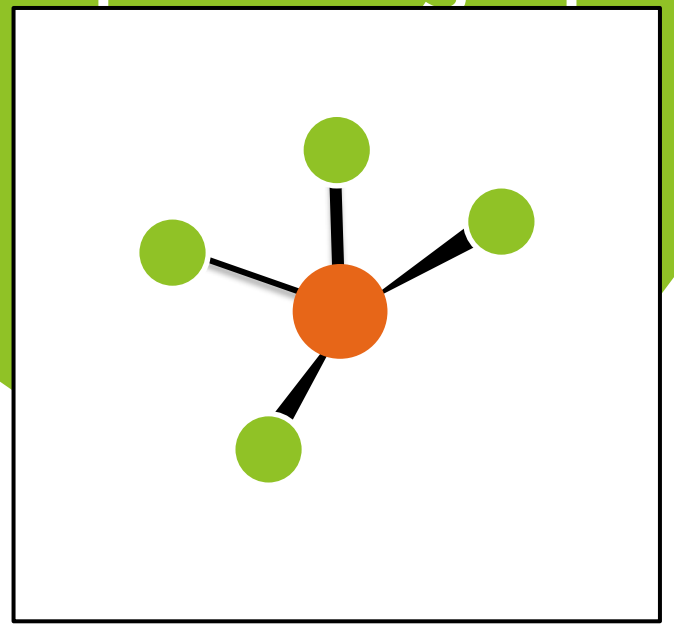


Predicting the properties of molecules

Marchwiany, M., Birowska, M., Popielski, M., Majewski, J. & Jastrzębska, A. Surface-Related Features Responsible for Cytotoxic Behavior of MXenes Layered Materials Predicted with Machine Learning Approach. Materials 13 (2020).

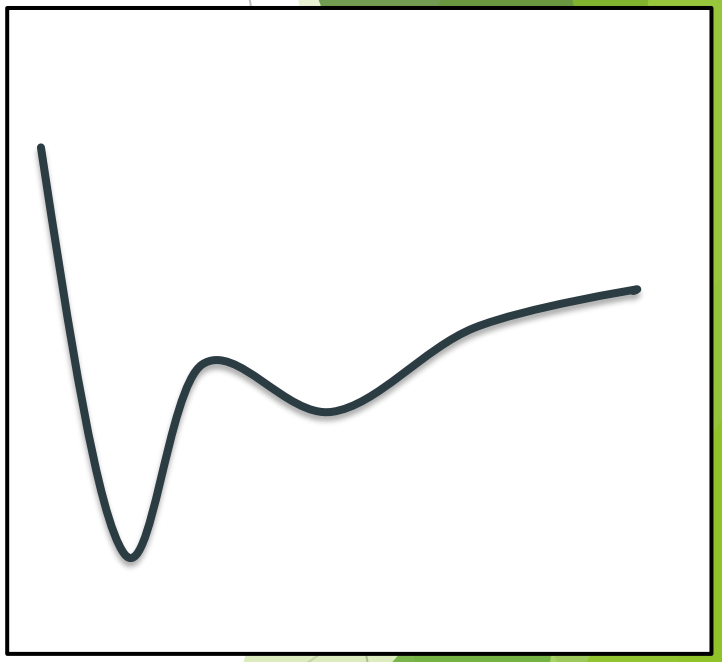


Speeding up

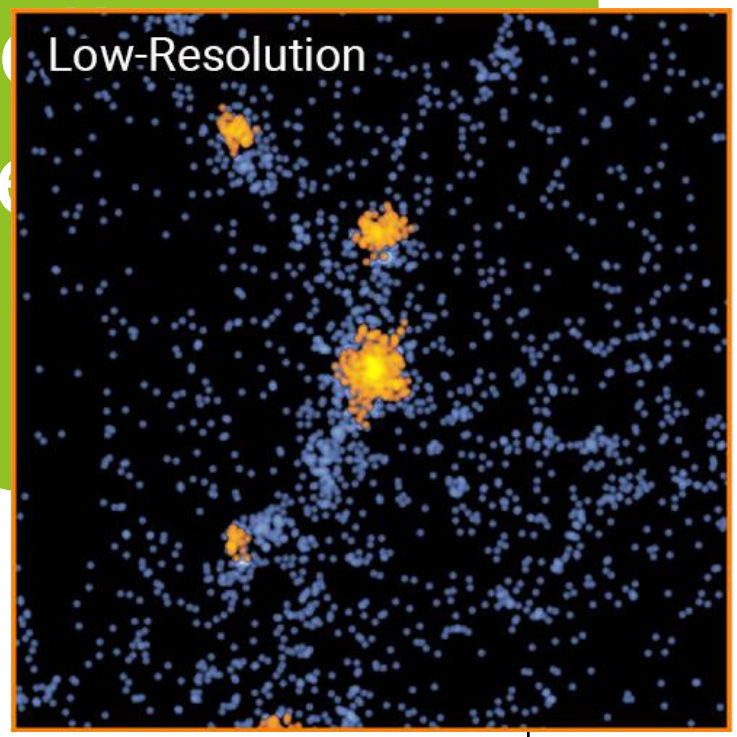


$$U(R) = \sum_{bonds} k_r \left(r - r_{eq}\right)^2$$
$$+ \sum_{at} k_\theta \left(\theta - \theta_{eq}\right)^2$$
$$+ \left[\phi - \gamma\right])$$
$$+ \quad ^2$$
$$+ \sum_{i<j}^{atoms} \varepsilon_{ij} \left[\left(\frac{r_m}{r_{ij}}\right)^{} - 2\left(\frac{r_m}{r_{ij}}\right)^6\right]$$
$$+ \sum_{i<j}^{atoms} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}$$

ML



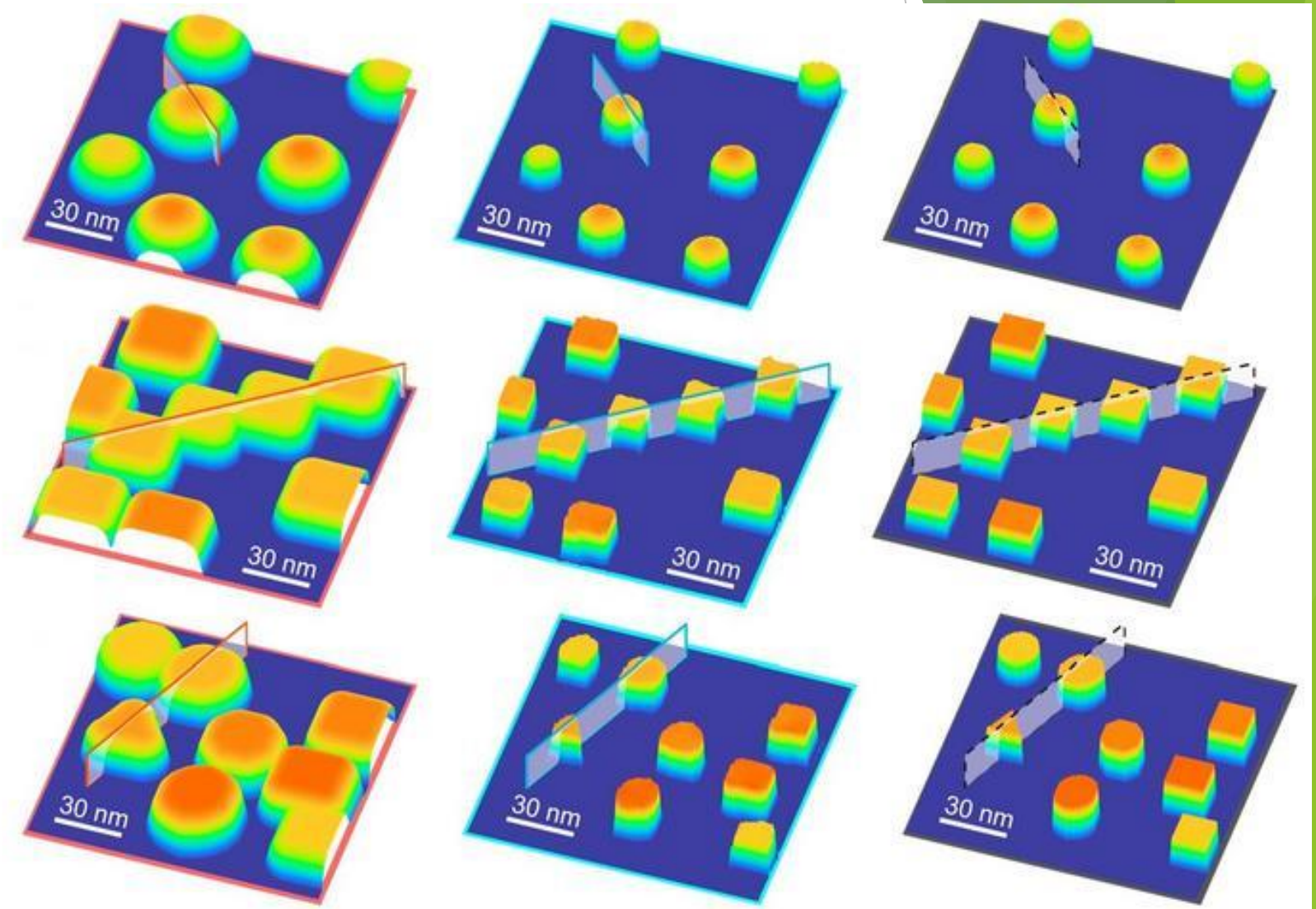
Artificial int sci



Li, Yin & Ni, Yueying & Croft, Rupert & Matteo, Tiziana & Bird, Simeon & Feng, Yu. (2021). AI-assisted superresolution cosmological simulations. Proceedings of the National Academy of Sciences of the United States of America. 118. 10.1073/pnas.2022038118.

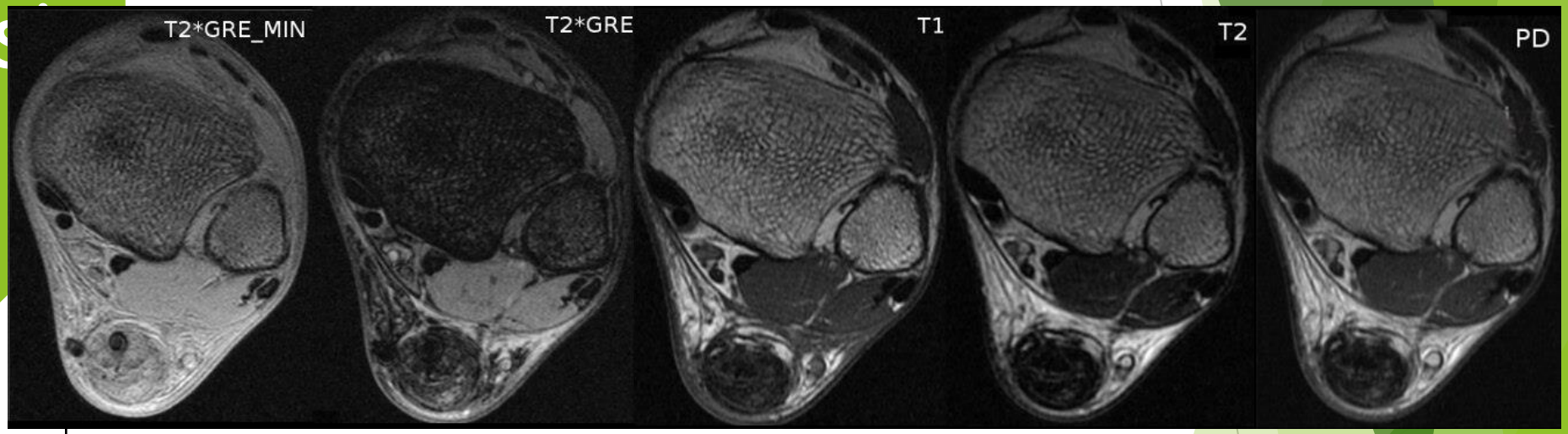
Atomic Force Microscopy

https://pubs.acs.org/doi/10.1021/acs.nanolett.3c04712



N. Kapiński, J. Nowosielski, M.E. Marchwiany, J. Zielinski, B. Ciszkowska-Lyson, B. Bartosz, T. Trzcinski, K. Nowiński, Krzysztof. (2019). Late fusion of deep learning and hand-crafted features for Achilles tendon healing monitoring.

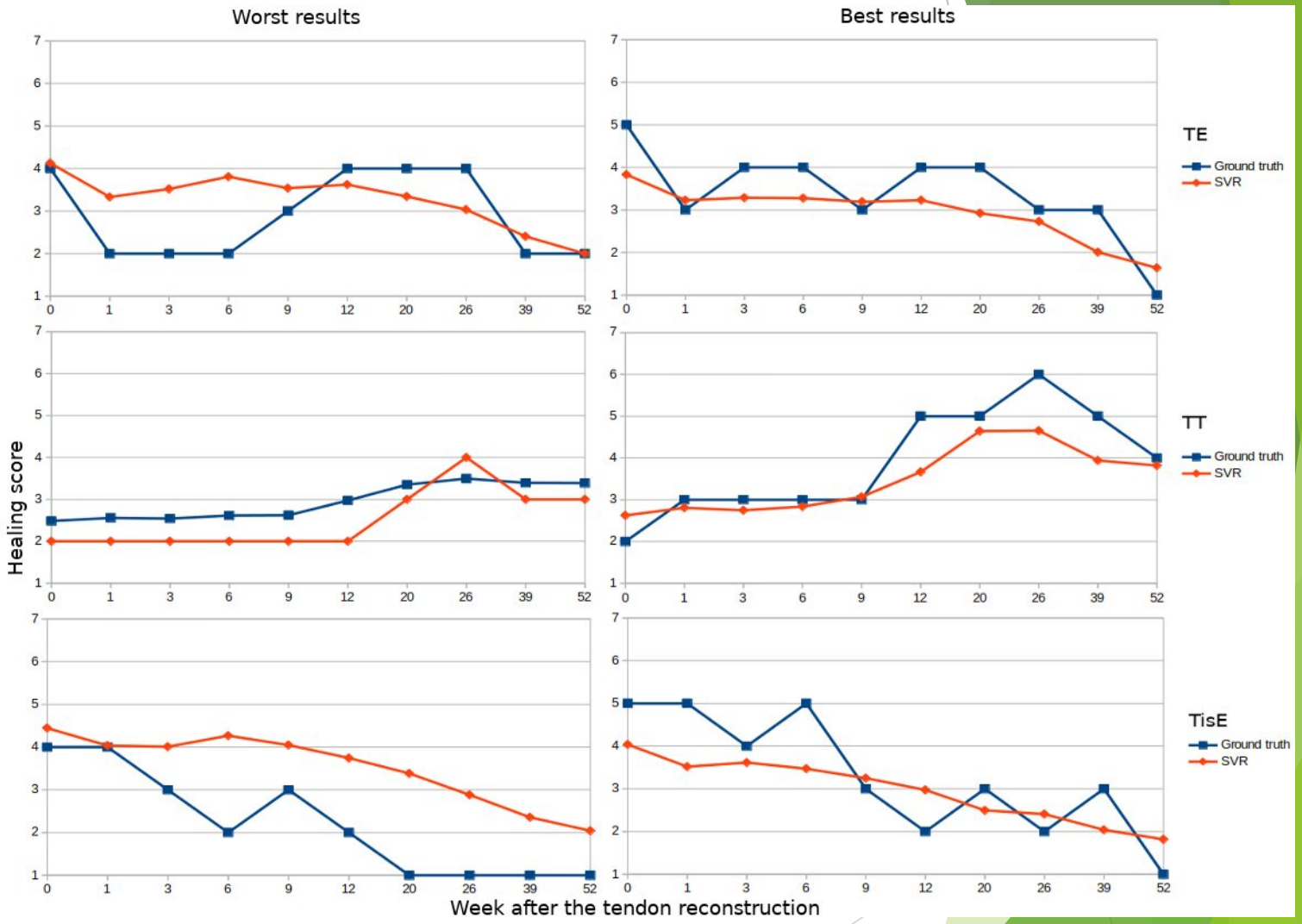
Image analys



Comparison of different MRI protocols illustrating the healing process of the Achilles tendon in the 12 week following reconstruction.

Image analysis

N. Kapiński, J. Nowosielski, M.E. Marchwiany, J. Zielinski, B. Ciszkowska-Lyson, B. Bartosz, T. Trzcinski, K. Nowiński, Krzysztof. (2019). Late fusion of deep learning and hand-crafted features for Achilles tendon healing monitoring.



Comparison of healing scores for TT, TE and TisE parameters.

Predicting chemical reactions

$CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O$

$Fe + S \rightarrow FeS$

$CaCO_3 \rightarrow CaO + CO_2$

$2H_2O \rightarrow 2H_2 + O_2$

$CaCO_3 + HCl \rightarrow CaCl_2 + CO_2 + H_2O$

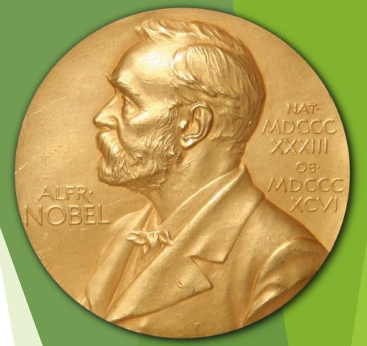
$C_4H_6O_6 + NaHCO_3 \rightarrow NaC_4H_5O_6 + H_2O + CO_2$

$CH_4 + 2O_2$ **Translation** $CO_2 + 2H_2O$

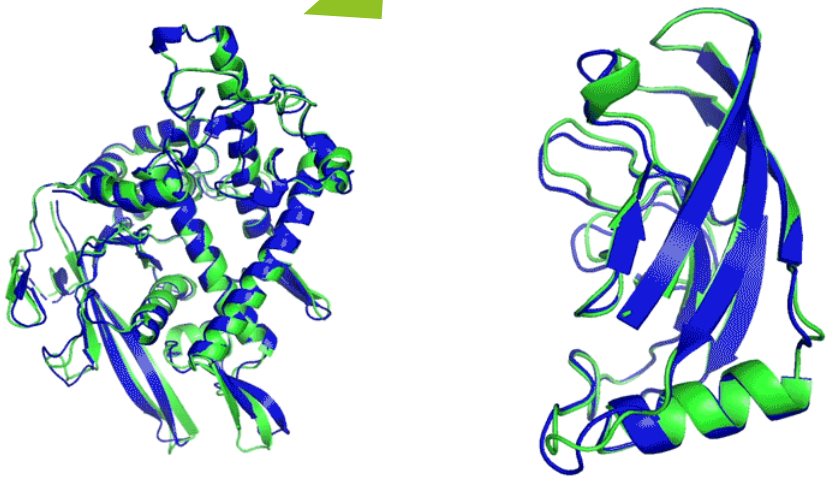
Protein folding

GAMGSEIEHIEEAIANAKTKADHERLVAHYEEEAKRLE
KKSEEYQELAKVYKKITDVYPNIRSYMVLHYQNLTRRY
KEAAEENRALAKLHHELAIVED

https://github.com/google-deepmind/alphafold



nobel prize in chemistry 2024



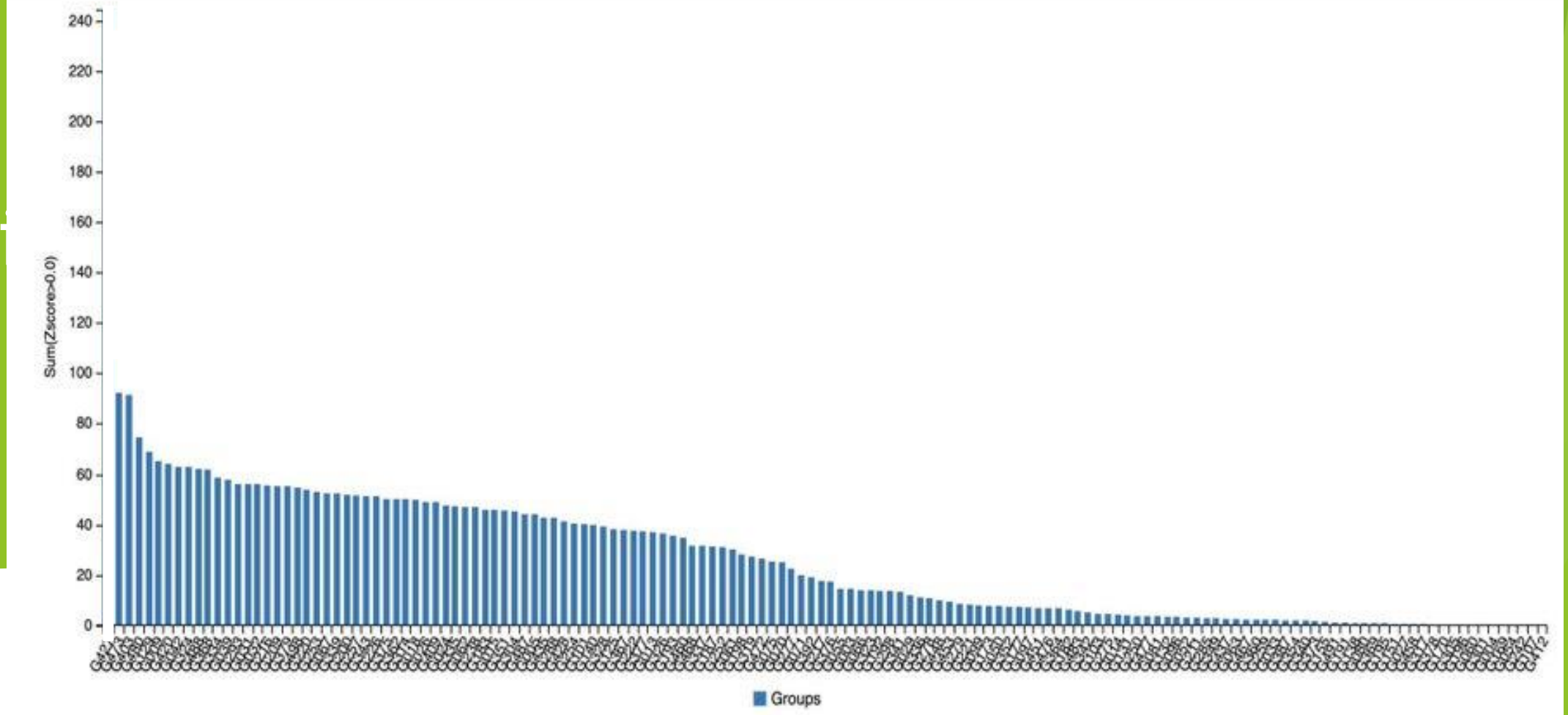
https://deepmind.google/technologies/alphafo

Protein folding

- **AlphaFold** was taught by showing the sequences and structures of around **100,000 known proteins**.
- It can now predict the **shape of a protein**, almost instantly, down to **atomic accuracy**.
- At **CASP13** (in 2018), AlphaFold came **first**. At CASP 14 (in 2020), we presented our latest version of AlphaFold – which displayed **a level of accuracy so high** that the community considered the **protein–folding problem solved**.
- Since then, the AlphaFold methods paper has received over 10,000 citations. This puts it in the top 100 most cited papers of the last decade, and in the 900 most cited papers of all time.

https://deepmind.google/technologies/alphafo

Protei



#	GR code	GR name	Domains Count	SUM Zscore (>-2.0)	Rank SUM Zscore (>-2.0)	AVG Zscore (>-2.0)	Rank AVG Zscore (>-2.0)	SUM Zscore (>0.0)	Rank SUM Zscore (>0.0)	AVG Zscore (>0.0)	Rank AVG Zscore (>0.0)
1	427	AlphaFold2	92	244.0217	1	2.6524	1	244.0217	1	2.6524	1
2	473	BAKER	92	90.8241	2	0.9872	2	92.1241	2	1.0013	2

AI for experiments planning

Experiments design:

- **Hypothesis Generation**: AI can analyze vast amounts of existing literature to generate novel hypotheses: identifying patterns and knowledge gaps
- **Methodology Recommendations**: AI can recommend appropriate experimental designs and methodologies**:** determining sample sizes and suggesting optimal settings for experiments
- **Resource Optimization**: AI tools can propose experimental setups that maximize resource efficiency, helping researchers save time and costs associated with traditional experimental approaches.

AI for experiments planning

Experiment Implementation

- **Real-Time Assistance**: During experiments, AI systems can function as interactive assistants, providing immediate support to participants: clarifying instructions, answering questions, and monitoring parameters
- **Automation of Tasks**: Advanced AI systems can automate routine laboratory tasks, such as controlling equipment or managing data collection processes.
- **Active Learning**: Some AI systems employ active learning techniques to suggest subsequent experimental parameters based on previous results. This iterative approach helps refine experiments by focusing on areas of uncertainty or confusion

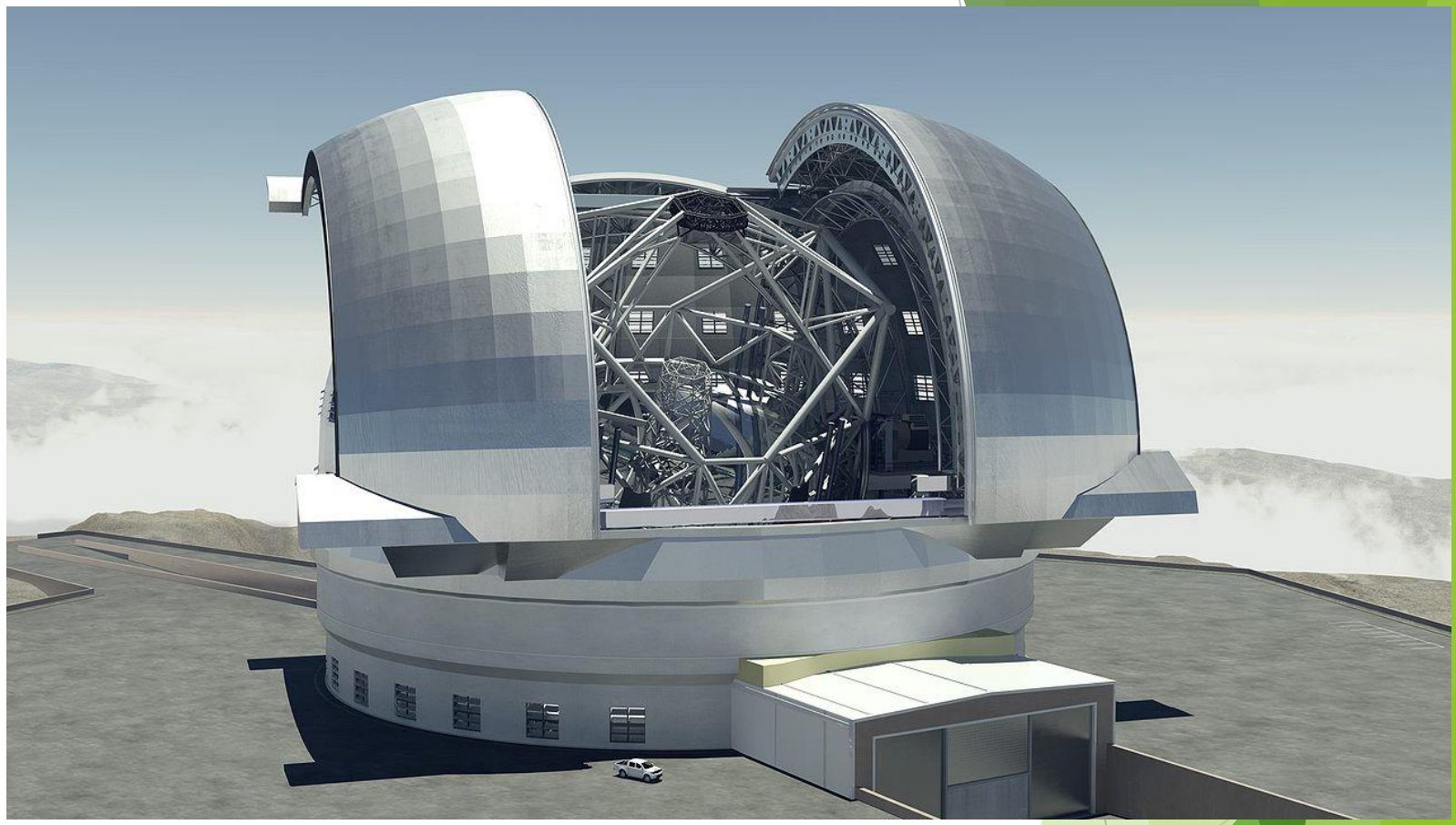
AI for experiments planning

Data Analysis

- **Data Processing and Interpretation**: AI excels at processing large datasets quickly and accurately: statistical analyses, identify new variables, and generate visualizations
- **Natural Language Processing**: By applying NLP techniques to data from experiments, AI can extract meaningful patterns related to participant sentiments or cognitive processes, enhancing the understanding of experimental results
- **Reproducibility and Simulation**: AI can assist in reproducing results through simulations based on the experimental data: validating findings and ensuring that they are robust across different contexts

Artificial intelligence in experiment

https://en.wikipedia.org/wiki/Extremely_Large_Telescope



Extremely Large Telescope (ELT)
Segmented main mirror :
diameter : 39.3 m
Number of segments: 798

LLM in Science

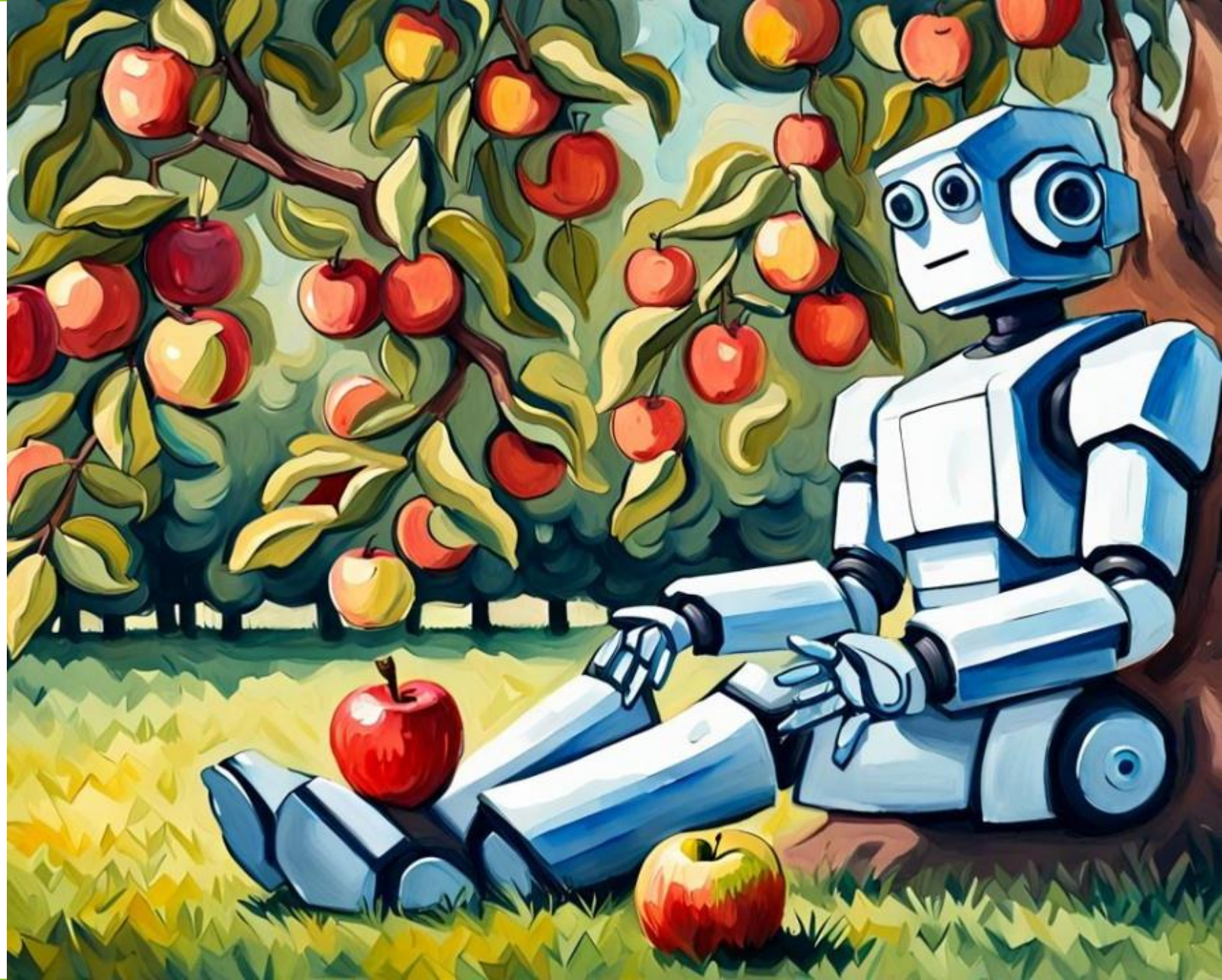
- Analyzes of **scientific articles**
  - Summaries Content search
  - Citation analysis
- **Information extraction** from the article database
- **Designing** molecules based on **description**
- Predicting **chemical reactions**

Other AI in science

- Documents Management
- Reports Generation
- Presentations creating
- Meetings Summarization
- Tasks Management
- Collaboration Tools

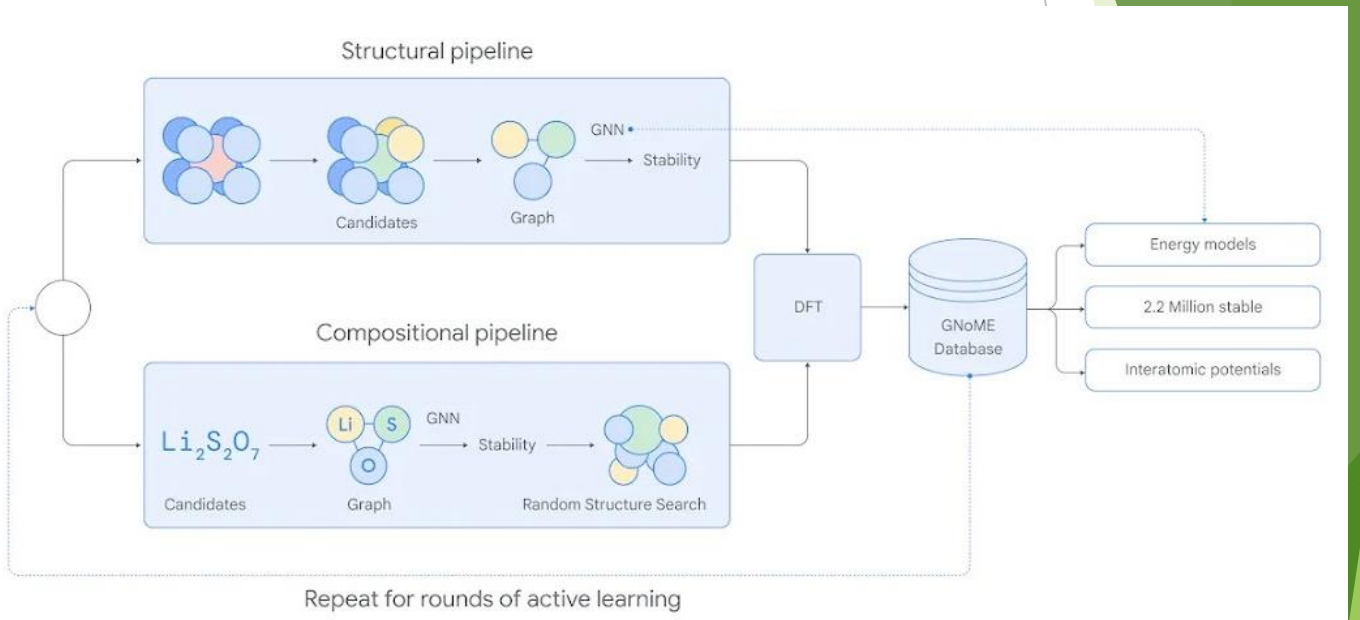
3

AI as scientist



GNOME

Merchant, A., Batzner, S., Schoenholz, S.S. *et al.* **Scaling deep learning for materials discovery**. *Nature* **624**, 80–85 (2023).



iLAB

Schneider, G. Automating drug discovery. *Nat Rev Drug Discov* **17**, 97–113 (2018). https://doi.org/10.1038/nrd.2017.232

iLAB

Schneider, G. Automating drug discovery. *Nat Rev Drug Discov* **17**, 97–113 (2018). https://doi.org/10.1038/nrd.2017.232

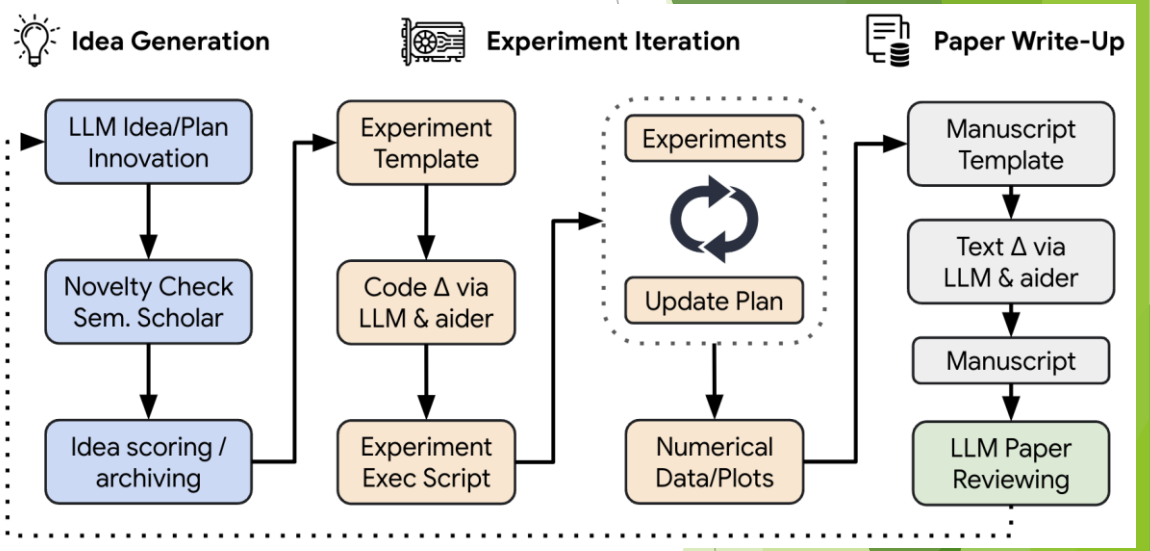
iLAB

Schneider, G. Automating drug discovery. *Nat Rev Drug Discov* **17**, 97–113 (2018). https://doi.org/10.1038/nrd.2017.232

AI Scientist

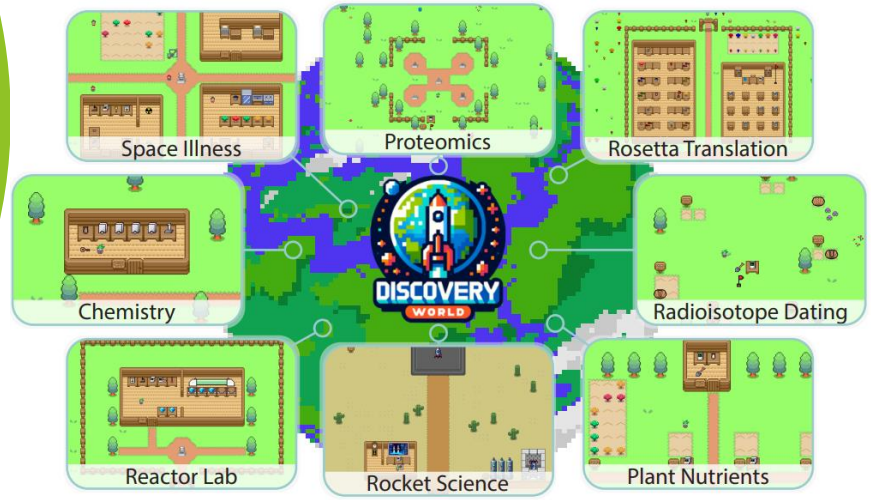
Lu, C., Lu, C., Lange, R. T. *et al.* **The AI Scientist: Towards Fully Automated Open-Ended Scientific Discovery.** arxiv (2024)

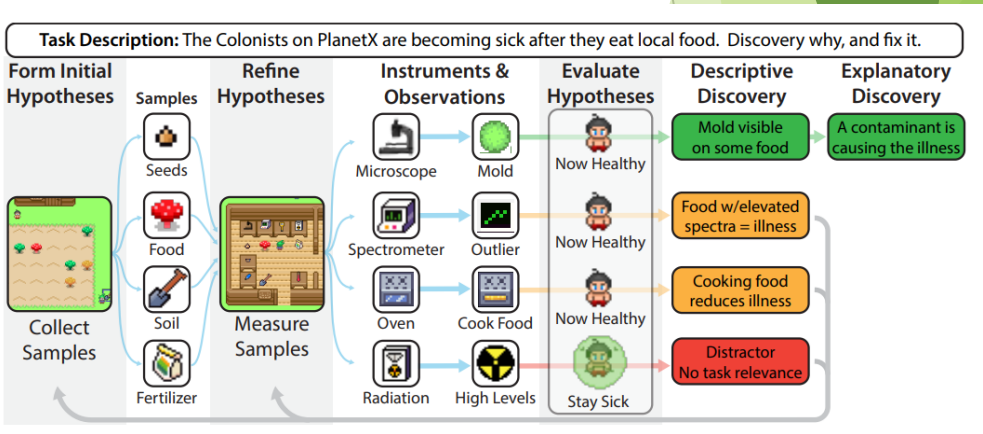
1. Idea Generation
2. Literature Review
3. Experimental Design and Planning
4. Execution of Experiments
5. Data Analysis
6. Manuscript Writing
7. Automated Peer Review
8. Iterative Improvement



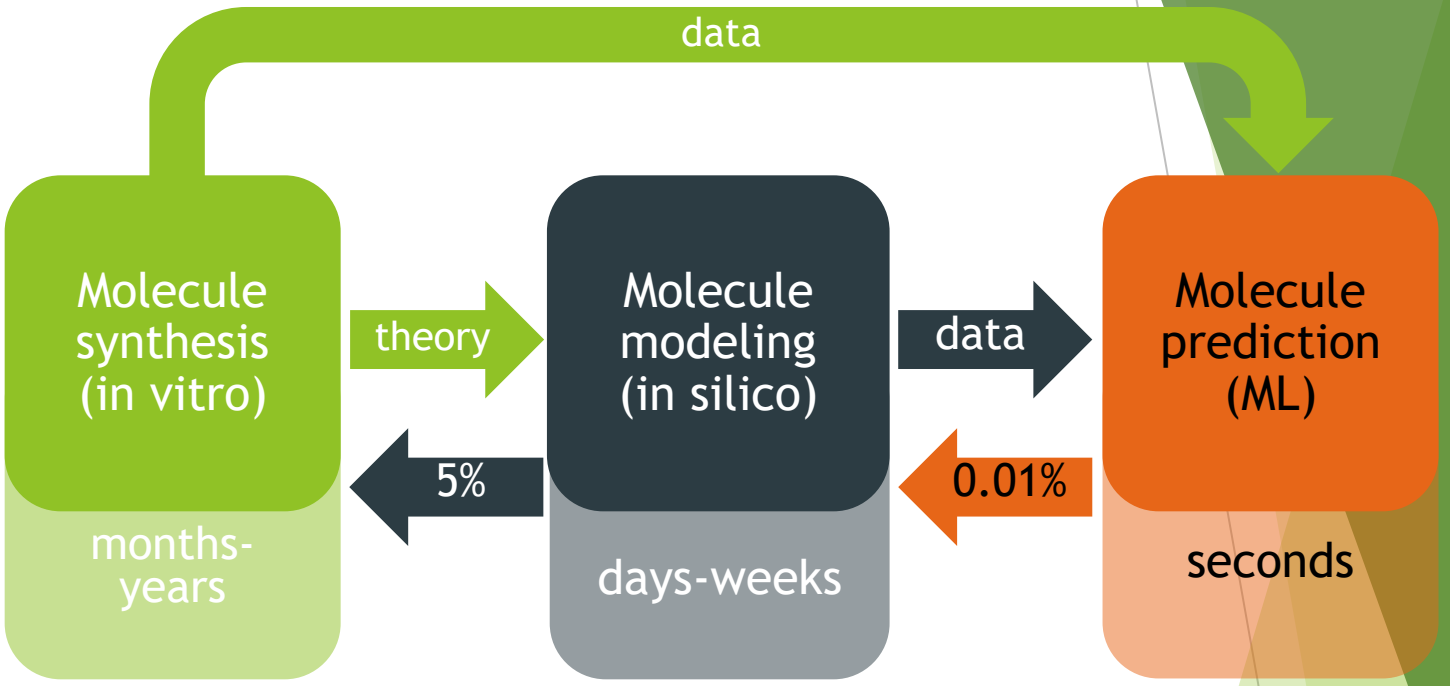
AI Scientist

Jansen, P., Côté, M.-A., Khot, T *et al.* **DISCOVERYWORLD: A virtual environment for developing and evaluating automated scientific discovery agents.** arxiv (2024)



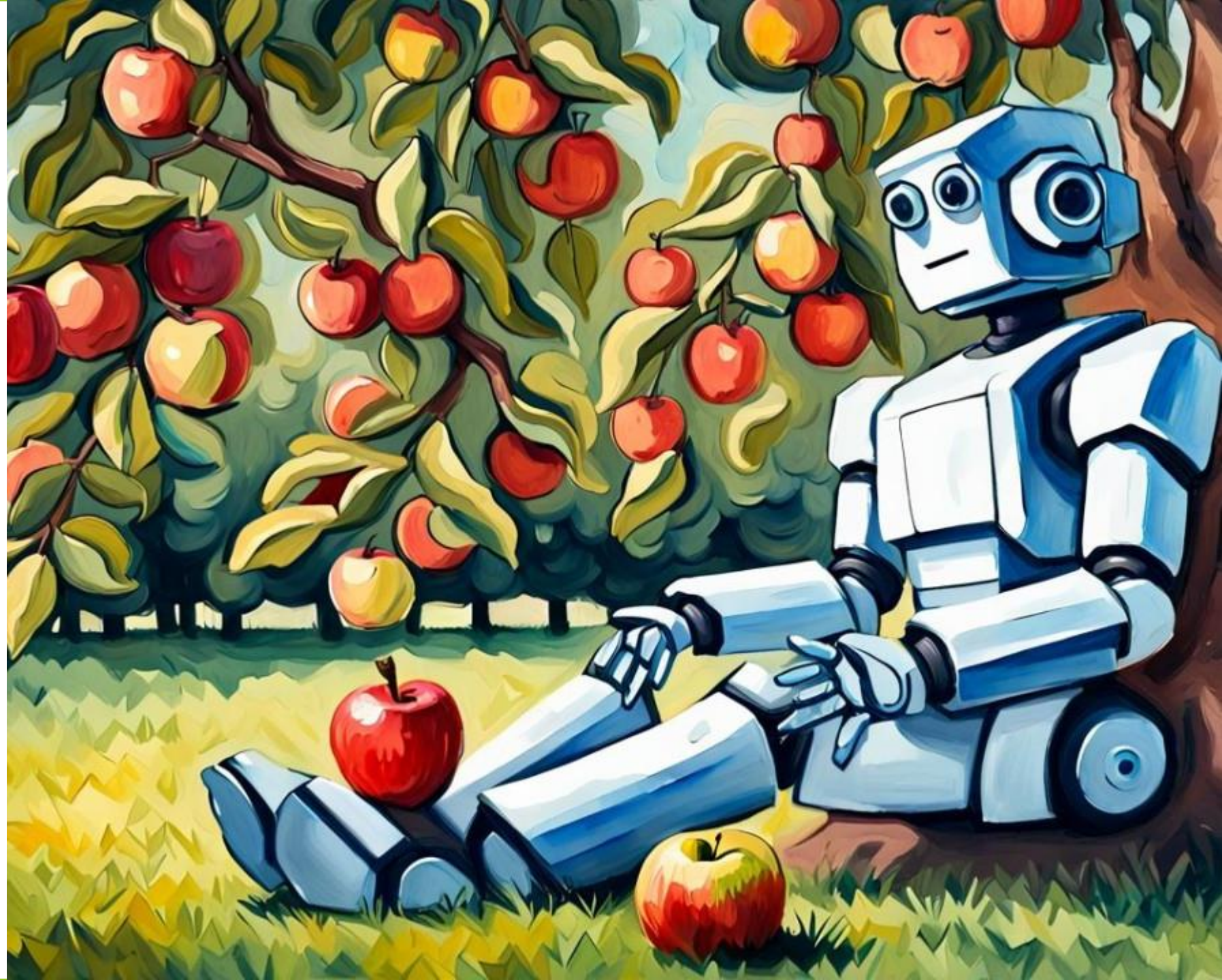


BUT …


data
Molecule synthesis (in vitro)
theory
Molecule modeling (in silico)
data
Molecule prediction (ML)
5%
0.01%
months-years
days-weeks
seconds

4

Why is this difficult?



Challenges

- AI **ethics**
- AI **security**
- **Data:**
  - Data sources
  - Data diversity
  - Data quality
- **Complexity of calculations**
- Model **validation**
- **Detecting false** predictions
- **Hallucinations**
- **Interpretability**
- **Knowledge generalization**

Questions?