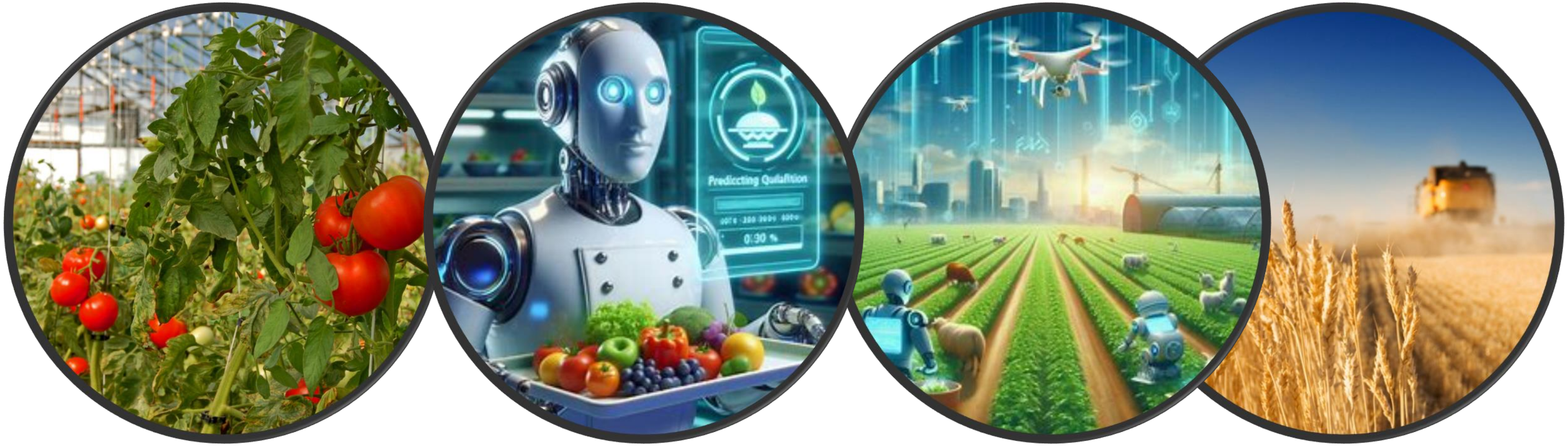
Scientific Machine Learning (SciML) in food-related domains

Dr. Xuezhen Guo and Prof.dr.ir. RGM (Ruud) van der Sman

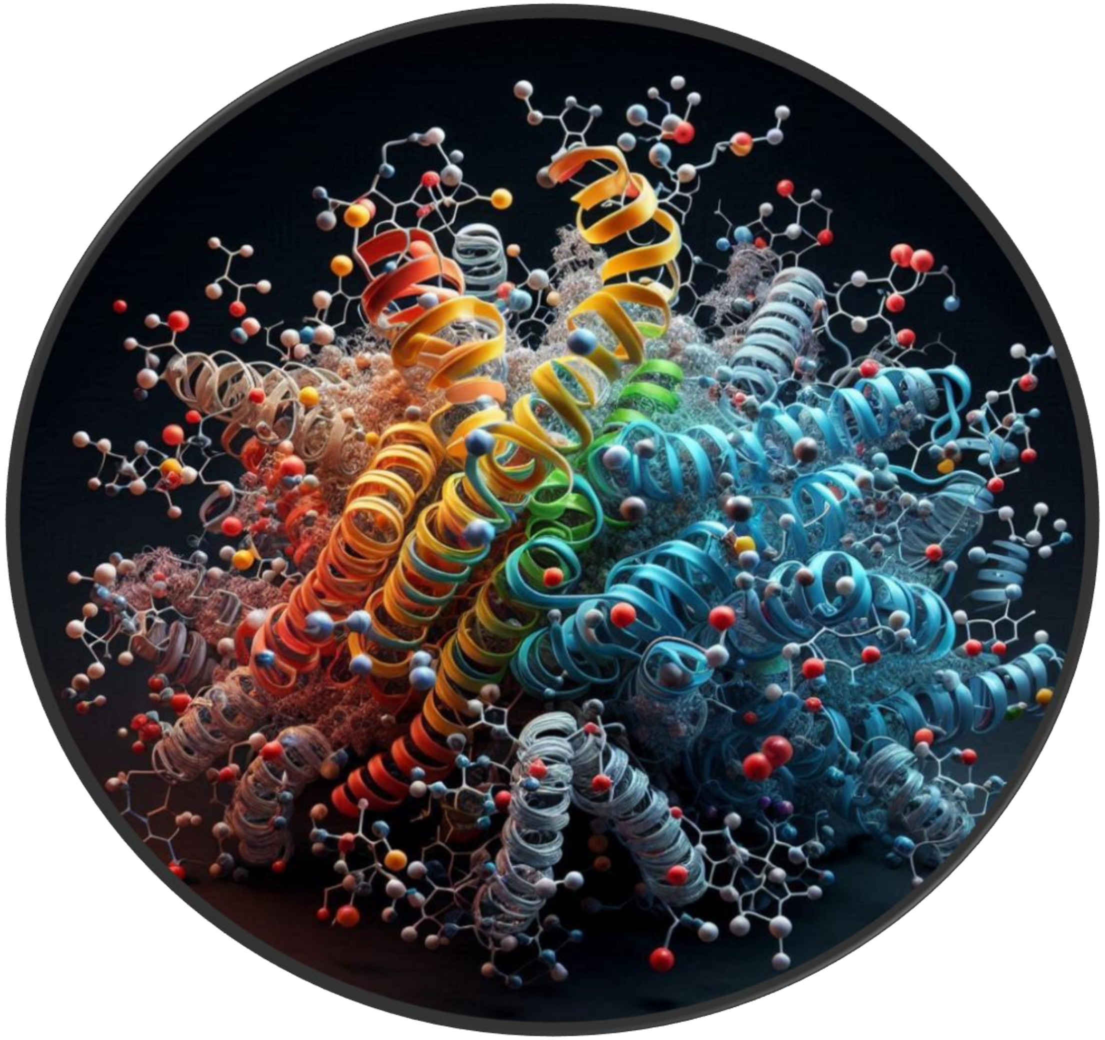
Wageningen Food & Biobased Research



WAGENINGEN
UNIVERSITY & RESEARCH

Content

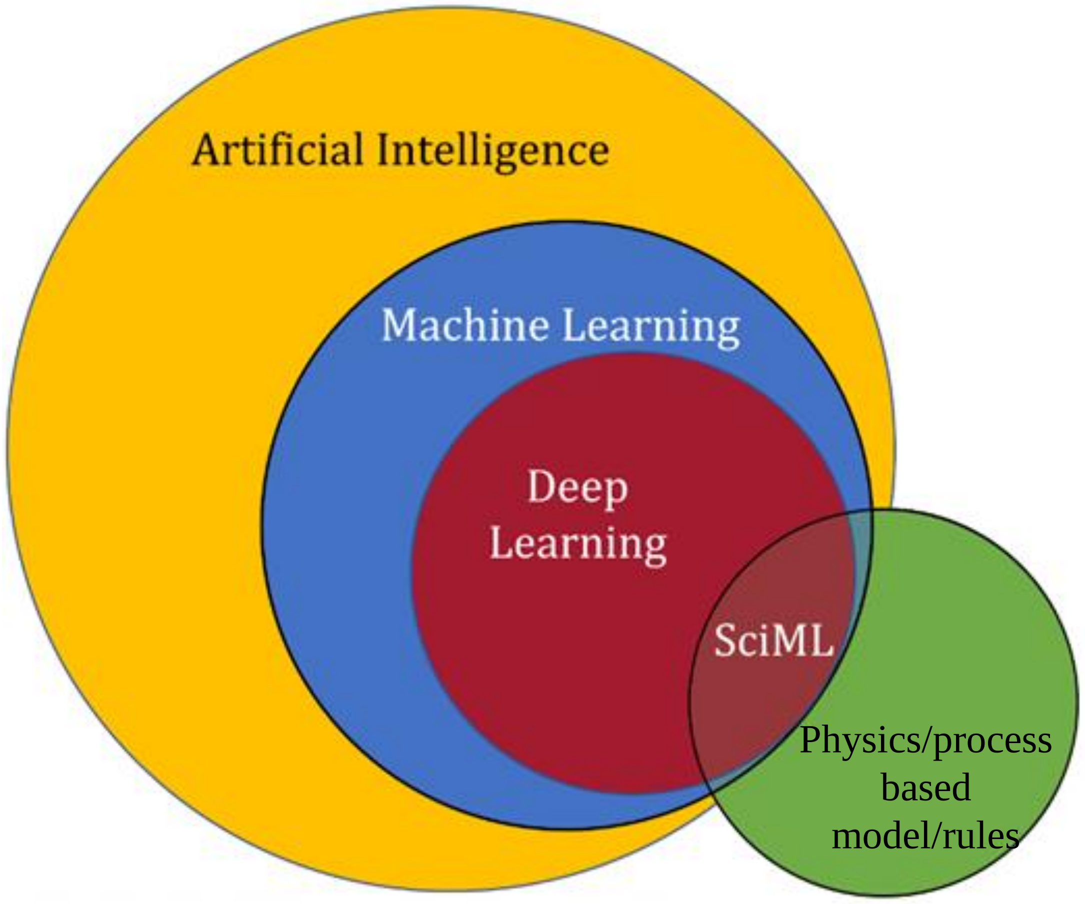
- Scientific machine learning (SciML)
- Example of SciML: Alphafold2
- SciML for WFBR topics



Part 1: Scientific machine learning (SciML)

What is scientific machine learning and why?

- Physics/process-based model/ rules + neural networks (deep learning).
- Lack of data
- Domain knowledge reduces data requirement for model training



https://sciml.wur.nl/reviews/sciml/sciml.html

Nobel prize 2024: the year of AI and SciML

The Nobel Prize in Physics 2024

Ill. Niklas Elmehed © Nobel Prize Outreach

John J. Hopfield

Prize share: 1/2

Ill. Niklas Elmehed © Nobel Prize Outreach

Geoffrey E. Hinton

Prize share: 1/2

The Nobel Prize in Chemistry 2024

Ill. Niklas Elmehed © Nobel Prize Outreach

David Baker

Prize share: 1/2

Ill. Niklas Elmehed © Nobel Prize Outreach

Demis Hassabis

Prize share: 1/4

Ill. Niklas Elmehed © Nobel Prize Outreach

John M. Jumper

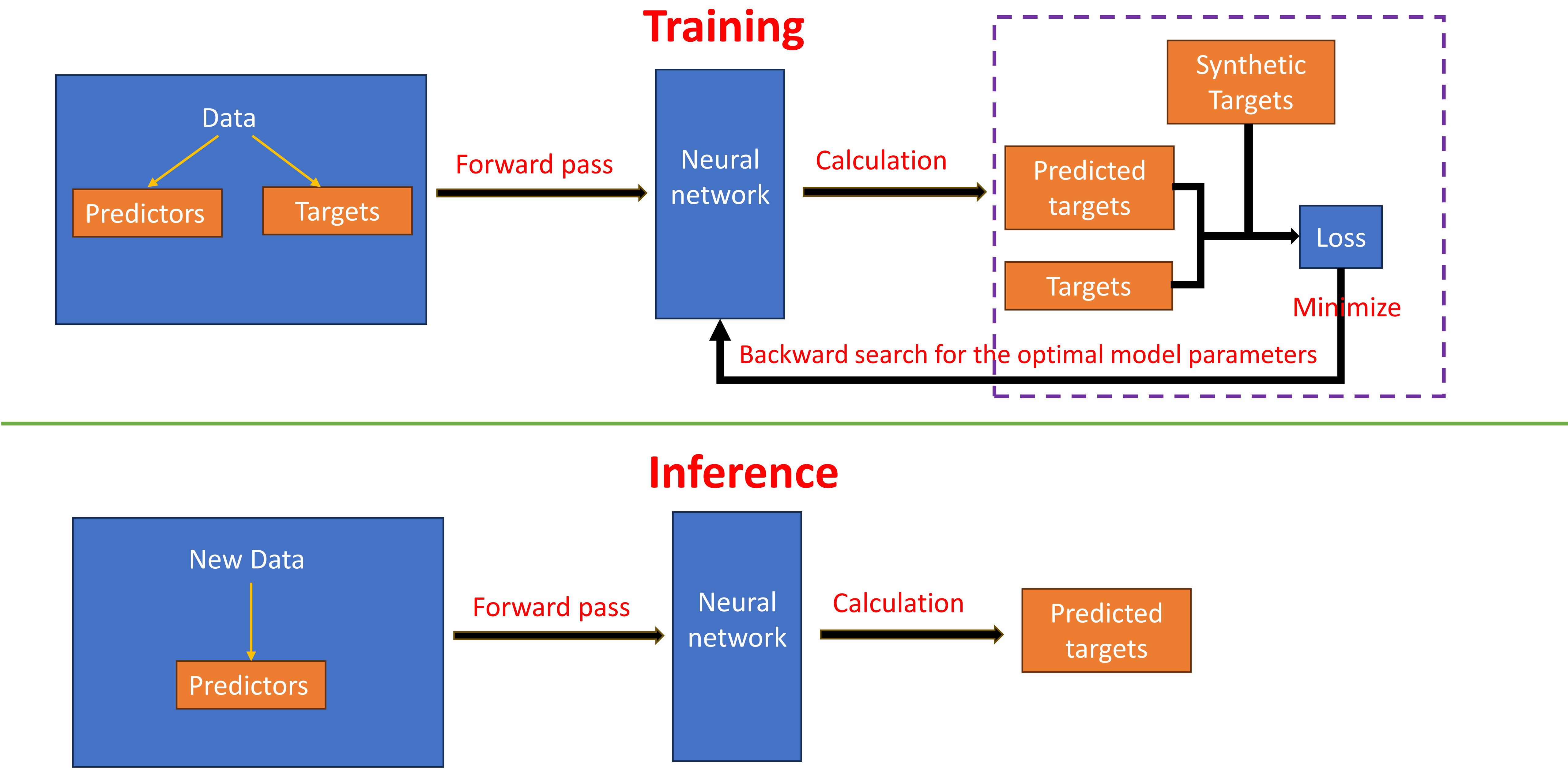
Prize share: 1/4

Neural networks

Domain-knowledge based Computational model

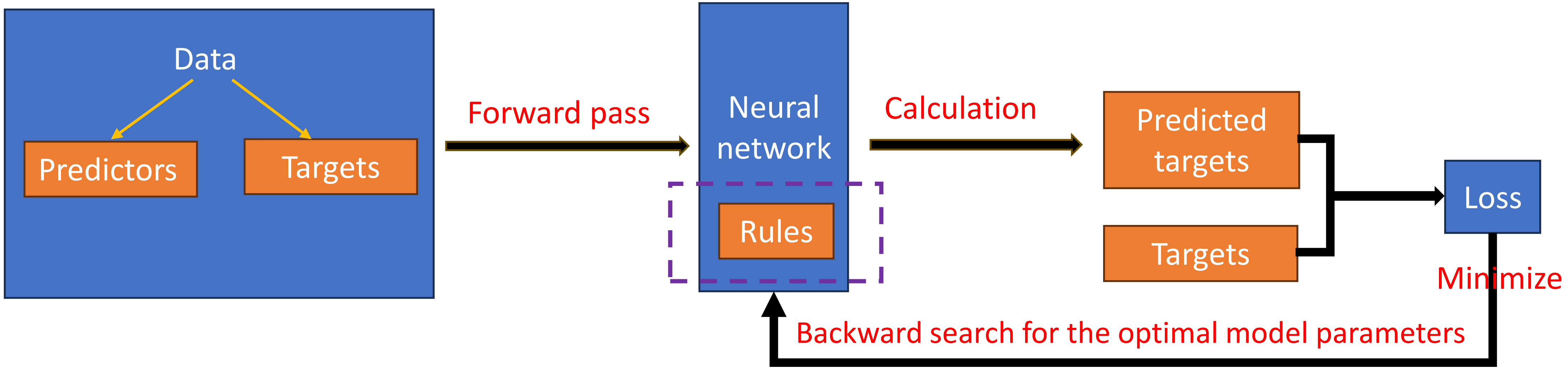
SciML: Alphafold2

SciML: 1. Physics-guided model

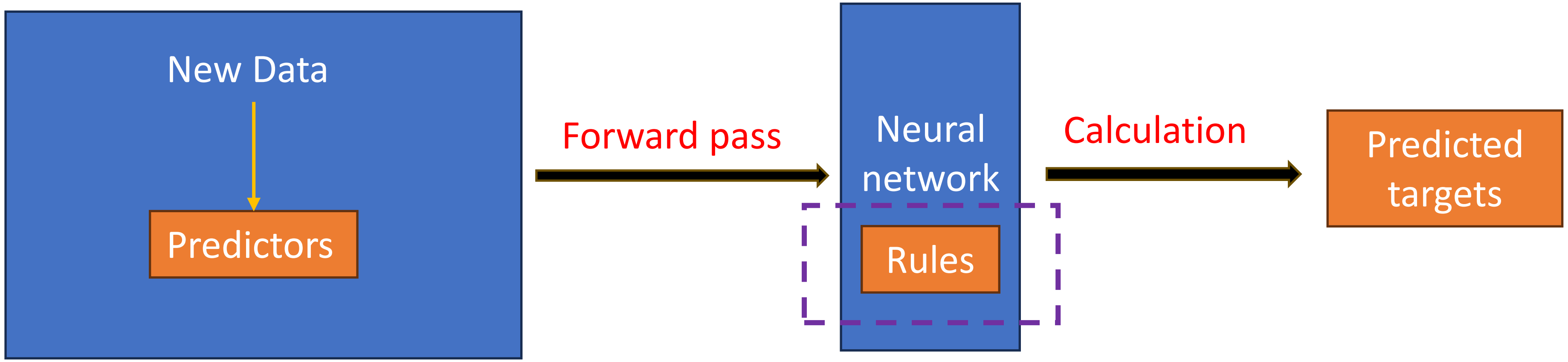


SciML: 2. Physics-encoded model

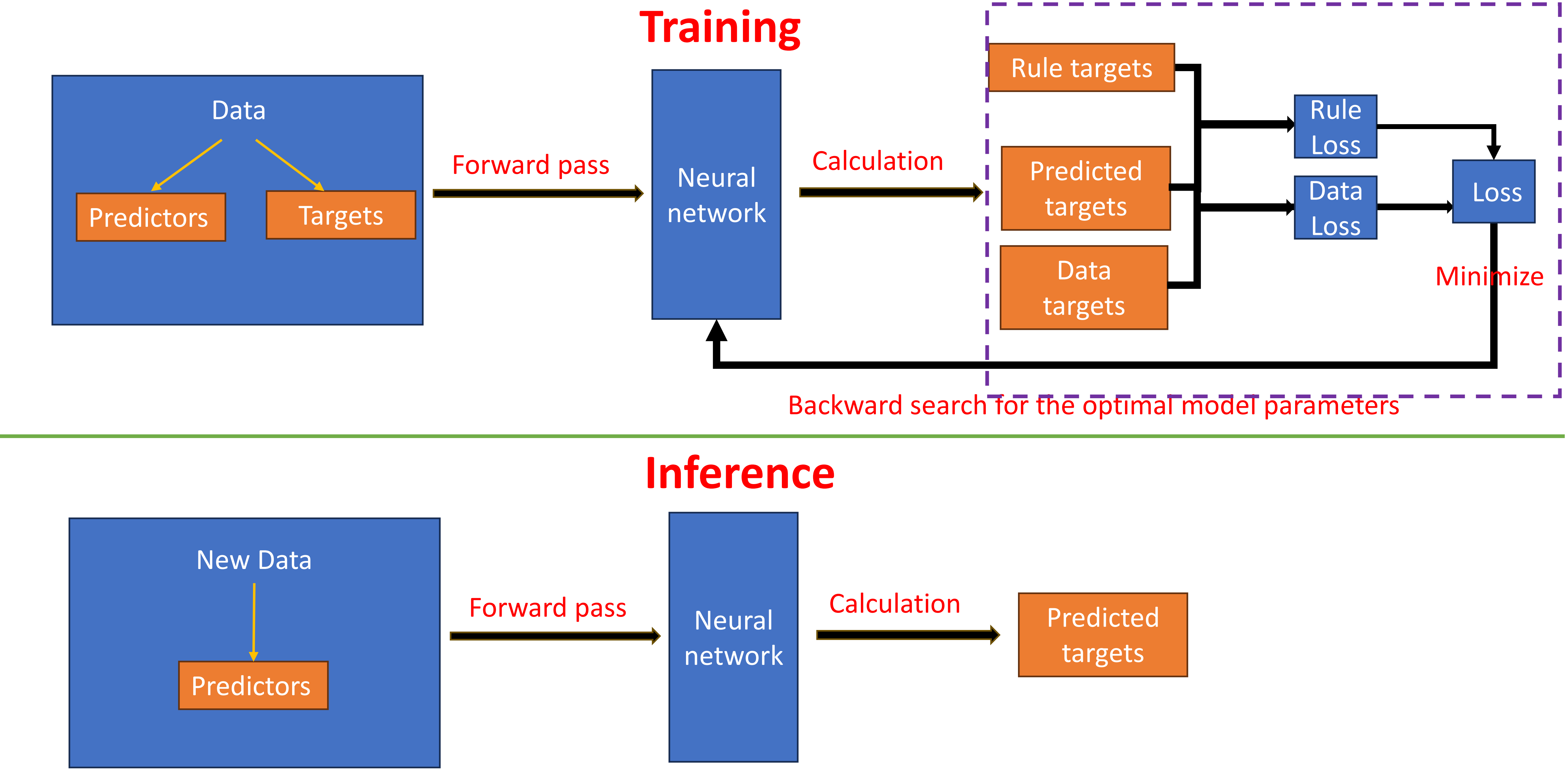
Training



Inference

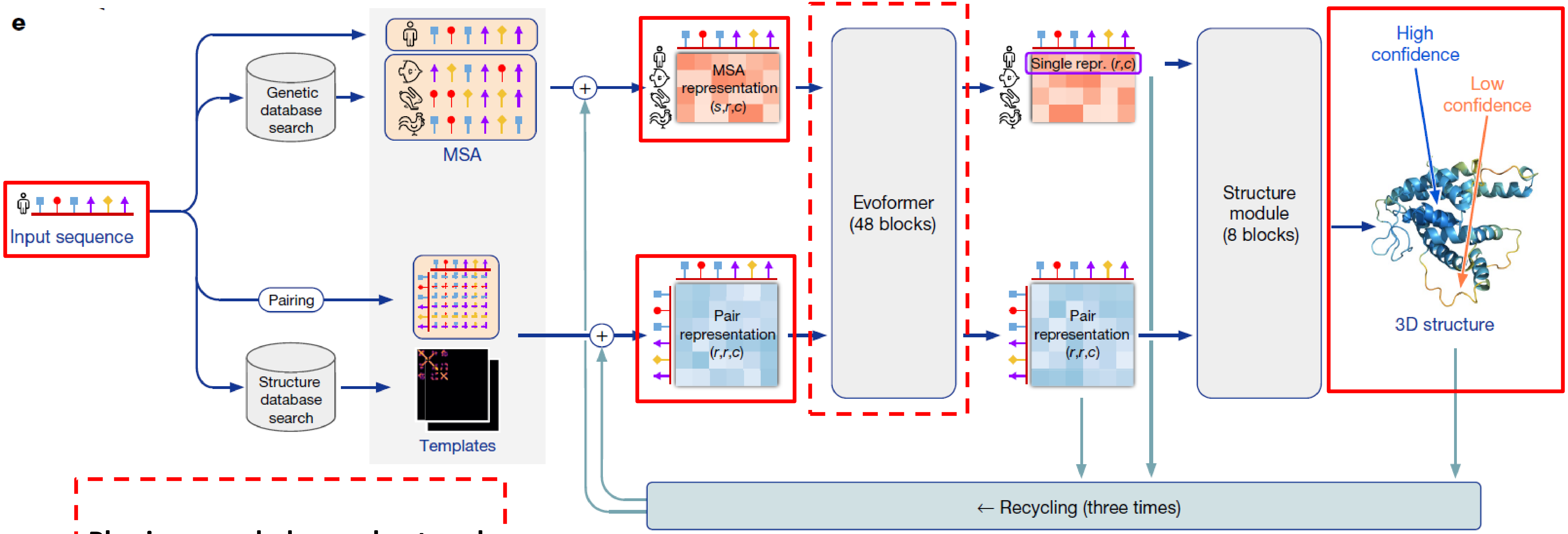


SciML: 3. Physics-informed model



Part 2: Example of SciML
Alphafold2

AlphaFold2: Model architecture



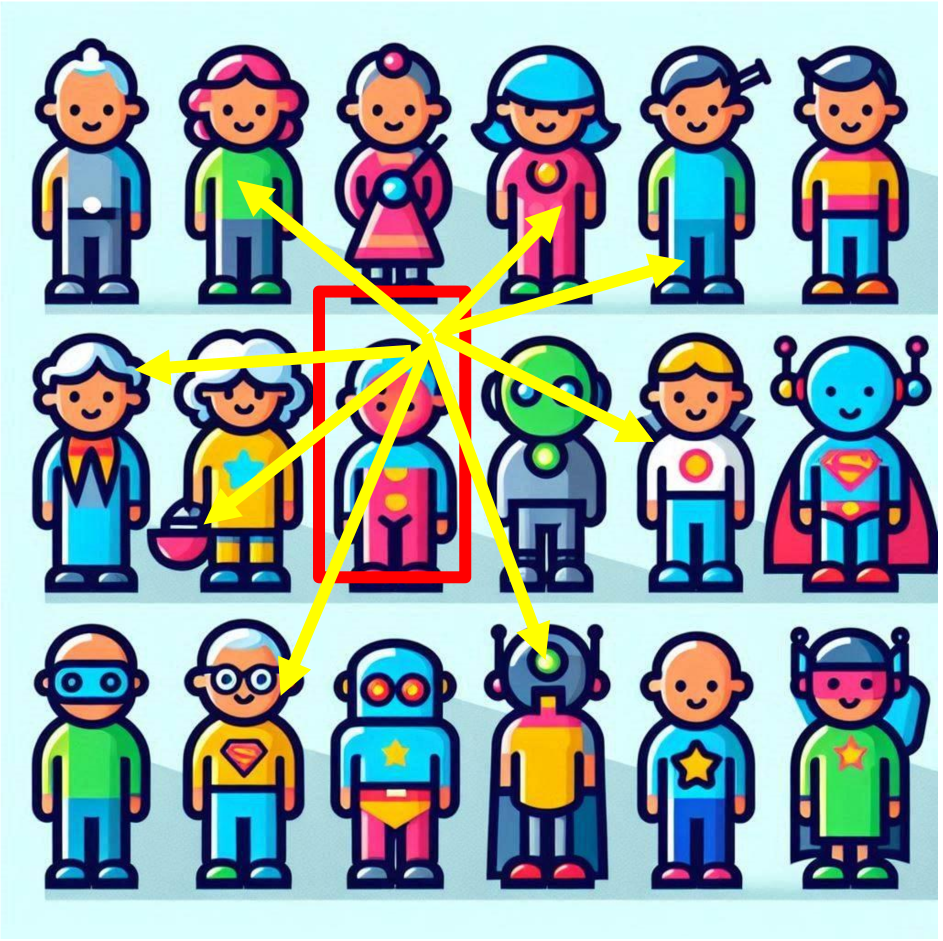
Physics-encoded neural networks

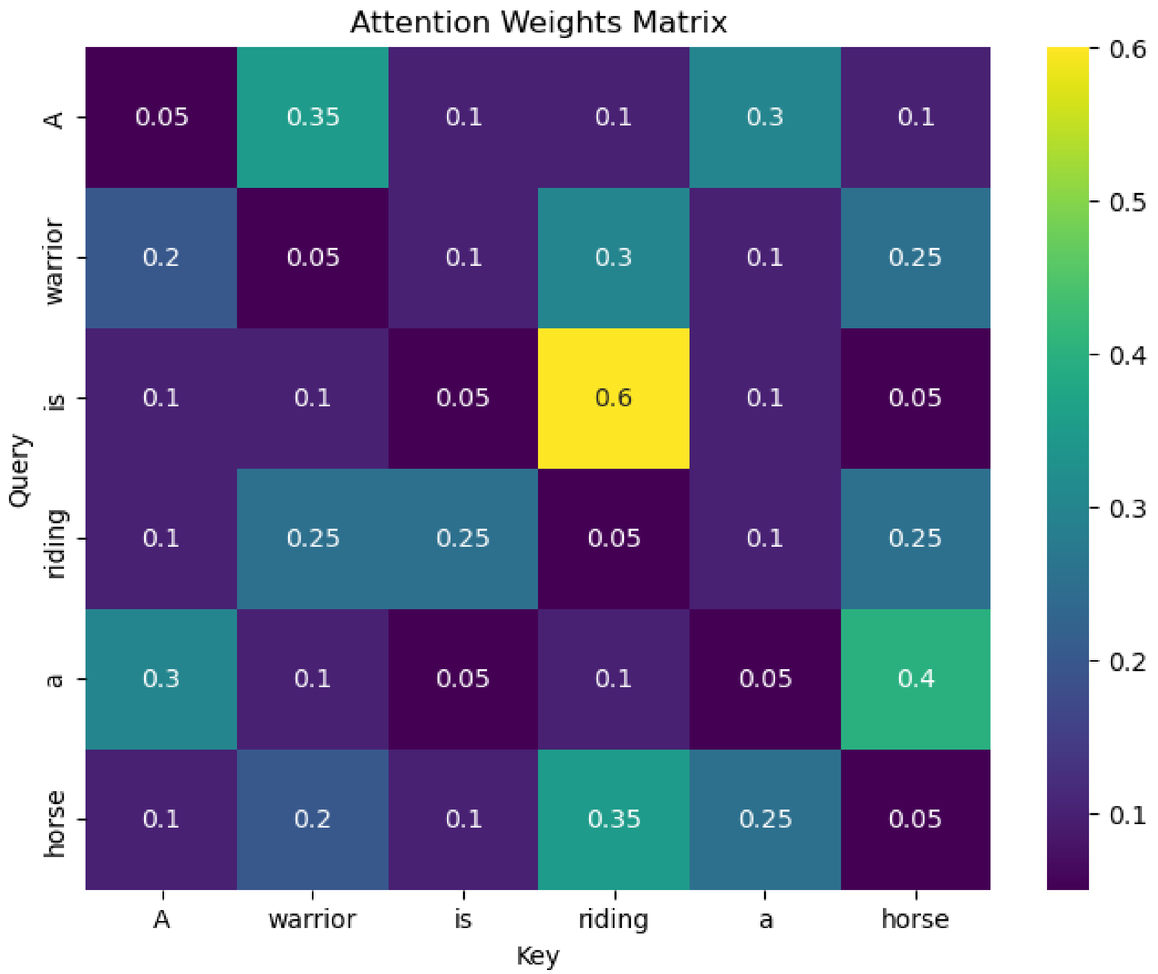
Self-Attention mechanism

(Self-) Attention

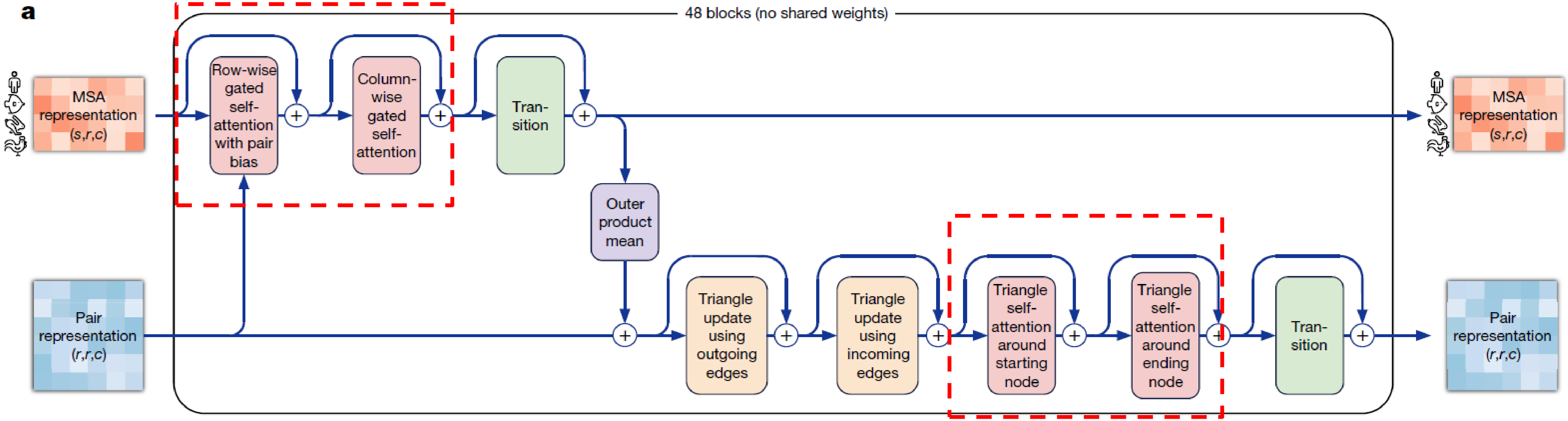
Positional encoding

A warrior is riding a horse





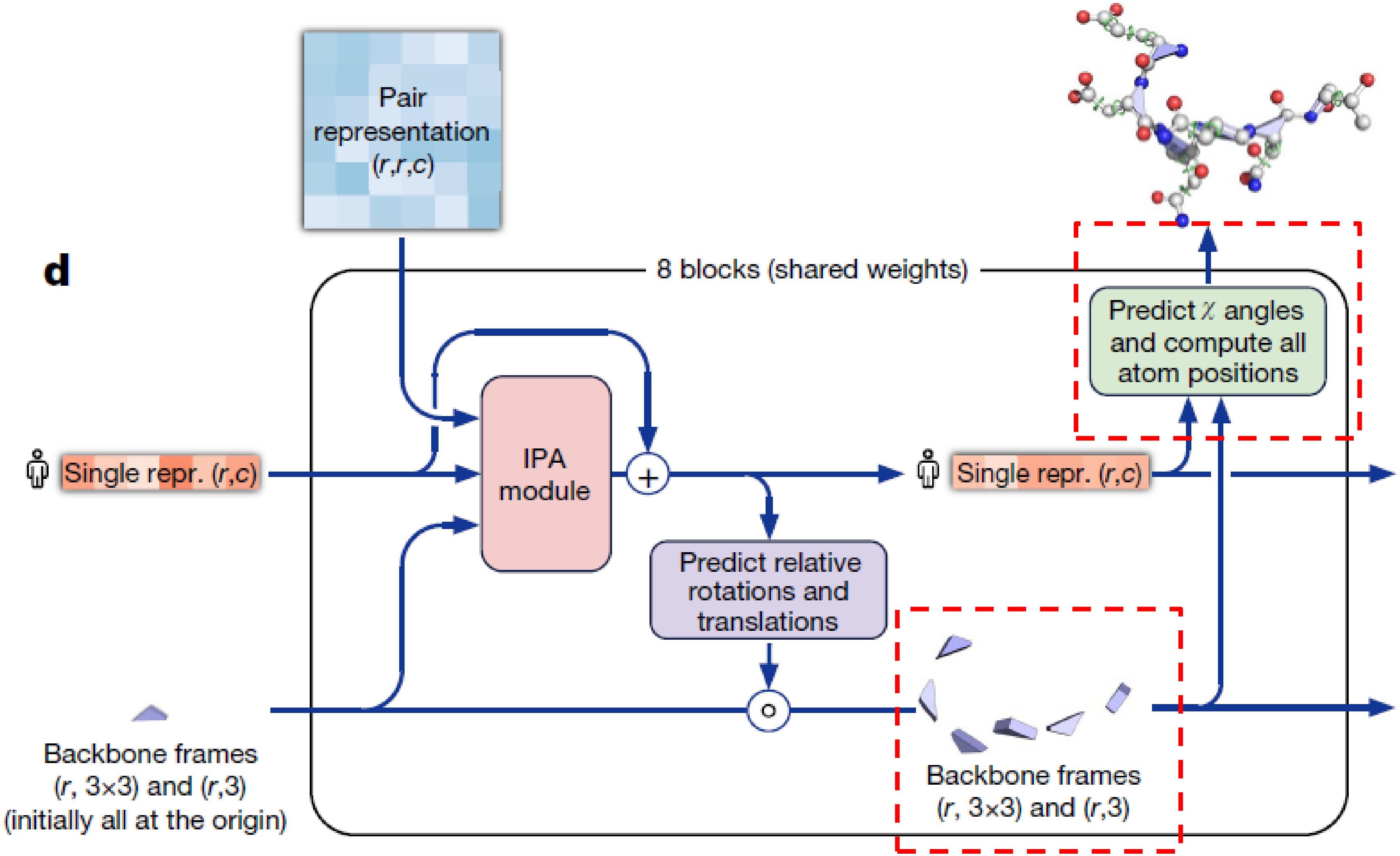
AlphaFold2: Evoformer



Physics-encoded neural networks

Self-Attention mechanism

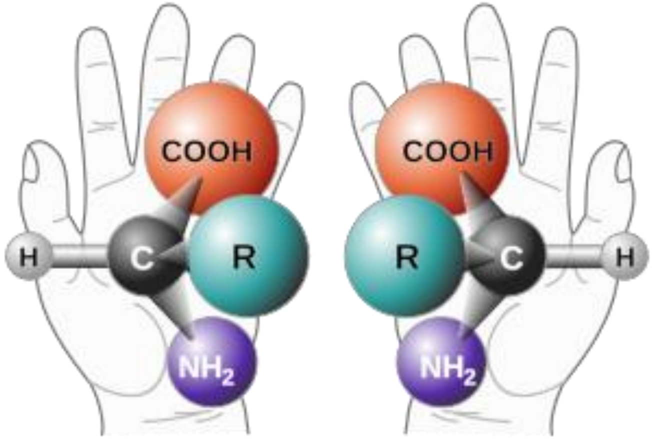
AlphaFold2: Structure module



AlphaFold2: physics-informed loss function

$$L = \begin{cases} 0.5L_{FAPE} + 0.5L_{aux} + 0.3L_{dist} + 2L_{msa} + 0.01L_{conf} & \text{training} \\ 0.5L_{FAPE} + 0.5L_{aux} + 0.3L_{dist} + 2L_{msa} + 0.01L_{conf} + 0.01L_{exp\ resolved} + 1.0L_{viol} & \text{fine tuning} \end{cases}$$

Frame aligned point error (FAPE)



Chirality

RMSD-like loss does not ensure chirality

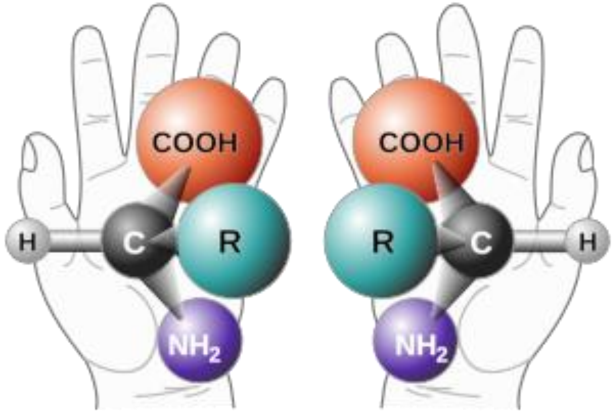
AlphaFold2: Frame aligned point error (FAPE)

$$1:\ \vec{\mathbf{x}}_{ij} = T_i^{-1} \circ \vec{\mathbf{x}}_j$$

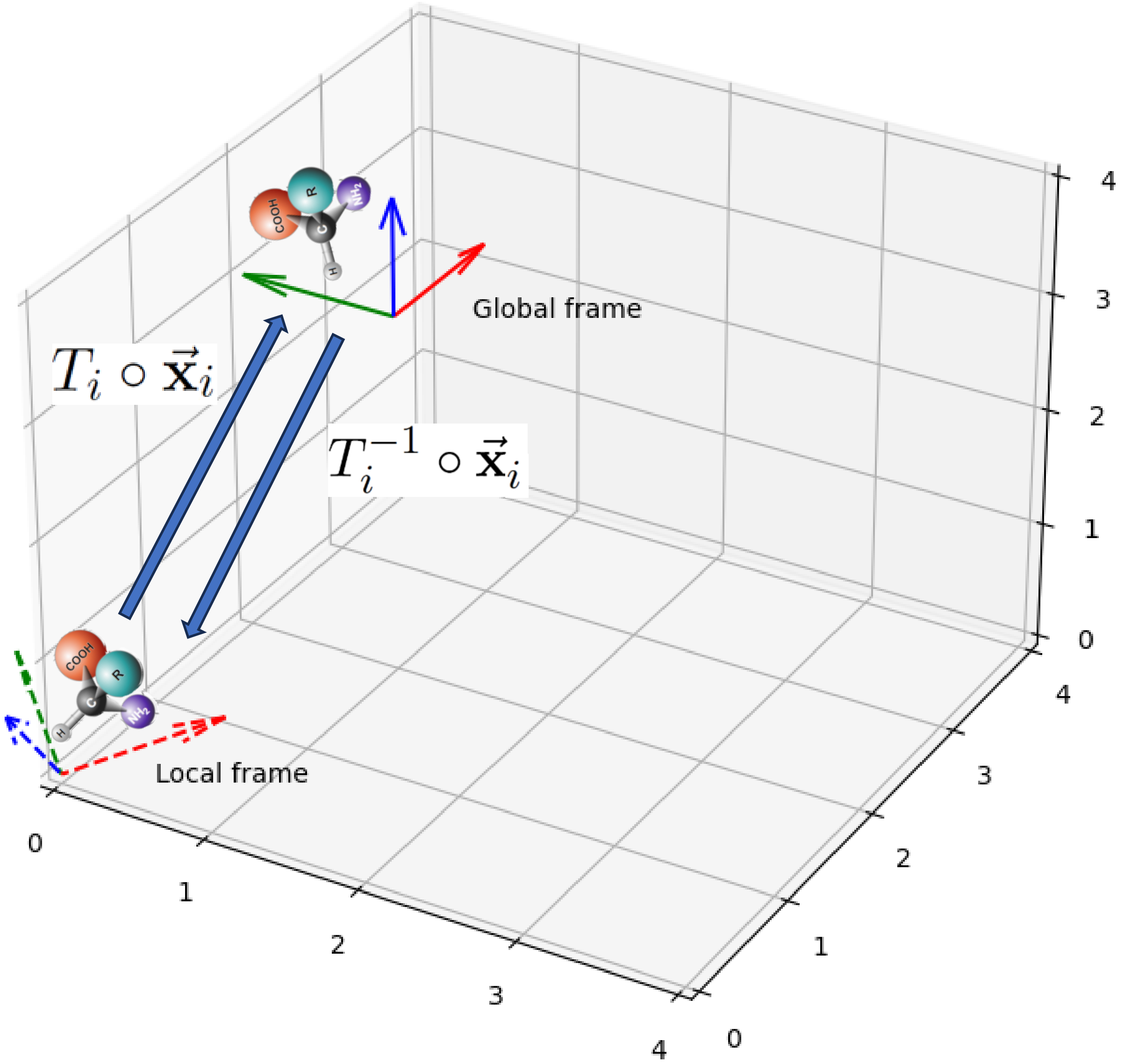
$$2:\ \vec{\mathbf{x}}_{ij}^{\text{true}} = T_i^{\text{true}-1} \circ \vec{\mathbf{x}}_j^{\text{true}}$$

$$3:\ d_{ij} = \sqrt{\|\vec{\mathbf{x}}_{ij} - \vec{\mathbf{x}}_{ij}^{\text{true}}\|^2 + \epsilon}$$

$$4:\ \mathcal{L}_{\text{FAPE}} = \frac{1}{Z}\,\text{mean}_{i,j}(\text{minimum}(d_{\text{clamp}}, d_{ij}))$$



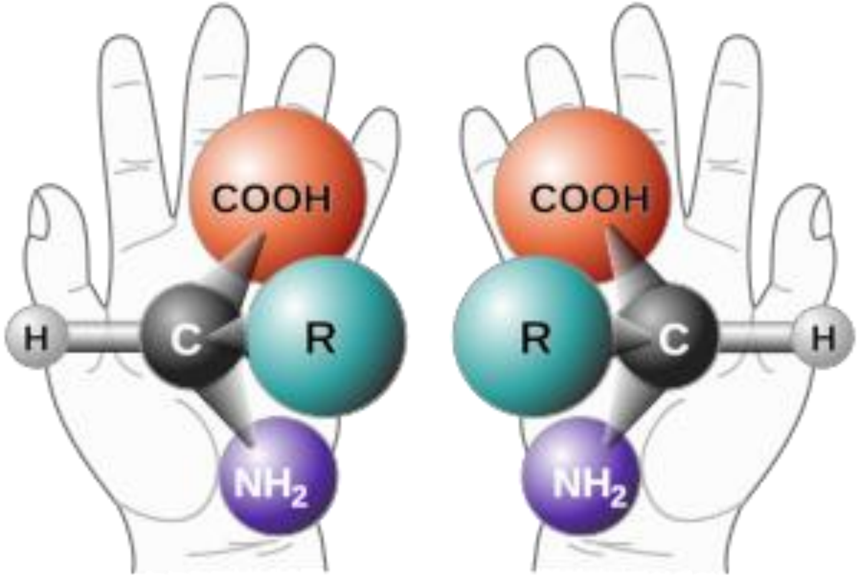
Chirality



AlphaFold2: Frame aligned point error (FAPE)

$$T_i^{-1} \circ \vec{\mathbf{x}}_j \overset{\text{reflection}}{\mapsto} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \left(T_i^{-1} \circ \vec{\mathbf{x}}_j\right)$$

$$\begin{aligned} \text{FAPE}(X, X_{\text{reflection}}) &= \sum_{ij} \left\| T_i^{-1} \circ \vec{\mathbf{x}}_j - \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \left(T_i^{-1} \circ \vec{\mathbf{x}}_j\right) \right\| \\ &= 2 \sum_{ij} \left| \left(T_i^{-1} \circ \vec{\mathbf{x}}_j\right)_z \right| \end{aligned}$$



Chirality

~~RMSD-like loss~~

Part 3: SciML for WFBR topics

(Potential) SciML applications in WFBR domains

- Food/biomass processing models (Neural Networks + e.g., Kinetic/Thermodynamic models)

- Molecule Design (Neural Networks + Molecular Dynamic Simulation)

- Rheology modelling for 3D printing (Neural Networks + Viscoelastic Models)

- Kinetic quality decay (Neural Networks + Kinetic models)

- Cold storage volatile mapping (Neural Networks + Computational Fluid Dynamics Models)

- Food supply chain parameters (FLW) prediction (Neural Networks + knowledge-based rules)

Take away

SciML can help WUR researchers to employ AI to improve the efficiency and quality of their research in food-related domains (and beyond), even when data availability is relatively limited.