

Digital Twins for Biomedical Research & Health Sciences

Overview & Synthesis

Contents

- What is a Digital Twin? Definitions, dreams, reality.
- Scientific machine learning for Digital Twins.
- Digital Twins in biomedical research and health sciences.
- LLMs today and tomorrow.
- Ethics and moral issues.
- Where do we stand, where should we go, and how do we get there?

Digital Twin

A good definition

- “A digital twin couples **computational** models with a **physical** counterpart to create a system that is dynamically updated through bidirectional **data** flows as conditions change.” [NASEM]
- “A digital twin is a **virtual** representation of an object or system designed to accurately reflect a **physical** object.” [IBM]
- “A digital twin is an integrated **data-driven** virtual representation of real-world entities and processes, with synchronized **interaction** at a specified frequency and fidelity.” [DT consortium]

Digital Twin

A good definition

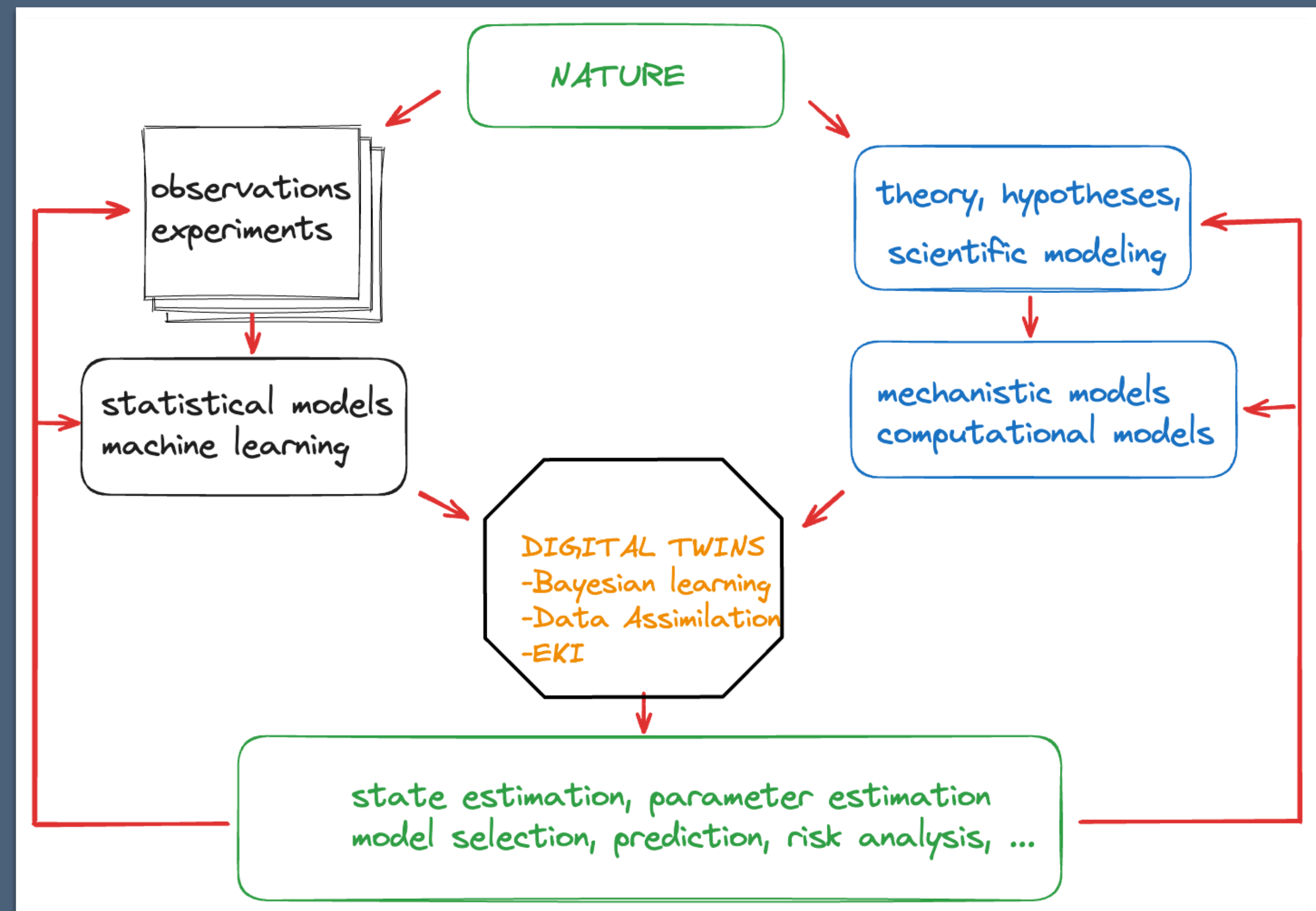
- “A digital twin couples **computational** models with a **physical** counterpart to create a system that is dynamically updated through bidirectional **data** flows as conditions change.” [NASEM]
- “A digital twin is a **virtual** representation of an object or system designed to accurately reflect a **physical** object.” [IBM]
- “A digital twin is an integrated **data-driven** virtual representation of real-world entities and processes, with synchronized **interaction** at a specified frequency and fidelity.” [DT consortium]



Digital Twin

A good definition

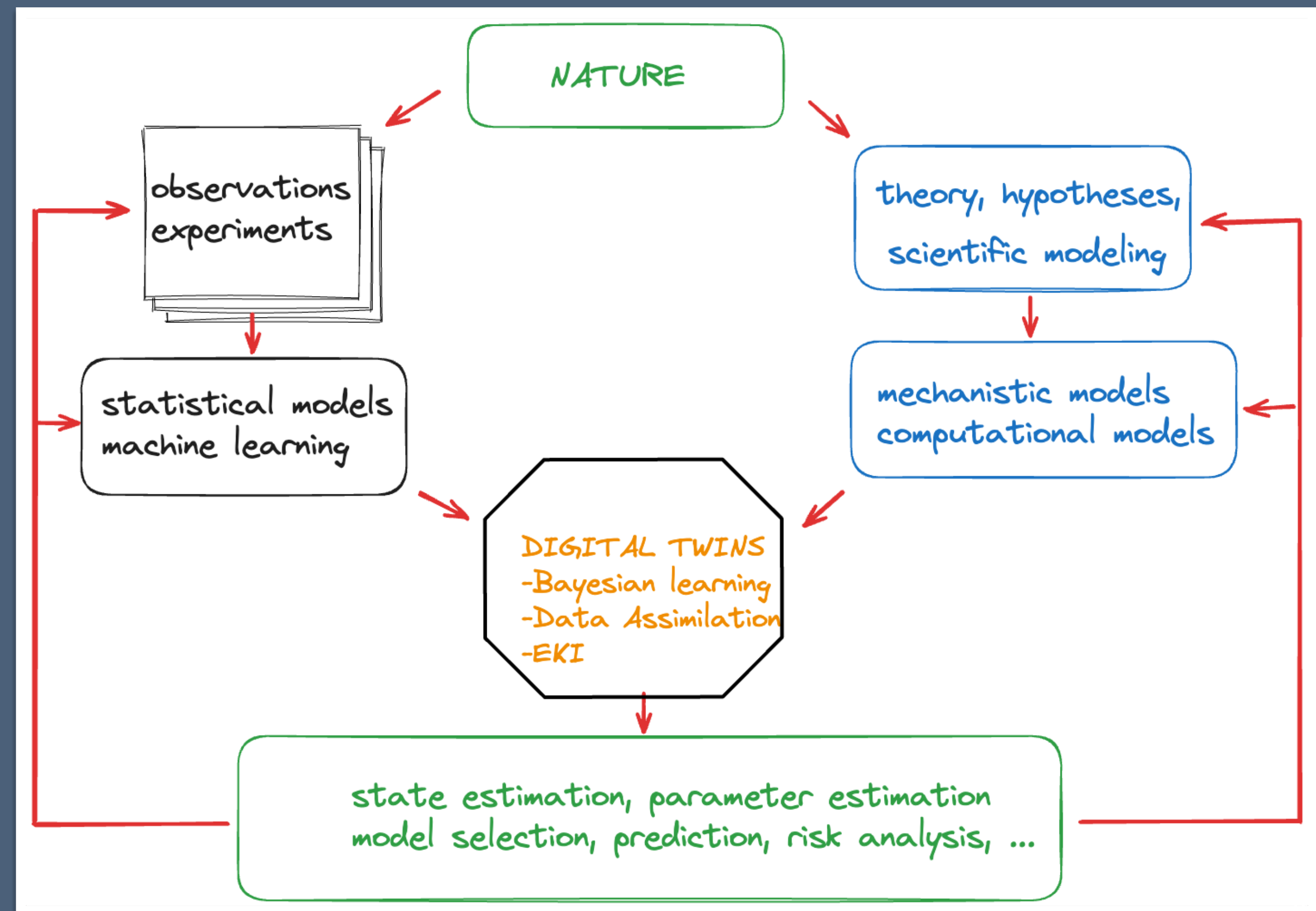
- “A digital twin couples **computational** models with a **physical** counterpart to create a system that is dynamically updated through bidirectional **data** flows as conditions change.” [NASEM]
- “A digital twin is a **virtual** representation of an object or system designed to accurately reflect a **physical** object.” [IBM]
- “A digital twin is an integrated **data-driven** virtual representation of real-world entities and processes, with synchronized **interaction** at a specified frequency and fidelity.” [DT consortium]



Digital Twin

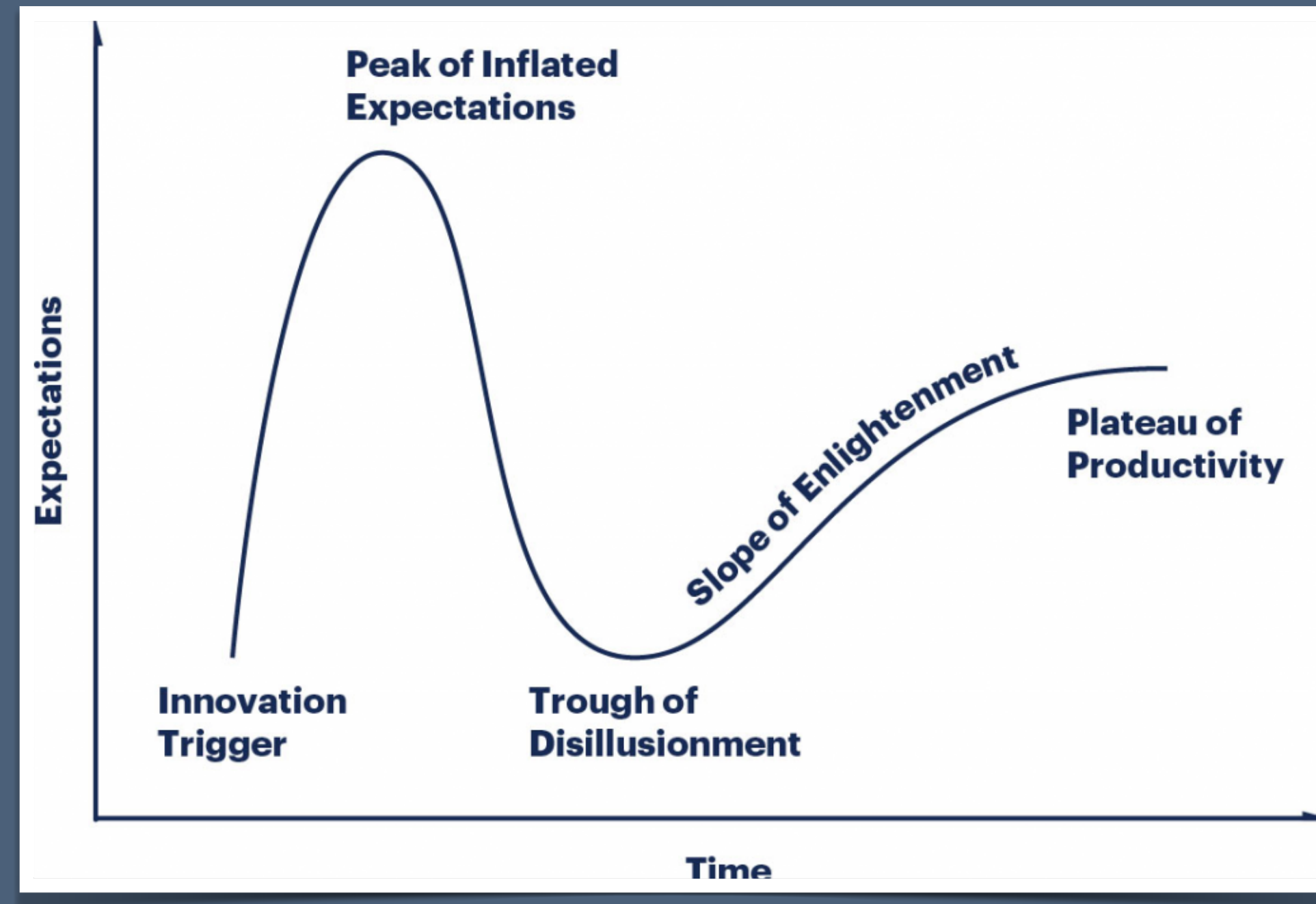
A good definition

- “A digital twin couples **computational** models with a **physical** counterpart to create a system that is dynamically updated through bidirectional **data** flows as conditions change.” [NASEM]
- “A digital twin is a **virtual** representation of an object or system designed to accurately reflect a **physical** object.” [IBM]
- “A digital twin is an integrated **data-driven** virtual representation of real-world entities and processes, with synchronized **interaction** at a specified frequency and fidelity.” [DT consortium]



Myths

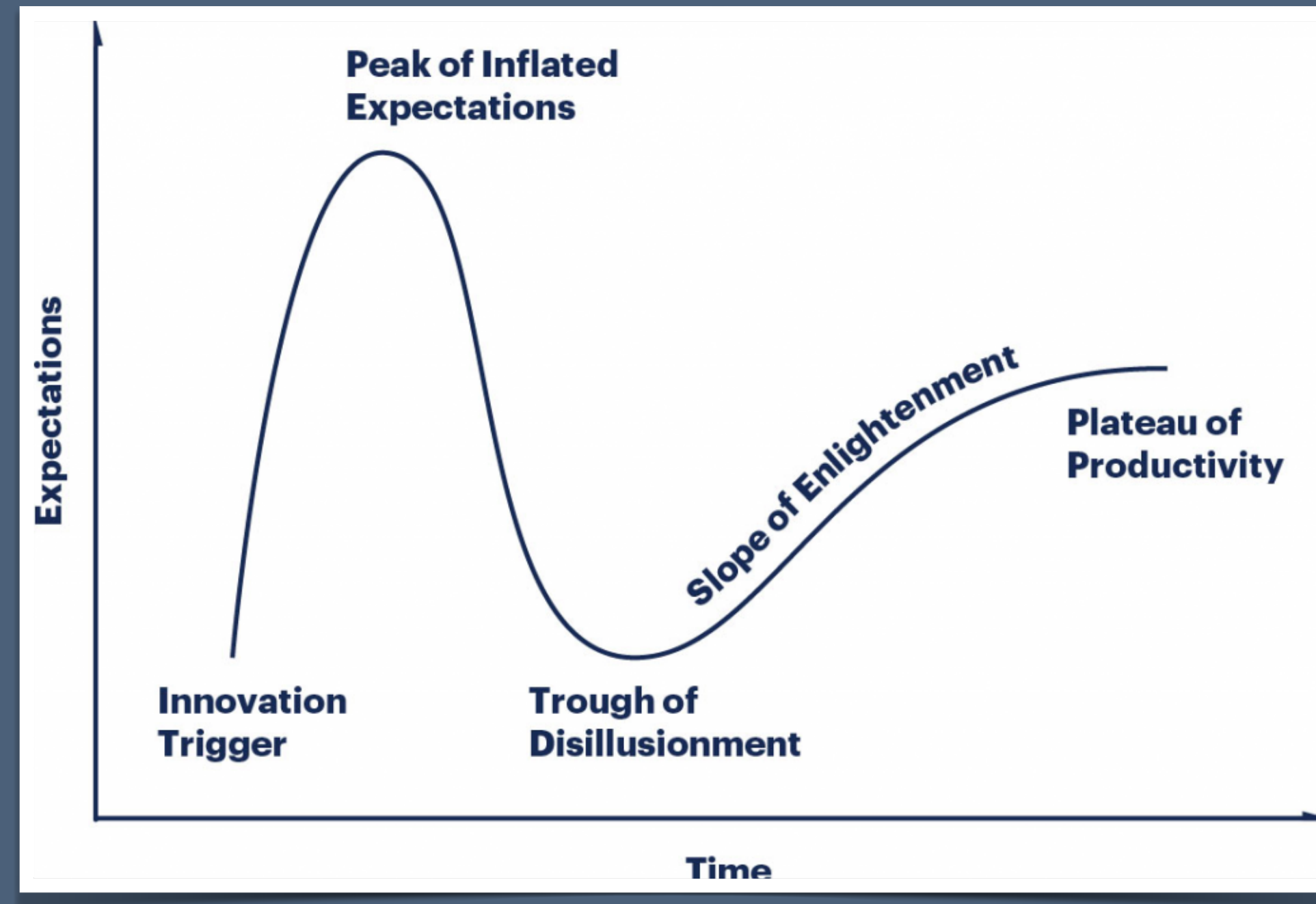
Hype ...



Myths

Hype ...

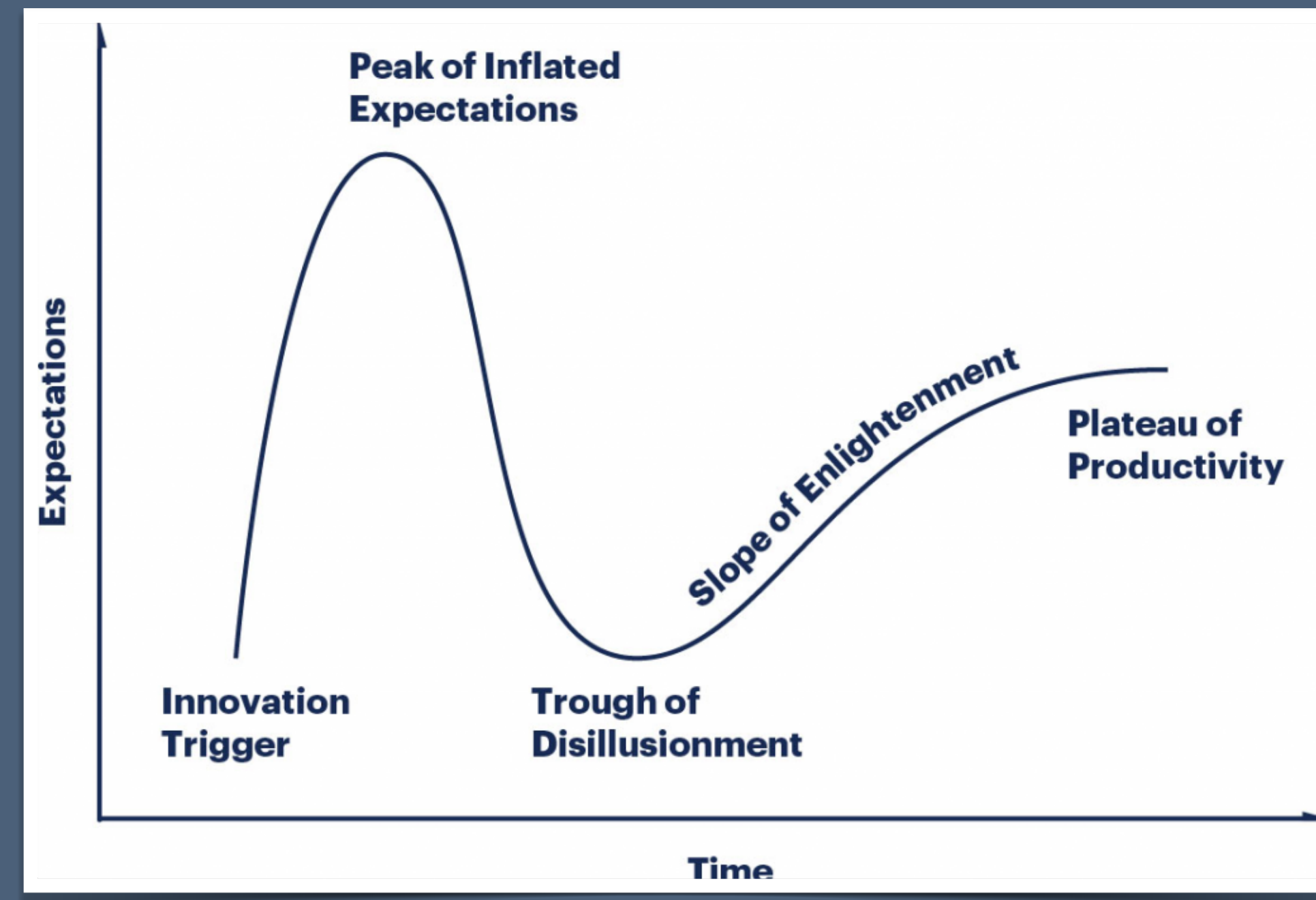
- You can “twin” anything!



Myths

Hype ...

- You can “twin” **anything**!
- But:
 - **Not** transferable.
 - **Not** certifiable.
 - **High** costs.
 - **High** complexity.
 - Deployment is **hard**.



Dreams

- Personalized medicine

REAL WORLD PATIENT

The patient and the tumor from which data is gathered using various clinical assessments to inform the digital twin.

VVUQ

Verification, validation, and uncertainty quantification

As the patient and tumor are constantly evolving and the data collection can also change over time, VVUQ must occur continually for digital twins.

Uncertainty quantification needs to be addressed for all aspects of the digital twin, including the patient's data, modeling and simulation, and decision making.

DIGITAL TWIN

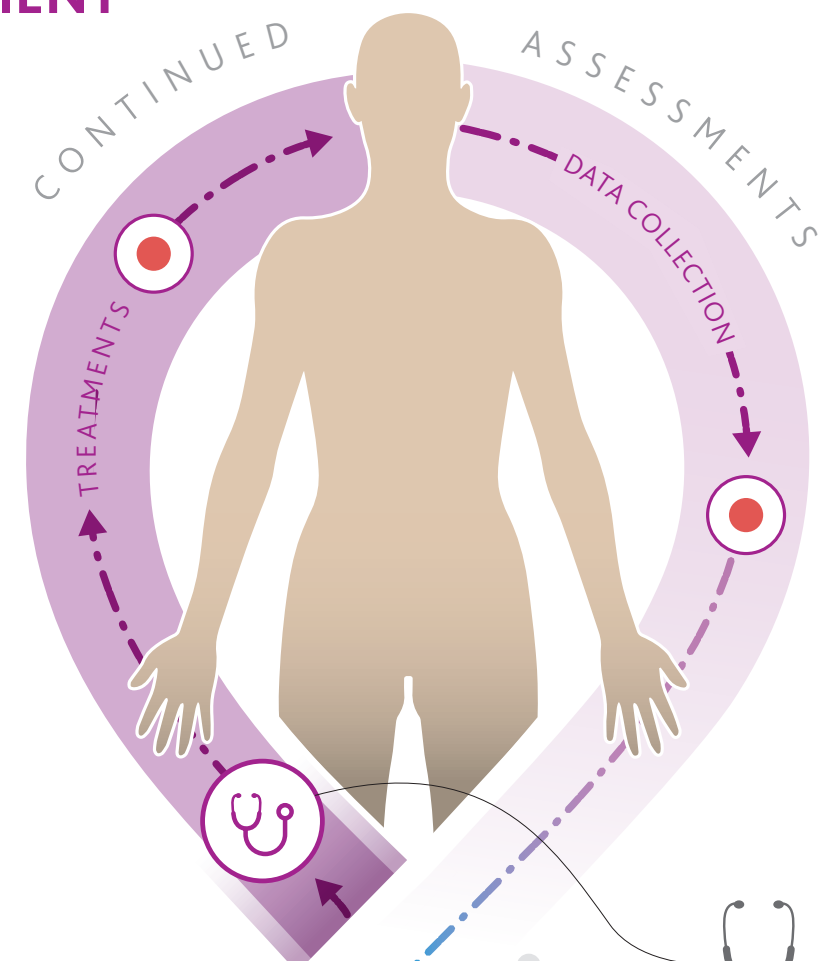
The virtual representation comprised of models describing temporal and spatial characteristics of the patient and tumor with dynamic updates using data from the real world patient.



Modeling

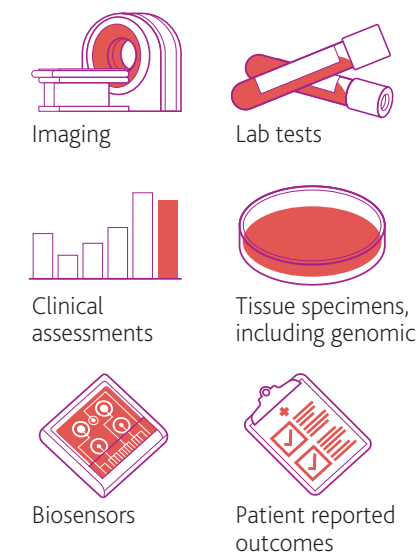
Models spanning a range of fidelities and resolutions may be utilized and potentially integrated together.

As new observed data are acquired, the data are assimilated and the models are calibrated, updated, and estimated.



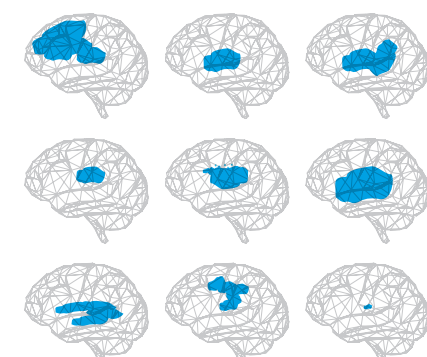
Clinical assessments

Data are collected in many ways:



Human and digital twin interaction

Utilizing the simulated predictions and related uncertainties, the clinician and patient can make informed clinical-decisions around treatment and also the clinical assessments, which affect the data informing the digital twin.



Simulations & predictions

Simulations of potential treatments can generate predictions of outcome and in turn can be optimized to determine the most favorable treatment options.

- Smart Cities



- Destination Earth



Reality

Present and future

Reality

Present and future

- Mathematics of Inverse Problems and DA - see next lecture.

Reality

Present and future

- Mathematics of Inverse Problems and DA - see next lecture.
- Scientific Machine Learning - blending science with data.

Reality

Present and future

- Mathematics of Inverse Problems and DA - see next lecture.
- Scientific Machine Learning - blending science with data.
- Generative AI... - the real game-changer?

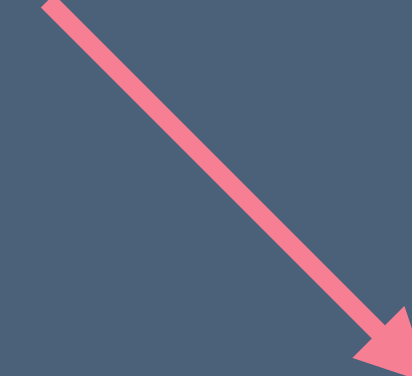
ML and SciML

ML/AI in a Nutshell

Finding a pattern.

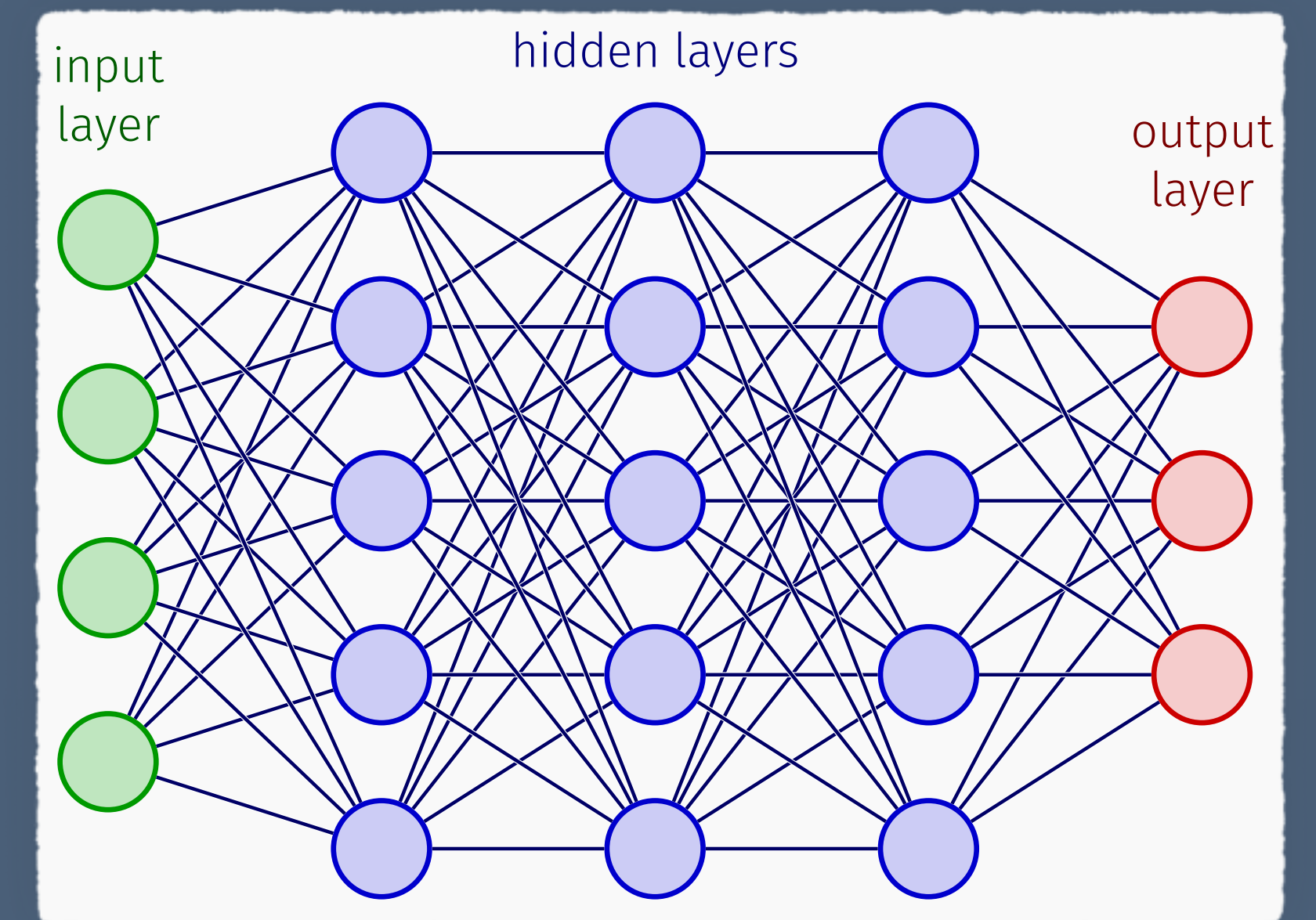
$$f : x \rightarrow y$$

- 
- Input
 - Features
 - Parameters
 - Causes

- 
- Output
 - Observations
 - Outcomes
 - Effects

ML/AI in a Nutshell

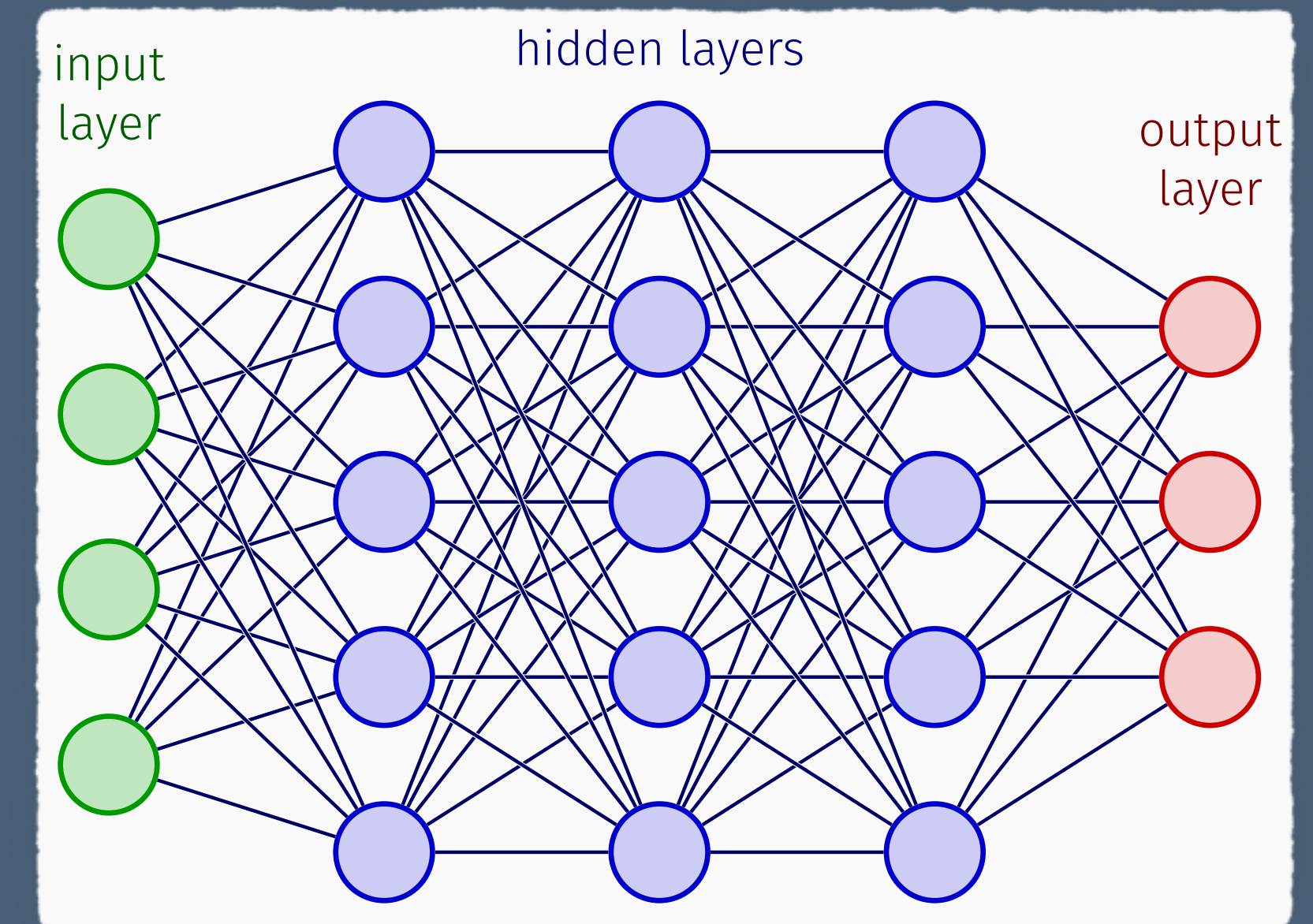
How is this done?



ML/AI in a Nutshell

How is this done?

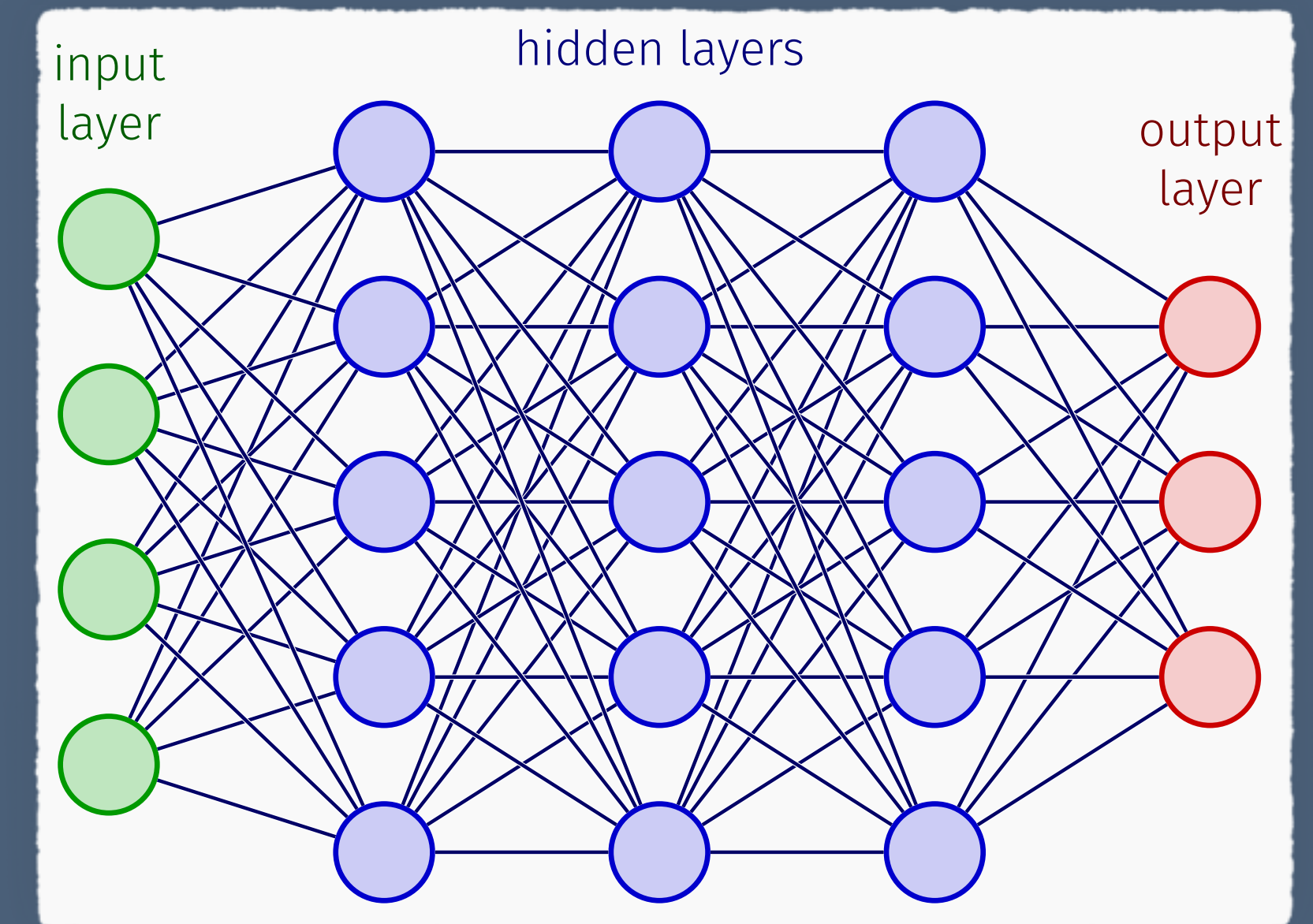
1. Suppose the ML/AI model depends on **parameters**, θ .



ML/AI in a Nutshell

How is this done?

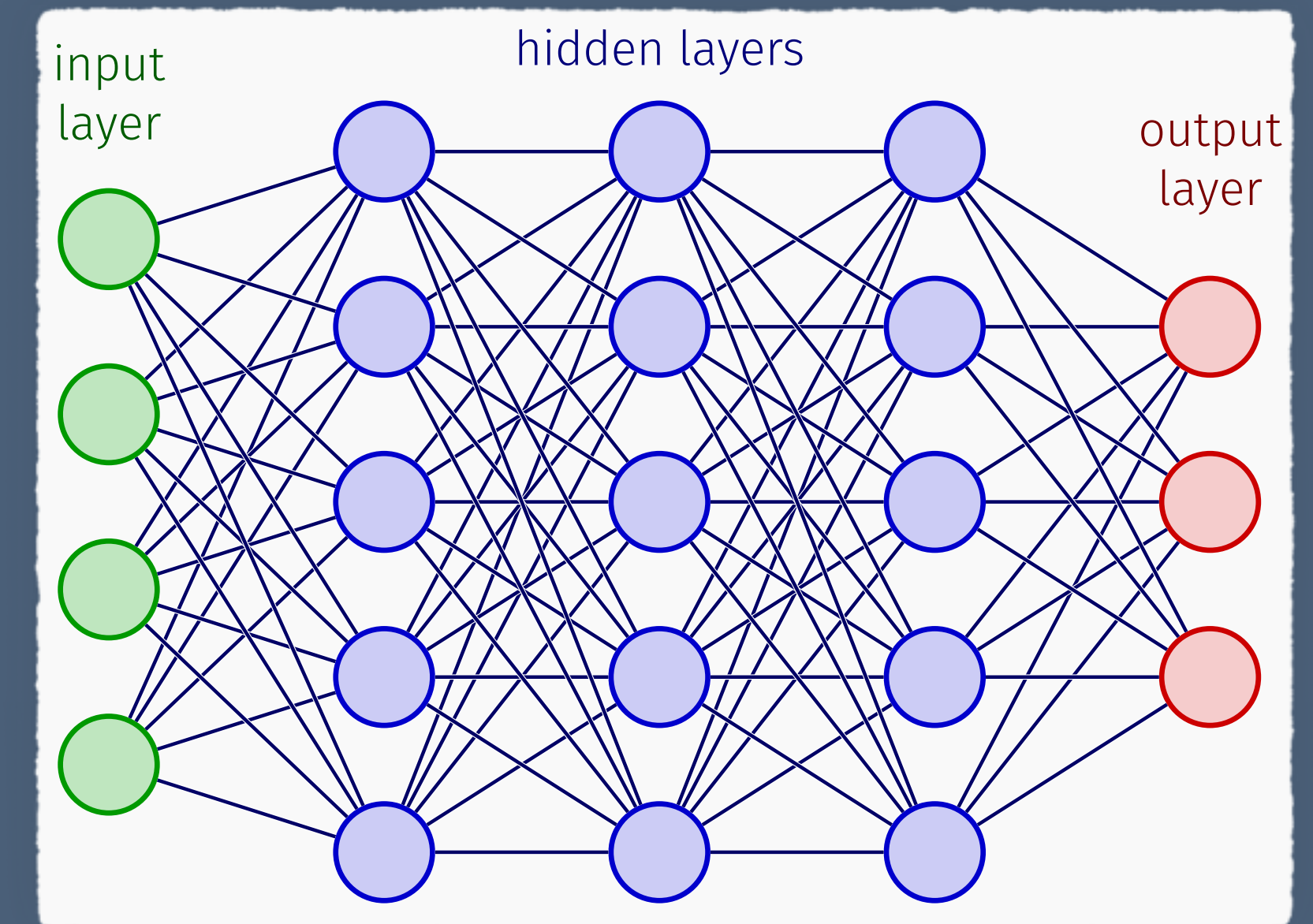
1. Suppose the ML/AI model depends on **parameters**, θ .
2. Define a **loss function** $\mathcal{L}$ that measures the difference between model predictions and measured observations.



ML/AI in a Nutshell

How is this done?

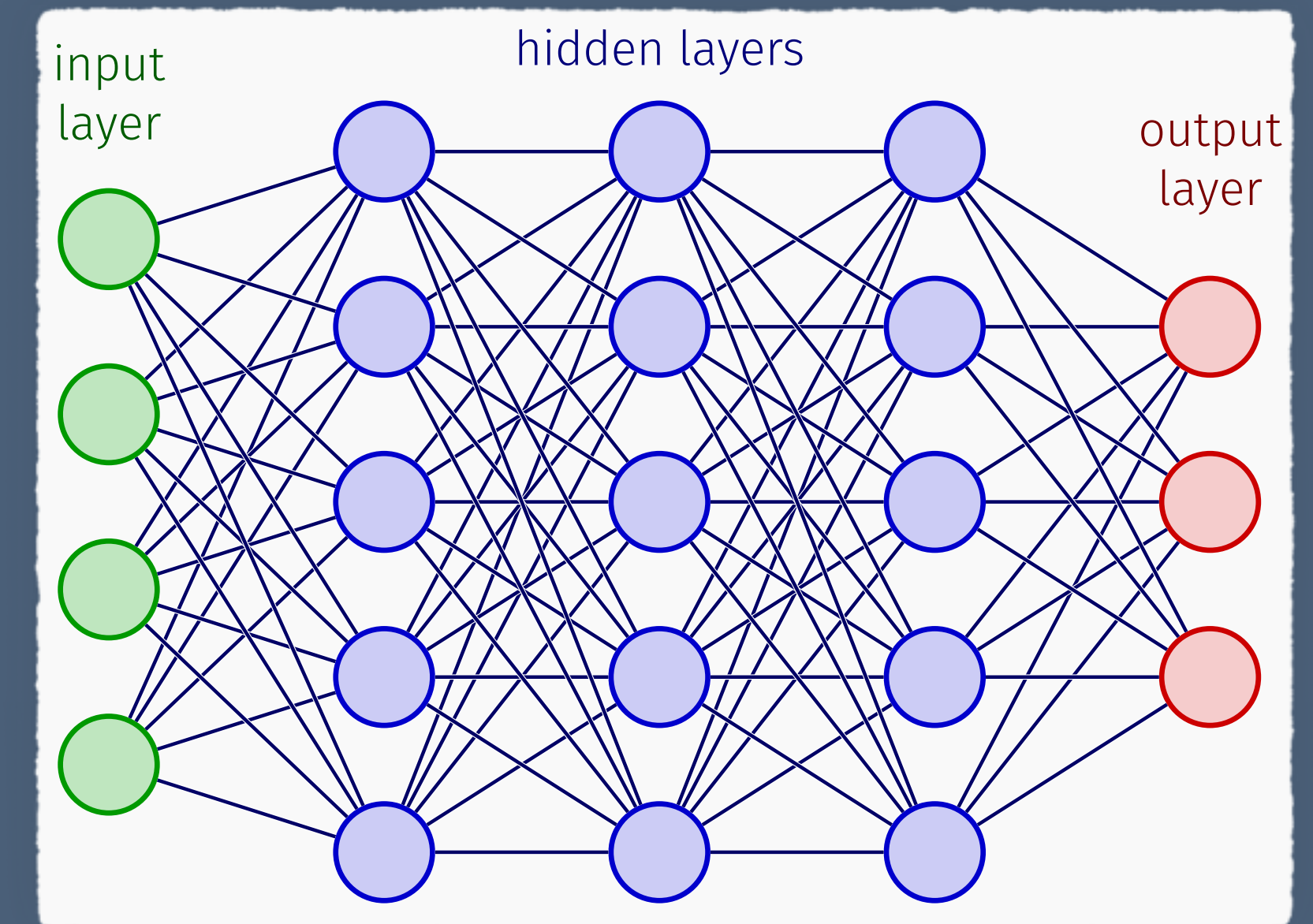
1. Suppose the ML/AI model depends on **parameters**, θ .
2. Define a **loss function** $\mathcal{L}$ that measures the difference between model predictions and measured observations.
3. Find the optimal parameters that **minimize** the loss function.



ML/AI in a Nutshell

How is this done?

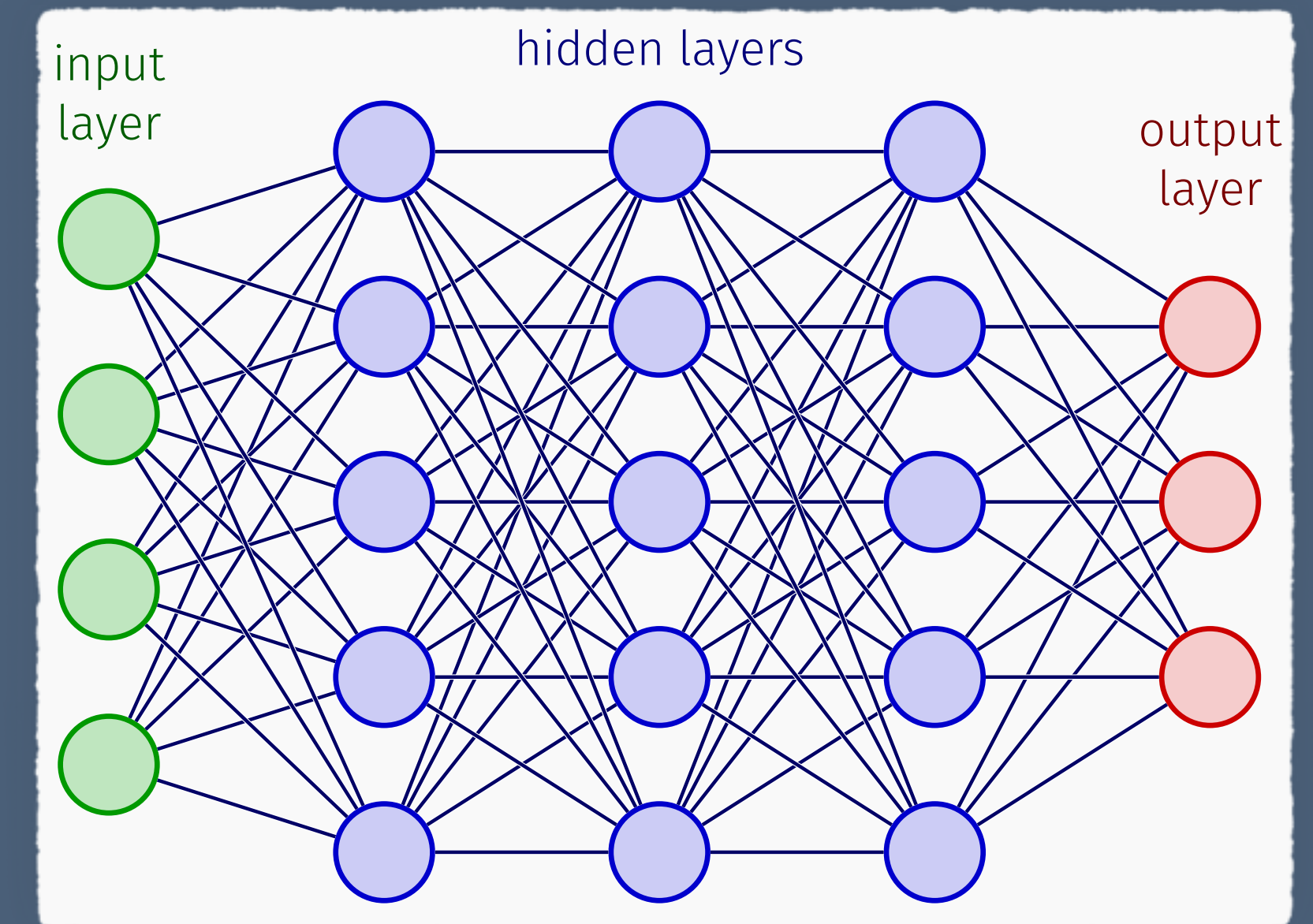
1. Suppose the ML/AI model depends on **parameters**, θ .
2. Define a **loss function** $\mathcal{L}$ that measures the difference between model predictions and measured observations.
3. Find the optimal parameters that **minimize** the loss function.
 - *This is exactly how we solve an **inverse problem**.*



ML/AI in a Nutshell

How is this done?

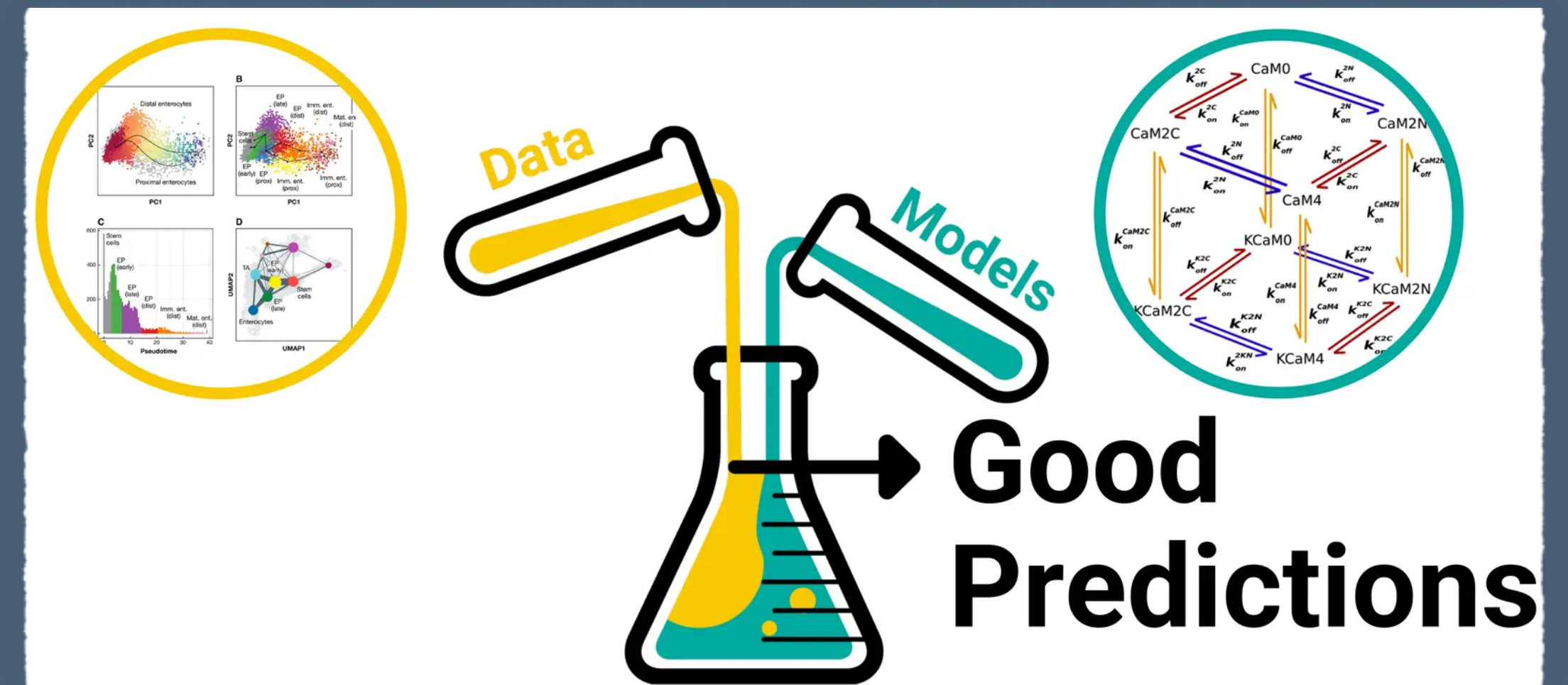
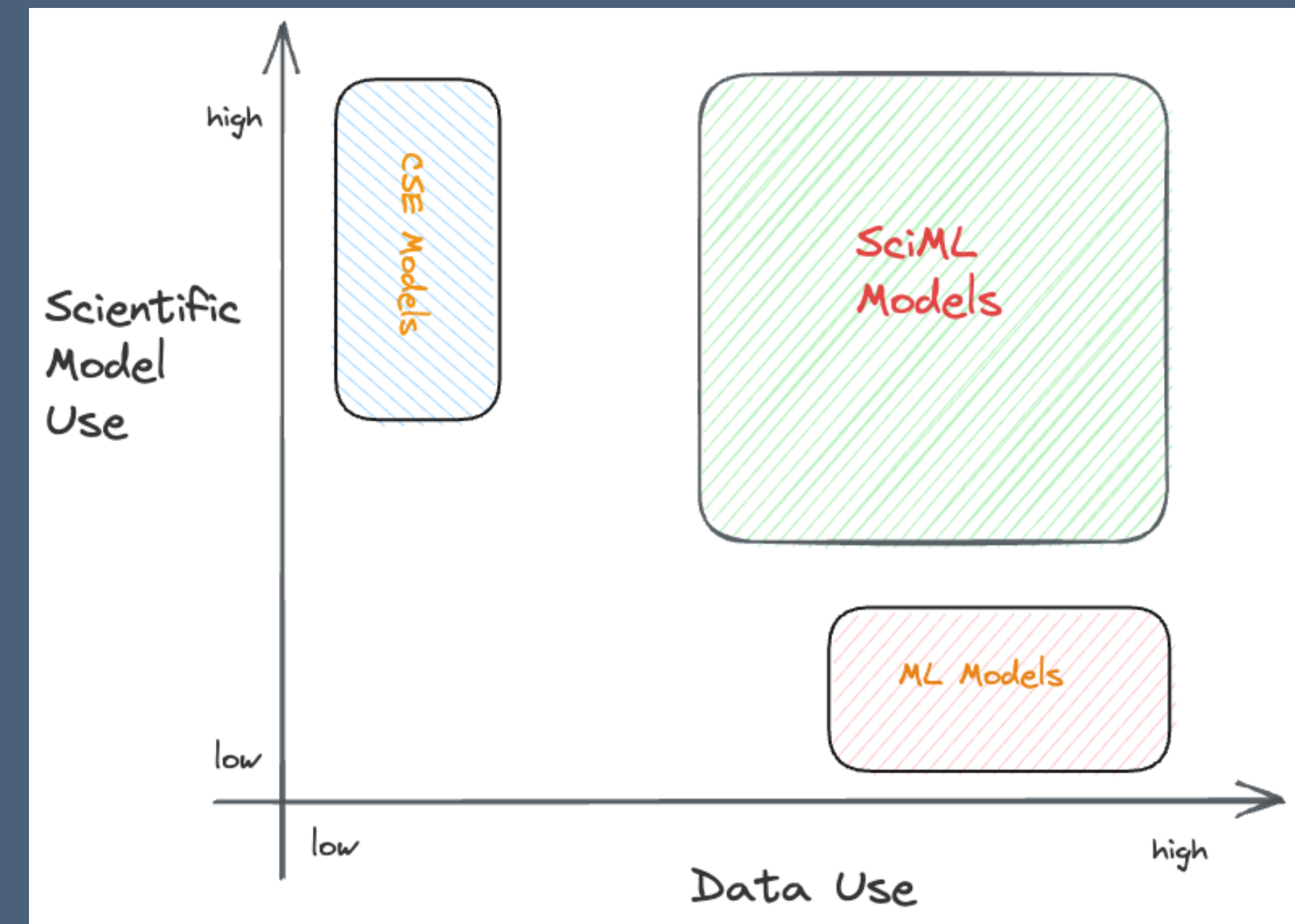
1. Suppose the ML/AI model depends on **parameters**, θ .
2. Define a **loss function** $\mathcal{L}$ that measures the difference between model predictions and measured observations.
3. Find the optimal parameters that **minimize** the loss function.
 - *This is exactly how we solve an **inverse problem**.*
 - *Inverse problems are at the core of **Digital Twins**.*



Scientific Machine Learning

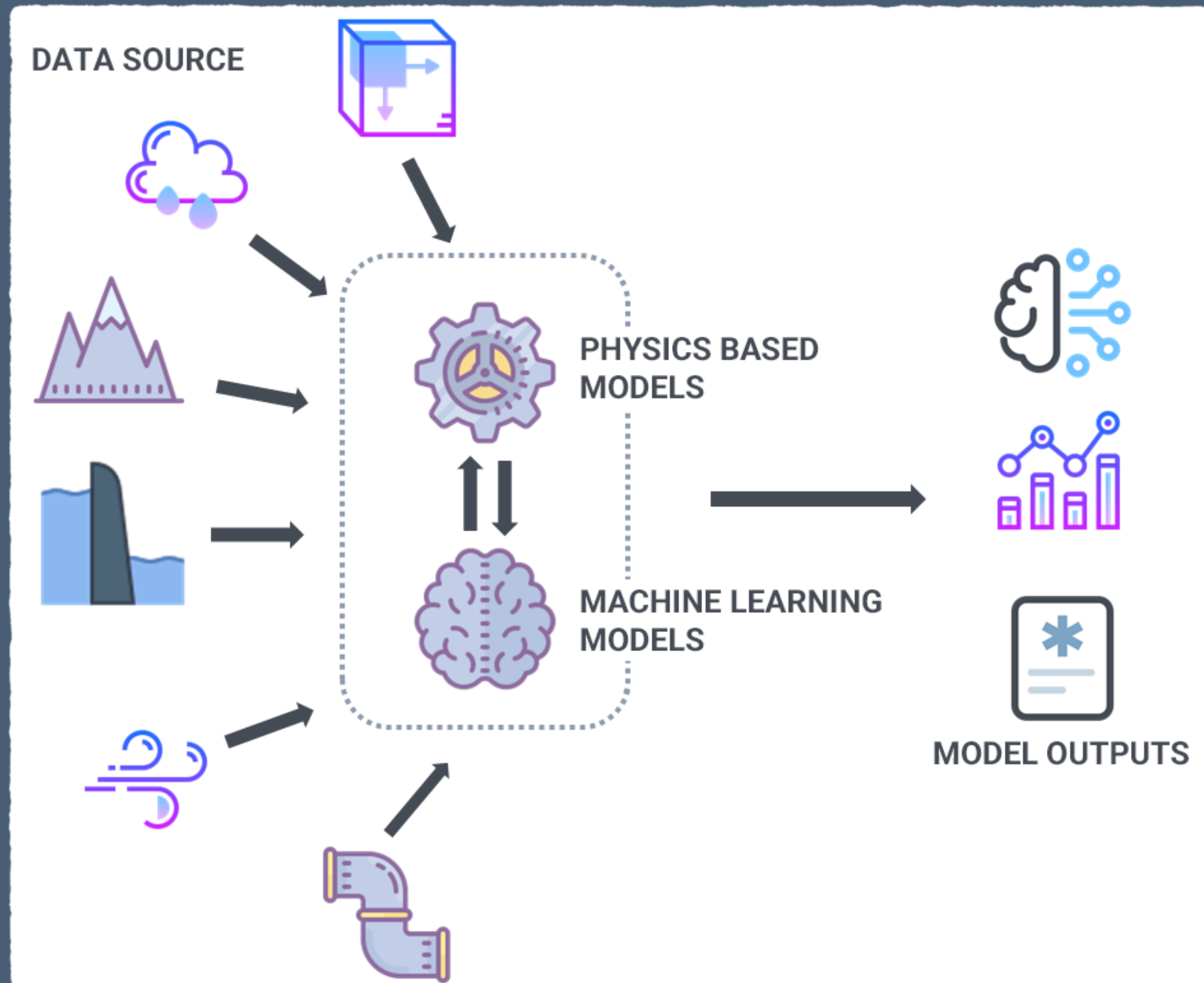
Blending data and science

- ML/AI is 100% **data-driven**.
- ML/AI produces black boxes.
- There is NO **guarantee** that the results are scientifically valid.
- SciML seeks to **blend** science with ML/AI.
- There are many, many **techniques** for this... Research is ongoing.



Scientific Machine Learning

Blending physics-based and ML models

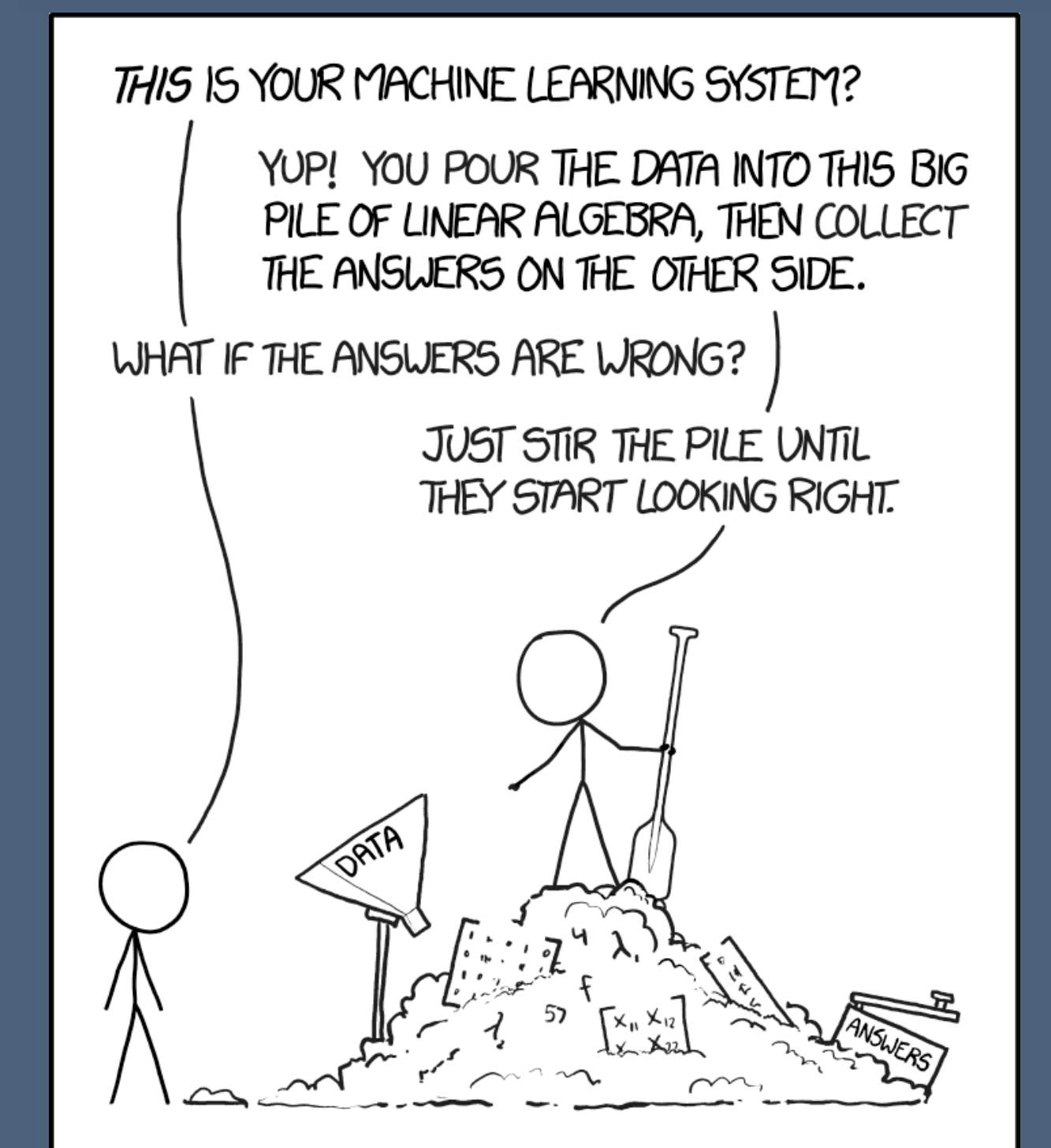


- PINN and variants, based on [autograd](#).
- Operator learning.
- Feature engineering.
- Layer engineering.
- Bayesian neural networks.
- Gaussian process regression.
- Generative models (LLMs).

Why does AI/ML need science?

Trust

- Data-driven black-box methods do fail:
 - 1. Output is physically inconsistent.
 - 2. ML methods/algorithms can (and do) easily learn spurious (local) relationships that look very good on training and test data.
 - 3. ML cannot distinguish between correlation (what?) and causation (why?).
- This leads to:
 - 1. Poor generalization.
 - 2. Lack of Interpretability.
 - 3. Lack of trustworthiness and certifiability.

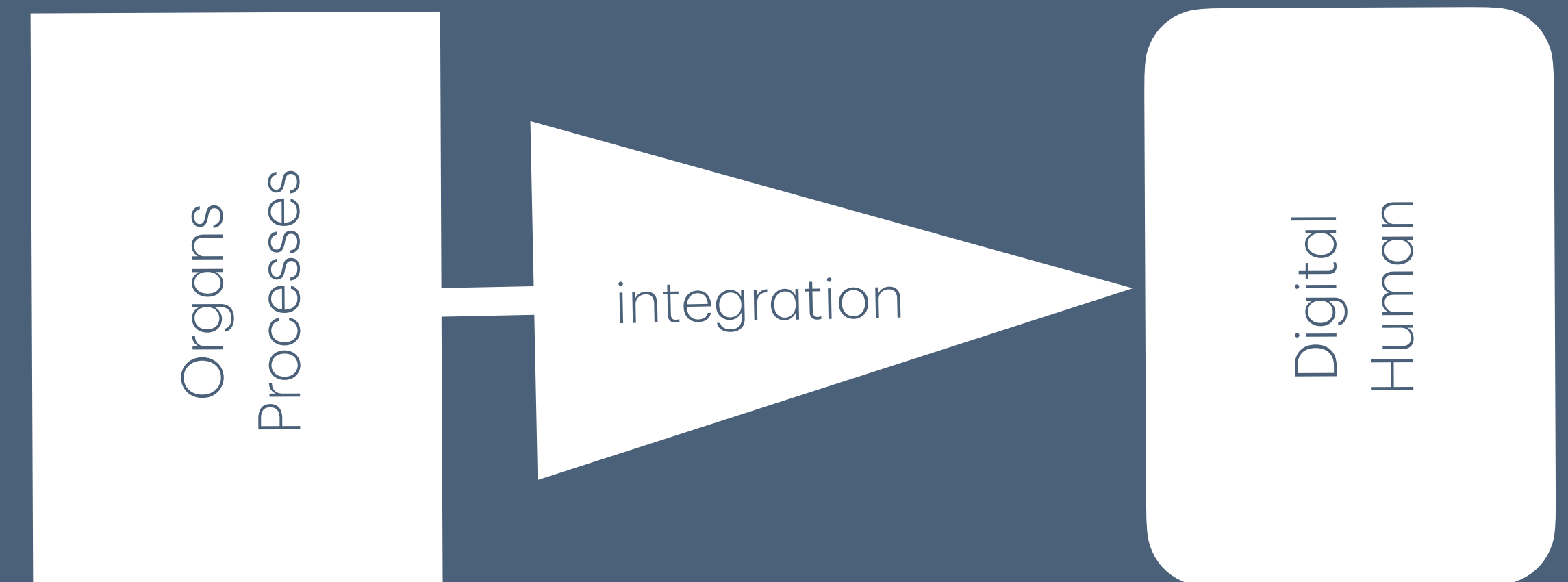
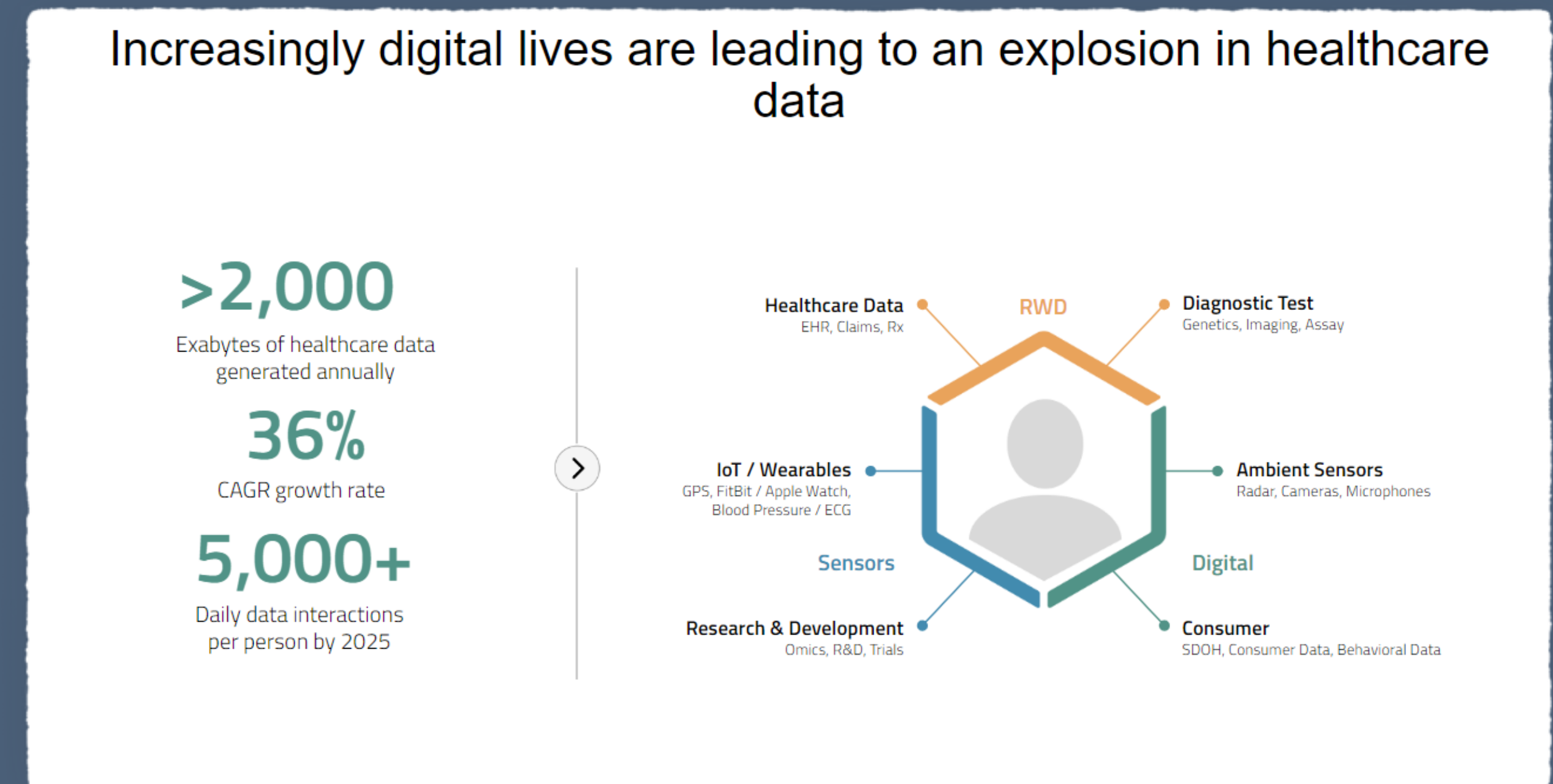


Digital Twins

Digital Twins in biomedical research and health sciences

Where do we stand?

- Lots of **data**.
- Credible process-level and organ-level **twins** exist already.
- Need to integrate these into a **system-of-systems**.
- **Digital Human** will then become possible, thanks to Exascale HPC—no longer sci-fi



Examples

Many, many, many...

- Omics: proteomics, metabolomics, transcriptomics, multi-omics.
- Genomics.
- Imaging, radiomics.
- Hospital management.
- Surgery scenario modelling.

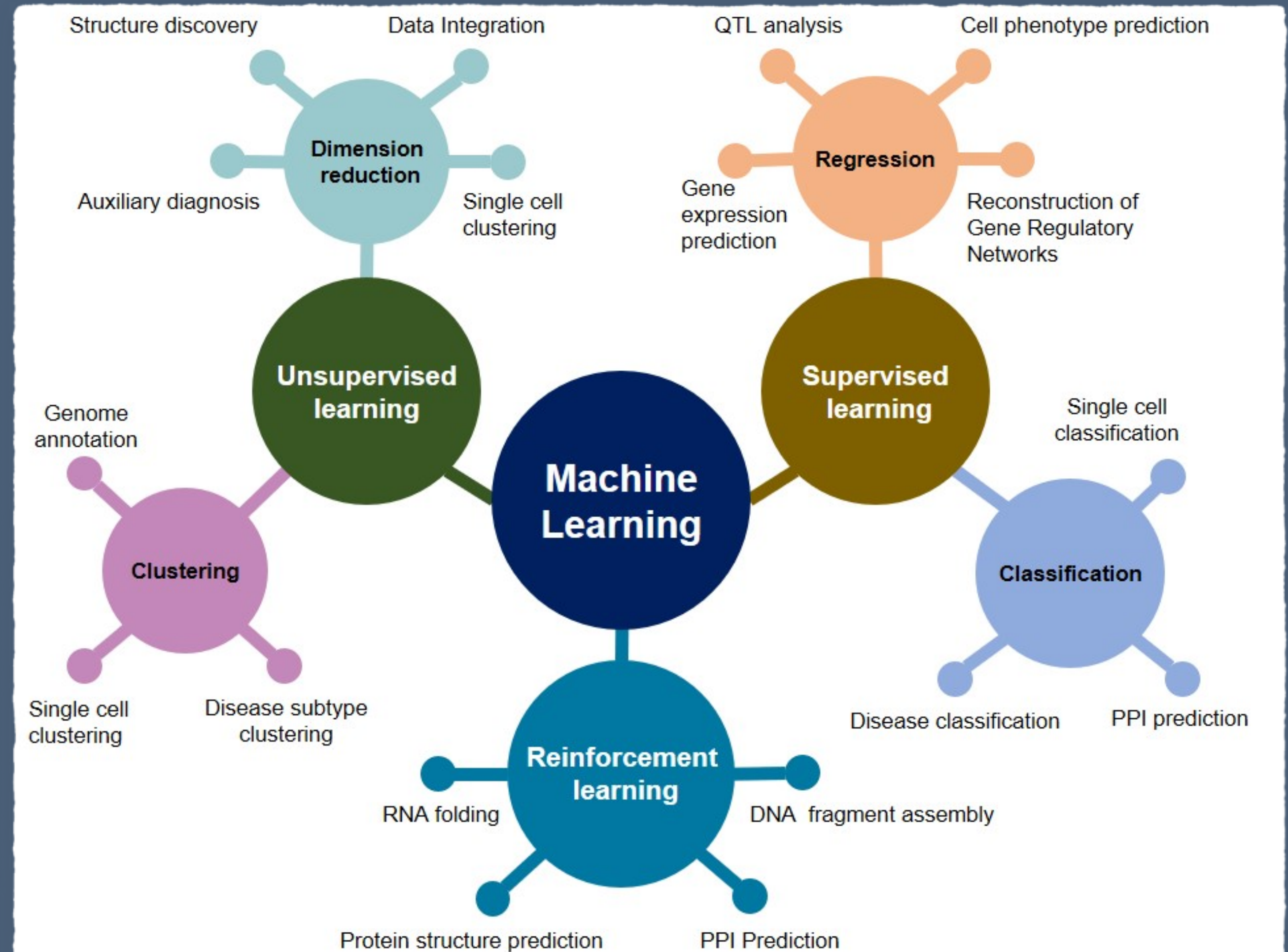


Figure S1. The hierarchy of machine learning approaches and their applications in omics studies.

LLMs for Health Sciences

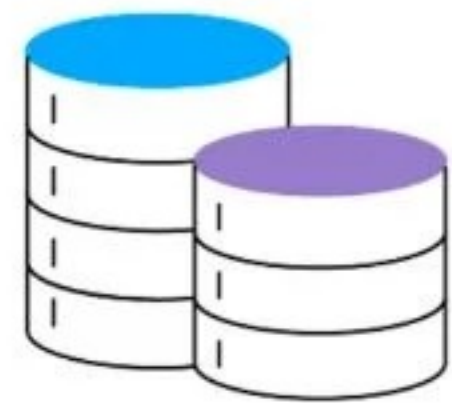
The next frontier!

- What is an LLM?
- How can LLMs become domain-specific?
- Examples in biomedical and health sciences.

Large Language Model

Components

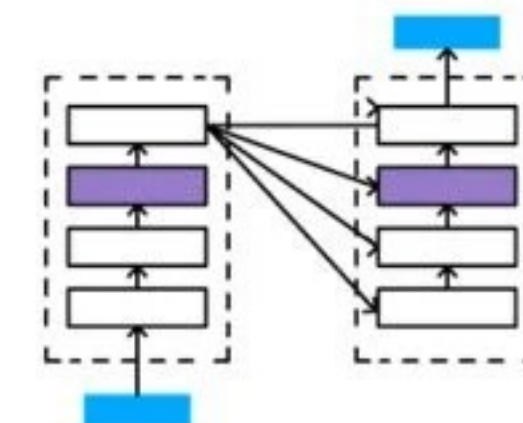
Large Language Models (LLM)



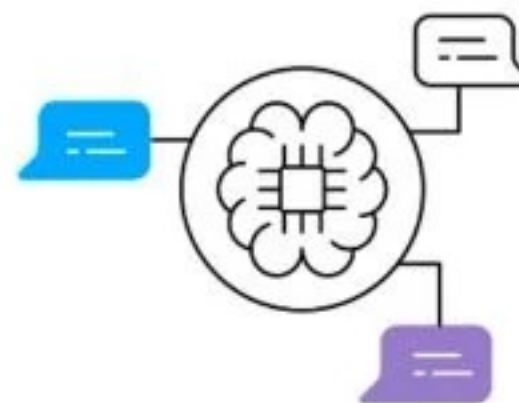
Massive Dataset



Deep Learning



Transformer Architecture



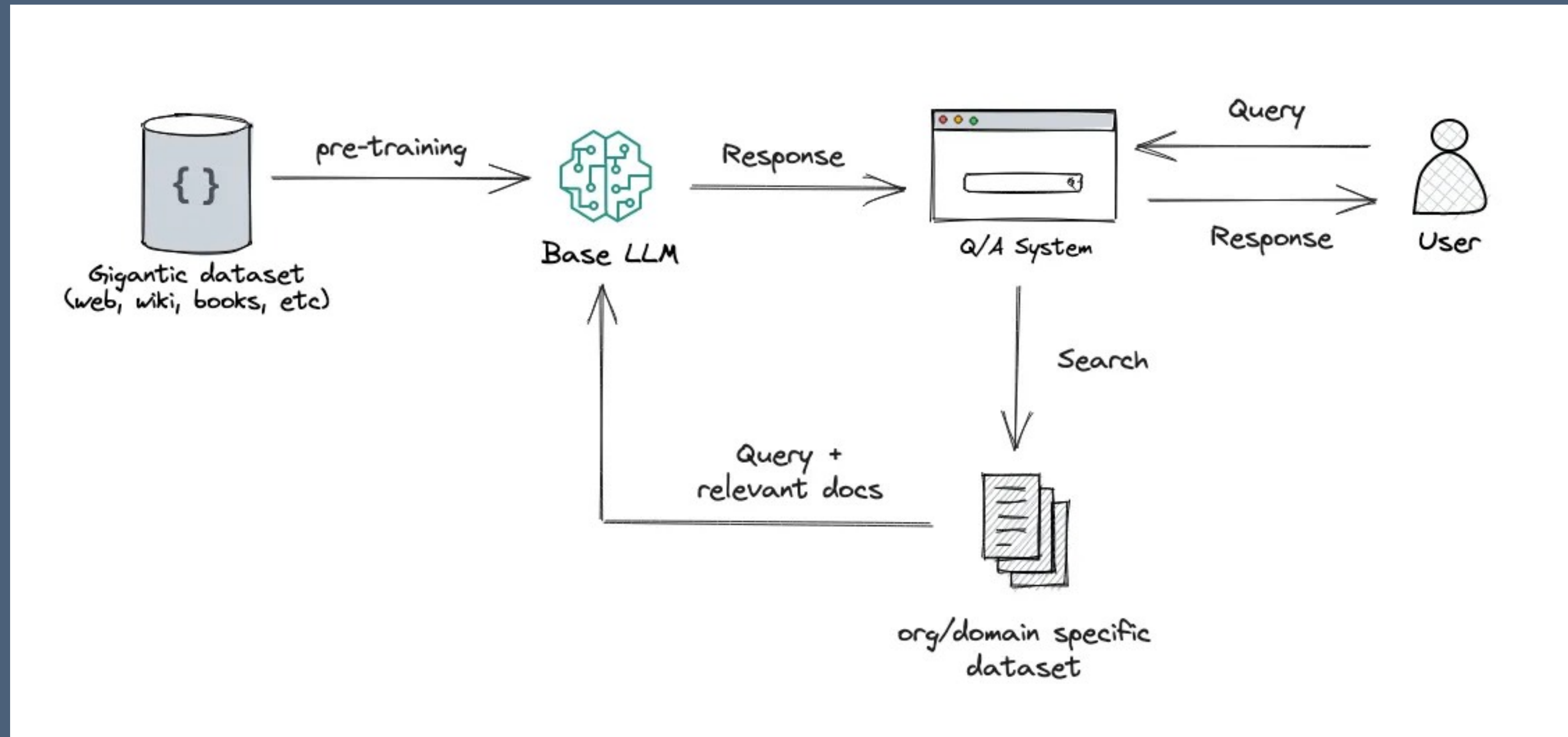
Self-supervised Learning



Fine-tuning

Large Language Model

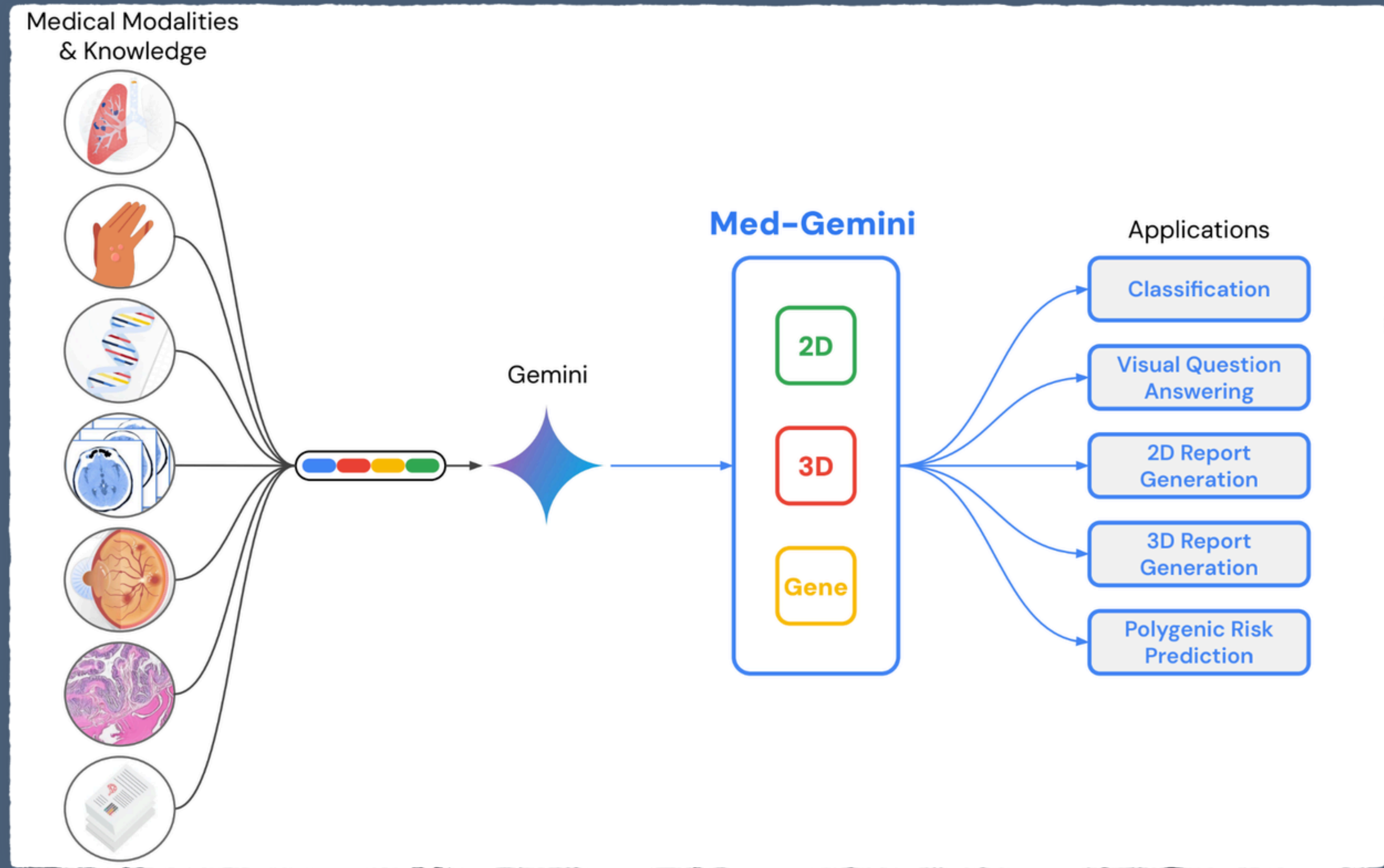
Specialization—RAG



LLM for Health

Med-Gemini-Google

- Family of multimodal models built upon Gemini specifically designed for the healthcare industry.
- “Groundbreaking Family of AI Models Revolutionizing Medical Diagnosis and Clinical Reasoning”
- Models excel in multimodal tasks, with substantial improvements in analyzing medical images and videos and accurately retrieving information from long health records.



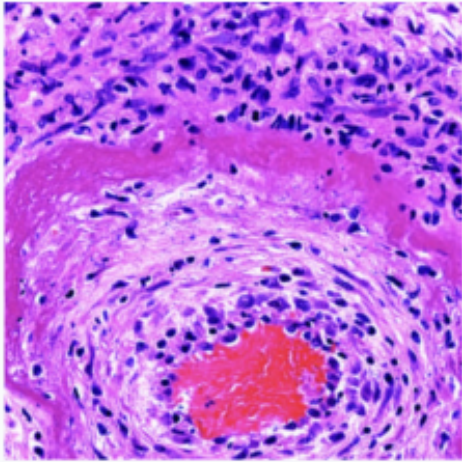
LLM for Health

Med-Gemini-Google

- Family of multimodal models built upon Gemini specifically designed for the healthcare industry.
- “Groundbreaking Family of AI Models Revolutionizing Medical Diagnosis and Clinical Reasoning”
- Models excel in multimodal tasks, with substantial improvements in analyzing medical images and videos and accurately retrieving information from long health records.

Open-ended Visual QA (Path-VQA)

Visual input



Instruction
You are a helpful medical assistant. The following are questions about medical knowledge. Solve them in a step-by-step fashion, referring to authoritative sources as needed.
Question: What does the wall of the artery show with protein deposition and inflammation?

Response
a circumferential bright pink area of necrosis

Image Classification (PAD-UFES-20 6-condition classification)

Visual input



Instruction
You are a helpful dermatology assistant. The following are questions about skin lesions. Categorize the skin lesions into the most likely class given the patient history. Output a single option letter from the provided options as the final answer.
Patient History: Age: 51, Gender: female, Smoke: false, Drink: false, Family skin cancer history: true, Family any cancer history: false, Lesion region: back, Lesion itch: false, Lesion grew: false, Lesion bled: false, Lesion elevation: false, Fitzpatrick scale: 1.0, Diameters (mm): [12.0, 8.0].
Question: Which of the following is the most likely diagnosis of the patient's skin lesion? (A) Nevus (B) Basal Cell Carcinoma (C) Squamous Cell Carcinoma (D) Actinic Keratosis (E) Seborrheic Keratosis (F) Melanoma.

Response
(A)

Open-ended Visual QA in Chinese (Slake-VQA)

Visual input



Instruction
You are a helpful medical assistant. The following are questions about medical knowledge. Solve them in a step-by-step fashion, referring to authoritative sources as needed.
Question: 图像里包含的区域属于身体哪个部分?

Response
腹部

Image Classification (MIMIC-CXR 13-condition classification)

Visual input



Instruction
You are a helpful radiology assistant. The following are questions about findings in chest X-ray in the frontal view. Identify if a specific type of abnormality is shown in the X-ray.
Given the <VIEW> X-ray image,
Question: Which of the following abnormalities are indicated by the image? (A) Atelectasis (B) Cardiomegaly (C) Consolidation (D) Edema (E) Enlarged Cardiomeastinum (F) Fracture (G) Lung Lesion (H) Lung Opacity (I) Pleural Effusion (J) Pleural Other (K) Pneumonia (L) Pneumothorax (M) Support Devices

Response
(A)

Close-ended Visual QA (NEJM Image Challenge, USMLE-MM)

Visual input



Instruction
You are a medical expert. Only output the final {diagnosis, answer}. Do not output the reasoning or explanation. Output the final diagnosis in the format "Final {Diagnosis, Answer}: X" where X is the most {possible medical diagnosis, correct letter choice}.
Question: Infection with which one of the following organisms is the most likely cause of this rash? (A) Coxsackie virus type A16 (B) Echovirus type 16 (C) Group A streptococcus (D) Herpes simplex virus type 1 (E) Norwalk virus

Response
Final Answer: (A)

Image Classification (MIMIC-CXR normal vs abnormal classification)

Visual input



Instruction
You are a helpful radiology assistant. The following are questions about findings in chest X-ray in the frontal view. Identify if a specific type of abnormality is shown in the X-ray.
Given the <VIEW> X-ray image,
Question: are there any abnormalities indicated by the image? (A) Yes (B) No.

Response
(A)

Waveform Signal Visual QA (ECG-QA)

Raw sensor input*



Instruction
Given this ECG sequence, please answer the following question. From the provided options, select all that apply. List your selections alphabetically, separated by commas.
Question: What signs of a rhythm-related disorder can be found in this ECG recording?
Options: atrial fibrillation, atrial flutter, bigeminal pattern, normal functioning artificial pacemaker, sinus arrhythmia, sinus bradycardia, sinus rhythm, sinus tachycardia, supraventricular tachycardia

Response
atrial fibrillation, atrial flutter

Text Report Generation (MIMIC-CXR)

Visual input



Instruction
You are a helpful radiology assistant.
Question: Describe the findings and impression in the chest X-ray following the instructions.

Response
FINDINGS: Low lung volumes are present. This accentuates the size of the cardiac silhouette which is likely mildly enlarged. Mediastinal and hilar contours are likely within normal limits. A right brachiocephalic venous stent is re-demonstrated. There is crowding of the bronchovascular structures with probable mild pulmonary vascular congestion. No pleural effusion or pneumothorax is identified.
IMPRESSION: Low lung volumes with mild pulmonary vascular congestion.

AlphaFold

Protein structure

Article | [Open access](#) | Published: 08 May 2024

Accurate structure prediction of biomolecular interactions with AlphaFold 3

[Josh Abramson](#), [Jonas Adler](#), [Jack Dunger](#), [Richard Evans](#), [Tim Green](#), [Alexander Pritzel](#), [Olaf Ronneberger](#), [Lindsay Willmore](#), [Andrew J. Ballard](#), [Joshua Bambrick](#), [Sebastian W. Bodenstein](#), [David A. Evans](#), [Chia-Chun Hung](#), [Michael O'Neill](#), [David Reiman](#), [Kathryn Tunyasuvunakool](#), [Zachary Wu](#), [Akvilė Žemgulytė](#), [Eirini Arvaniti](#), [Charles Beattie](#), [Ottavia Bertolli](#), [Alex Bridgland](#), [Alexey Cherepanov](#), [Miles Congreve](#), ... [John M. Jumper](#)  [+ Show authors](#)

[Nature](#) **630**, 493–500 (2024) | [Cite this article](#)

874k Accesses | **2100** Altmetric | [Metrics](#)

CASE STUDY

Accelerating the fight against malaria

[Learn more](#)



CASE STUDY

Paving the way for potential Parkinson's treatments

[Learn more](#)



CASE STUDY

Racing against drug-resistant bacteria

[Learn more](#)



CASE STUDY

Breaking down plastic pollution

[Learn more](#)



CASE STUDY

Designing more effective drugs

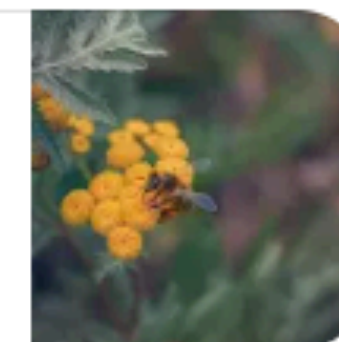
[Learn more](#)



CASE STUDY

Increasing honeybees' chances of survival

[Learn more](#)



AlphaFold is not an LLM, but it uses related transformer technology. You can think of it as a specialized geometric transformer for biology.

Ethics and Trust

ETHICS, BIAS, RESPONSIBILITY, TRUST

DEFINITIONS

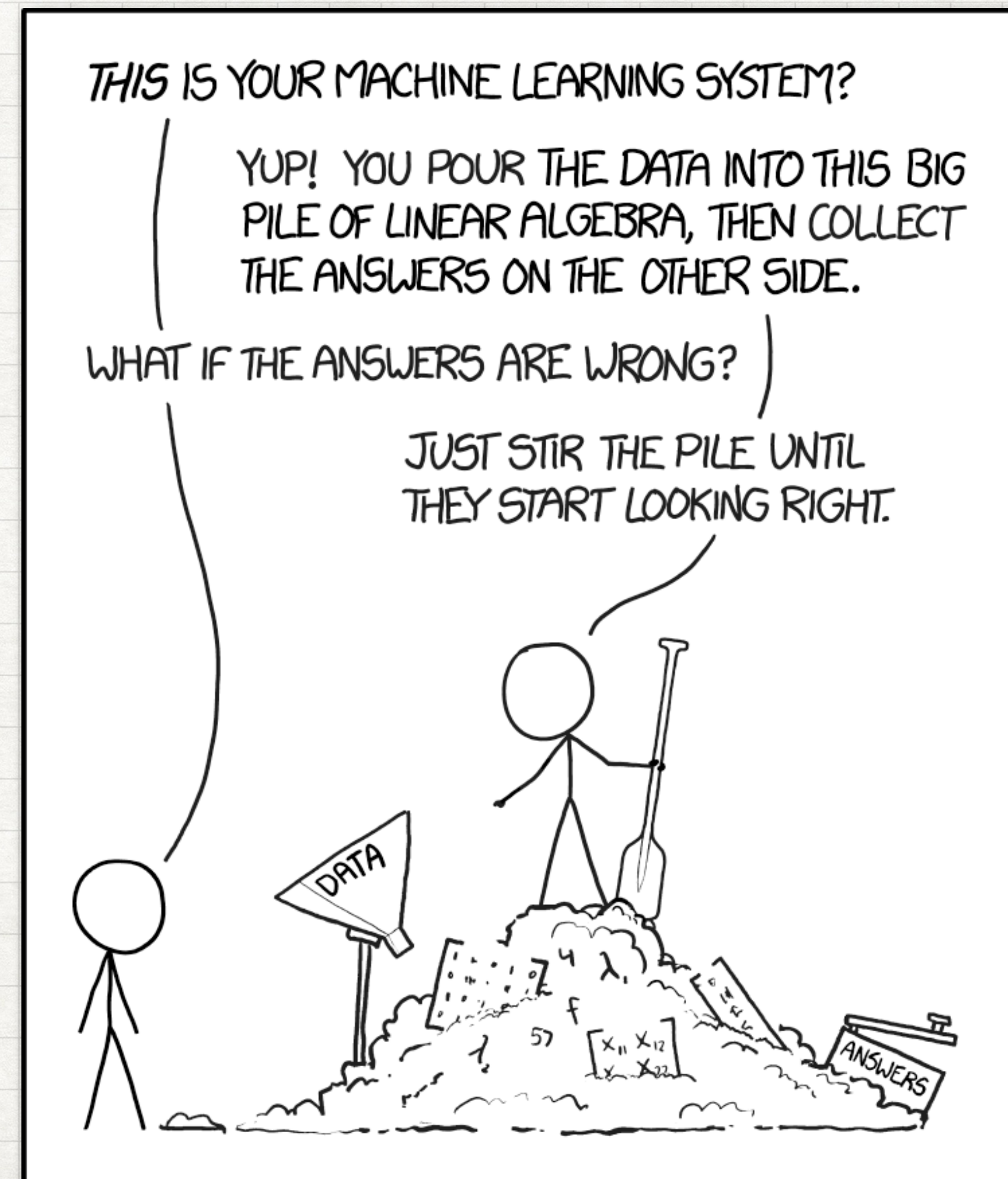
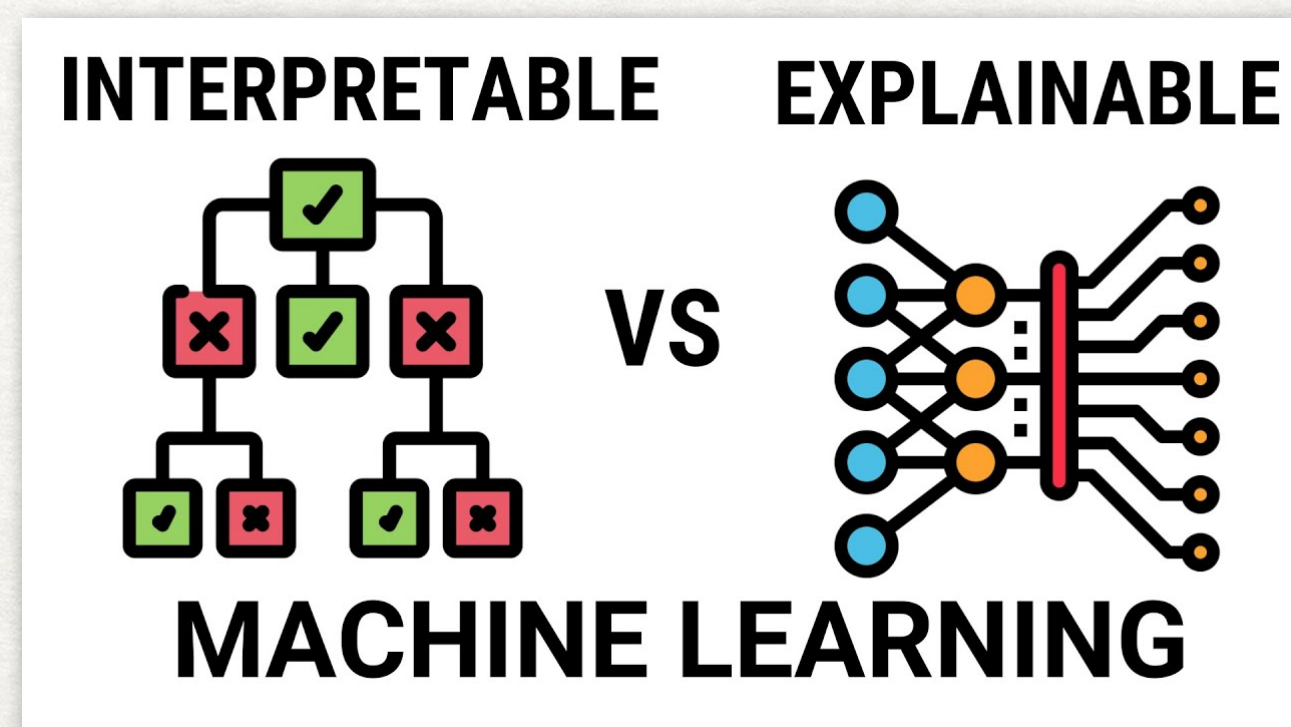
- **ETHICS** = what is morally good or bad, right or wrong? (norms)
- **BIAS** = prejudice against a person, object, position.
- **TRUST** = willingness to assume risk by relying on, or believing in, the actions of another party
- **TRUSTWORTHY AI** should be lawful, ethical, unbiased.



EXPLAINABILITY & INTERPRETABILITY

DEFINITIONS

- **INTERPRETABLE AI** = can be understood, line-by-line, without additional explanation = permits a decision of trust = not a black box
- **EXPLAINABLE AI** = can be explained post hoc, after training, in a way that makes it understandable = transparency in black boxes = feature importance, effects, interactions



HOW CAN AI GO WRONG?

TRAINING, MODELS, SOCIETY

- Training data issues:

- Non-representative, lack of geodiversity.
- Faulty, biased training labels.
- Adversarial effects.

- AI model issues:

- Learn faulty strategies.
- Fake something plausible.
- Non-trustworthy, lack of robustness.

- Societal issues:

- Lack of consent on data collection.
- Disenfranchise scientists.
- CO2 emissions.
- Globally applicable AI approaches may stymie burgeoning efforts in developing countries.

Conclusions

Where do we stand?

AI & HPC Infrastructure Investments

Annual Investment Comparison for 2025



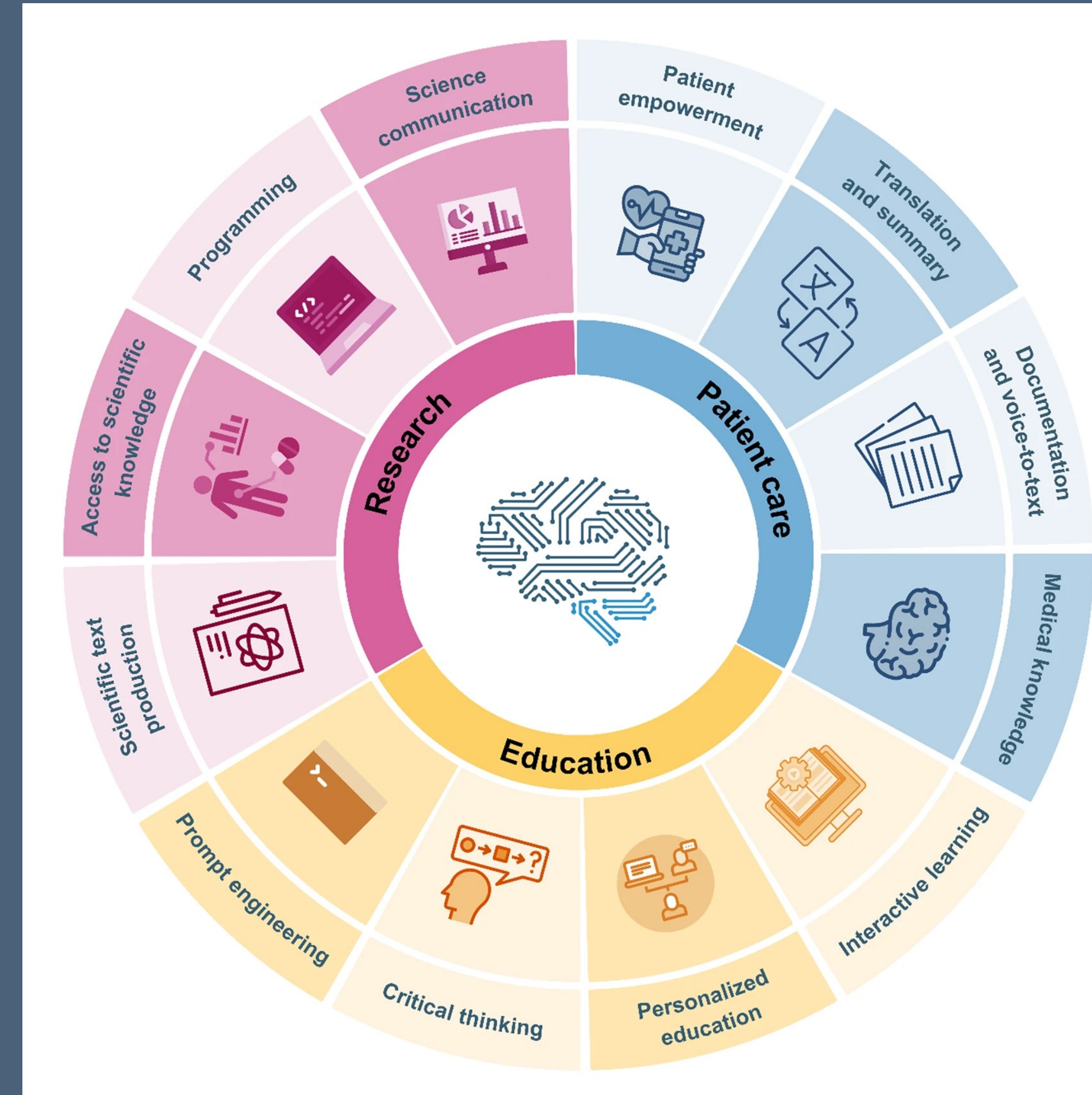
Key Strategic Insights

- Private AI investment exceeds entire national economies
- Research institutions face unprecedented resource disparities
- European HPC funding gap highlights strategic vulnerability
- Need for coordinated European digital sovereignty initiative

LARGE LANGUAGE MODELS

Tomorrow?

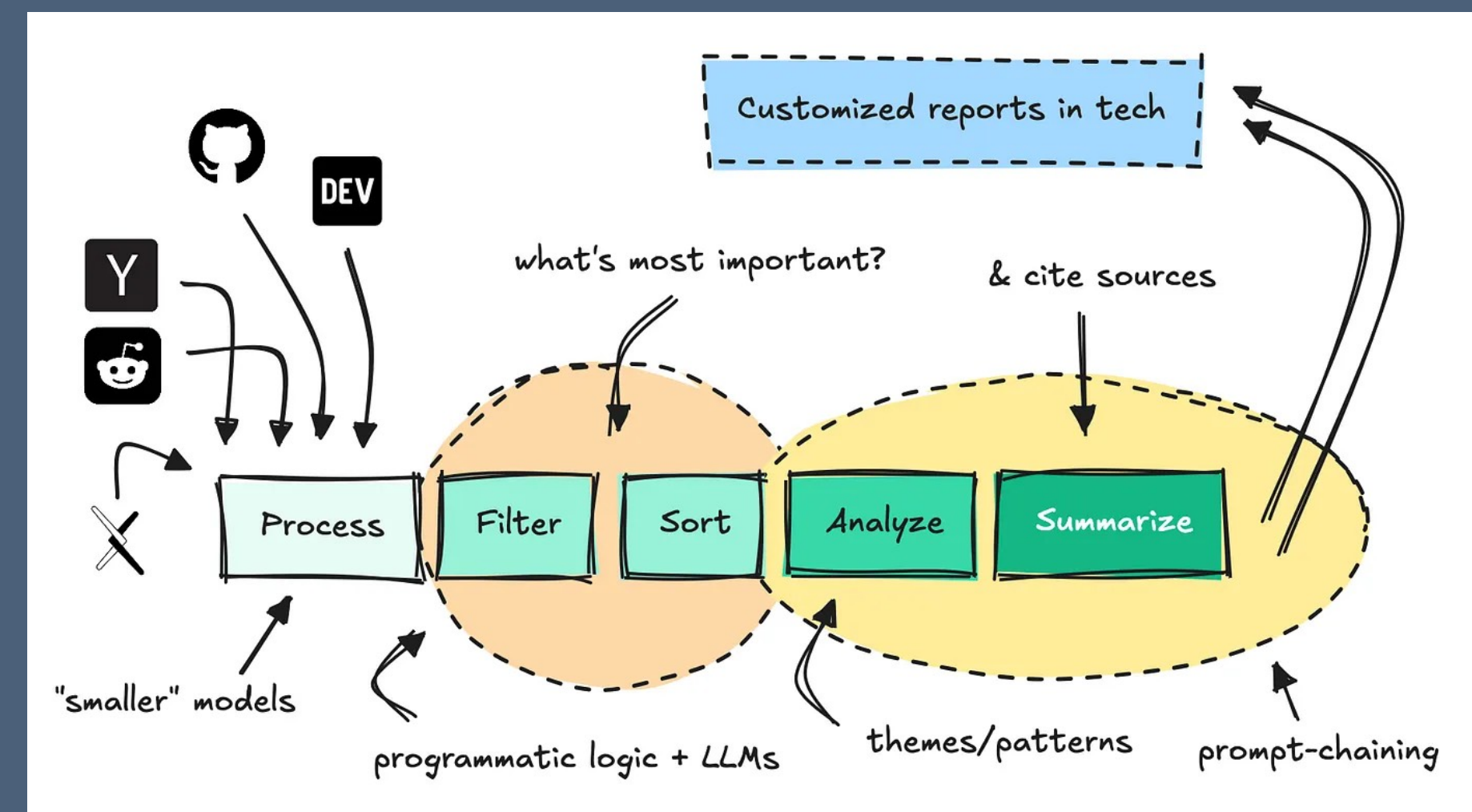
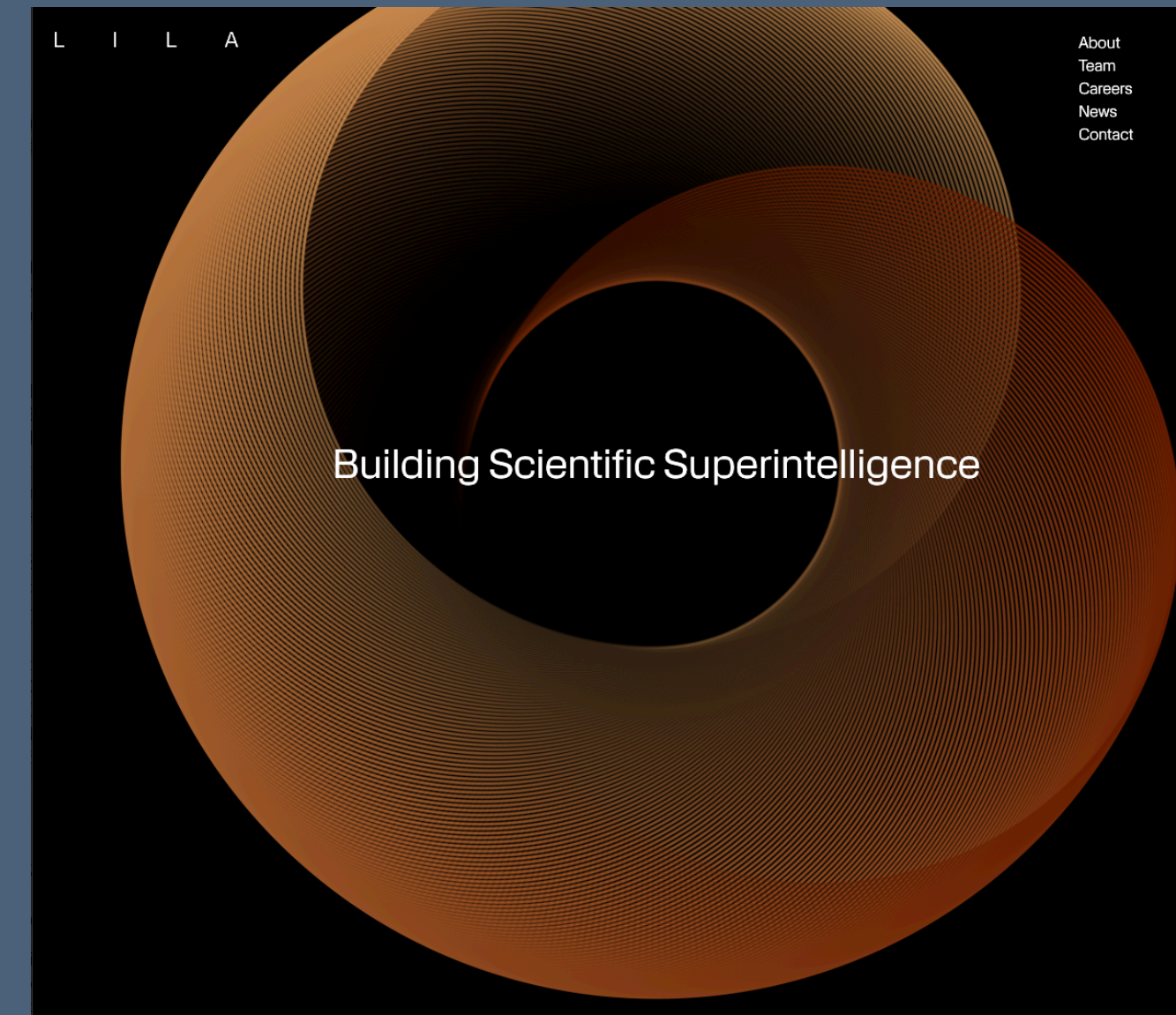
- Small number of very large, intelligent LLMs
- Many smaller, specialized LLMs
 - Healthcare, legal, finance, etc.
- Personal LLMs (your “story”)
- Uses in education:
 - How to educate?
 - What to teach?



The future of scientific research?

Where are we heading?

- Superintelligence platforms and **autonomous labs** for life, chemical, and materials sciences (eg. LILA = \$1.3bn)
- **Research AI agents** to execute complete workflows. Eg. agent that can explore all research on a given subject, aggregate millions of texts, filter data based on a persona, and find patterns and themes that you can act on.



Thank You!

Contact:

- mark.asch@u-picardie.fr
- <https://www.linkedin.com/in/mark-asch-8a257130/>
- <https://github.com/markasch>
- <https://markasch.github.io/DT-tbx-v1/>

