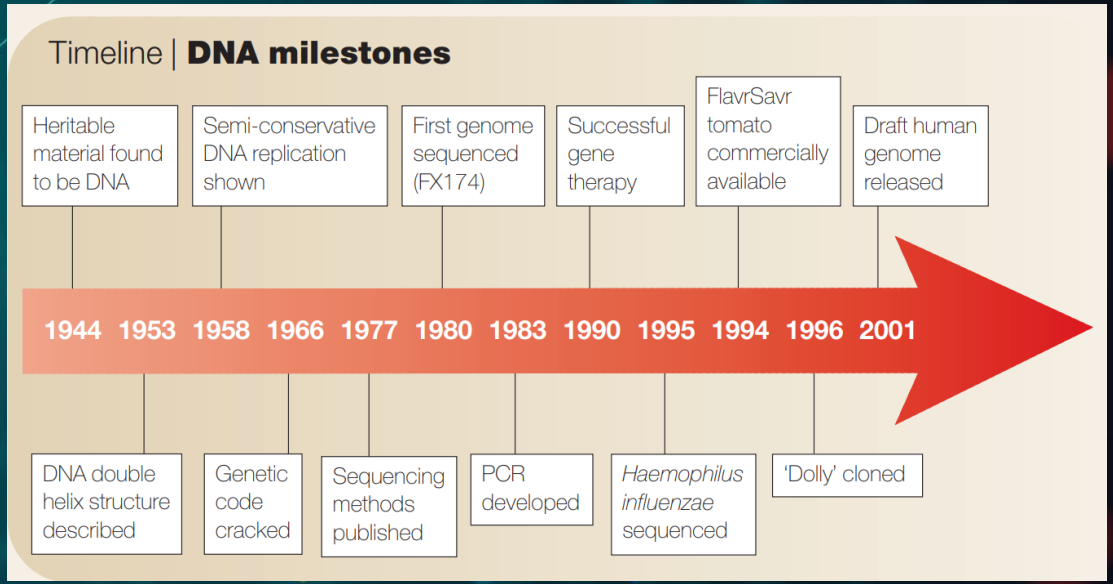
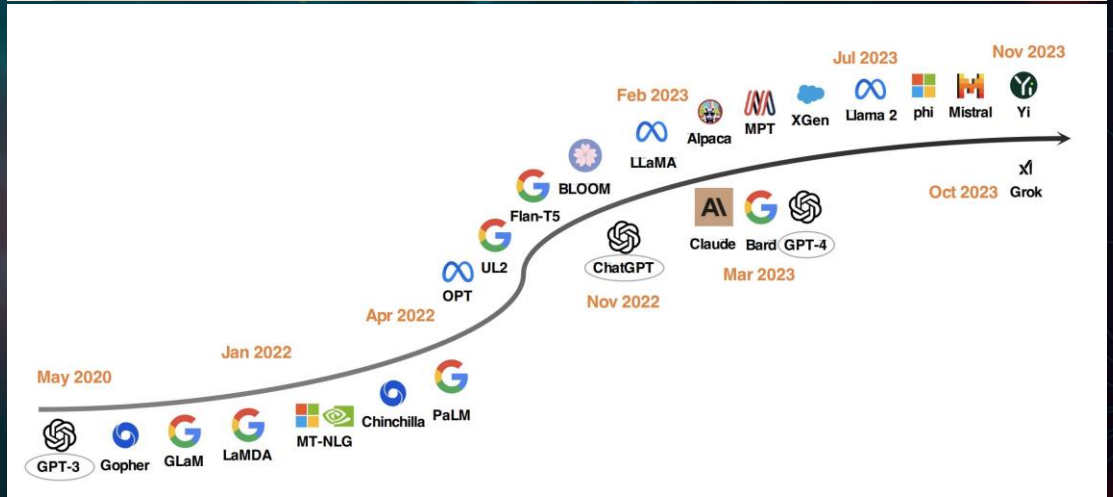
بوت کمپ هوش مصنوعی در تحقیقات علوم پزشکی

Dr. Sina Arabi

Genomics Vs. LLMs

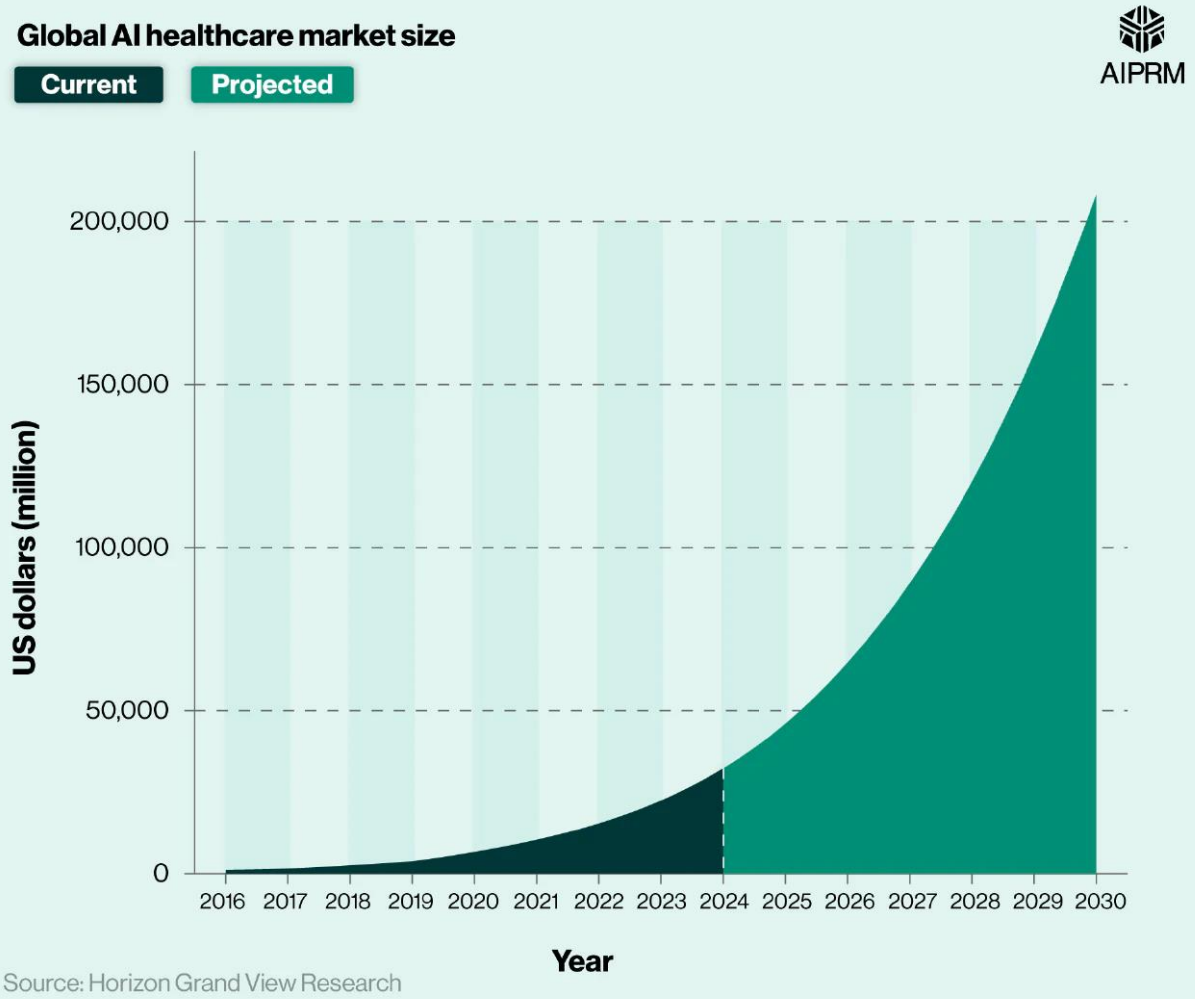




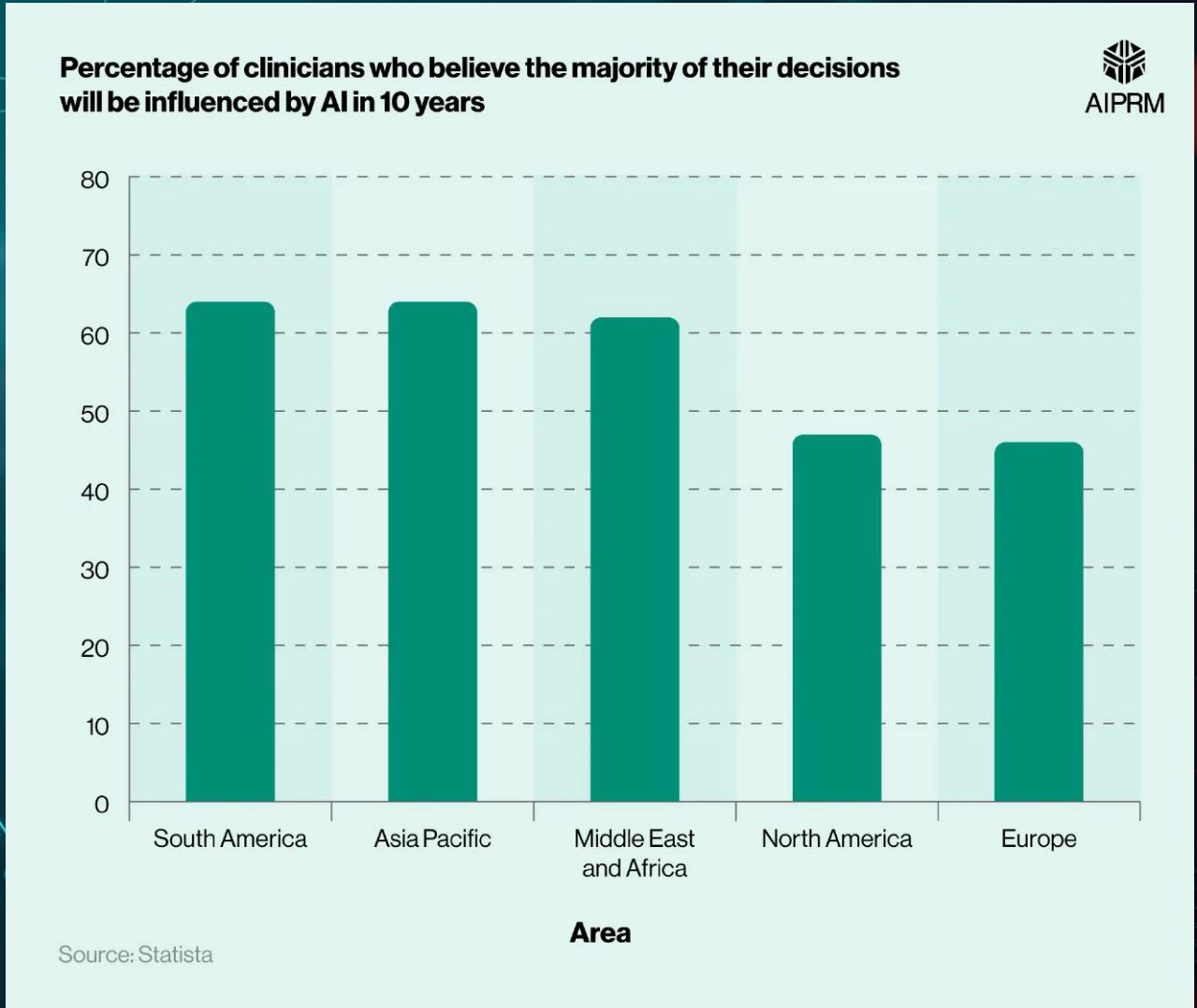
10 interesting facts about AI in healthcare

- Between 2020 and 2023, the market size of AI in healthcare grew by 233%, from $6.7 billion to $22.4 billion.
- The USA is forecast to generate $102.2 billion in revenue by 2030 through AI in the health sector.
- Nearly two-thirds (64%) of clinicians in South America believe the majority of their decisions will be benefitted by AI.
- More than half (53%) of EU healthcare organizations plan to use medical robotics by the end of 2024.
- A quarter (25%) of US hospitals already use predictive analysis driven by AI.
- Four in five (80%) pathologists believe AI will boost life expectancy.

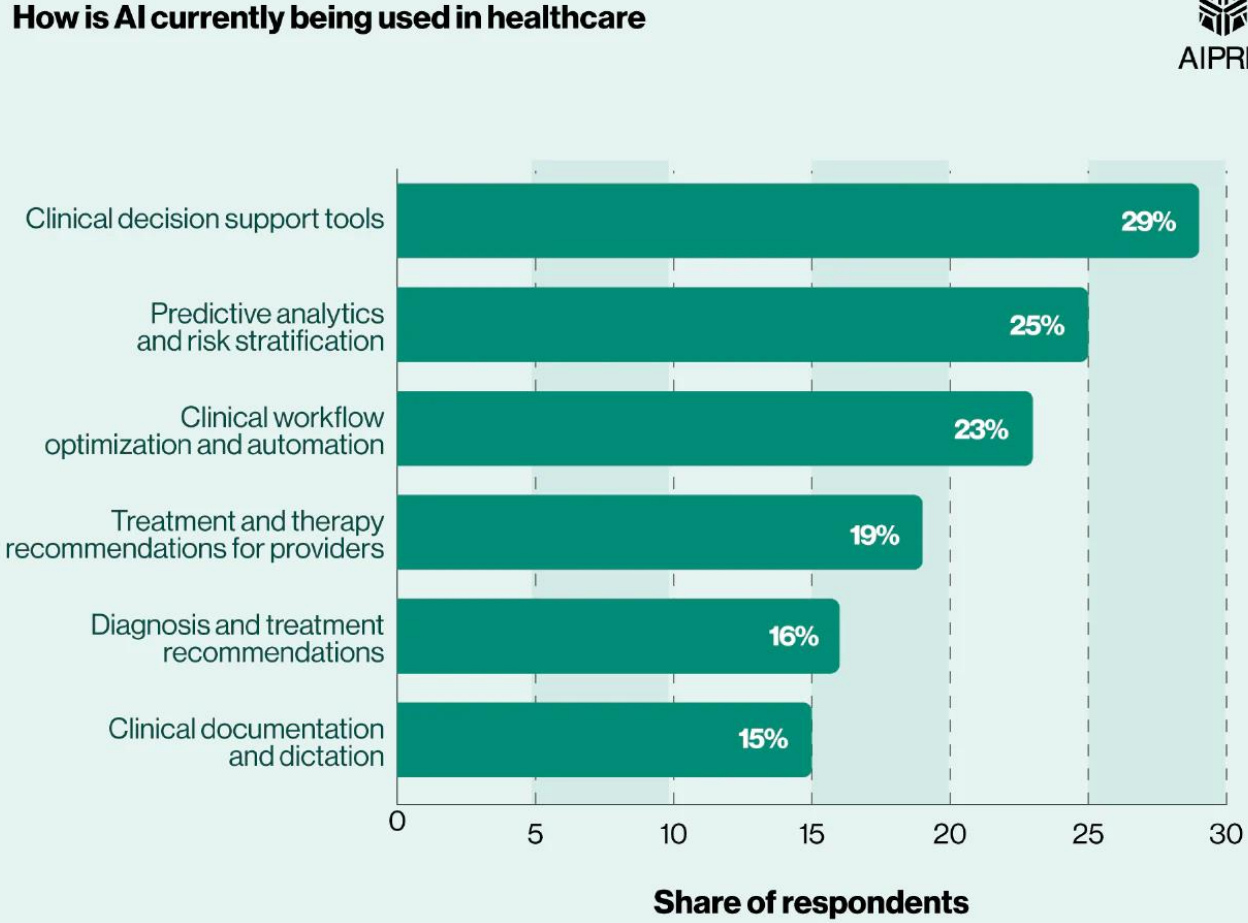
Global AI healthcare market size



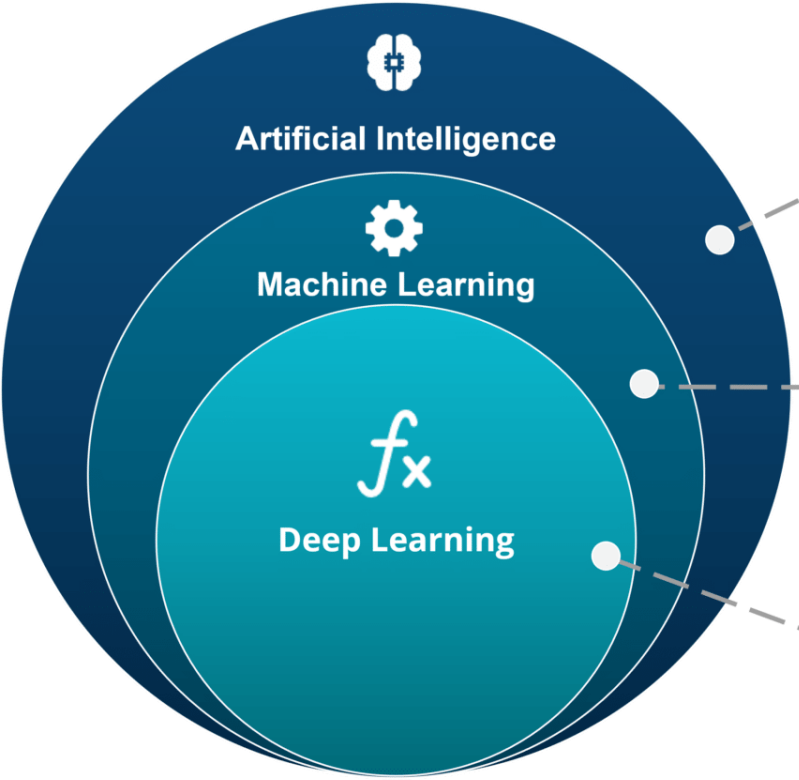
Professional consensus



AI usage in Clinics



Source: Statista



ARTIFICIAL INTELLIGENCE

A technique which enables machines to mimic human behaviour

هوش مصنوعی

MACHINE LEARNING

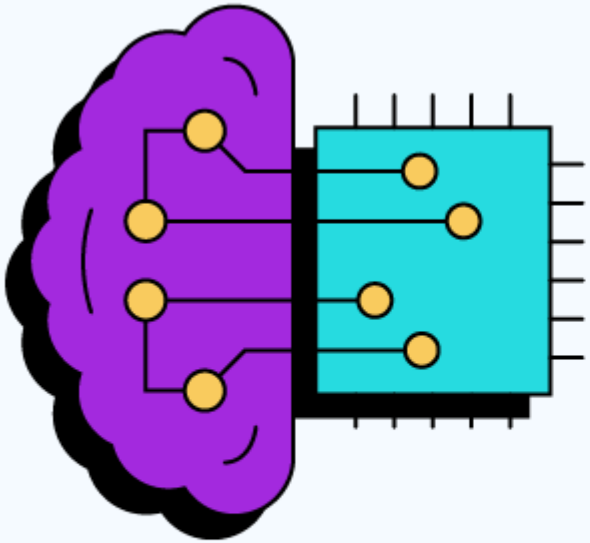
Subset of AI technique which use statistical methods to enable machines to improve with experience

یادگیری ماشین

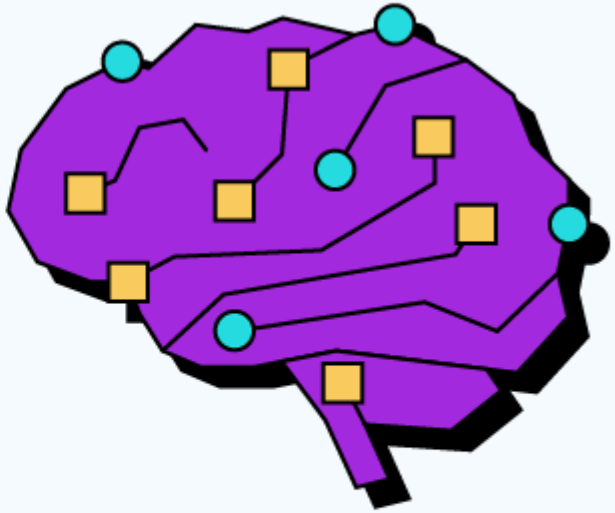
DEEP LEARNING

Subset of ML which make the computation of multi-layer neural network feasible

یادگیری عمیق



VS



MACHINE LEARNING

DEEP LEARNING

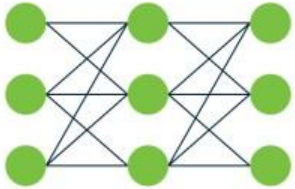
یادگیری ماشین

یادگیری عمیق

Machine Learning

Input

Feature extraction



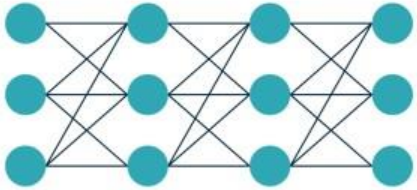
Classification

CAR
NOT CAR

Output

Deep Learning

Input

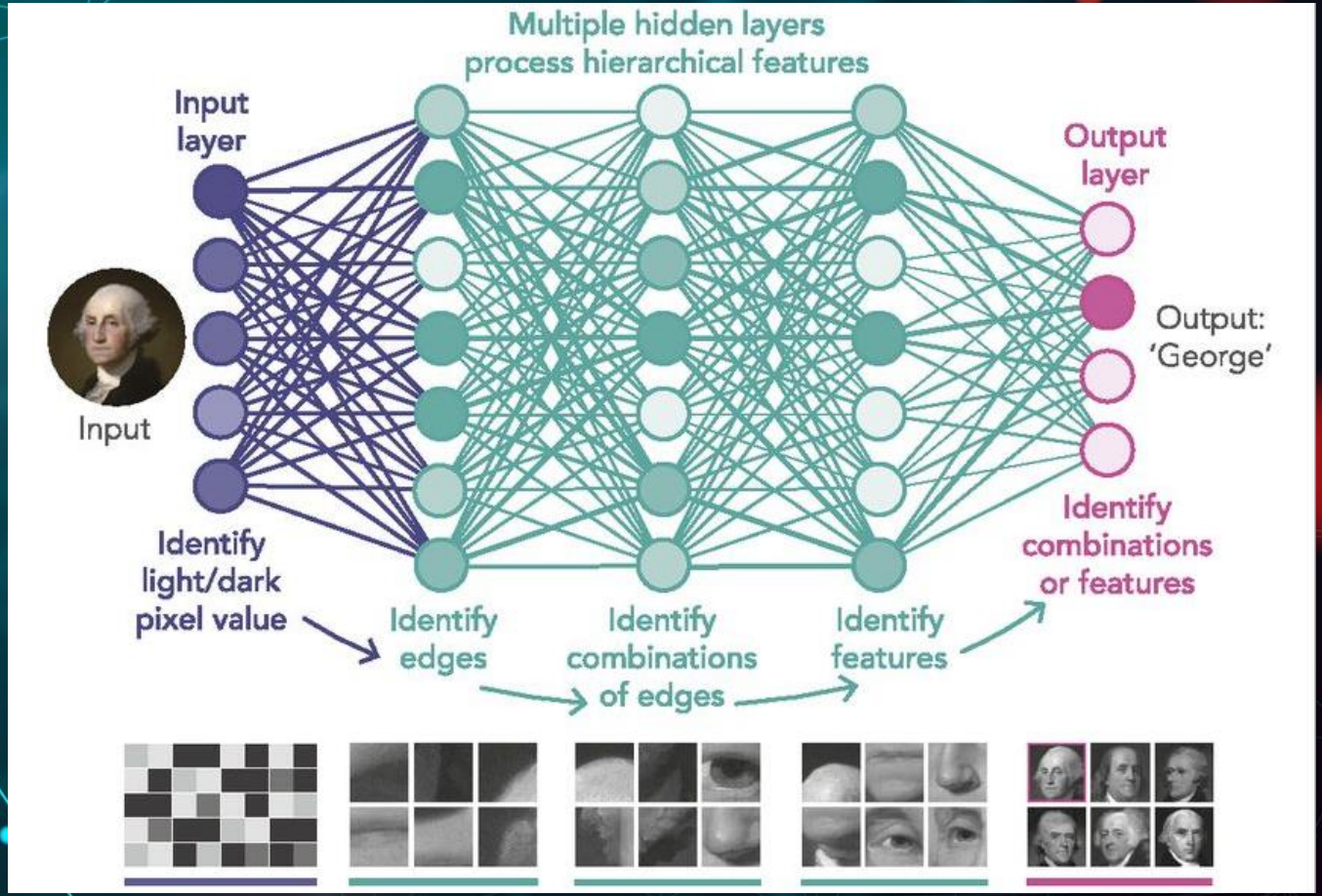


Feature extraction + Classification

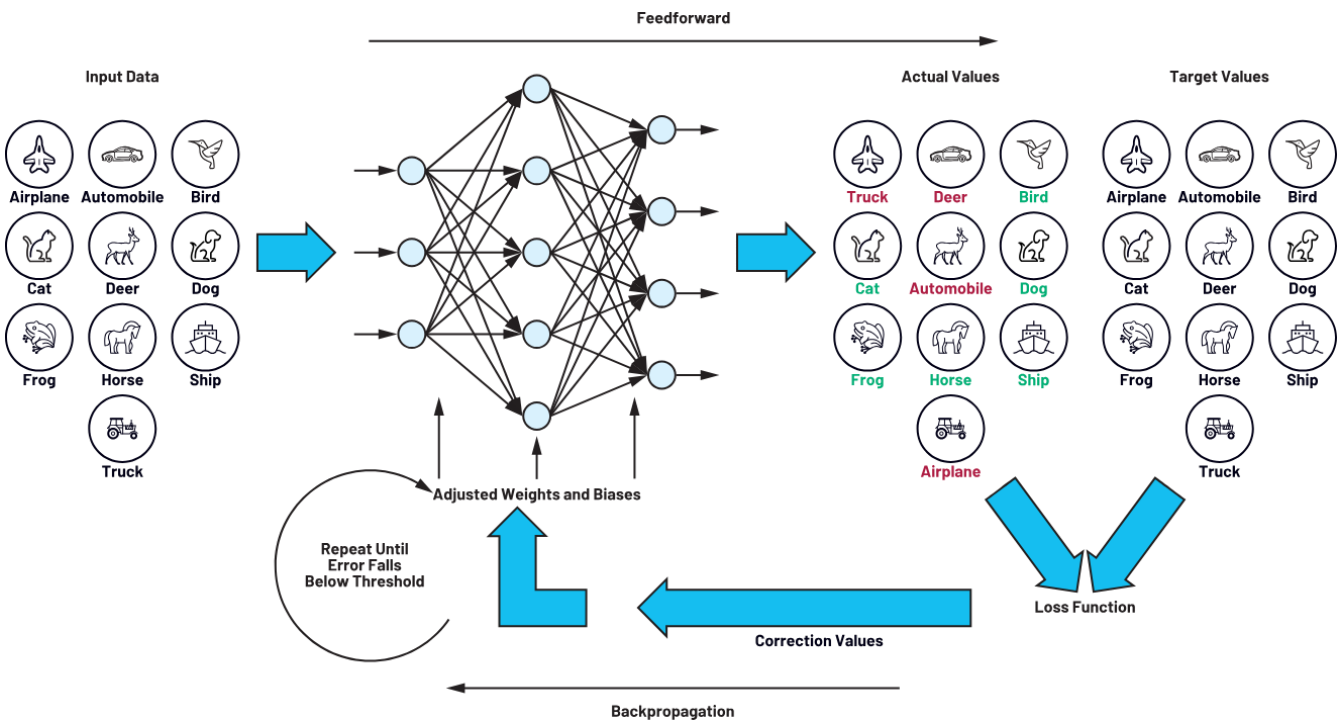
CAR
NOT CAR

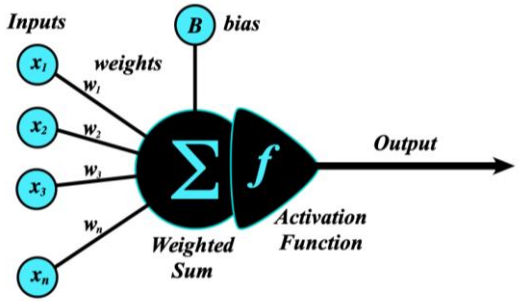
Output

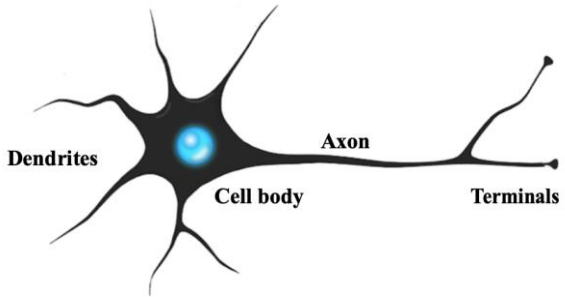
Deep learning neural network



*pimeyes.com/en


Feedforward
Input Data
Airplane
Automobile
Bird
Cat
Deer
Dog
Frog
Horse
Ship
Truck
Actual Values
Truck
Deer
Bird
Cat
Automobile
Dog
Frog
Horse
Ship
Airplane
Target Values
Airplane
Automobile
Bird
Cat
Deer
Dog
Frog
Horse
Ship
Truck
Adjusted Weights and Biases
Repeat Until
Error Falls
Below Threshold
Loss Function
Correction Values
Backpropagation


Inputs
bias
weights
Output
Activation
Function
Weighted
Sum


Dendrites
Axon
Cell body
Terminals

Popular Deep Learning Architectures

01

Convolutional Neural Networks (CNNs)

02

Recurrent Neural Networks (RNNs)

03

Generative Adversarial Networks (GANs)

04

Transformers: natural language processing (NLP), (e.g., GPT, BERT): Generative Pre-trained Transformer

What is an LLM?

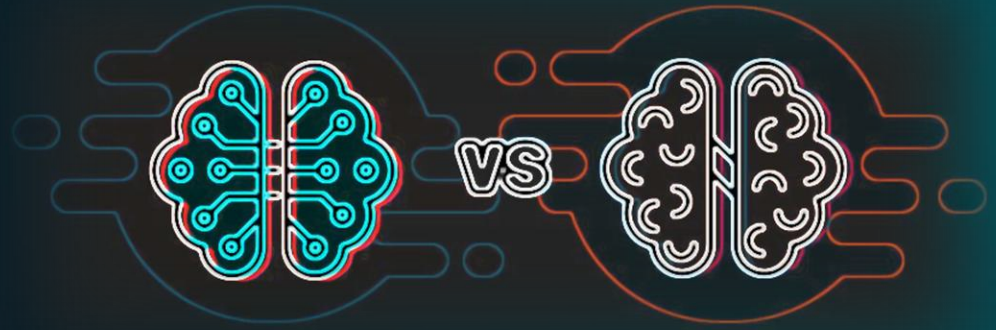
Large Language Models

are advanced machine learning models designed to understand, generate, and process human text by using vast amounts of data and powerful neural network architectures. They are a specialized application of deep learning, primarily in the domain of natural language processing (NLP).

Large Language Models (LLMs) are trained on diverse and massive datasets

1. Text from the Internet
2. Books
3. Scientific Literature
   -Research Papers: Articles from repositories like PubMed, arXiv, and other academic archives.
4. Wikipedia
5. News Articles
6. Social Media
7. Code Repositories (GitHub)

As of November 15, 2024, ChatGPT's training data includes information up to 2023.



2017

Transformers

Advanced features
Decreased training time

2018

GPT-1

117 million parameters
Revolutionary model

2018

BERT

340 million parameters
Bidirectional text processing

2019 - 2020

LLM Scaling

GPT-2 (1.5 billion parameters)
GPT-3 (175 billion parameters)

2023

GPT-4

1.76 trillion parameters
Multimodal capabilities

GPT Models (e.g., GPT-3, GPT-4 by OpenAI)

Data Resources:

- Common Crawl: A large dataset of web pages covering diverse topics.
- WebText: A dataset of web content curated by OpenAI, focused on high-quality text.
- Wikipedia: Provides structured general knowledge.
- Books: Includes fiction and non-fiction books for a diverse understanding of language.
- Code Repositories: Publicly available code for coding-related tasks (used in GPT models like Codex).

Web Documents: A vast collection of publicly available web pages, encompassing a wide range of topics and writing styles, to provide a comprehensive understanding of human language and knowledge.

Books: Inclusion of both fiction and non-fiction books to enhance the model's grasp of narrative structures, factual information, and varied linguistic expressions. A dataset of over 11,000 books for rich textual diversity.

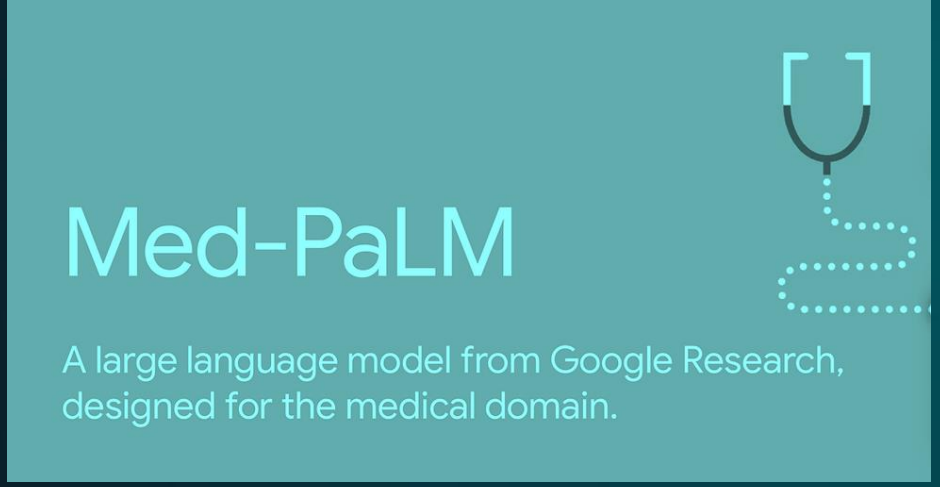
Code: Datasets comprising programming code from various languages and domains, enabling Gemini to assist in code generation, debugging, and understanding software development contexts.

Multimodal Data:

Images: A diverse set of images to train the model in visual recognition and description tasks.

Audio: Audio recordings, including speech and environmental sounds, to facilitate tasks like transcription and audio analysis.

Video: Video content to enable understanding and generation of information that combines visual and auditory data.



PubMed: A repository of biomedical literature.
Clinical Notes: Anonymized medical notes.
Medical Datasets: Tailored for healthcare applications (UpToDate).

Summary

Using ChatGPT for data analysis

HOW WE DO IT

01

Data preparation

02

Basic Data Analysis

03

Statistical tests

04

Visualization

Prepare Your Data

1. Ensure your data is in a format that ChatGPT can work with.
2. CSV, Excel, or other common data formats.
3. Keep sensitive data anonymized to maintain privacy.
4. Identify missing or inconsistent data.
5. Labelling

Basic Data Analysis

- Summarizing data:

  - mean
  - median
  - mode
  - standard deviation

Statistical tests

- Choose the right statistical test (e.g., t-test, ANOVA, chi-square).
- Explain how to interpret p-values or confidence intervals.
- Generate code for performing statistical tests.

Summary

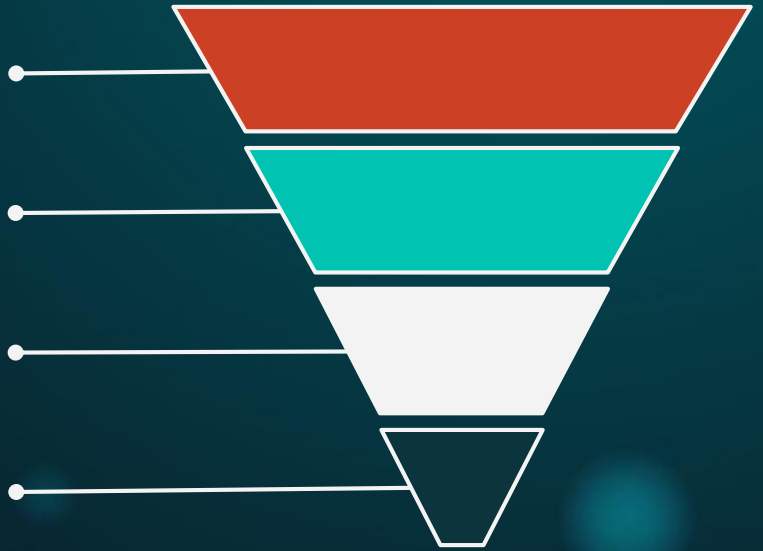
Use LLMs for writing a manuscript

Outline structure

Draft text; Improve clarity; Paraphrasing; Data Interpretation

Summarizing papers; Finding gaps

Title and abstract suggestions



Outline structure:

Ask the model to help you draft an outline based on your manuscript's purpose (e.g., introduction, methods, results, discussion, and conclusion).

D Draft text

I Improve clarity

P Paraphrasing

D Data Interpretation

Summarizing papers; Finding gaps

Title and abstract suggestions

Using ChatGPT for Submission

HOW WE DO IT

01

AI-assisted journal selection

02

Cover Letter Preparation

03

pre-submission evaluation

04

Responding to Reviewers

AI-assisted journal selection

Title, Abstract, Keywords

take impact factor into consideration

Author Guidelines

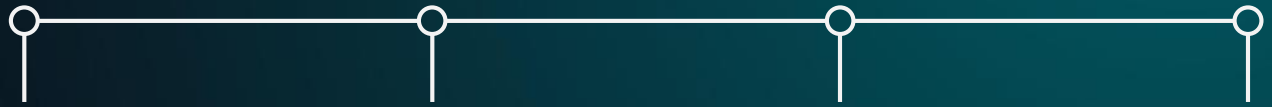
Check Compatibility

Cover Letter Preparation

- How do I write a good cover letter?
- Use the abstract and title for writing a cover letter.
- Revision of cover letter.

pre-submission evaluation
AI as a Reviewer!

responding to reviewers' comments



Understand Reviewer Comments

Before using ChatGPT, ensure you thoroughly understand the reviewers' comments and identify the key points they are raising.

Provide Context to ChatGPT

When prompting ChatGPT, provide sufficient context. Include:
The reviewer’s comment.
Your draft response (if you have one).
Relevant parts of your manuscript (or summarize them).

ChatGPT can help formulate a thorough, balanced response.

Refining the Final Document

While ChatGPT is a powerful tool, it’s crucial to ensure that the responses reflect your genuine scientific understanding and intentions. **Always review the generated text** to ensure accuracy and alignment with your manuscript.

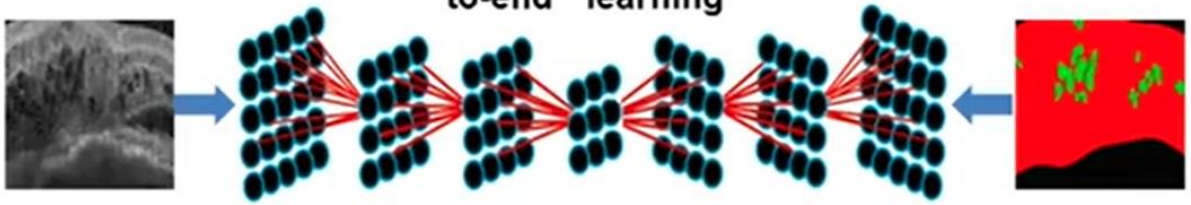
artificial intelligence in medical imaging research

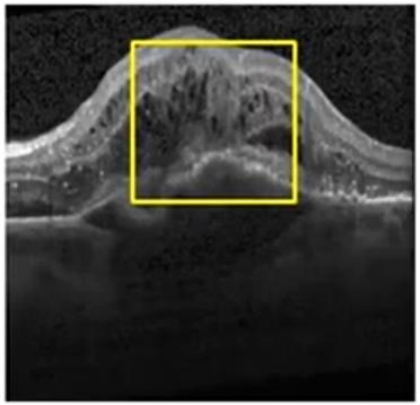
Segmentation in OCT: Deep learning

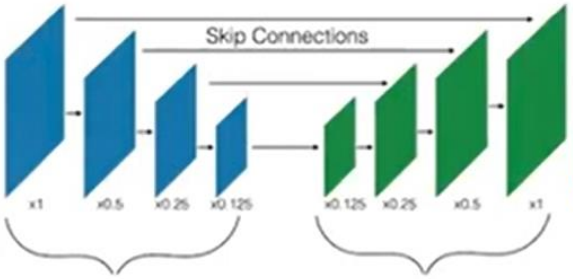
OPTIMA
Ophthalmic Image Analysis

End-to-end learning

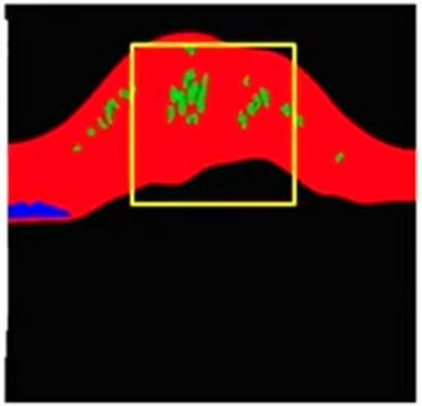
Instead of learning to classify an input image from hand-engineered features, deep learning models are learning to do both, extract features and classify them from the input directly, hence allowing for fully „end-to-end“ learning







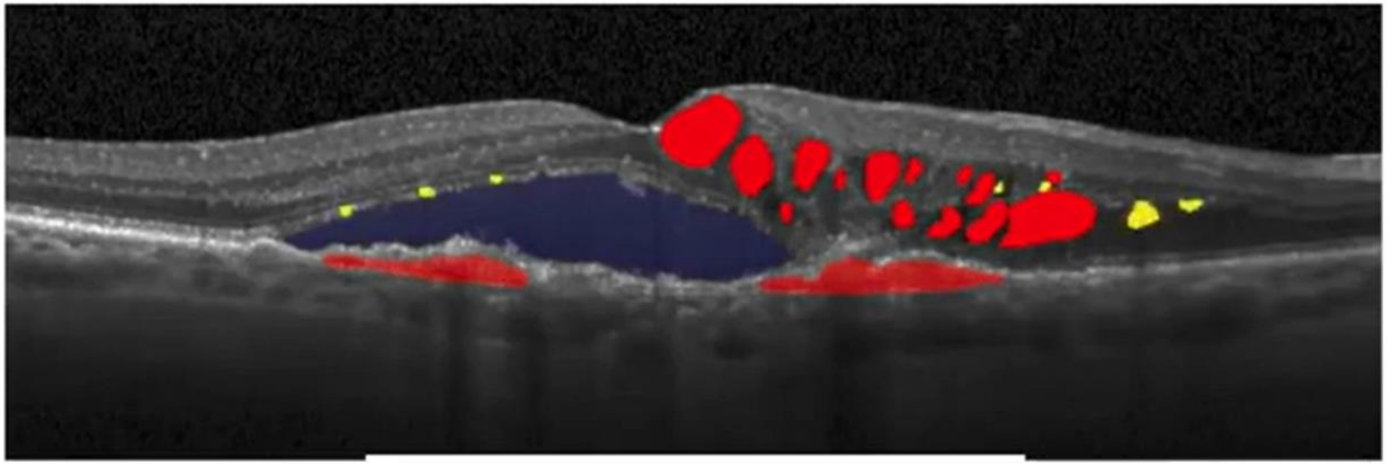
encoder-decoder architecture

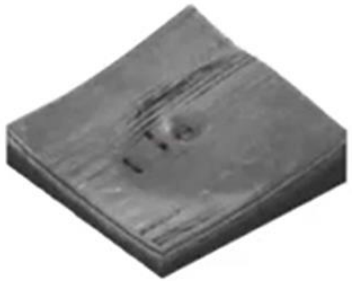


Schlegl et al. Ophthalmology, 2018

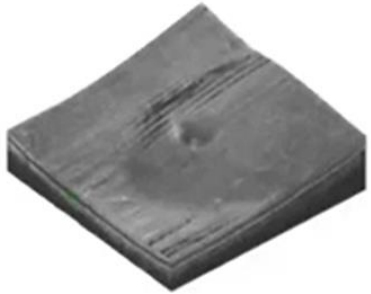
Automated segmentation/quantification of all features, locations and time points

OPTIMA
Ophthalmic Image Analysis

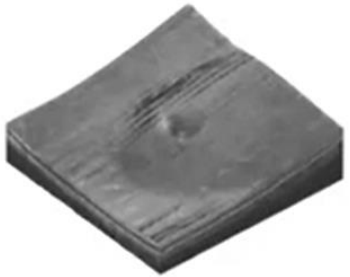




Intraretinal cysts



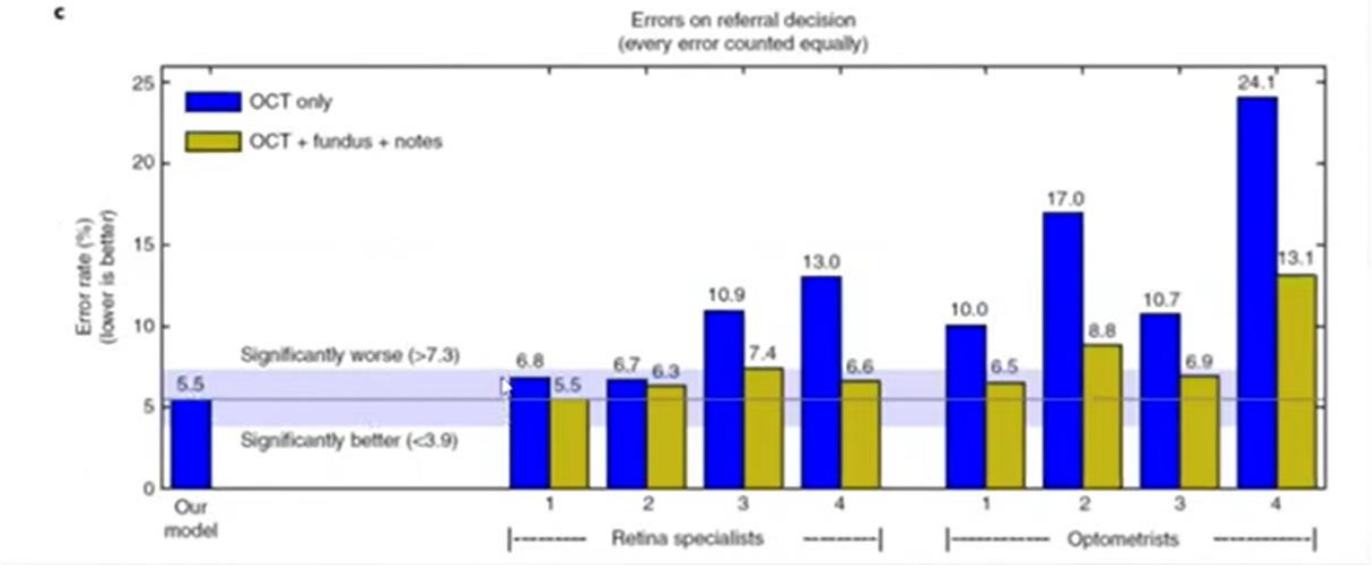
Subretinal fluid



PED

Is AI-based imaging superhuman?

OPTIMA
Ophthalmic Image Analysis



c, Total error rate (1 − accuracy) on referral decision. Values outside the light-blue area (3.9–7.3%) are significantly different (95% confidence interval, using a two-sided exact binomial test) to the framework performance (5.5%). AUC, area under curve.