Artificial Intelligence and Big Data

in Health Care and Research

Alex Sanchez-Pla
Departament de Genètica, Microbiologia i Estadística.
Universitat de Barcelona
2025-12-10

Outline

1) AI is Everywhere

2) Data Science, Machine Learning and AI

3) Data, Health Data, Big Data

4) AI in Health is not future, it's now

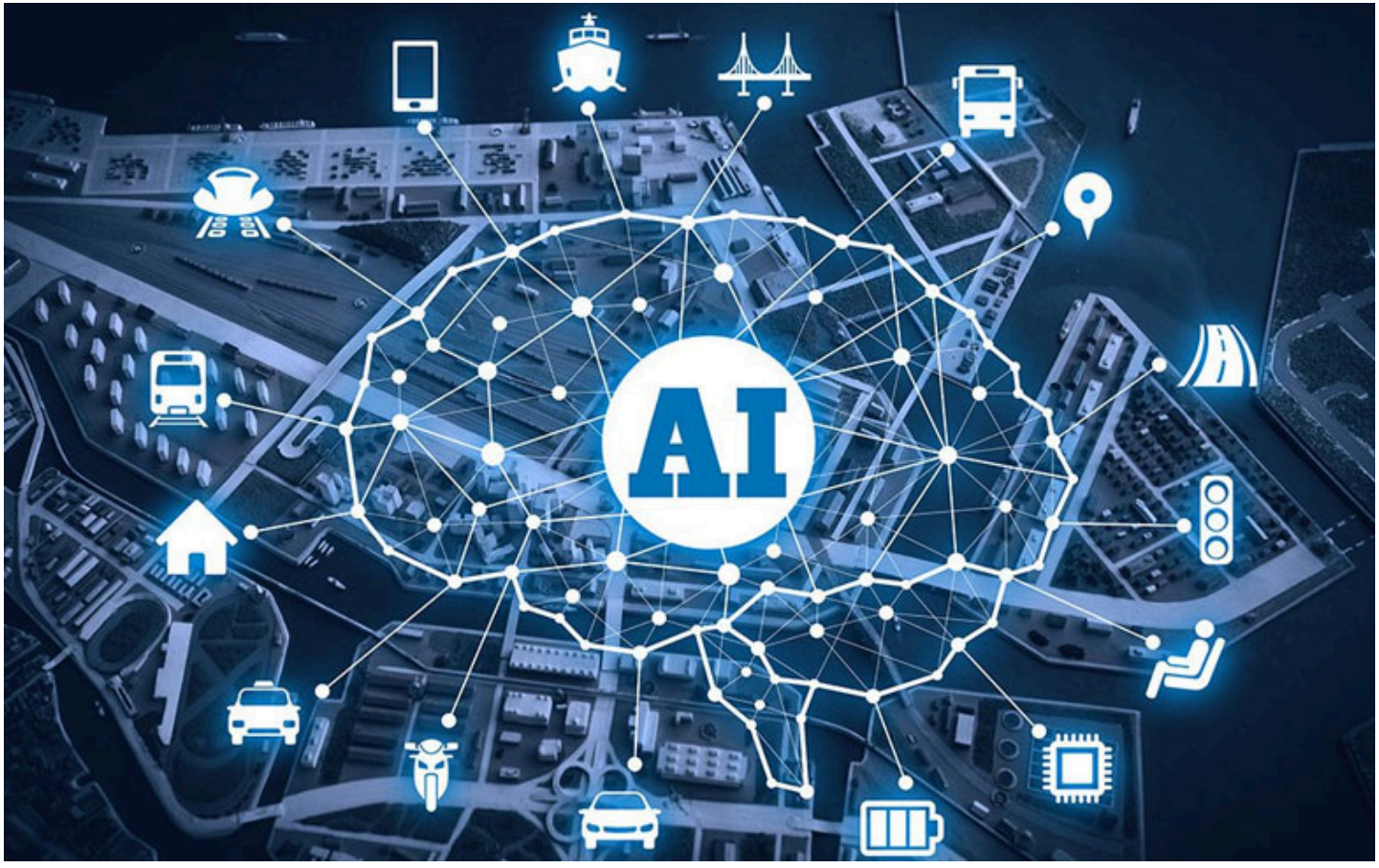
5) What to say about Ethics?

6) Summary

7) References

Artificial Intelligence Everywhere

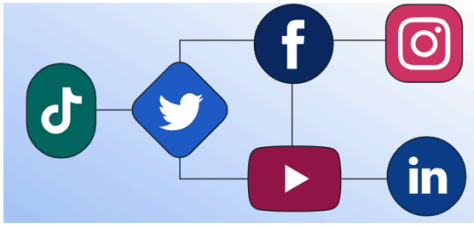
AI : Everyone talks about it

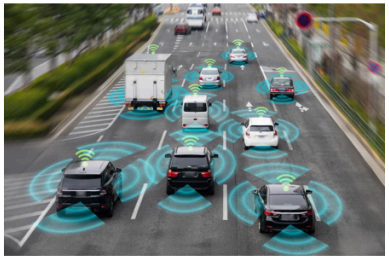


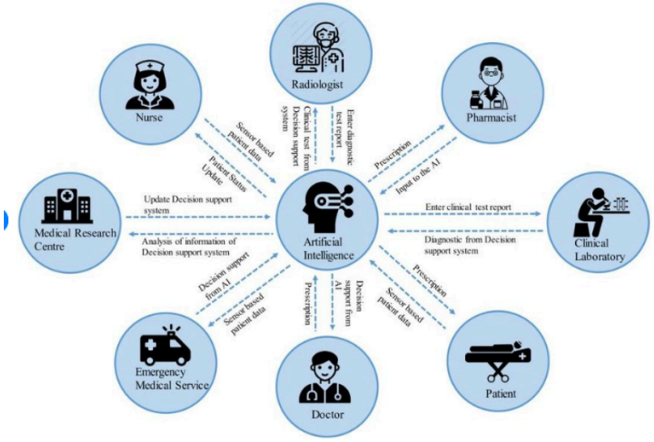
AI : Everyone talks about it

- The term "Artificial Intelligence", AI, refers to *any technology that enables a machine to perform tasks that would typically require human intelligence*, such as **learning**, **problem-solving**, and **decision-making**.
- AI is present in many different technologies and applications that we use every day, such as **voice recognition** in smartphones, **personal assistants** like Siri and Alexa or **recommendation systems** as in Netflix, Amazon or Google.
- Not to say **LLM-based chatbots** like *ChatGPT*, *Gemini* or *Claude* that aim at being game-changers in knowledge management

Examples of AI in everyday life







Data Science, Machine Learning and Artificial Intelligence

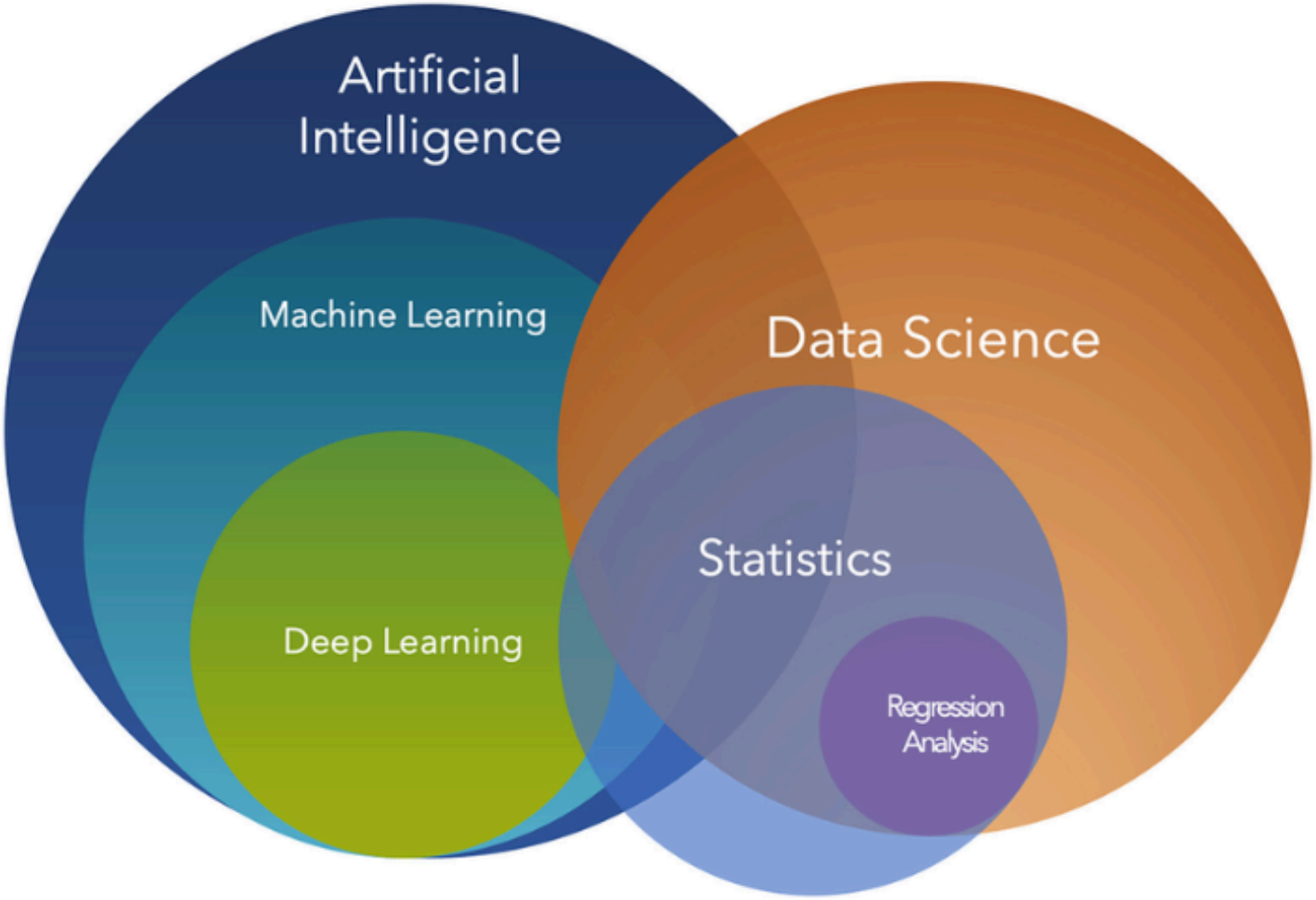
Data Sc., Machine Learning and AI

- Artificial intelligence encompasses divers areas and related disciplines. Two of these are particularly relevant for our goals
  - **Data Science:** is an interdisciplinary field that utilizes scientific methods, processes, algorithms, and systems to extract knowledge and insights from structured and unstructured data.
  - **Machine Learning:** is a discipline that focuses on developing algorithms and models that enable computers to learn from data and make predictions or decisions without being explicitly programmed.

Data Sc., Machine Learning and AI

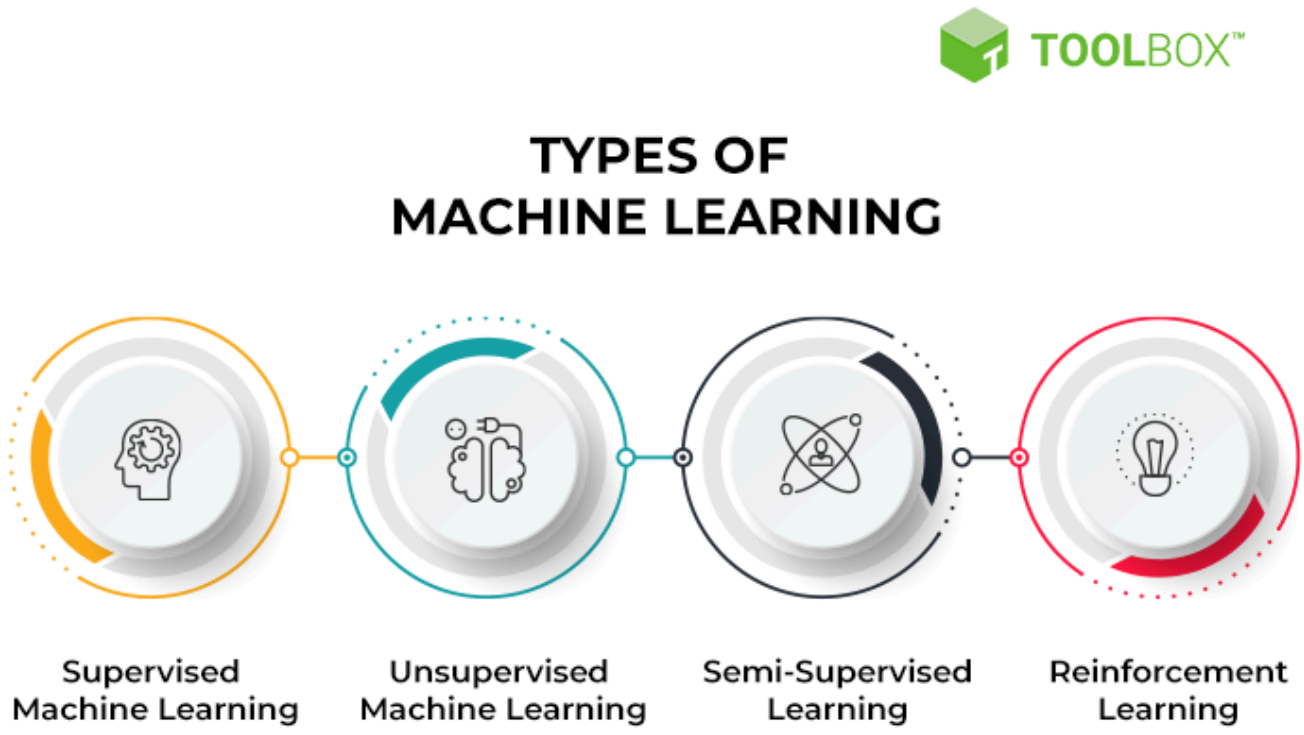
- One can view Data Science and machine learning working together in Artificial Intelligence:
  - Data Science acts as **the foundation**, providing the methodologies to collect, process, and analyze data.
  - Machine Learning **serves as the engine**, employing algorithms to recognize patterns and derive insights from the data.
- Artificial Intelligence integrates the *Learnings from Data* performed by Machine Learning, enabling systems to make intelligent decisions and exhibit adaptive behavior.

Data Sc., Machine Learning and AI

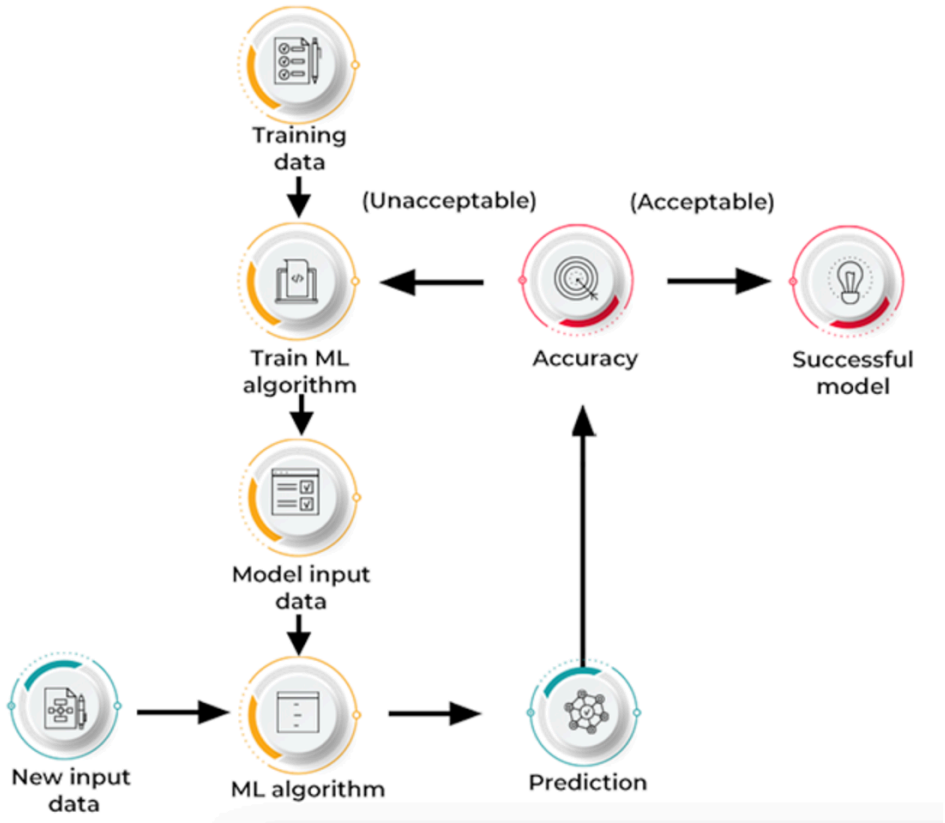


Machine learning

Machine learning (ML) is a discipline that *provides machines with the ability to automatically learn from data and identify patterns to make predictions*



How Machine Learning works



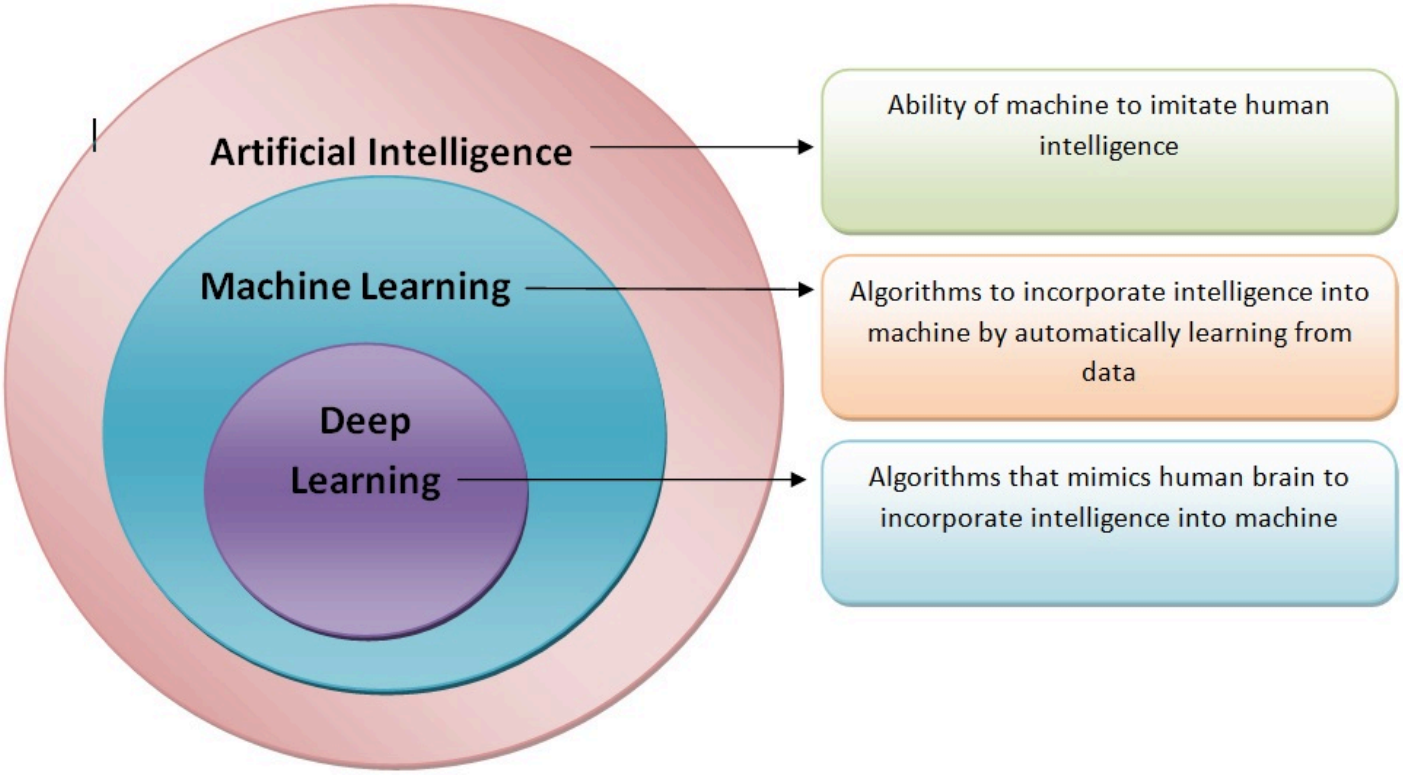
An algorithm is trained on input data that has been labelled for a particular output.

The model is fed with labelled data in the training phase, which instructs the model on what output is related to each specific input value.

The trained model is then presented with test data: data that has been labelled, but the labels have not been revealed to the model before.

The aim of the testing data is to measure how accurately the model will perform with unseen data.

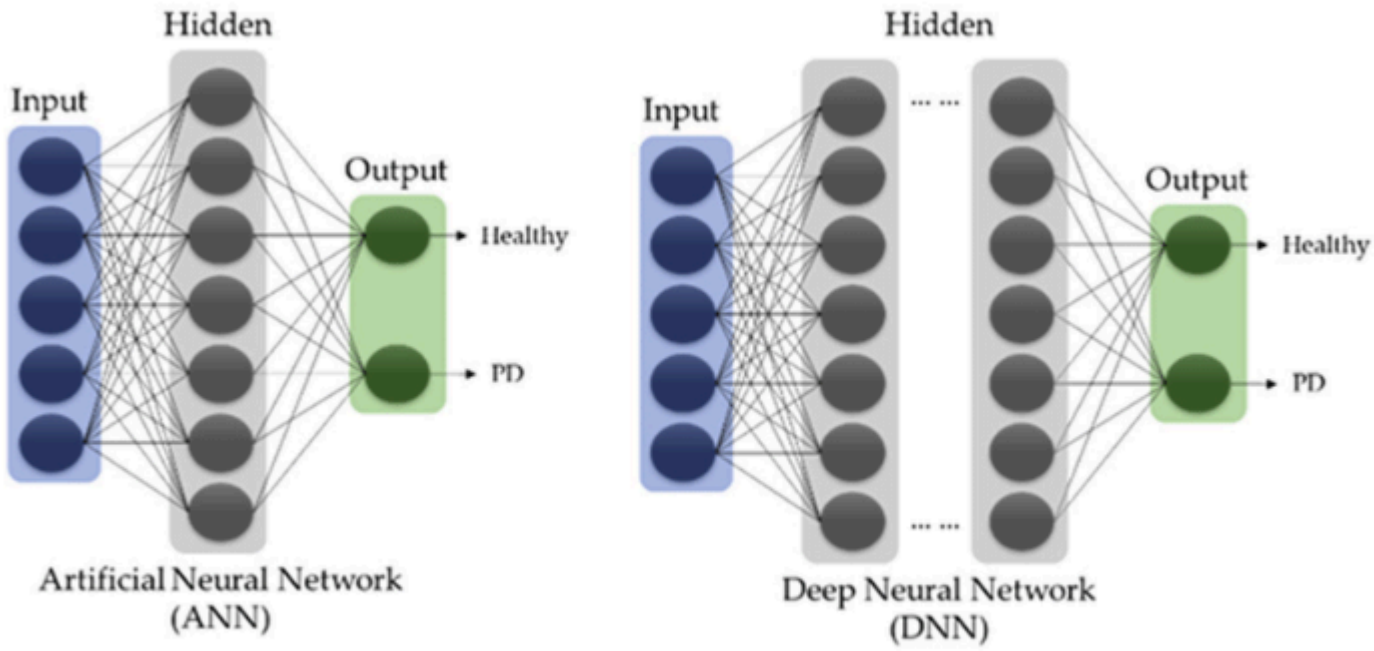
Machine Learning and AI



Source: Deep Learning vs. Machine Learning: A Beginner's Guide

Neural networks and Deep NN

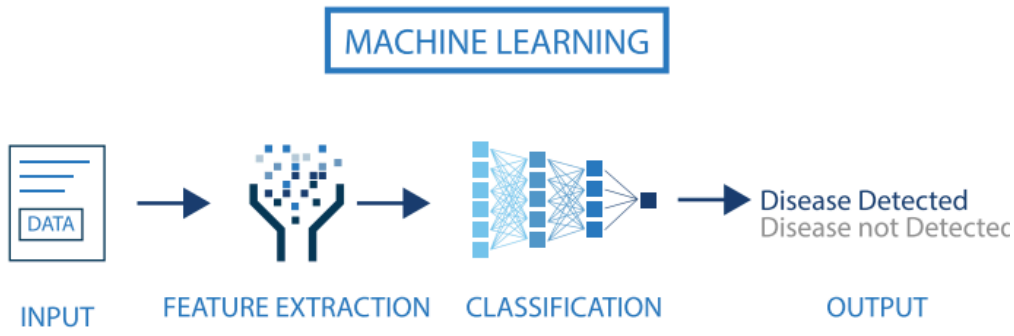
A big number of AI systems rely on *Deep Neural Networks*.

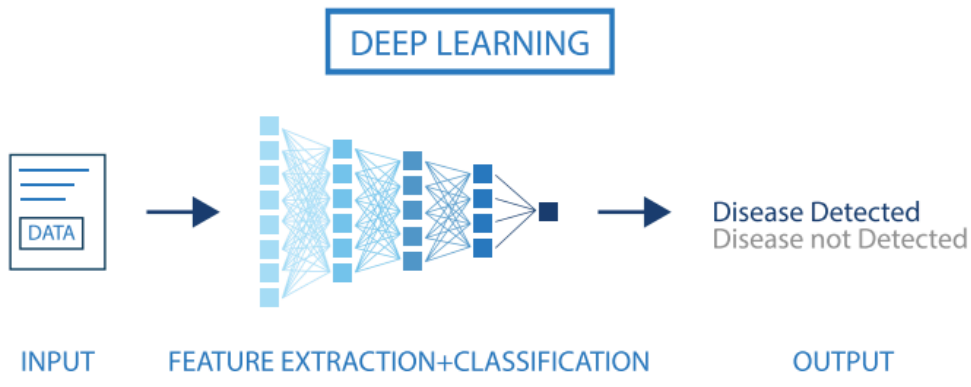


Basic architecture of ANN and DNN models.

Source: Application of Deep Learning Models for Automated Identification of Parkinson's Disease: A Review (2011-2021)

How does DL improve





- DL is a clear improvement of ML for many applications.
- Buut Deep learning requires large, well-annotated datasets, *which are not always available in healthcare.*

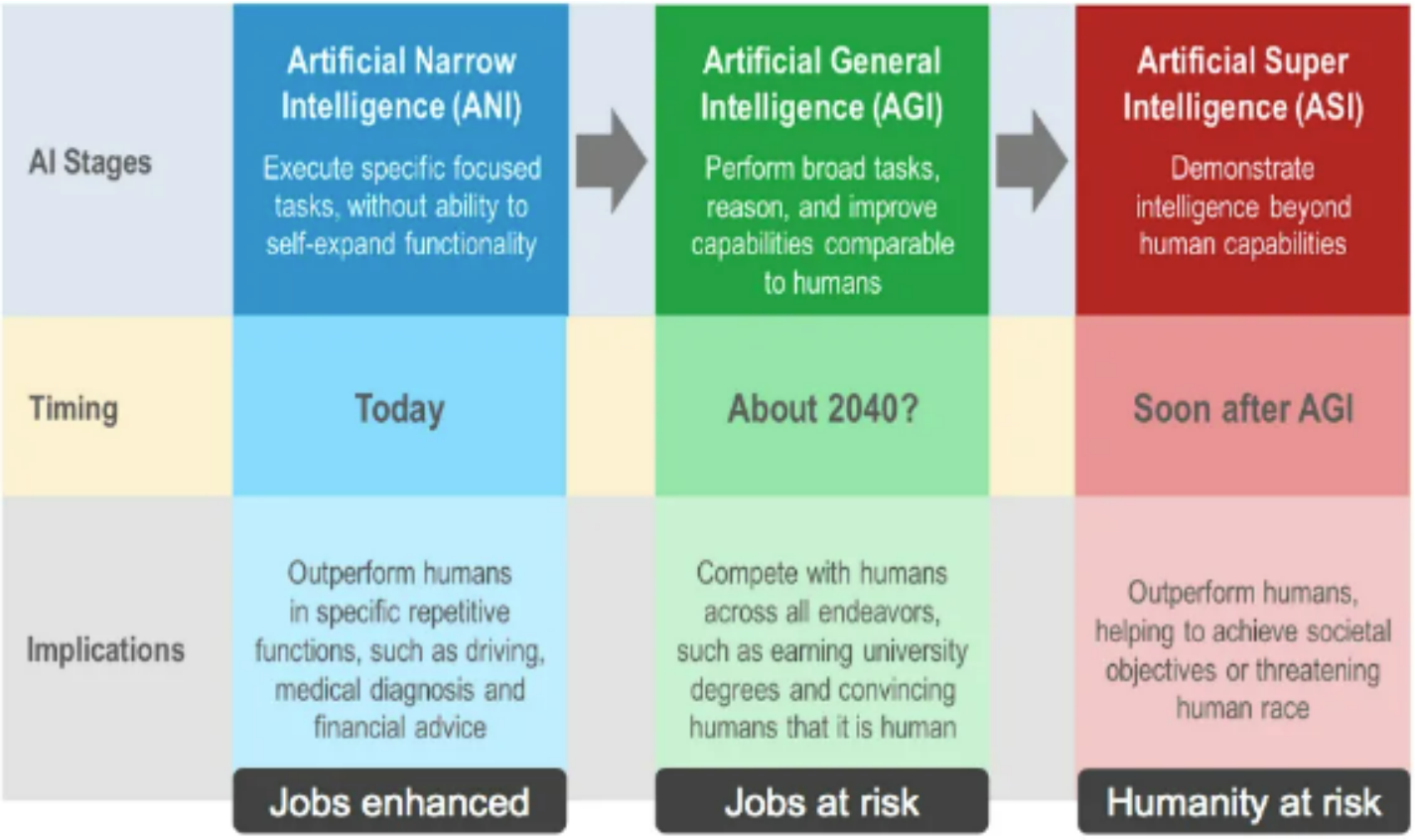
The impact of Deep learning

- Near-human-level image classification
- Near-human-level speech transcription
- Near-human-level handwriting transcription
- Dramatically improved machine translation
- Dramatically improved text-to-speech conversion
- Digital assistants such as Google Assistant and Amazon Alexa
- *Significant progress in autonomous driving, though far from fully generalizable*
- Improved ad targeting, as used by Google, Baidu, or Bing
- Improved search results on the web
- Ability to answer natural language questions

Where to now?

Artificial Intelligence stages

- AI can be expected to evolve through three stages.



source: AI for Legal: ANI, AGI and ASI

AI in summary

- Deep Learning, (big) data availability and increasing computing capacity will be at the basis of important changes.
- This exciting, but how concerned should we be?
- According to F. Chollet, the developer of Keras,
  - "*We shouldn't believe the short-term hype, but should believe in the long-term vision.*
  - *It may take a while for AI to be deployed to its true potential, the extent of which none has yet dared to dream*
  - *But AI is coming, and it'll transform our world in fantastic way*".

Generative AI: A New Paradigm

What is Generative AI?

- Models that *create* new content: text, images, clinical notes, molecules, synthetic data.
- Goes beyond prediction → **generation** of meaningful domain-specific outputs.
- Based on large-scale **foundation models** trained on multi-terabyte corpora.

Generative AI in Health

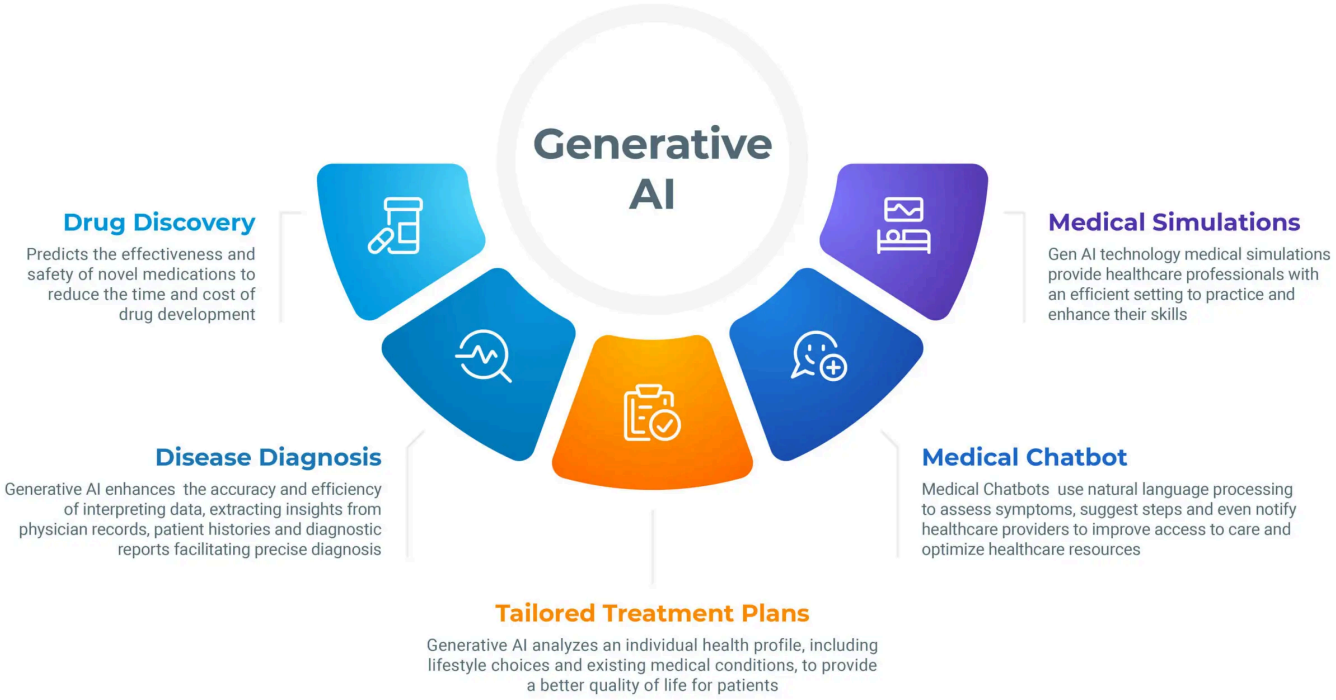
Why does it matter in health?

- Automates clinical documentation and reporting.
- Enhances medical imaging (reconstruction, denoising, anomaly detection).
- Accelerates drug discovery by generating new chemical structures.
- Improves patient communication and medical education.

Key risks

- Hallucinations, bias amplification, low explainability, regulatory constraints.

Generative AI use cases



Source: Generative AI in Healthcare: Use Cases, Benefits, and Challenges

Large Language Models (LLMs)

- Neural networks trained on trillions of words.
- Acquire **emergent abilities**: reasoning, summarization, translation, coding.
- Increasingly **multimodal**: integrate text + images + signals.

LLMs in medicine

Why are they important in medicine?

- Provide a flexible interface to medical knowledge.
- Adaptable to new tasks with minimal supervision.
- Support rapid assimilation of complex biomedical information.

Key message

> LLMs are not databases — they are probabilistic reasoning engines shaped by data.

LLM Applications in Healthcare

Clinical practice

- **Documentation:** automatic visit notes, discharge summaries, radiology reports
- **Decision support (non-critical):** guideline summarization, medication checks.
- ***Clinical triage:*** *symptom classification and urgency prediction.*
- **ICD/SNOMED coding** from free-text notes.

Biomedical research

- Literature summarization and question answering.
- Extraction of structured data from unstructured EHR.
- Hypothesis generation and experimental design assistance.

Limitations of LLMs

CHALLENGES AND LIMITATIONS OF LLMs IN HEALTHCARE

TECHNICAL CHALLENGES AND MODEL LIMITATIONS

>Hallucinations: generating plausible but factually incorrect information

>Limited contextual understanding and knowledge cutoffs
>Computational requirements and energy consumptions
>Latency of model responses, problematic in urgent cases

ETHICAL CONSIDERATIONS AND GOVERNANCE

>Patient privacy concerns and compliance with regulations

>Informed consent and transparency
> Accountability and responsibility for LLM deployment

BIAS, FAIRNESS, AND HEALTH EQUITY

> Risk of perpetuating biases from historical medical data
> Exacerbate health disparity

INTEGRATION WITH CLINICAL WORKFLOW

> Disruption to existing clinical routines and environments

> Interpretability with electronic health record systems

ChatGPT list of limitations (1)

1. Hallucinations

- LLMs may generate *confident but incorrect* statements.
- Especially dangerous in clinical reasoning, or advice.
- Hallucinations often look plausible and are hard to detect.

2. Factual inconsistency

- Lack of truthfulness: models do not "know", they **predict**.
- Can mix outdated and current evidence, or merge unrelated concepts.

3. Dataset shift & brittleness

- Performance drops when input data differ from training data
- Sensitive to small changes in wording, formatting, or context.

ChatGPT list of limitations (2)

4. Bias amplification

- If trained on biased data → may perpetuate or amplify inequalities (e.g., race, gender, socioeconomic factors).

5. Privacy & security concerns

- Risk of leaking sensitive information through model outputs.
- **EU AI Act**, where clinical LLMs = *high-risk* systems.

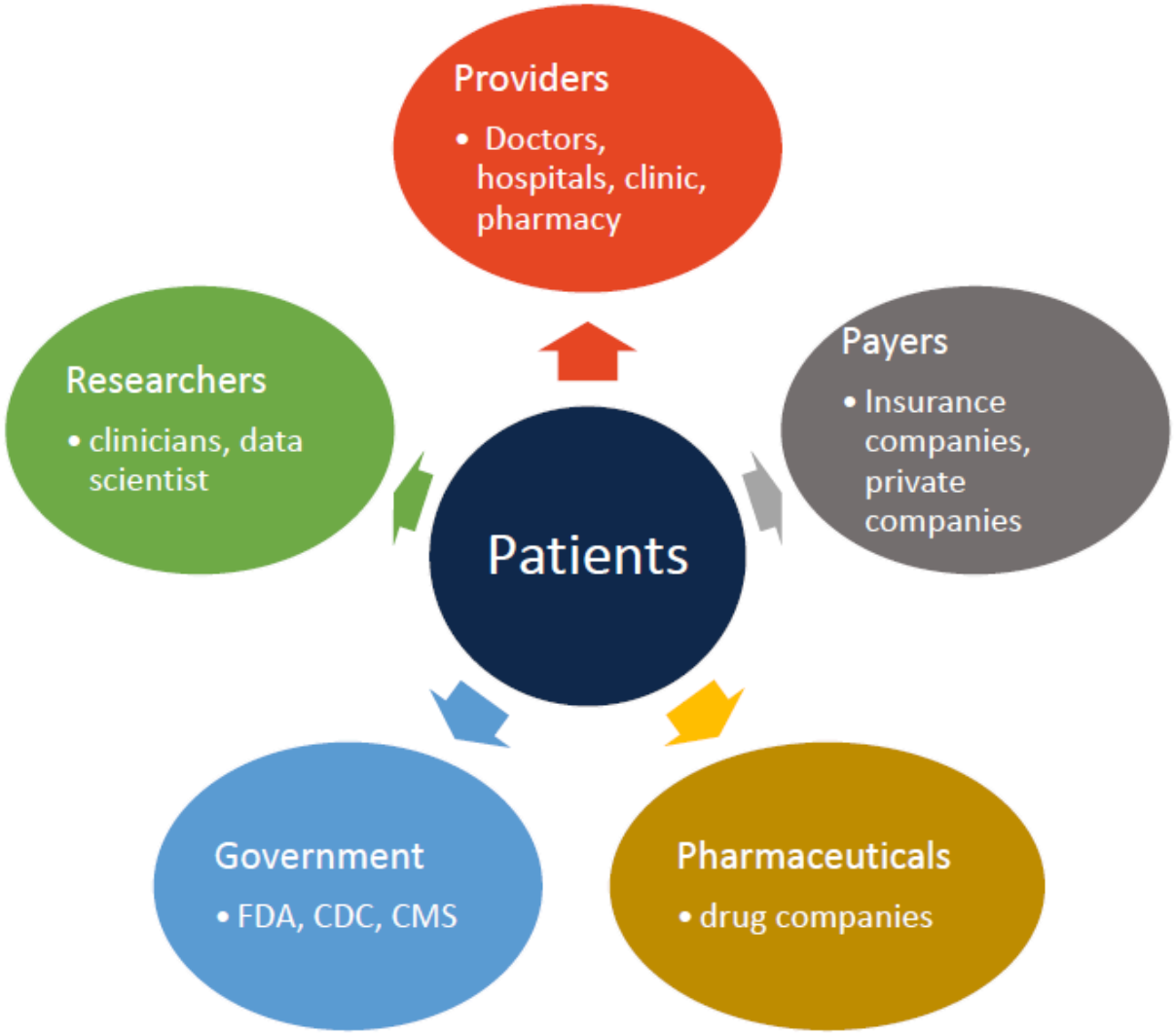
6. Lack of interpretability

- Explanations are often post-hoc and unreliable.
- Difficult to trace why a model produced a recommendation.

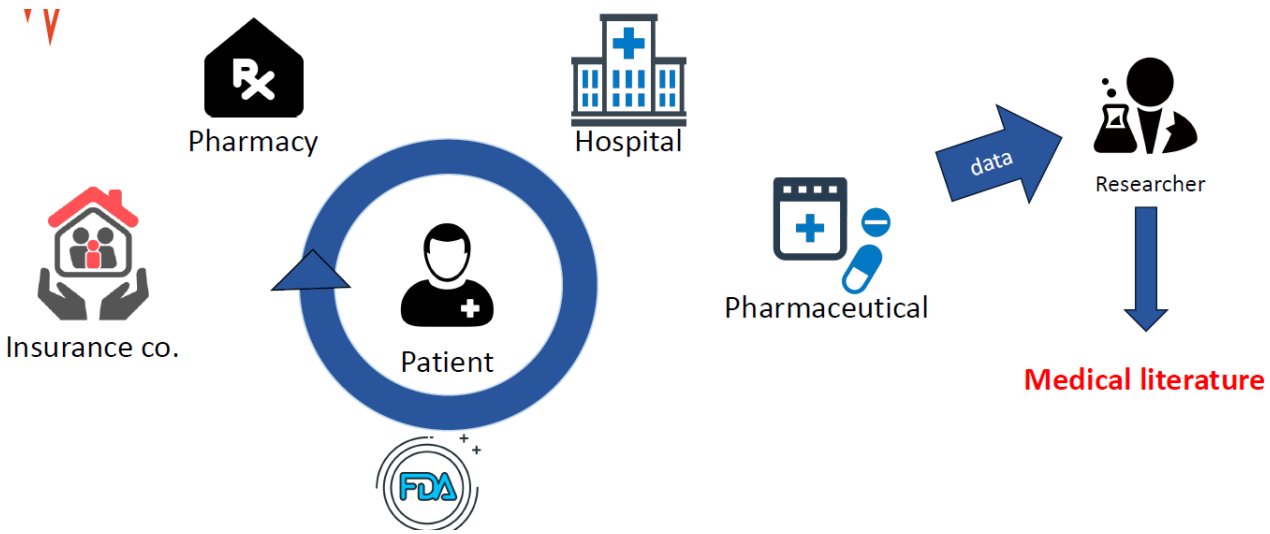
7. Regulatory & validation challenges

Data, Health Data and Big Data

The players of healthcare



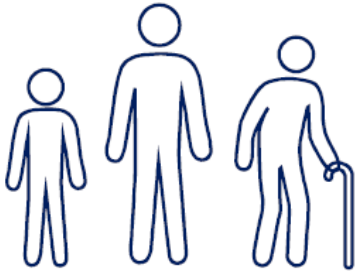
The health data life cycle



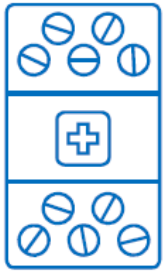
Health data types

- Electronic health records
- Medical claims
- Clinical notes
- Medical literature
- Continuous signals
- Imaging data
- Medical ontology
- Clinical trial data
- Drug discovery data

Electronic Health Records (EHR)



Demographics



Medications

Clinical Notes and Reports

Continuous Monitoring Data

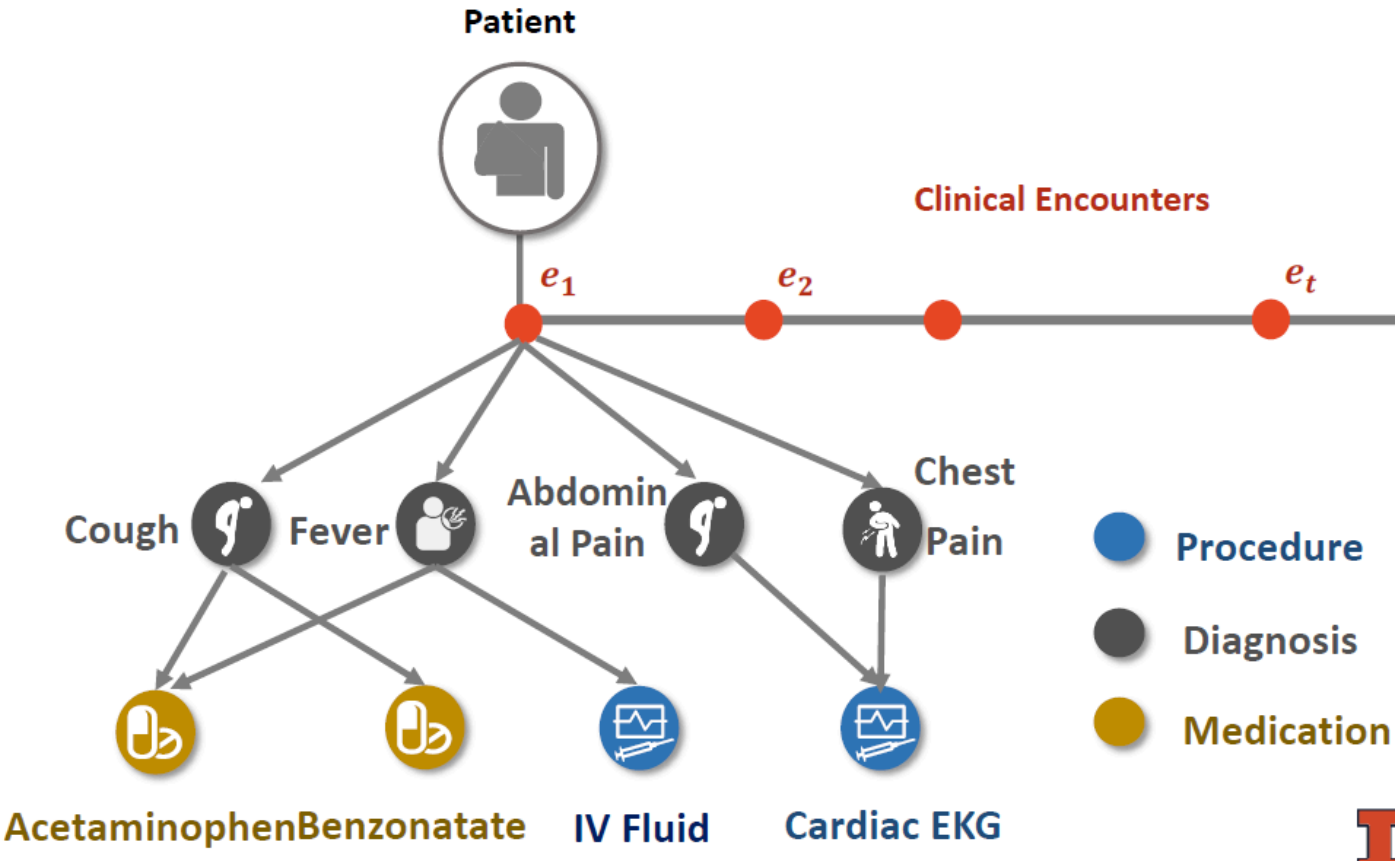
Multi-typed Medical Codes

Medical Images

The EHR workflow

- Outpatient visit/encounter
  - Symptom, diagnosis, lab tests, medication prescriptions
- Inpatient visit/encounter
  - Admission: chief complaints
  - Continuous monitoring data, imaging data
  - Discharge: discharge summary

Longitudinal EHR data



Properties of EHR

Pros

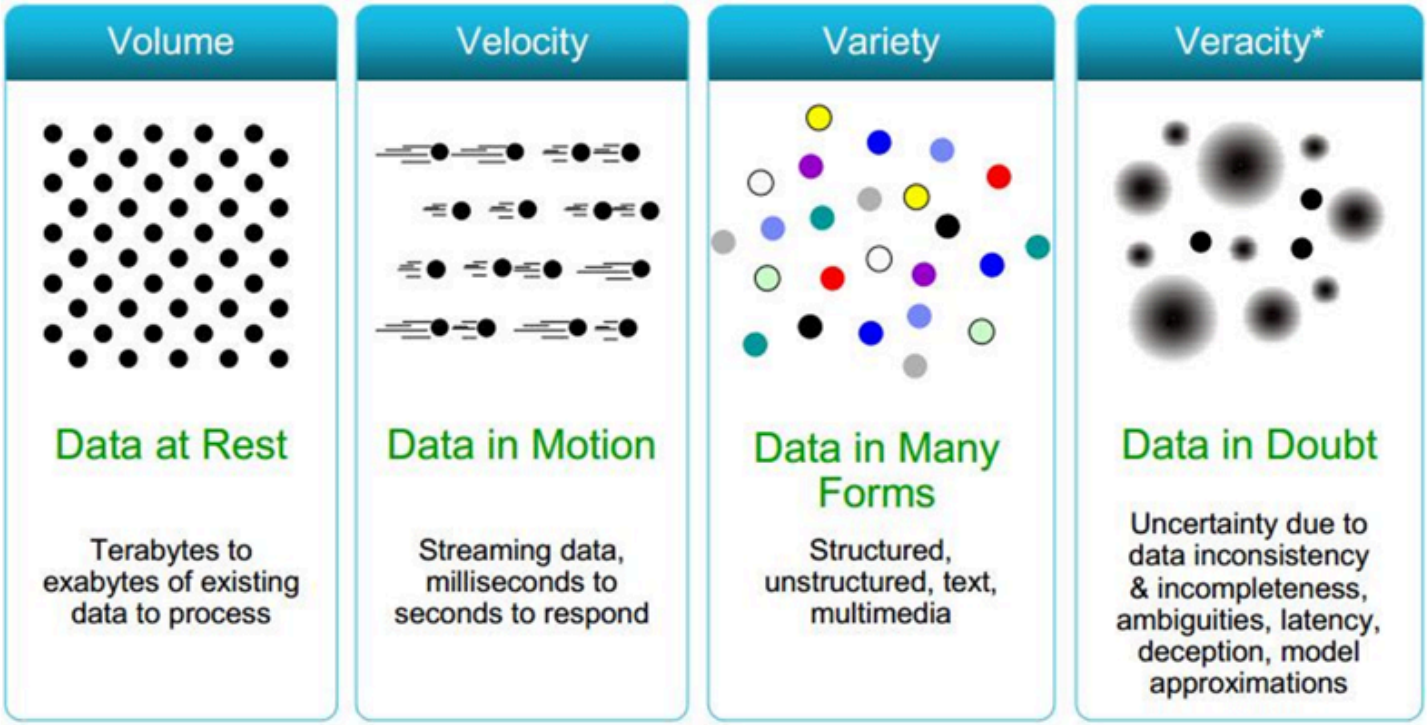
- **Rich data types**: structured + unstructured data
- **Longitudinal**: multiple visits over time

Cons

- **Complex**: multi-modality data
- **Siloed data**: each hospital has its own EHR
- **Privacy concern**: difficult to share/access

- Electronic Health Records are the main source of information in Health Care
- While potentially very usefultheir complexity limits rapid and efficient analysis.

The Vs that define Big Data



Health Care Data is Big

Volume

- Genomic data
- Medical imaging

Variety

- Electronic health records
- Medical knowledge base

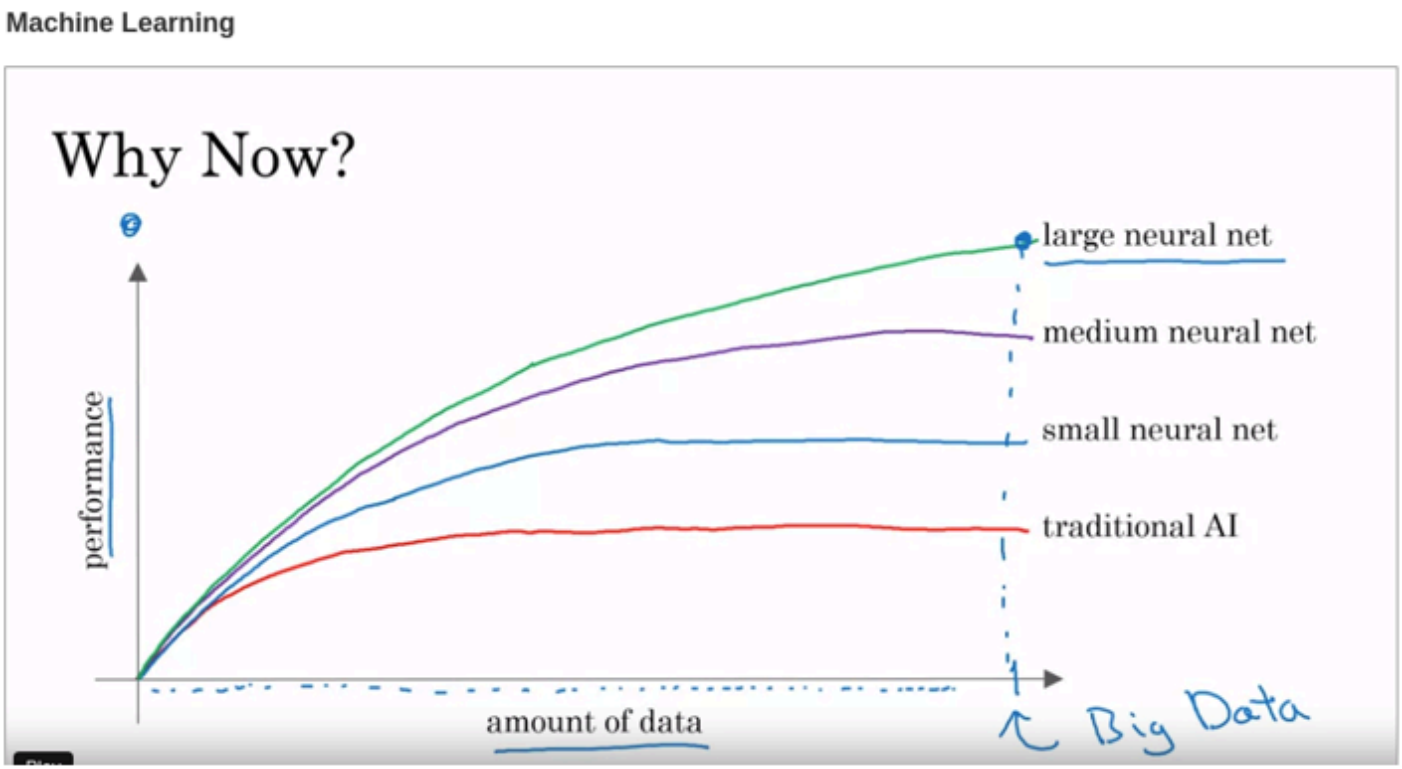
Velocity

- Real-time monitoring in ICU
- mHealth

Veracity

- Noises are everywhere
  - Missing data, errors in the data, false alarms

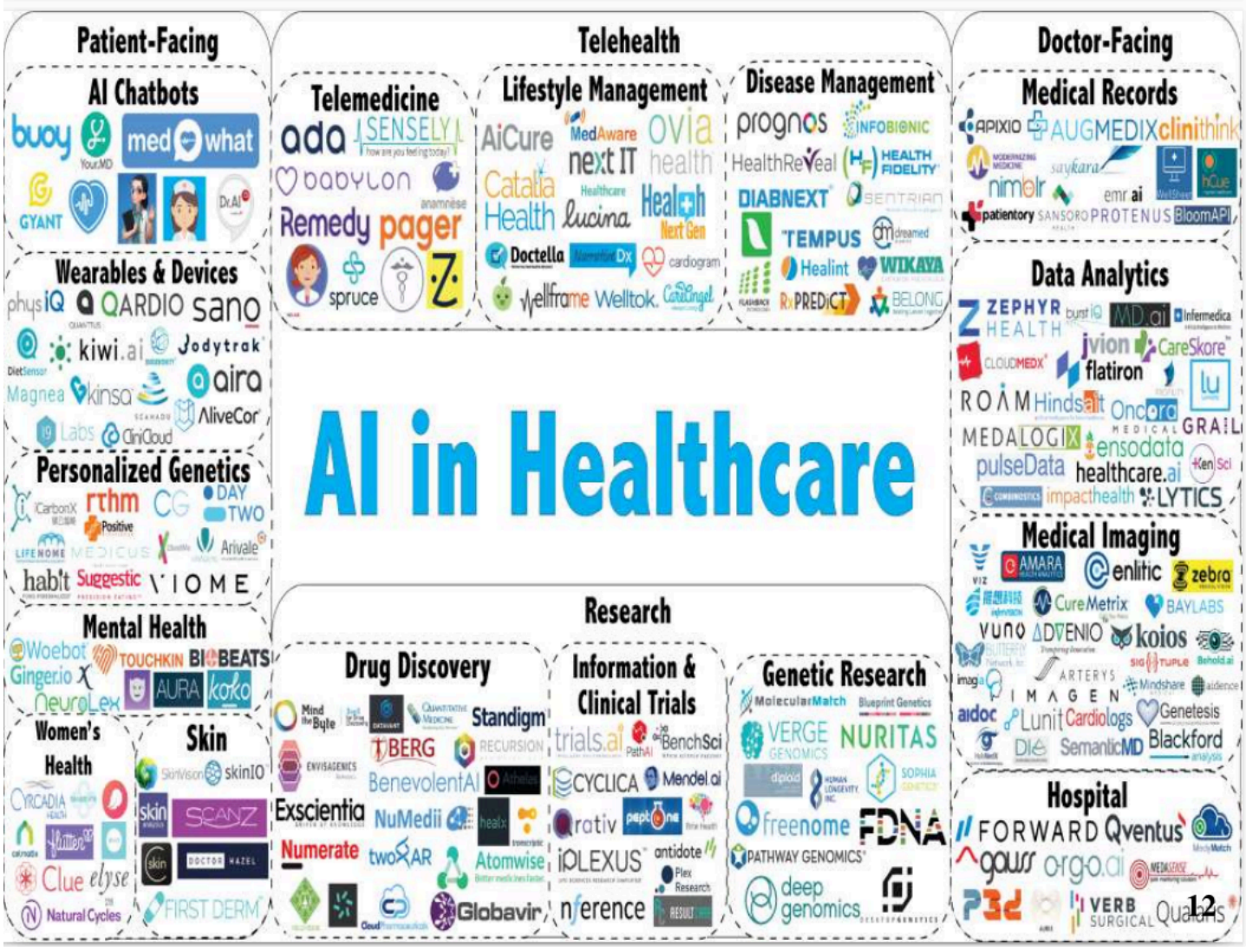
AI for Big Data (or Big Data 4 AI?)



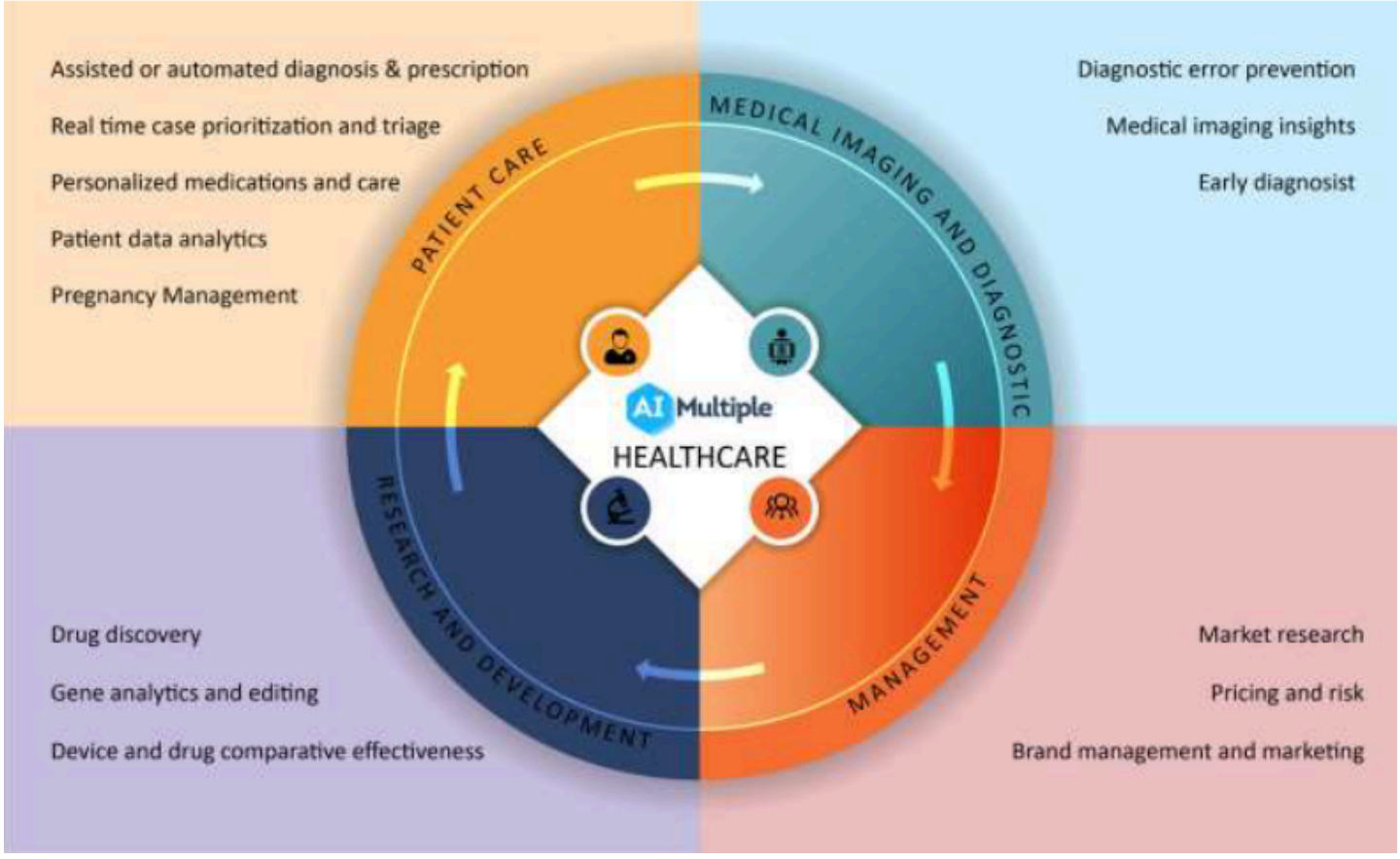
source: AI for everyone (Coursera)

AI in Health ain't future. It's today!

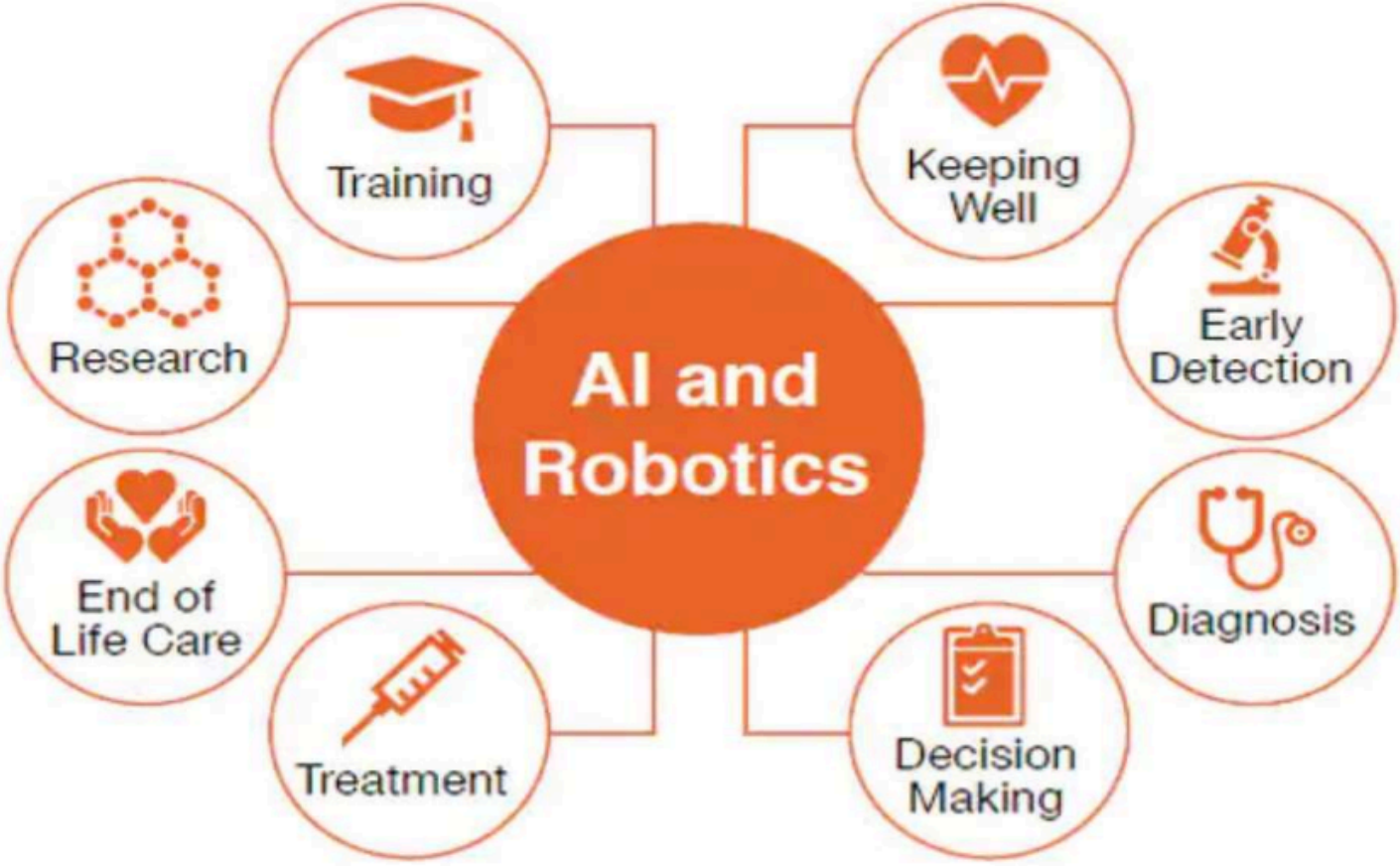
AI in Health care



AI in Health care



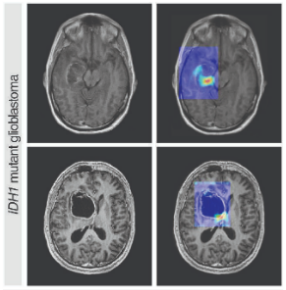
AI in Health ain't future. It's today!



AI and well being

- AI powered monitoring gadgets, such as bracelets, smart watches and many other have become common.
- Some, such as the Apple Watch allow for continuous heart-rate monitoring and are intended to detect potential CV incidents.
- Many potential benefits:
  - Motivation for exercise
  - Real time data for detection / prevention
- But not free from problems
  - Lack of standarization and credibility
  - Potential data protection issues when sharing data

AI in Health for early detection



A deep learning algorithm trained to analyze images from MRI scans predicts the presence of an *IDH1* gene mutation in brain tumors.
Credit: CA Cancer J Clin March/April 2019. doi: 10.3322/caac.21552. CC BY 4.0.

- AI, particularly through image analysis with deep learning, is being used to detect diseases, such as cancer, more accurately in early stages.
- Automatic detection can be faster and highly accurate, but *beware of biases and overfitting*
- It can reduce false positives and unnecessary expensive/dangerous tests s.a. biopsies.
- Not free from problems
  - Gender or race biases exist
  - Some tumours that would never progress are included causing unnecessary alert

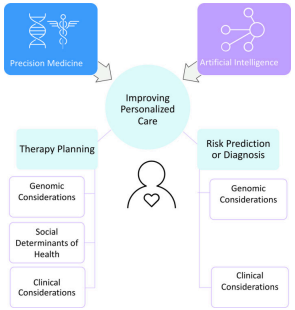
Can Artificial Intelligence Help See Cancer in New, and Better, Ways?

AI for Diagnosis and Prognosis

- Machine learning algorithms are being used in many fields for classification (diagnosis) and prediction (prognosis).
- Some may work like black-boxes (less intuitive than some statistical models),
- Under the appropriate circumstances, enough data, and *correctly used* may be very accurate.
- Typical scenario is analysis of medical imaging data or real time monitoring signals.

Dermatologist-level classification of skin cancer with deep neural networks

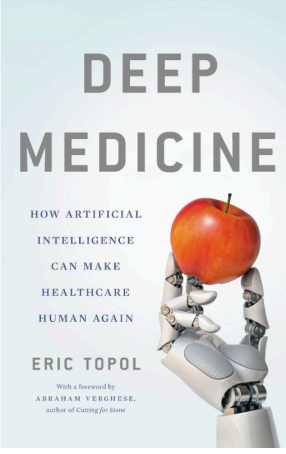
AI for Treatment design



Precision Medicine, AI, and the Future of Personalized Health Care

- AI empowered systems may be used to analyze EHR, Images and reports from a patient's history to help select the correct, individually customized treatment path.
- Data availability, combined with AI tools allows shifting from:
  - *Treatments for populations* where medical decisions are taken based on a few similar physical characteristics among patients, to
  - *Preventive, Personalized, Precision* medicine to provide the specific treatment for a specific patient

AI will (mostly) help, not replace



- A common concern is *Will AI replace humans?*
- AI advocates don't think so, instead they talk of *complement*, *enhance*, support...
- This concern also exists in medicine
  - Some AI experts claim that in 3-5 years radiologist will not be necessary anymore.
- Dr. Eric Topol in his book Deep Medicine:
  - Introducing AI in Healthcare is good
  - It lets the machine do the machine work
    - Fill forms, Scan images
  - With more efficiency and new workflows.
  - And the doctor has more time for patients.

What to say about Ethics?

Ethics, Data, Risks and Biases

When one considers possible issues related with AI use, distinct aspects must be addressed.

- **Ethical Considerations:** AI systems need to be made transparent, accountable, fair, and require informed consent to align with moral principles and prevent harm.
- **Data related:** User privacy must be safeguarded, data must be secured and data management and processing must comply with privacy regulations in AI.
- **AI Risks:** Risks must be anticipated so mitigating potential adverse outcomes, such as unintended consequences, job displacement, and security vulnerabilities.
- **AI Biases:** Unfair or discriminatory patterns in AI decision-making, arising from biased training data or algorithms must be addressed for

Ethical considerations

1. Transparency:

- Algorithms & decision-making processes should be transparent to users.
- Clear explanations on how AI systems arrive at specific outcomes should be available to anyone involved with the outcome.

2. Accountability:

- Responsible parties for AI outcomes and decisions should be identifiable.
- Mechanisms for addressing errors or unintended consequences needed.

3. Privacy:

- User data should be safeguarded with privacy regulations.
- Privacy-preserving techniques, e.g. *federated learning*, should be implemented

4. Informed Consent:

- Users must be informed about how their data will be used by AI systems.
- Explicit consent must be obtained for data collection and processing.

Data Protection in AI

1. **Data Collection:**
   - Fair sourcing of training data: avoid biased/discriminatory datasets.
   - Minimize the collection of unnecessary personal information.
2. **Data Security:**
   - Implement robust security measures to protect models and training data.
   - Prevent unauthorized access and data breaches.
3. **Data Ownership:**
   - Clarify ownership rights/responsibilities regarding the data used.
   - Respect intellectual property and data licensing agreements.
4. **Data Retention:**
   - Establish appropriate time frames for retaining user data.
   - Delete/anonymize data no longer necessary for AI model training.

The risks of AI

1. **Unintended Consequences:**
   - Anticipate and mitigate potential negative impacts of AI on society.
   - Prepare for unexpected outcomes and unintended uses of AI.
2. **Job Displacement:**
   - Address the potential impact of AI on employment.
   - Implement strategies to retrain/upskill affected workers.
3. **Security Risks:**
   - Guard against malicious uses of AI for cyber attacks.
   - Ensure robustness and resilience of AI systems to adversarial attacks.

AI Biases

1. **Training Data Bias:**
   - Identify and address biases present in training datasets.
   - Implement techniques like data augmentation to enhance diversity.
2. **Algorithmic Bias:**
   - Regularly audit and evaluate AI algorithms for fairness.
   - Adjust algorithms to minimize disparate impacts on different demographic groups.
3. **Explainability:**
   - Ensure that AI models are interpretable and that decision-making processes are explainable.
   - Facilitate external scrutiny to identify and rectify biases.

In conclusion ...

AI in healthcare holds immense promise,

But responsible and ethical deployment is crucial

- to maximizing benefits
- while mitigating risks
- and ensuring equitable access.

References and Resources

References and Resources

- Books and Reports
  - *Deep Medicine*. Eric Topol
  - *Machine Learning and AI for Healthcare* Arjun Panesar
  - *Artificial intelligence in healthcare Applications, risks, and ethical and societal impacts* European Parliament.
  - *Inteligencia Artificial en la medicina del 3er milenio* Lopez Farré et alt.
- Courses
  - AI for everyone
  - Introduction to AI