Beyond Images: How Multimodal AI is Transforming Healthcare

Sara El-Ateif, 30-31 Oct 2024

IeHEALTHF 2024

About Me

Sara El-Ateif

Dr. in Computer Science
Google Ph.D. Fellow
Google Developer Expert in ML

- Certified Business Coach
- NVIDIA DLI Instructor/University Ambassador
- Founder of Anajia
- Co-Founder of AI Wonder Girls

Agenda

- Introduction
- Applications of Multimodal AI in Healthcare
- Methods and Techniques of Multimodal AI
- Generative AI in Multimodality
- Ethical and Regulatory Considerations
- Conclusion and Future Directions

Introduction

Introduction

What is Multimodal AI?

A type of input or output, such as video, image, audio, text, etc [1].

[1] Kiros, Ryan; Salakhutdinov, Ruslan; Zemel, Rich (2014-06-18). "Multimodal Neural Language Models". Proceedings of the 31st International Conference on Machine Learning. PMLR: 595–603.
[2] Benzebouchi NE, Azizi N, Ashour AS, et al (2019) Multi-modal classifier fusion with feature cooperation for glaucoma diagnosis. J Exp Theor Artif Intell 31:841–874. https://doi.org/10.1080/0952813X.2019.1653383

Introduction

What is Multimodal AI?

Multimodality refers to the principle of gathering several complementary* modalities and joining them into one to provide a complete view of a subject or disease. We can refer to this process as diversity as well [2].

*Complementarity means that a modality can provide a certain type of information that cannot be deduced from another modality.

[1] Kiros, Ryan; Salakhutdinov, Ruslan; Zemel, Rich (2014-06-18). "Multimodal Neural Language Models". Proceedings of the 31st International Conference on Machine Learning. PMLR: 595–603.
[2] Benzebouchi NE, Azizi N, Ashour AS, et al (2019) Multi-modal classifier fusion with feature cooperation for glaucoma diagnosis. J Exp Theor Artif Intell 31:841–874. https://doi.org/10.1080/0952813X.2019.1653383

Introduction

What is Multimodal AI?

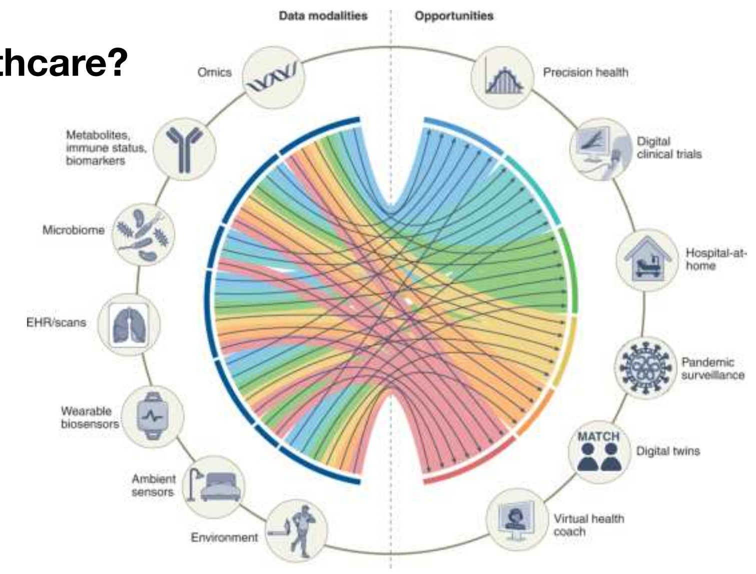
AI systems that process and integrate information from multiple data sources (modalities) like text, images, videos, and sensor data.

[1] Kiros, Ryan; Salakhutdinov, Ruslan; Zemel, Rich (2014-06-18). "Multimodal Neural Language Models". Proceedings of the 31st International Conference on Machine Learning. PMLR: 595–603.
[2] Benzebouchi NE, Azizi N, Ashour AS, et al (2019) Multi-modal classifier fusion with feature cooperation for glaucoma diagnosis. J Exp Theor Artif Intell 31:841–874. https://doi.org/10.1080/0952813X.2019.1653383

Introduction
Why is it Essential in Healthcare?

Healthcare data is inherently multimodal (e.g., patient history, imaging, genomic data).

Multimodal AI unlocks insights hidden within individual data silos, offering a holistic patient view.



From: Multimodal biomedical AI

Applications of Multimodal AI in Healthcare

Applications of Multimodal AI in Healthcare

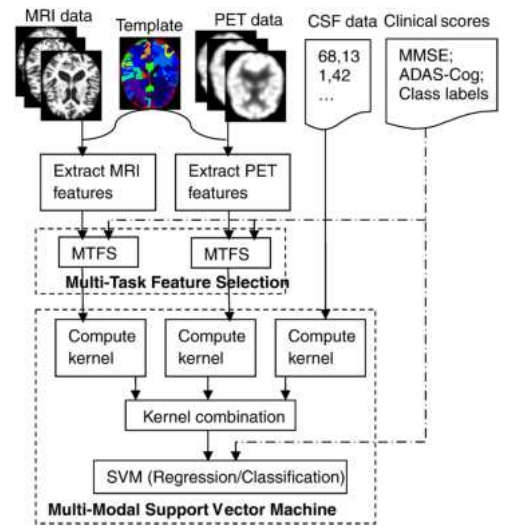
Disease Diagnosis

Combining imaging data with genetic information and patient history for earlier and more accurate disease detection.

Applications of Multimodal AI in Healthcare

Disease Diagnosis

Zhang et al. [3] conducted a study where they integrated structural MRI, PET, and cognitive tests to develop a multivariate classification system for predicting the diagnosis and progression of Alzheimer's disease.

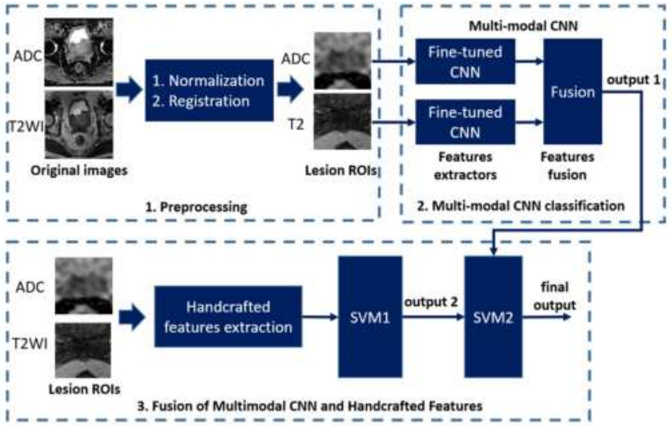


[3] Zhang, Daoqiang, Dinggang Shen, and Alzheimer's Disease Neuroimaging Initiative. "Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease." NeuroImage 59.2 (2012): 895-907.

Applications of Multimodal AI in Healthcare

Disease Diagnosis

Le et al. [4] proposed a multimodal CNN model to diagnose prostate cancer in multi-parametric MRI, using apparent diffusion coefficient and T2-weighted images.



[4] Le, Minh Hung, et al. "Automated diagnosis of prostate cancer in multi-parametric MRI based on multimodal convolutional neural networks." Physics in Medicine & Biology 62.16 (2017): 6497.

Applications of Multimodal AI in Healthcare

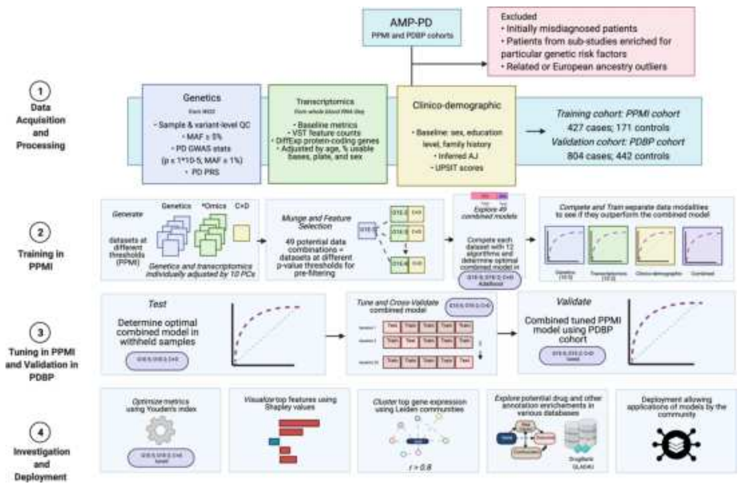
Risk Prediction

Identifying individuals at high risk of developing certain diseases based on lifestyle factors, genetic predispositions, and medical history.

Applications of Multimodal AI in Healthcare

Risk Prediction

Makarious et al. [5] developed a machine learning model that integrates genomics, transcriptomics, and clinical data to make improved predictions of Parkinson's disease risk, which were validated in an external cohort.



From: Multi-modality machine learning predicting Parkinson's disease

[5] M.B. Makarious, et al., Multi-modality machine learning predicting Parkinson's disease, NPJ Parkinsons Dis. 8 (1) (2022), https://doi.org/10.1038/s41531-022- 00288-w. Dec.

Applications of Multimodal AI in Healthcare

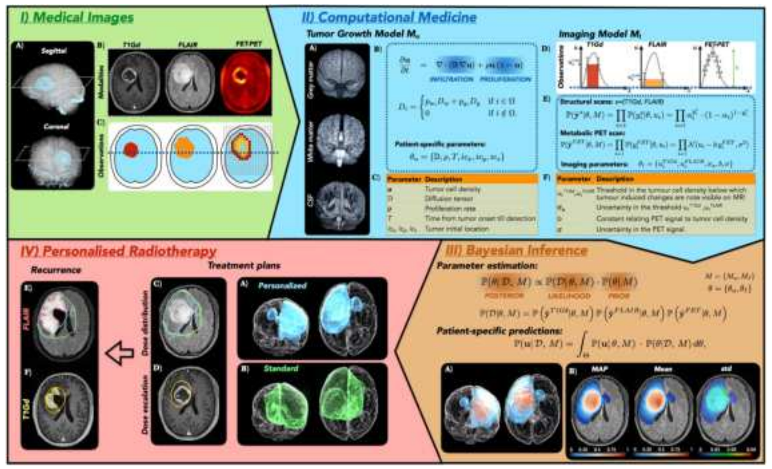
Treatment Planning and Personalization

Tailoring treatment plans to individual patients by integrating information from imaging, genomic data, and treatment response history.

Applications of Multimodal AI in Healthcare

Treatment Planning and Personalization

Lipková et al. [6] developed a Bayesian machine learning framework to personalize radiotherapy plans for glioblastoma patients by integrating patient-specific multimodal MRI and FET-PET scans with a mathematical tumor growth model to predict tumor cell density beyond visible lesions.



[6] Lipková, Jana, et al. "Personalized radiotherapy design for glioblastoma: integrating mathematical tumor models, multimodal scans, and Bayesian inference." IEEE transactions on medical imaging 38.8 (2019): 1875-1884.

Applications of Multimodal AI in Healthcare

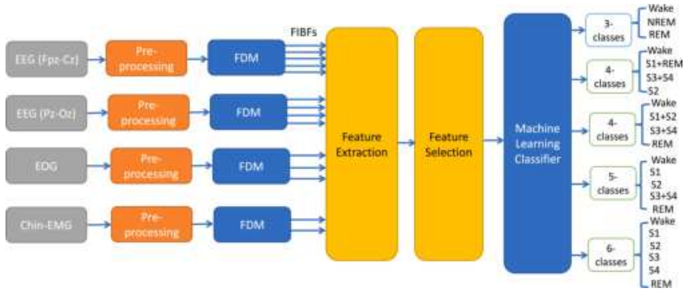
Patient Monitoring and Follow-up

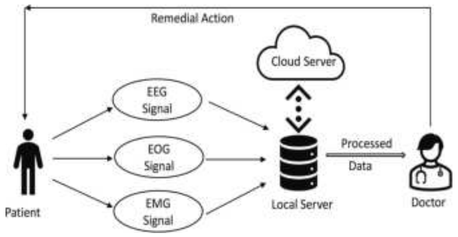
Monitoring patient recovery and treatment response using data from wearable sensors, electronic health records, and patient-reported outcomes.

Applications of Multimodal AI in Healthcare

Treatment Planning and Personalization

Fatimah et al. [7] utilized electroencephalogram (EEG), electromyogram (EMG), and electrooculogram (EOG) data to improve the classification of sleep stages. Their method highlighted the potential for low-cost sensor-based setups for continuous patient monitoring and feedback.





[7] Fatimah, Binish, Amit Singhal, and Pushpendra Singh. "A multi-modal assessment of sleep stages using adaptive Fourier decomposition and machine learning." Computers in Biology and Medicine 148 (2022): 105877.

Methods and Techniques of Multimodal AI

Multimodal Fusion

Multimodal Fusion
Intro

Early Fusion

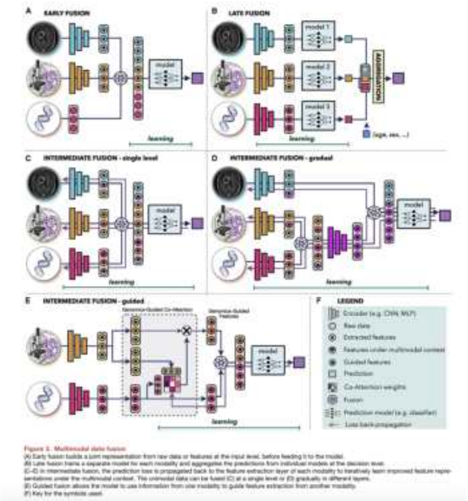
Combines data at the input level

Joint Fusion

Processes all modalities simultaneously.

Late Fusion

Combines outputs from single-modal models.



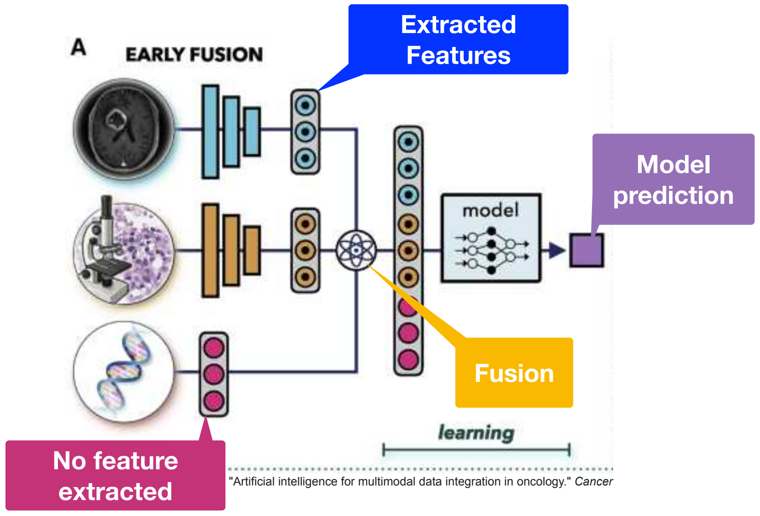
Source: Lipkova, Jana, et al. "Artificial intelligence for multimodal data integration in oncology." Cancer cell 40.10 (2022): 1095-1110.

Multimodal Fusion

Early Fusion

Figure 3. Multimodal data fusion

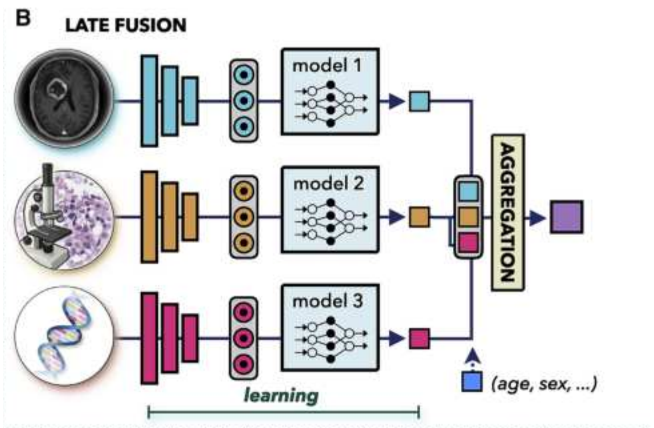
(A) Early fusion builds a joint representation from raw data or features at the input level, before feeding it to the model.



Multimodal Fusion
Late Fusion

Figure 3. Multimodal data fusion

(B) Late fusion trains a separate model for each modality and aggregates the predictions from individual models at the decision level.



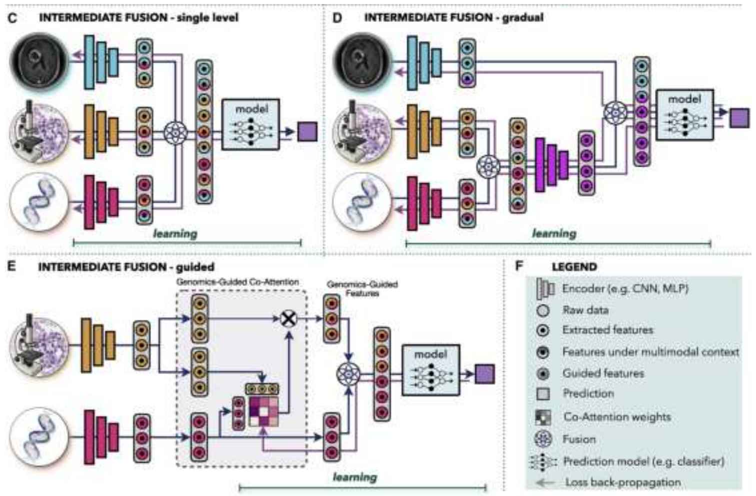
Source: Lipkova, Jana, et al. "Artificial intelligence for multimodal data integration in oncology." *Cancer cell* 40.10 (2022): 1095-1110.

Multimodal Fusion

Intermediate Fusion

Figure 3. Multimodal data fusion

(C–E) In intermediate fusion, the prediction loss is propagated back to the feature extraction layer of each modality to iteratively learn improved feature representations under the multimodal context.

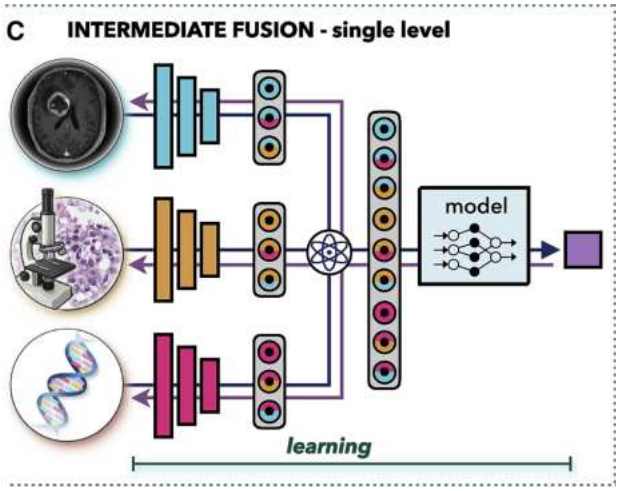


Source: Lipkova, Jana, et al. "Artificial intelligence for multimodal data integration in oncology." *Cancer cell* 40.10 (2022): 1095-1110.

Multimodal Fusion
Intermediate Fusion

Figure 3. Multimodal data fusion

The unimodal data can be fused (C) at a ***single level*** or (D) gradually in different layers.

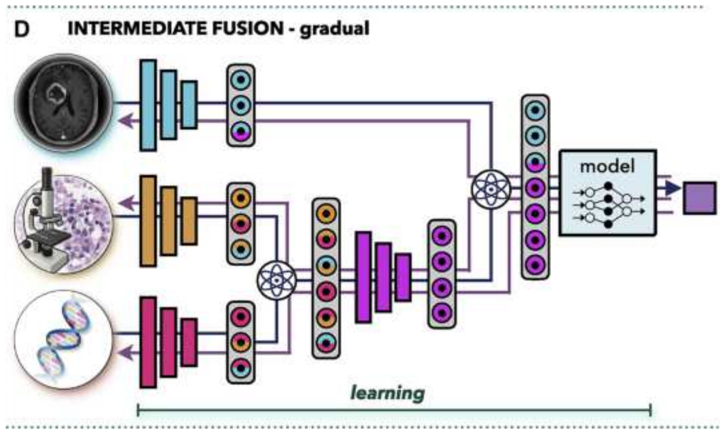


Source: Lipkova, Jana, et al. "Artificial intelligence for multimodal data integration in oncology." *Cancer cell* 40.10 (2022): 1095-1110.

Multimodal Fusion
Intermediate Fusion

Figure 3. Multimodal data fusion

The unimodal data can be fused (C) at a single level or (D) ***gradually*** in different layers.



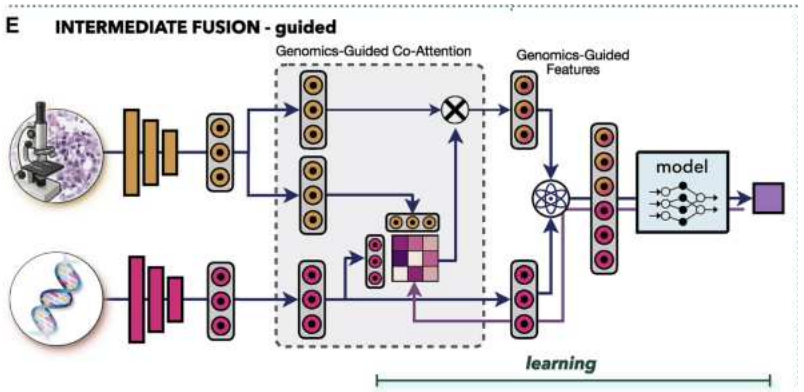
Source: Lipkova, Jana, et al. "Artificial intelligence for multimodal data integration in oncology." *Cancer cell* 40.10 (2022): 1095-1110.

Multimodal Fusion

Intermediate Fusion

Figure 3. Multimodal data fusion

(E) Guided fusion allows the model to use information from one modality to guide feature extraction from another modality.



Source: Lipkova, Jana, et al. "Artificial intelligence for multimodal data integration in oncology." *Cancer cell* 40.10 (2022): 1095-1110.

Challenges

Challenges of Multimodal AI
Intro

Challenges related to:

1. **Combining data** with different scales, resolutions, and noise levels
2. **Missing Data:** In real-world healthcare settings, it's common to have incomplete datasets with missing modalities for some patients.

Challenges of Multimodal AI

Combining data

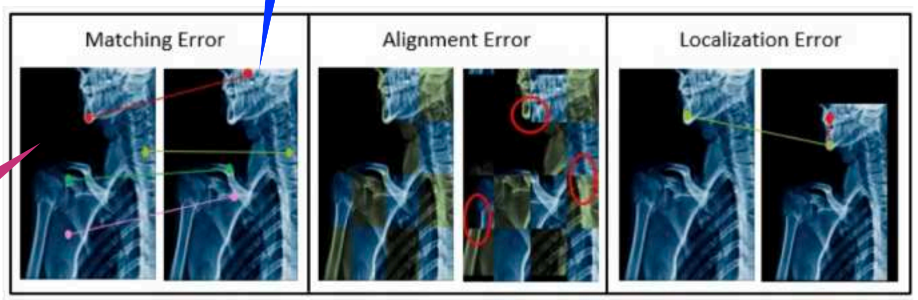
Scale and Resolution Variations	Noise and Artifacts	Computational Complexity
Misalignment & Registration Errors	Variable Noise Characteristics	High Data Dimensionality
Information Loss during Resampling	Artifact Interference	Algorithm Development
Feature Inconsistency		

Challenges of Multimodal AI

Scale and Resolution Variations

Misalignment & Registration Errors

Different modalities often have vary izes, making direct overlay and comparison difficult. Misalignment can lead to inaccurate inferences and treatment planning.



Source: Darzi, Fatemehzahra, and Thomas Bocklitz. "A Review of Medical Image Registration for Different Modalities." Bioengineering 11.8 (2024): 786.

Methods and Techniques of Multimodal AI

Scale and Resolution Variations

Misalignment & Registration Errors

Different modalities often have varying pixel/voxel sizes, making direct overlay and comparison difficult. Misalignment can lead to inaccurate inferences and treatment planning.

Information Loss during Resampling

Resampling data to a common scale can lead to information loss, particularly when downsampling from higher resolution (e.g., MRI) to lower resolution (e.g., PET).

Feature Inconsistency

Features extracted at different scales might not correspond accurately, making it difficult to establish meaningful relationships across modalities.

Methods and Techniques of Multimodal AI

Scale and Resolution Variations

Misalignment & Registration Errors

Different modalities often have varying pixel/voxel sizes, making direct overlay and comparison difficult. Misalignment can lead to inaccurate inferences and treatment planning.

Information Loss during Resampling

Resampling data to a common scale can lead to information loss, particularly when downsampling from higher resolution (e.g., MRI) to lower resolution (e.g., PET).

Feature Inconsistency

Features extracted at different scales might not correspond accurately, making it difficult to establish meaningful relationships across modalities.

Challenges of Multimodal AI

Noise and Artifacts

Variable Noise Characteristics

Each modality has its own noise profile (e.g., Gaussian noise in MRI, Poisson noise in PET). This complicates joint analysis as noise from one modality can be misinterpreted as signal in another.

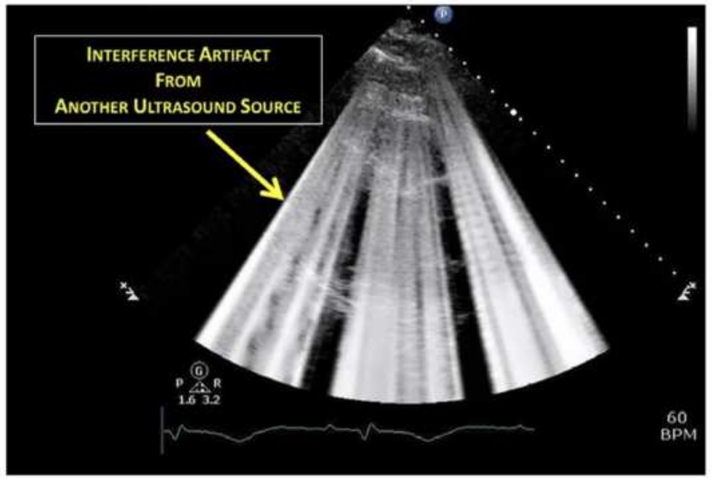
Artifact Interference

Artifacts specific to each modality (e.g., motion artifacts in MRI, attenuation artifacts in PET) can further complicate data interpretation and the ability to distinguish true signal from noise.

Challenges of Multimodal AI
Noise and Artifacts

Artifact Interference

Artifacts specific to each modality (e.g., motion artifacts in MRI, attenuation artifacts in PET) can further complicate data interpretation and the ability to distinguish true signal from noise.



Source: https://x.com/MDBeni/status/1425330258857893895/photo/1

Challenges of Multimodal AI

Computational Complexity

High Data Dimensionality

Multimodal datasets are inherently high-dimensional, increasing memory requirements and computational time for analysis.

Algorithm Development

Developing robust algorithms that can handle the complexities of multimodal data while efficiently extracting relevant information is challenging.

Challenges of Multimodal AI

Strategies to Address These Challenges

Advanced Registration Techniques: Employ robust deformable registration methods to align multimodal images accurately.

Multi-Resolution Analysis: Develop methods that can analyze data at multiple resolutions, leveraging information from each scale.

Noise Reduction and Artifact Correction: Apply appropriate denoising and artifact correction techniques tailored to each modality.

Feature Fusion Strategies: Explore different feature fusion approaches (early, late, or hybrid fusion) to effectively combine information from different modalities.

Deep Learning Methods: Utilize deep learning models that can learn complex relationships and handle multimodal data effectively.

Challenges of Multimodal AI
Missing Modality

Techniques:

- **Imputation:** Estimate missing data based on available information.
- **Late Fusion Architectures:** Design models that can handle varying input combinations, allowing predictions even with missing modalities.
- **Robust Training Procedures:** Develop training strategies that are less sensitive to missing data, such as using loss functions that account for partial input.


Deep Learning

Deep Learning for Multimodal AI
Intro

Different neural network architectures can be used for specific modalities and integration tasks.

Here's how different neural network architectures can be tailored for specific modalities and their integration:

Deep Learning for Multimodal AI

Modality-Specific Architectures

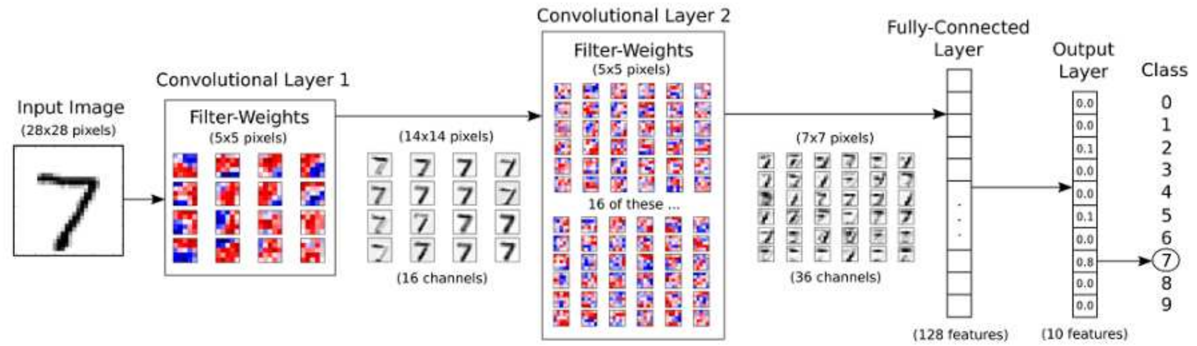
Convolutional Neural Networks (CNNs) for Images

CNNs excel at capturing spatial hierarchies and features in medical images due to their convolutional and pooling layers.

Applications in Multimodal Medical AI:

- **Tumor Segmentation:** Identifying tumor boundaries in MRI or CT scans.
- **Disease Classification:** Classifying diseases from X-ray images, like pneumonia or tuberculosis.
- **Fracture Detection:** Automatically detecting bone fractures in X-ray images.
- **Anatomical Landmark Detection:** Locating specific points in medical images for surgery planning or image registration.

Deep Learning for Multimodal AI
CNNs



Source: https://tensorflownet.readthedocs.io/en/latest/ConvolutionNeuralNetwork.html

Deep Learning for Multimodal AI

Modality-Specific Architectures

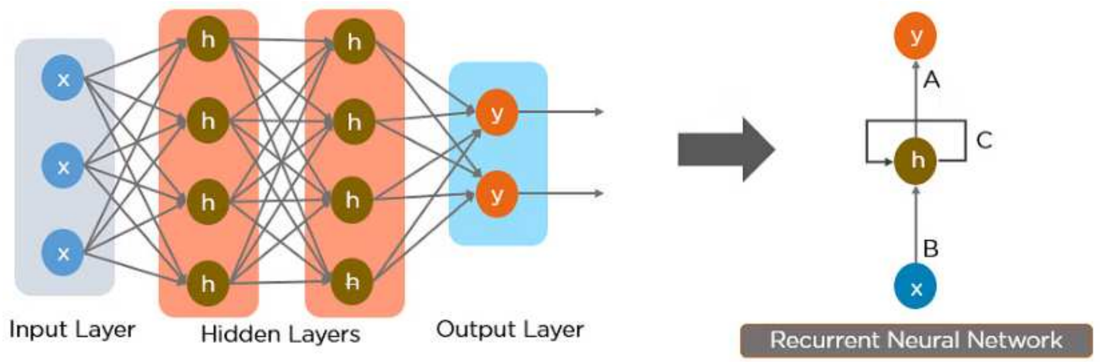
Recurrent Neural Networks (RNNs) for Sequential Data

RNNs are well-suited to process time-series data common in healthcare.

Applications in Multimodal Medical AI:

- **Patient Monitoring:** Predicting patient deterioration or risk of complications using data from electronic health records (EHRs) and vital signs.
- **Disease Progression Modeling:** Tracking the progression of diseases like Alzheimer's or Parkinson's using longitudinal imaging data and clinical notes.
- **Genomic Sequence Analysis:** Analyzing DNA and RNA sequences to identify genetic mutations associated with diseases.
- **EEG/ECG Analysis:** Processing electroencephalogram (EEG) or electrocardiogram (ECG) signals to detect abnormalities and diagnose conditions.

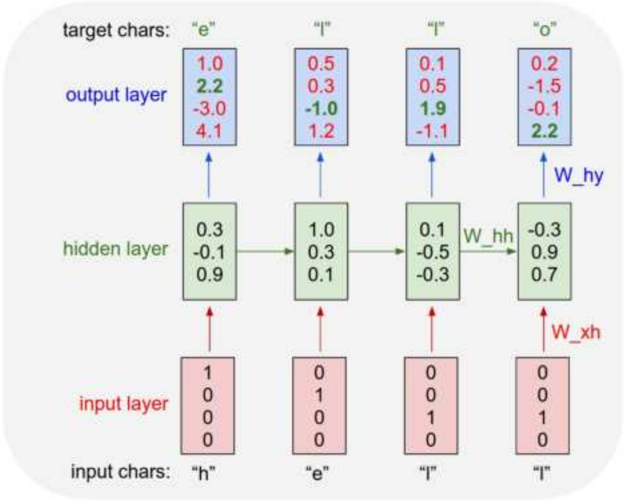
Deep Learning for Multimodal AI
RNNs



Source: https://www.simplilearn.com/tutorials/deep-learning-tutorial/rnn

Deep Learning for Multimodal AI

RNNs



Source: https://github.com/udacity/deep-learning-v2-pytorch/blob/master/recurrent-neural-networks/char-rnn/Character_Level_RNN_Solution.ipynb

Deep Learning for Multimodal AI

Modality-Specific Architectures

Autoencoders for Dimensionality Reduction and Feature Learning

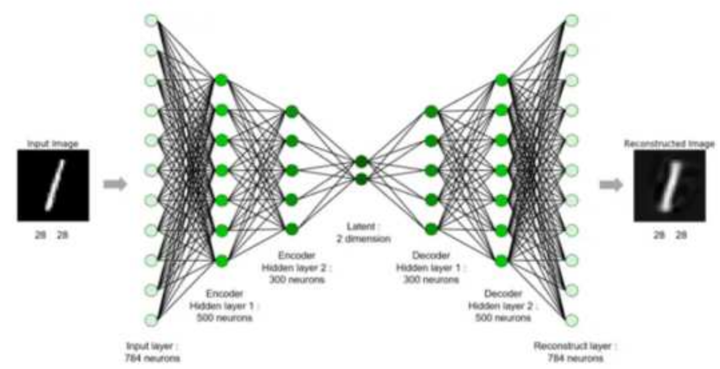
Autoencoders can learn efficient representations of complex medical data.

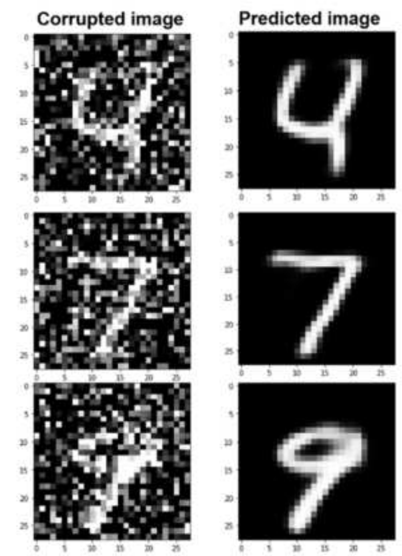
Applications in Multimodal Medical AI:

- **Medical Image Fusion:** Combining information from multiple imaging modalities (e.g., PET-CT, MRI-DTI) to improve diagnostic accuracy.
- **Patient Risk Stratification:** Identifying patients at high risk of developing a specific disease based on their medical history, genetic data, and lifestyle factors.
- **Drug Discovery:** Learning representations of molecules to accelerate drug discovery processes.

Deep Learning for Multimodal AI

Autoencoders

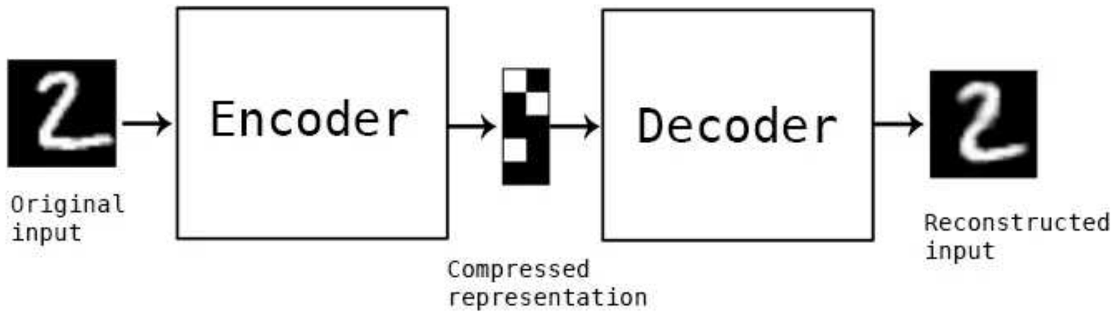




Source: https://wikidocs.net/193827

Deep Learning for Multimodal AI

Autoencoders



Source: https://towardsdatascience.com/auto-encoder-what-is-it-and-what-is-it-used-for-part-1-3e5c6f017726

Deep Learning for Multimodal AI

Multimodal Integration Strategies

Early Fusion (Feature-Level Fusion)

- **Concept:** Features from different modalities are combined at the input level.
- **Advantages:** Can capture low-level interactions between modalities.
- **Challenges:** Requires careful alignment of modalities and handling different data scales.

Low-Level Interactions: The Basics

Low-Level refers to features extracted from the raw data early in the model's processing. These features are often basic visual patterns (for images), acoustic properties (for audio), or simple statistical summaries (for other data types).

Interactions mean how information from different modalities is combined to potentially create new, more informative features.

Example: Medical Image Fusion (PET-CT)

Imagine you want to detect a tumor using both Positron Emission Tomography (PET) and Computed Tomography (CT) scans.

PET Scan: Good at showing metabolic activity. A tumor might show as a bright spot due to high glucose uptake.

CT Scan: Excellent for revealing anatomical structures. It provides detailed images of bones and organs.

Low-Level Interaction in Early Fusion:

Feature Extraction: A CNN might extract the following low-level features:

PET: Areas of high intensity (indicating metabolic activity).

CT: Edges, shapes, and textures of organs.

Early Fusion: These features are combined directly. Now, the model can learn that:

An area with both high intensity in PET and irregular shapes in CT is more likely to be a tumor than an area with only high PET intensity (which could be a benign growth).

Why This Matters: Early fusion allows the model to identify subtle patterns that emerge only when considering both modalities together from the very beginning of the learning process. These patterns might be missed if the modalities were processed separately first (late fusion).

Easy to Remember Analogy: Think of it like baking a cake. Early fusion is like mixing all the ingredients (modalities) at the start. The flavors blend during baking, creating a richer taste. Late fusion is like baking separate cakes and then putting them together – you get the individual flavors, but not the same level of interaction.

Deep Learning for Multimodal AI

Multimodal Integration Strategies

Late Fusion (Decision-Level Fusion)

- **Concept:** Individual models are trained for each modality, and their predictions are combined at the end.
- **Advantages:** Simpler to implement, handles missing modalities gracefully.
- **Challenges:** May not capture rich interactions between modalities.

Deep Learning for Multimodal AI

Multimodal Integration Strategies

Hybrid Fusion

- **Concept:** Combines elements of both early and late fusion for a more flexible approach.
- **Example:** Early fusion for some modalities, late fusion for others.

In case you’re wondering. Here’s an explanation of when joint (intermediate) fusion might be preferred, along with a concrete example:

Joint/Intermediate Fusion: The Middle Ground

The Idea: Instead of fusing at the very beginning (early) or the very end (late), joint fusion combines representations from different modalities at an intermediate stage within the model.

When It's Useful:

Complex Relationships: When the most informative interactions between modalities are not immediately apparent from low-level features but emerge at a higher level of abstraction.

Modality-Specific Processing: When you want to first extract meaningful features from each modality independently before fusion.

Example: Diagnosing Depression (Text + Audio)

Imagine you're building a system to help diagnose depression using both:

Patient Interviews (Audio): Speech patterns, tone of voice, and pauses can contain clues.

Patient Questionnaires (Text): Self-reported symptoms and feelings provide valuable context.

Joint Fusion in Action:

Modality-Specific Processing:

Audio: An RNN processes speech to extract features like pitch, speaking rate, and pauses.

Text: A separate RNN or Transformer model analyzes text for sentiment, keywords related to depression (e.g., "hopeless," "fatigue"), and linguistic patterns.

Joint Fusion: The higher-level features from the audio and text models are combined at an intermediate layer. This allows the model to learn that:

A patient speaking slowly with a flat tone of voice (audio features) combined with frequent use of negative sentiment words in their questionnaire (text features) is a stronger indicator of depression than either modality alone.

Final Classification: The fused representation is then used for a final classification of whether the patient is likely to be depressed.

Easy to Remember Analogy: Think of making a smoothie.

Early fusion: Blending all the fruits together immediately (some flavors might get lost).

Late fusion: Tasting each fruit separately and then deciding what combination you like (misses interactions).

Joint fusion: Juicing some fruits (e.g., apples, oranges) separately to extract their flavors more effectively and then blending those juices with other ingredients (e.g., yogurt, spinach) for a more balanced and flavorful smoothie!

In essence, joint fusion gives you more control over how and when to combine information from different modalities, which is often beneficial for tasks with intricate relationships between modalities.

Deep Learning for Multimodal AI
Multimodal Integration Strategies

Attention Mechanisms

- **Concept:** Allow the model to focus on the most relevant modalities or parts of the input data dynamically.
- **Advantages:** Improves performance by selectively attending to important information.

Deep Learning for Multimodal AI
Advanced Architectures

Transformers (for Multimodal Sequence Modeling)

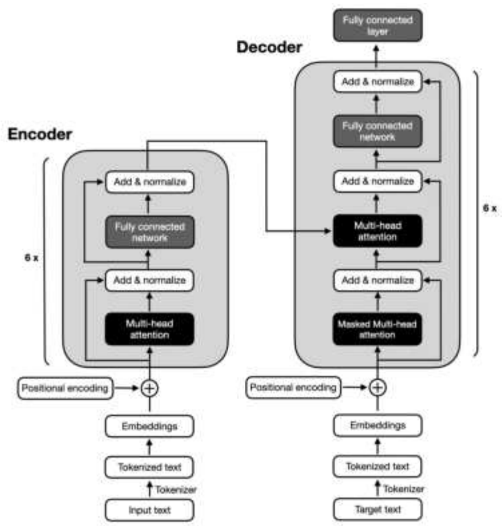
Transformers, with their powerful attention mechanisms, can effectively process and integrate sequential medical data from various sources.

Applications in Multimodal Medical AI:

- **Multimodal Medical Report Generation:** Generating comprehensive medical reports by combining information from images (e.g., X-rays, pathology slides), textual notes (e.g., clinical findings, history), and structured data (e.g., lab results).
- **Drug-Drug Interaction Prediction:** Predicting potential adverse drug-drug interactions by analyzing sequences of medications prescribed to a patient, along with their medical history and genomic information.
- **Patient Time-Series Prediction:** Forecasting patient health outcomes (e.g., risk of readmission, disease progression) using sequences of vital signs, lab tests, and clinical events.

Deep Learning for Multimodal AI

Transformers



Source: https://magazine.sebastianraschka.com/p/understanding-encoder-and-decoder

Deep Learning for Multimodal AI
Advanced Architectures

Graph Neural Networks (GNNs) for Relationships

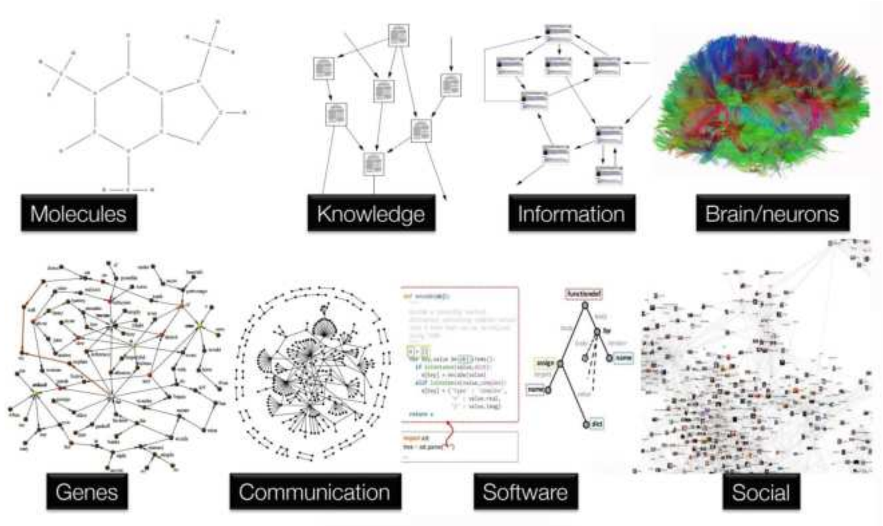
GNNs are well-suited for analyzing complex relationships within the human body or between medical entities.

Applications in Multimodal Medical AI:

- **Drug Repurposing:** Identifying new uses for existing drugs by analyzing relationships between drug molecular structures, disease pathways, and patient symptoms.
- **Disease Pathway Analysis:** Understanding complex disease mechanisms by modeling interactions between genes, proteins, and other biological entities involved in a particular disease.
- **Personalized Treatment Recommendations:** Developing personalized treatment plans by analyzing patient-specific factors (e.g., medical history, genetic profile, lifestyle) and their relationships to treatment options and potential outcomes.
- **Medical Knowledge Graph Completion:** Predicting missing links in medical knowledge graphs, such as relationships between diseases, symptoms, treatments, and genes.

Deep Learning for Multimodal AI

Graphs



Source: https://blogs.nvidia.com/blog/what-are-graph-neural-networks/

Deep Learning for Multimodal AI

Key Considerations

Data Alignment and Preprocessing

Careful alignment and preprocessing of multimodal data are essential to handle variations in scale, resolution, and noise.

Model Complexity and Interpretability

Balancing model complexity with interpretability is important for understanding the model's decisions.

Missing Modality Handling

Multimodal AI systems should be robust to missing data from certain modalities.

Generative AI in Multimodality

Generative Models

Generative Models
Introduction

Imagine a world where AI can help us overcome data scarcity in medical imaging, design new drugs, and even create personalized treatment plans.

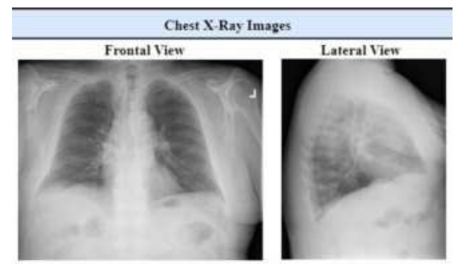
This is the promise of generative AI – a powerful new paradigm in artificial intelligence.

Generative Models

Introduction

Generative models learn the underlyi
data. They use that knowledge to crea   d
behaves like the real thing.

specific terms, sentence structures, and ways of expressing findings



Radiology Report

Indication: *XXXX-year-old with osteoarthritis of the hip scheduled for total hip replacement. Preoperative evaluation.*

Findings:
The heart, pulmonary XXXX and mediastinum are within normal limits. There is no pleural effusion or pneumothorax. There is no focal air space opacity to suggest a pneumonia. There are degenerative changes of the thoracic spine. There is a calcified granuloma identified in the right suprahilar region. The aorta is mildly tortuous and ectatic. There is asymmetric right apical smooth pleural thickening. There are severe degenerative changes of the XXXX.

Impression: *No acute cardiopulmonary disease*

Source: https://www.researchgate.net/publication/371989300_Cross-modal_Deep_Learning-based_Clinical_Recommendation_System_for_Radiology_Report_Generation_from_Chest_X-rays/figures?lo=1

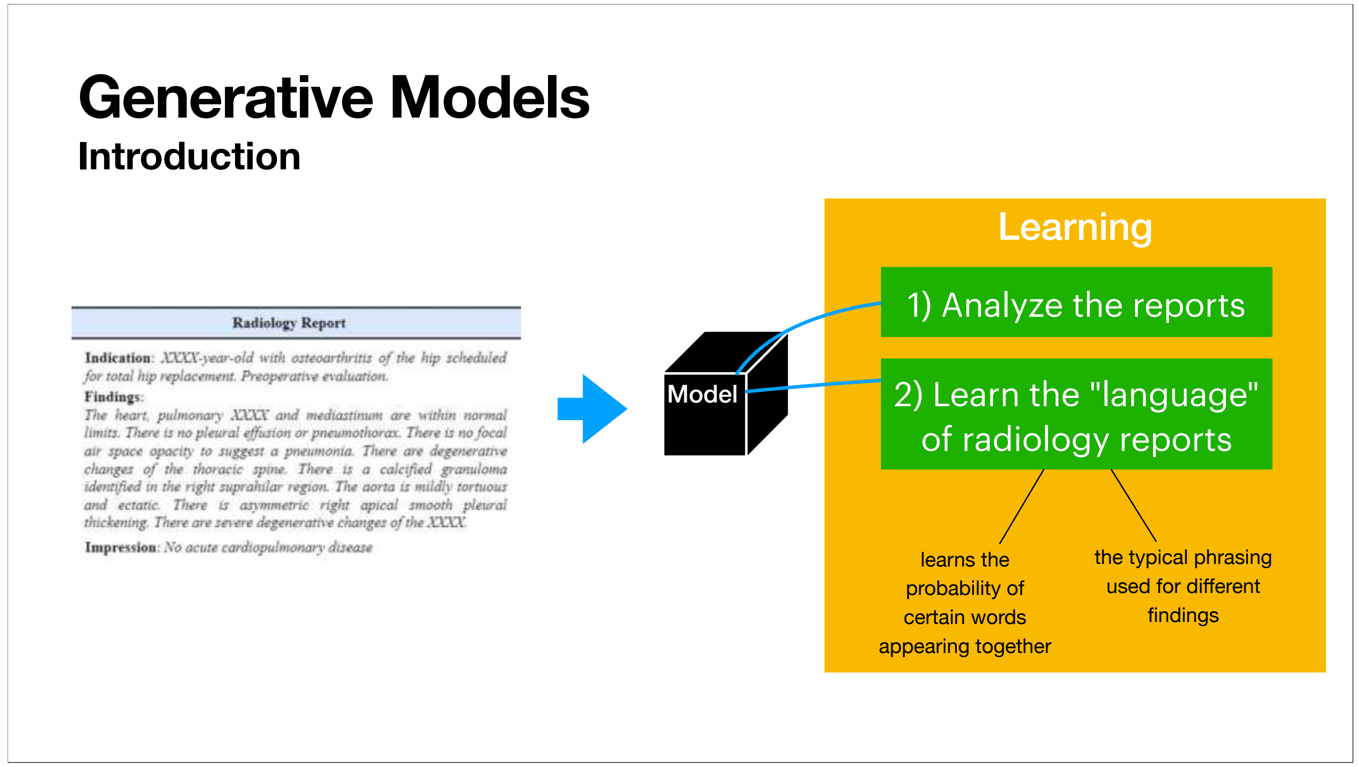
Analogy 1: Learning to Write Medical Reports Like a Doctor

Imagine you have a large collection of X-ray images and their corresponding reports written by expert radiologists.

Think of the reports as a language the doctors use to describe the images – they have specific terms, sentence structures, and ways of expressing findings.

A generative model, like an LLM, can analyze all these reports to learn the "language" of radiology reports. It learns the probability of certain words appearing together, the typical phrasing used for different findings, and more.

Once it understands this language, it can generate new, realistic-sounding reports even for X-rays it has never seen before.



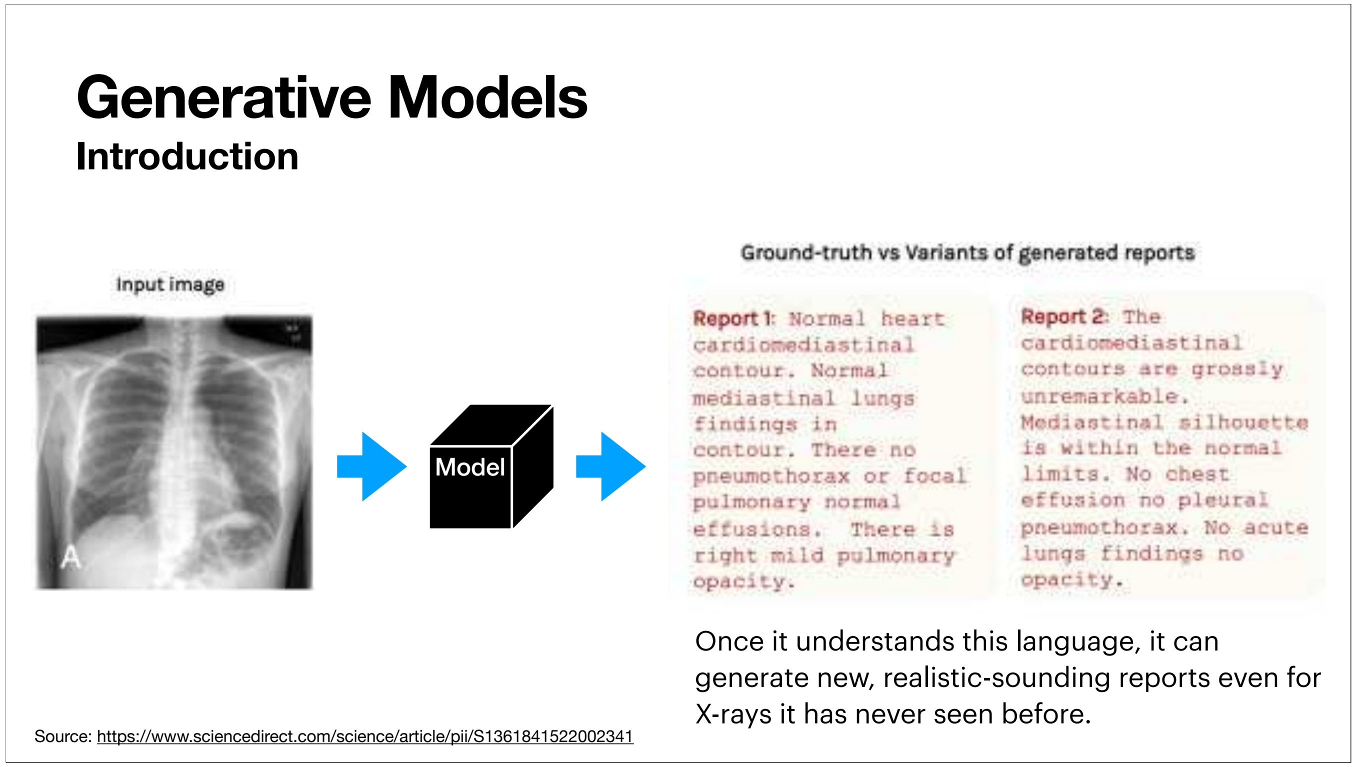
Analogy 1: Learning to Write Medical Reports Like a Doctor

Imagine you have a large collection of X-ray images and their corresponding reports written by expert radiologists.

Think of the reports as a language the doctors use to describe the images – they have specific terms, sentence structures, and ways of expressing findings.

A generative model, like an LLM, can analyze all these reports to learn the "language" of radiology reports. It learns the probability of certain words appearing together, the typical phrasing used for different findings, and more.

Once it understands this language, it can generate new, realistic-sounding reports even for X-rays it has never seen before.



Analogy 1: Learning to Write Medical Reports Like a Doctor

Imagine you have a large collection of X-ray images and their corresponding reports written by expert radiologists.

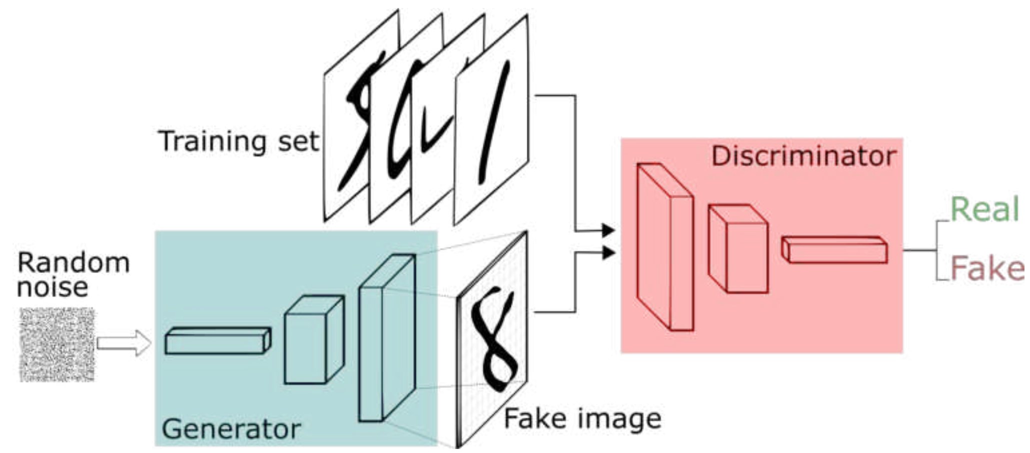
Think of the reports as a language the doctors use to describe the images – they have specific terms, sentence structures, and ways of expressing findings.

A generative model, like an LLM, can analyze all these reports to learn the "language" of radiology reports. It learns the probability of certain words appearing together, the typical phrasing used for different findings, and more.

Once it understands this language, it can generate new, realistic-sounding reports even for X-rays it has never seen before.

Generative Models

Generative Adversarial Networks (GANs)

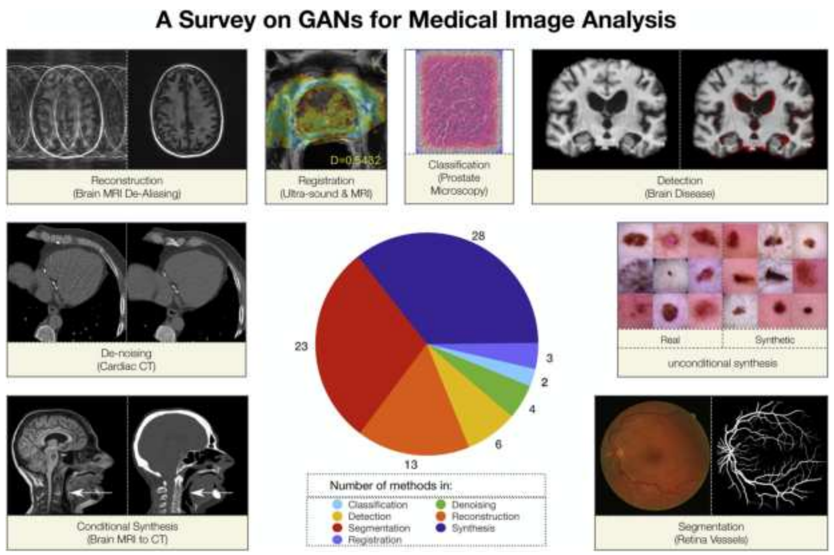


Source: https://sthalles.github.io/intro-to-gans/

GANs work like a competition between two AI systems:a Generator that tries to create realistic data and a Discriminator that learns to tell real data from the fake data the Generator produces. They get better through this adversarial process, like a doctor honing their diagnostic skills by encountering diverse and challenging patient cases.

Generative AI in Multimodality

Generative Adversarial Networks (GANs)



Source:https://www.sciencedirect.com/science/article/pii/S0933365719311510#fig0005


LLMs in Healthcare

LLMs in Healthcare

Bridging the Gap

While images are crucial in healthcare, so is textual information – patient records, medical reports, research papers.

Large Language Models (LLMs) are a type of generative AI specifically designed to understand and generate human-like text.

LLMs in Healthcare

Building block of LLMs: Transformers

Imagine you're a doctor reading a patient's medical history to make a diagnosis. You don't just focus on each word individually, right? You pay attention to:

- **Key words:** "Chest pain," "shortness of breath" stand out more than "went for a walk."
- **Relationships between words:** "Chest pain after exercise" is different from "chest pain relieved by rest."
- **Context of the whole history:** Knowing the patient's age, past medical conditions, etc., helps you connect the dots.

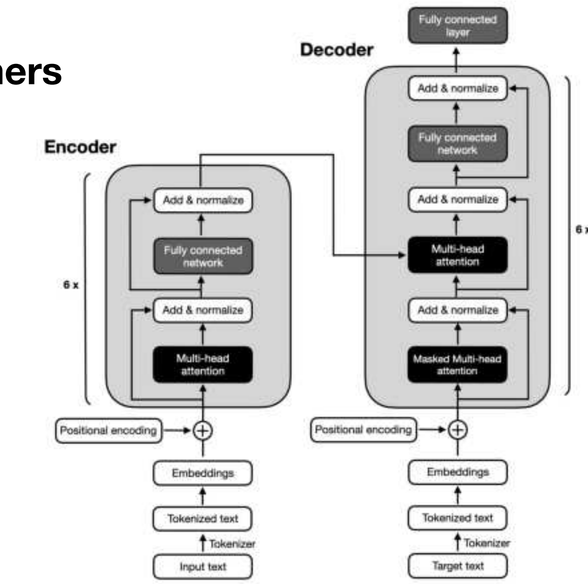


Illustration of the original transformer architecture proposed in Attention Is All You Need, 2017

https://magazine.sebastianraschka.com/p/understanding-encoder-and-decoder

LLMs in Healthcare

Building block of LLMs: : Transformers

That's what Transformers do, but with data:

Attention is Key: Instead of analyzing data (like text or images) in a strict sequence, transformers use "attention" mechanisms to focus on the most important parts for the task at hand. It's like how your brain prioritizes certain words when reading.

Capturing Relationships: They excel at understanding complex relationships within the data. For example, in a medical report, they can link symptoms to diagnoses, treatments to outcomes, and so on.

Context Matters: Transformers process data "holistically," considering the entire context instead of just individual pieces. Think of it like you considering a patient's entire medical history, not just the most recent visit.

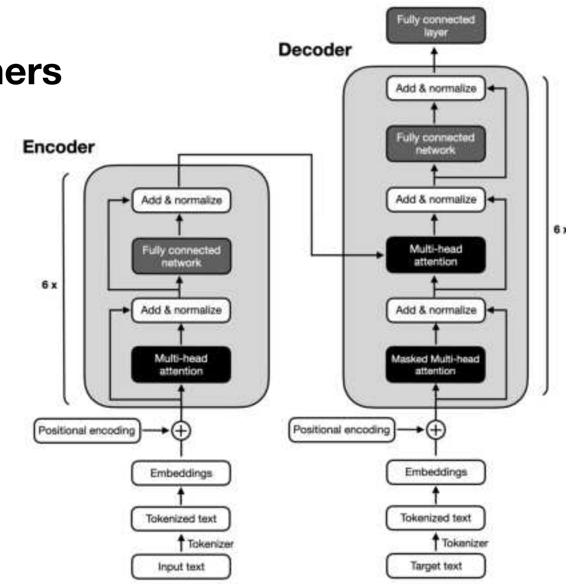


Illustration of the original transformer architecture proposed in Attention Is All You Need, 2017

https://magazine.sebastianraschka.com/p/understanding-encoder-and-decoder

LLMs in Healthcare

Large Language Models in Healthcare and Medical Domain: A Review

Source: Nazi, Zabir Al, and Wei Peng. "Large language models in healthcare and medical domain: A review." *Informatics*. Vol. 11. No. 3. MDPI, 2024.

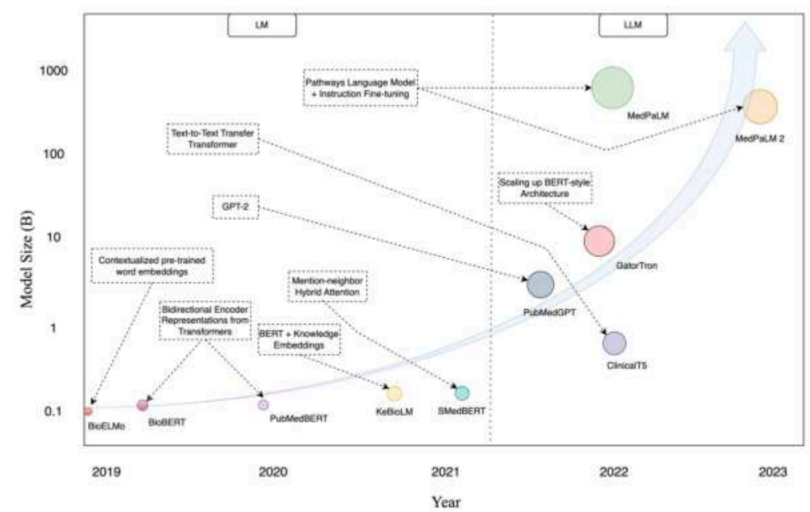


Figure 1: Scale of Medical Language Models: A Size Comparison

LLMs in Healthcare

Multimodal Large Language Model

Source: Nazi, Zabir Al, and Wei Peng. "Large language models in healthcare and medical domain: A review." *Informatics*. Vol. 11. No. 3. MDPI, 2024.

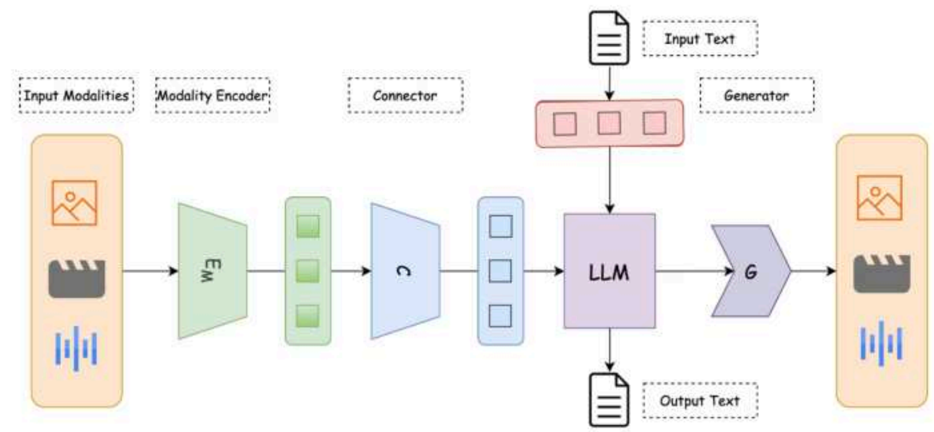


Figure 2: Schematic Representation of a Standard Multimodal Large Language Model (MLLM) Architecture

LLMs in Healthcare

Applications in Healthcare: Text-image retrieval

BiomedCLIP: a multimodal biomedical foundation model pretrained from fifteen million scientific image-text pairs [8], that can perform tasks from retrieval to classification to visual question-answering (VQA).

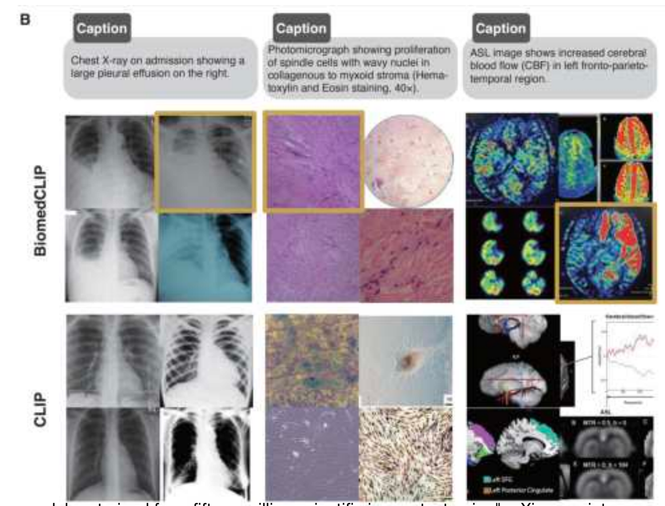
[8] Zhang, Sheng, et al. "BiomedCLIP: a multimodal biomedical foundation model pretrained from fifteen million scientific image-text pairs." arXiv preprint arXiv:2303.00915 (2023).

LLMs in Healthcare

Applications in Healthcare: Text-image retrieval

Figure 2: Comparison on cross-modal retrieval.

B: Three examples comparing BiomedCLIP and general-domain CLIP on text-to-image retrieval for sample PMC captions (top-4 predictions). Gold box indicates the ground truth figure for the caption [8]



[8] Zhang, Sheng, et al. "BiomedCLIP: a multimodal biomedical foundation model pretrained from fifteen million scientific image-text pairs." arXiv preprint arXiv:2303.00915 (2023).

LLMs in Healthcare

Applications in Healthcare: Text-image retrieval

Figure 4: Comparison on medical visual question answering.

B: ... BiomedCLIP correctly answers the questions in C and D. While not technically correct, its answer to B nevertheless correctly identifies liver as the metastatic focus (on the right side of the CT scan) [8].

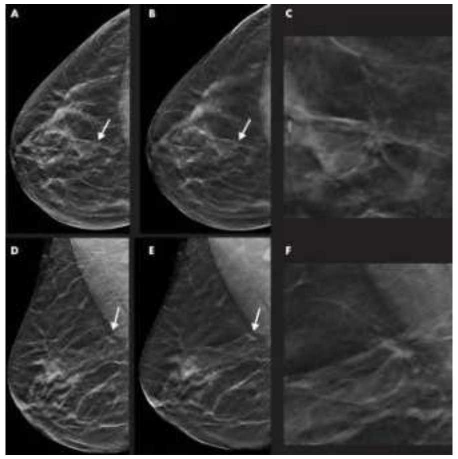
	B	C	D
Question:	Are there multiple or just 1 metastatic focus?	What are the hyperdensities on the periphery of the image?	What is the biological sex of the patient?
Answer:	one	ribs	female
MEVF:	right chest ✗	storage of urine ✗	inflammation ... ✗
QCR:	no ✗	intestine ✗	treat brain diseases ... ✗
PubMedCLIP:	yes ✗	spinal cord ✗	nodule ✗
BiomedCLIP:	right lobe of liver ✗	ribs ✓	female ✓

[8] Zhang, Sheng, et al. "BiomedCLIP: a multimodal biomedical foundation model pretrained from fifteen million scientific image-text pairs." arXiv preprint arXiv:2303.00915 (2023).

Warning

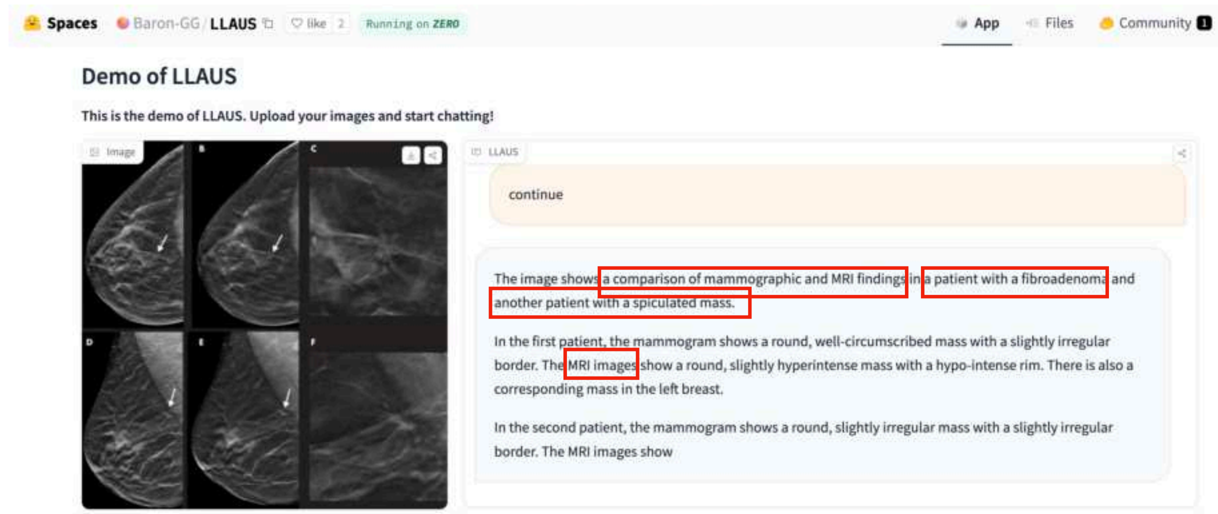
Images in a **55-year-old woman** with a spiculated mass localized in the upper central quadrant (arrow in A, B, D, and E) of **right breast detected with digital breast tomosynthesis (DBT) plus synthetic mammography (SM)**. Breast density was classified as category C with the Breast Imaging Reporting and Data System. Mass was invasive ductal carcinoma, stage I, and was estrogen and progesterone receptor positive and human epidermal growth factor receptor 2 negative. A, Image from SM in craniocaudal view. B, Single-slice DBT image in craniocaudal view. C, Magnification of the lesion depicted in B. D, Image from SM in mediolateral oblique view. E, Single-slice DBT image in mediolateral oblique view. F, Magnification of the lesion depicted in E. Images courtesy of Radiological Society of North America

Tomosynthesis With Synthetic Mammography Improves Breast Cancer Detection



Source: https://www.itnonline.com/content/tomosynthesis-synthetic-mammography-improves-breast-cancer-detection

Warning

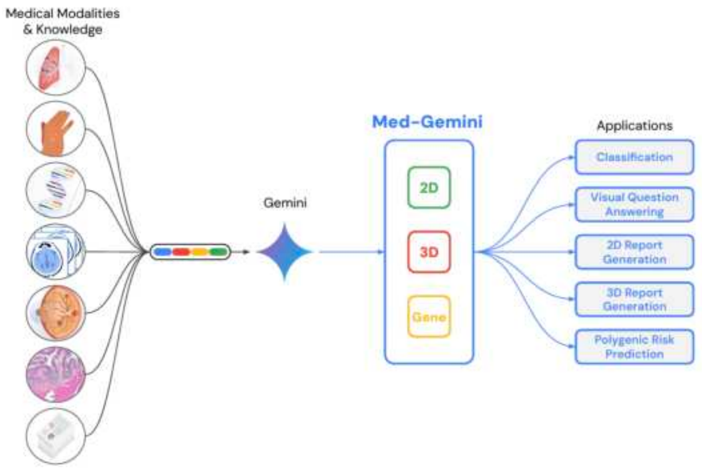


Source: https://huggingface.co/spaces/Baron-GG/LLAUS

MLLMs in Healthcare

Applications in Healthcare

Med-Gemini [9]: "a family of multimodal models from Google specifically designed to address complex medical tasks. Trained on a diverse range of medical data including radiology images, pathology slides, and genomic information, Med-Gemini models demonstrate impressive performance in tasks like generating medical reports, answering clinical questions from images, and predicting disease risk."



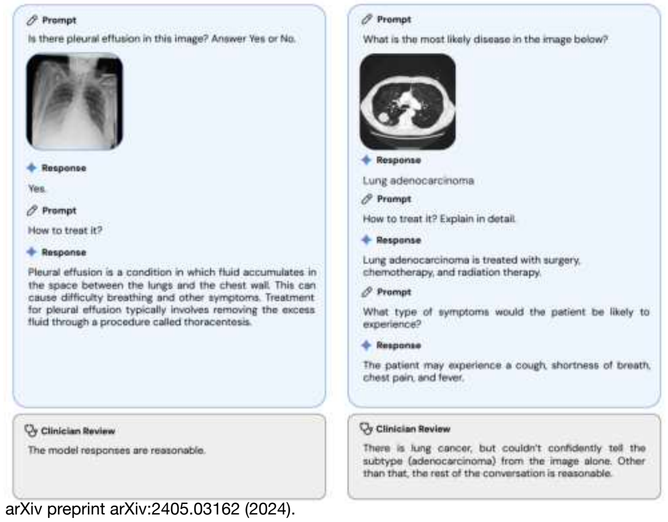
[9] Yang, Lin, et al. "Advancing multimodal medical capabilities of Gemini." arXiv preprint arXiv:2405.03162 (2024).

MLLMs in Healthcare

Applications in Healthcare: Open ended question answering

Figure 6 | Example of 2D medical image dialogue via open ended question answering.

For chest Xray (Johnson et al., 2019a), lung CT (Liu et al., 2021), fundus images (Cuadros and Bresnick, 2009), and skin lesion images (Pacheco et al., 2020) [9].



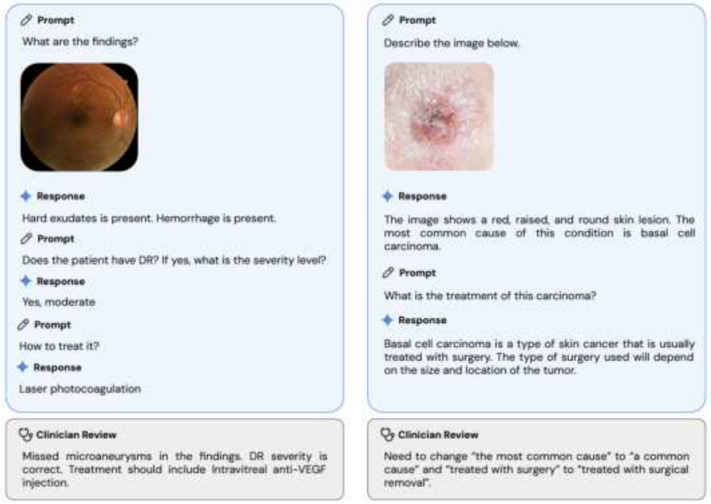
[9] Yang, Lin, et al. "Advancing multimodal medical capabilities of Gemini." arXiv preprint arXiv:2405.03162 (2024).

MLLMs in Healthcare

Applications in Healthcare: Open ended question answering

Figure 6 | Example of 2D medical image dialogue via open ended question answering.

For chest Xray (Johnson et al., 2019a), lung CT (Liu et al., 2021), fundus images (Cuadros and Bresnick, 2009), and skin lesion images (Pacheco et al., 2020) [9].



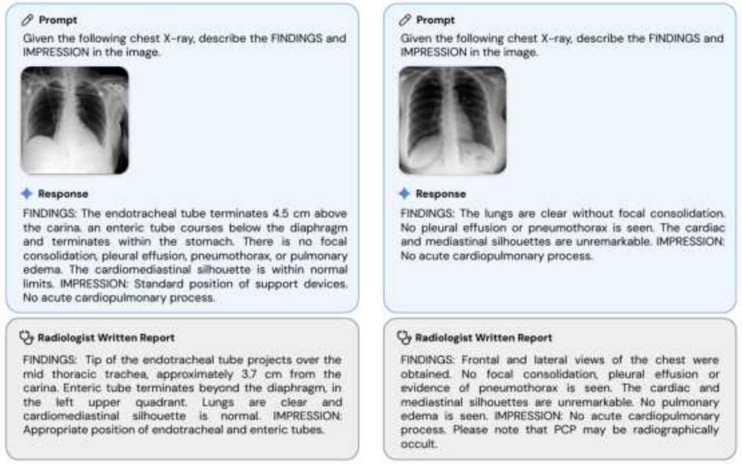
[9] Yang, Lin, et al. "Advancing multimodal medical capabilities of Gemini." arXiv preprint arXiv:2405.03162 (2024).

MLLMs in Healthcare

Applications in Healthcare: Report Generation

Figure 8 | Examples of chest X-ray report generation.

These examples demonstrate the capability of MedGemini for CXR report generation on various conditions. Top-left, support devices; Top-right, normal case; Bottom-left, acute abnormality; Bottom-right, chronic abnormality [9].



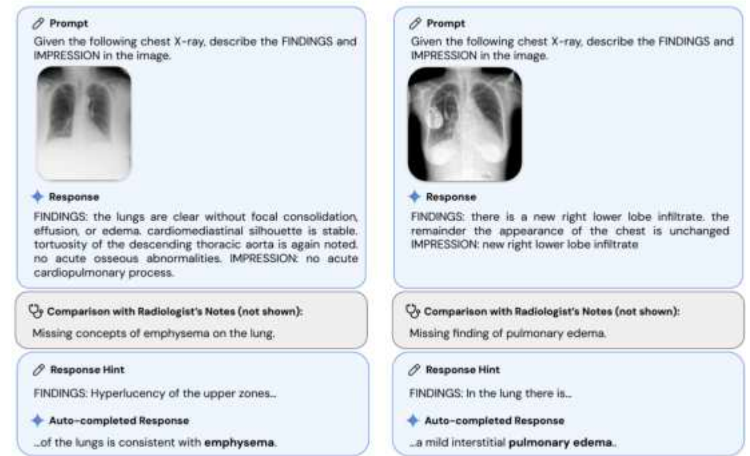
[9] Yang, Lin, et al. "Advancing multimodal medical capabilities of Gemini." arXiv preprint arXiv:2405.03162 (2024).

MLLMs in Healthcare

Applications in Healthcare: Report Generation

Figure 10 | Examples of chest X-ray report autocompletion.

In these examples, particular concepts were missing from the report generated without any hint, and were recovered with the autocomplete prefix hint. A) Emphysema, B) Pulmonary Edema [9].



[9] Yang, Lin, et al. "Advancing multimodal medical capabilities of Gemini." arXiv preprint arXiv:2405.03162 (2024).

Ethical and Regulatory Considerations

Ethical and Regulatory Considerations

Bias and Fairness

The Challenge

AI models are only as good as the data they are trained on. Biases in medical data can lead to unfair or inaccurate AI systems, exacerbating healthcare disparities.

Ethical and Regulatory Considerations

Bias and Fairness

Underrepresentation

If a model is primarily trained on data from a specific demographic group (e.g., Caucasian males), it may perform poorly for other groups (e.g., women, minorities).

Historical Bias

Data reflecting past disparities in healthcare access or treatment can be baked into AI systems, perpetuating those inequalities.

Ethical and Regulatory Considerations

Bias and Fairness

Diverse Datasets

It's crucial to use datasets that represent the diversity of the patient population we serve. This requires proactive efforts to collect data from underrepresented groups.

Bias Mitigation Techniques

Researchers are developing techniques to identify and mitigate bias in both data and algorithms. This is an ongoing area of research.

Ethical and Regulatory Considerations

Data Privacy and Security

Patient Data is Sensitive

Medical data is highly sensitive and personal. Protecting patient privacy is not only an ethical obligation but also a legal requirement.

HIPAA and Beyond

In the United States, HIPAA (Health Insurance Portability and Accountability Act) sets strict standards for protecting patient health information. But we must go beyond mere compliance.

Morocco does not have a direct equivalent to HIPAA (Health Insurance Portability and Accountability Act) in the US. However, there are laws and regulations that address data privacy and protection, particularly concerning health information:

Key Legislation and Regulations:

Law No. 09-08 on the protection of individuals with regard to the processing of personal data: This law, adopted in 2009, is the primary legislation on data privacy in Morocco. It applies to all sectors, including healthcare.

Law No. 13-09 on the organization and financing of the healthcare system: This law, passed in 2002, includes provisions related to the confidentiality of patient medical records.

Decree No. 2-14-290 implementing Law No. 09-08: This decree provides specific regulations on data processing, including the rights of individuals, security measures, and data transfers.

Key Principles for Health Data Protection in Morocco:

Confidentiality: Healthcare professionals have a legal and ethical obligation to protect patient confidentiality. Medical records are considered private and confidential information.

Consent: The collection and processing of personal health data require the individual's consent, except in specific circumstances (e.g., public health emergencies).

Security: Healthcare organizations must implement appropriate technical and organizational measures to secure personal health data and prevent unauthorized access, use, or disclosure.

Data Minimization: Only the necessary amount of personal data should be collected and processed for the specified purpose.

Data Retention: Personal health data should be retained only for as long as necessary for the intended purpose.

Enforcement and Oversight:

National Commission for the Control of the Protection of Personal Data (CNDP): This independent authority is responsible for overseeing the implementation of Law No. 09-08 and ensuring compliance with data protection regulations.

Challenges and Ongoing Efforts:

Specific Healthcare Data Protection Law: While the existing laws provide a framework for data protection, there is a growing need for a specific law dedicated to healthcare data privacy, addressing the unique challenges and sensitivities of this sector.

Enforcement and Implementation: Effective enforcement and implementation of existing data protection regulations remain crucial for ensuring patient privacy.

Digital Health Advancements: The rapid adoption of digital health technologies in Morocco raises new challenges for data security and privacy, requiring ongoing adaptation of regulations and practices.

In Summary:

While Morocco doesn't have a direct equivalent to HIPAA, the existing legal framework provides a basis for health data protection. Ongoing efforts are needed to strengthen regulations, enhance enforcement, and address the evolving landscape of digital health to better protect patient privacy.

Ethical and Regulatory Considerations

Data Privacy and Security

Robust Security

- **De-identification:** Removing identifying information from data used for training AI models is essential.
- **Encryption and Access Control:** Strong security measures are needed to prevent unauthorized access or data breaches. LLMs' vast data requirements amplify these risks.

Transparency with Patients

Patients have the right to know how their data is being used. Clear communication and informed consent are crucial for building trust.

Ethical and Regulatory Considerations

Explainability and Interpretability

Black Box Problem

Many AI models, especially deep learning models, are considered 'black boxes.' It's hard to understand why they make certain predictions.

Lack of Trust:

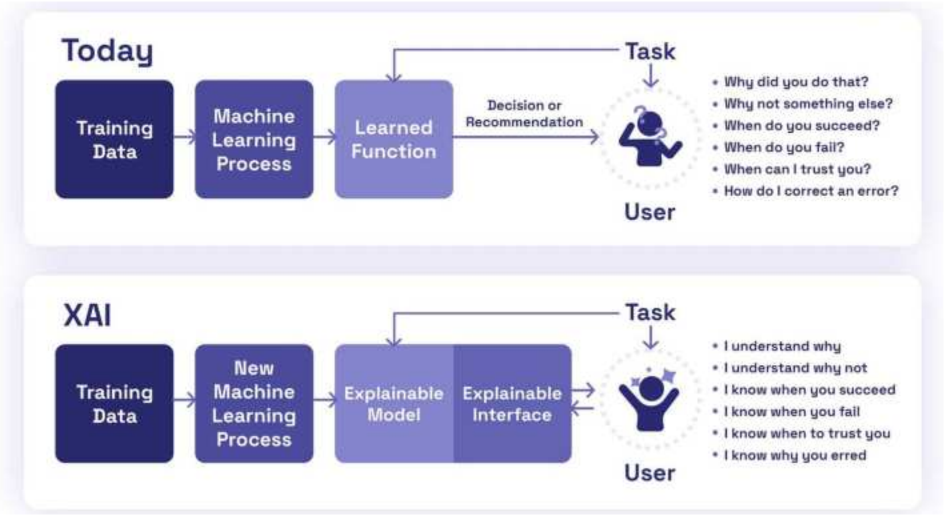
This lack of transparency can make it difficult for doctors to trust AI recommendations. Would you trust a treatment plan without understanding the reasoning behind it?

The Need for Explainability

We need AI systems that can explain their decision-making process in a way that is understandable to healthcare professionals.

Ethical and Regulatory Considerations

Explainability and Interpretability



Source: https://www.holisticai.com/blog/shap-values-game-theory-and-ai

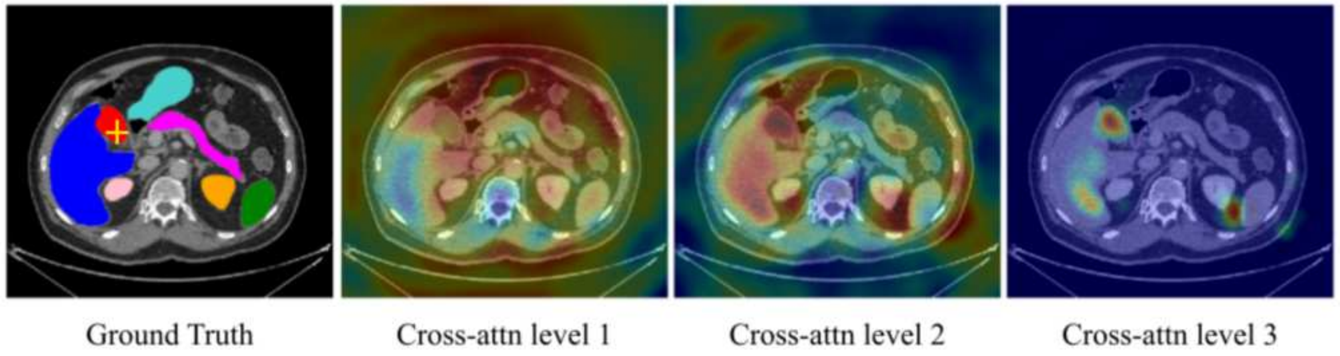
Ethical and Regulatory Considerations
Explainability and Interpretability

Methods for Interpretability

- **Attention Mechanisms (as in Transformers):** Some models, like transformers, offer insights into what parts of the data they are focusing on.
- **Explainable AI (XAI) Techniques:** Researchers are developing methods to visualize and interpret AI models, making them more transparent.

Ethical and Regulatory Considerations

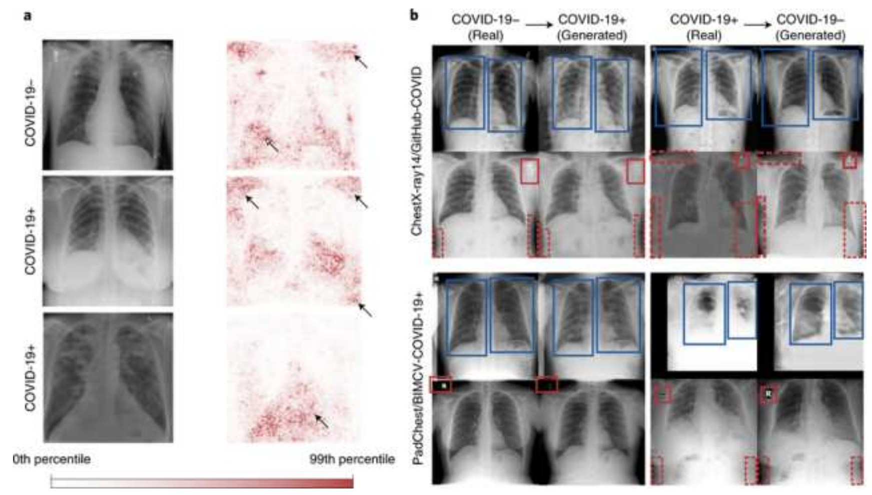
Explainability and Interpretability



Interpretable Medical Imagery Diagnosis with Self-Attentive Transformers: A Review of Explainable AI for Health Care
Source: https://www.mdpi.com/2673-7426/4/1/8

Ethical and Regulatory Considerations

Explainability and Interpretability



Source: https://www.researchgate.net/publication/352007877_AI_for_radiographic_COVID-19_detection_selects_shortcuts_over_signal/figures?lo=1

Ethical and Regulatory Considerations

Explainability and Interpretability

Benefits of Explainability:

- **Improved Trust:** Doctors are more likely to trust AI systems they understand.
- **Error Detection:** Explainability can help identify errors in AI models and improve their accuracy.
- **Patient Education:** Explanations can help patients understand their diagnoses and treatment options.

Conclusion and Future Directions

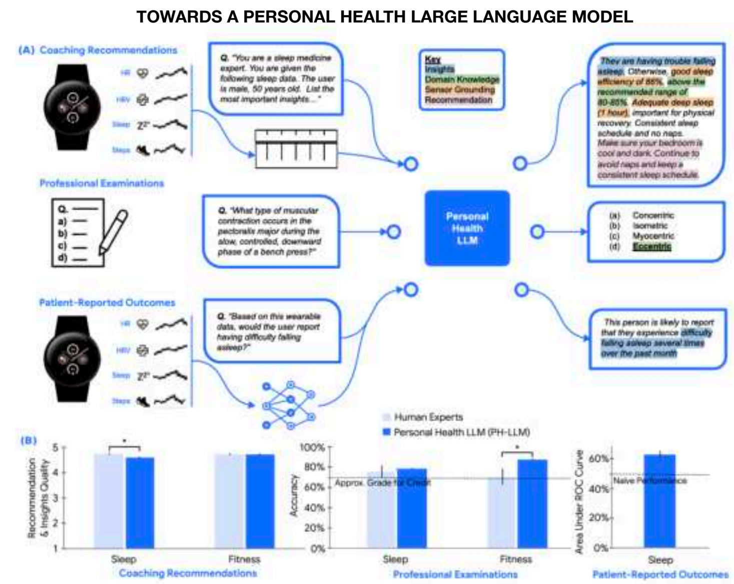
Conclusion

- Multimodal AI is transforming healthcare by unlocking insights hidden within diverse data sources.
- Applications span from disease diagnosis and risk prediction to personalized treatment and patient monitoring.
- Ethical considerations and responsible AI development are paramount for successful implementation in clinical settings.

Future Trends

Figure 1: PH-LLM: A Personal Health Large Language Model.

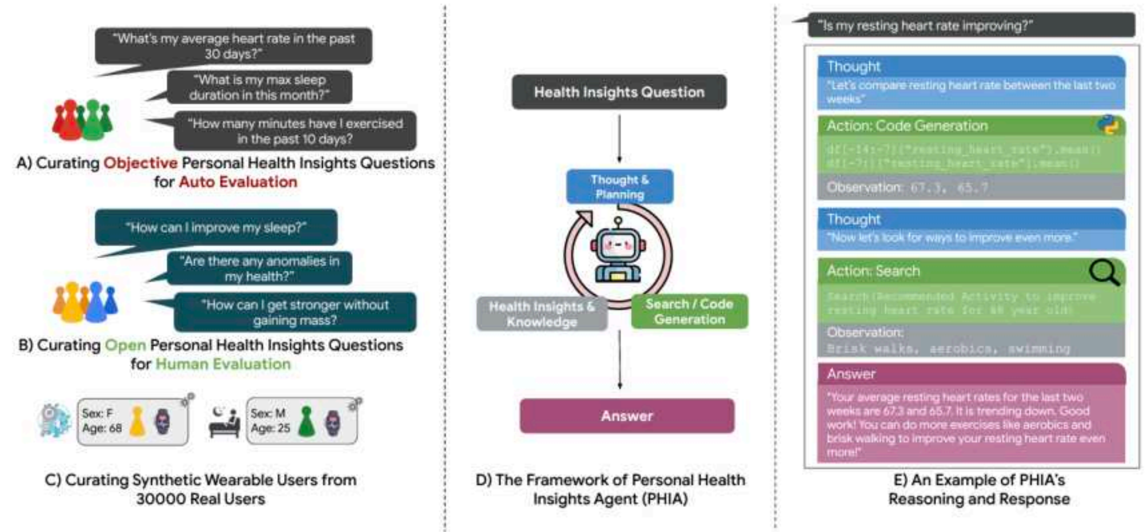
(A) We present PH-LLM, a version of Gemini fine-tuned for personal health and wellness. We evaluated PH-LLM on three aspects of personal health: generating personalized insights and recommendations for user goals in the domains of sleep and fitness... [10].



[10] Cosentino, Justin, et al. "Towards a Personal Health Large Language Model." *arXiv preprint arXiv:2406.06474* (2024).

Future Trends

Figure 1: An overview of our Personal Health Insights Agent (PHIA) [11]

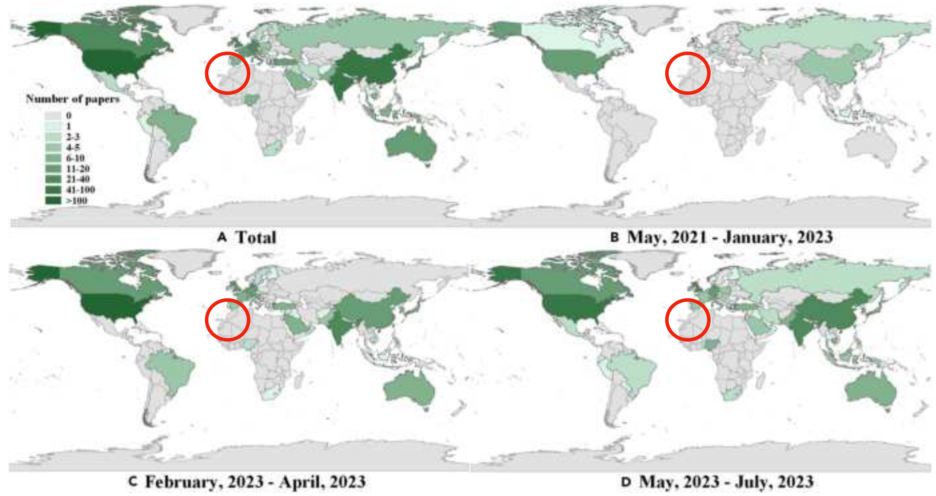


[11] Merrill, Mike A., et al. "Transforming wearable data into health insights using large language model agents." arXiv preprint arXiv:2406.06464 (2024).

Q&A

Sara El-Ateif, 30-31 Oct 2024

Call To Action



The application of large language models in medicine: A scoping review

Source: https://www.sciencedirect.com/science/article/pii/S2589004224009350#bib99

Code Resources

Brain Tumor 3D Multimodal CNN - All MRI Type

https://www.kaggle.com/code/michaelfumery/brain-tumor-3d-multimodal-cnn-all-mri-type

Multimodal Single-Cell Integration

https://www.kaggle.com/code/vslaykovsky/multi-67-cite-89-pytorch-swiss-army-knife/notebook

Thank you!

Sara El-Ateif, 30-31 Oct 2024

Reach Out!

@el_ateifSara

zaap.bio/saraelateif

researchgate.net/profile/Sara-El-Ateif-3