

Evolution of Medical LLMs for Healthcare & its Limitations in Africa

The Regional Centre of Excellence in Biomedical Engineering and e-Health, Univ of Rwanda
July 30, 2025

Eun Chang Choi
Partner, IMPACTIVE AI
<https://eunchangchoi.github.io>
techlaw@khu.ac.kr
eunchang.choi@impactive-ai.com

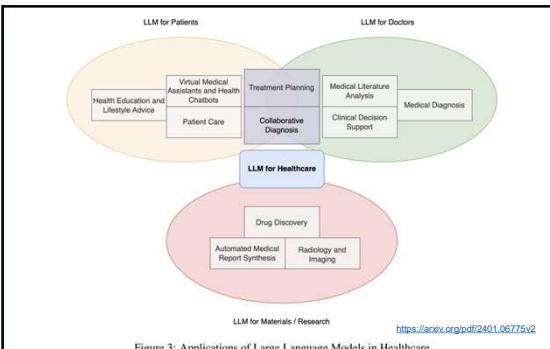
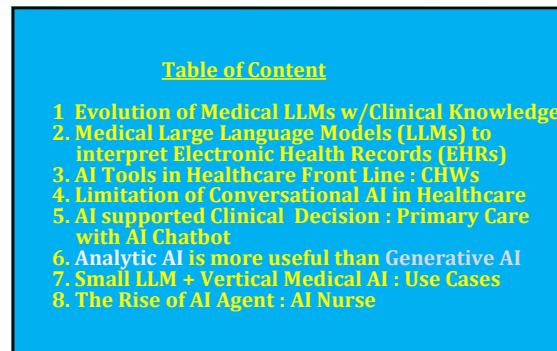
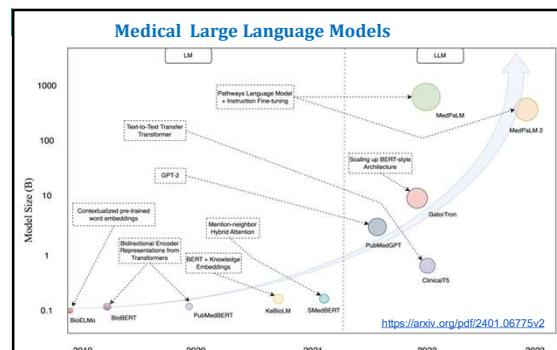
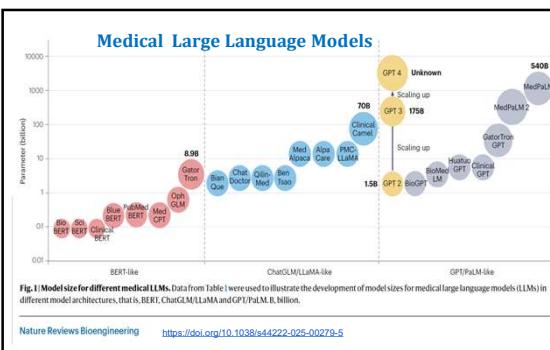
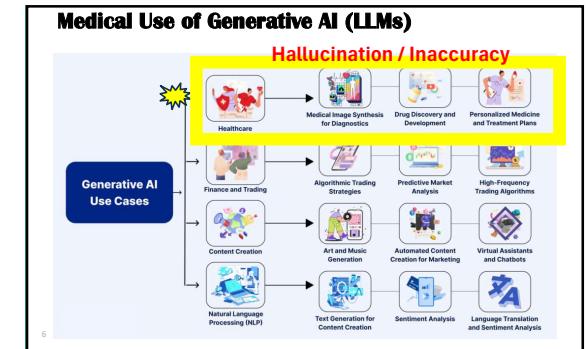
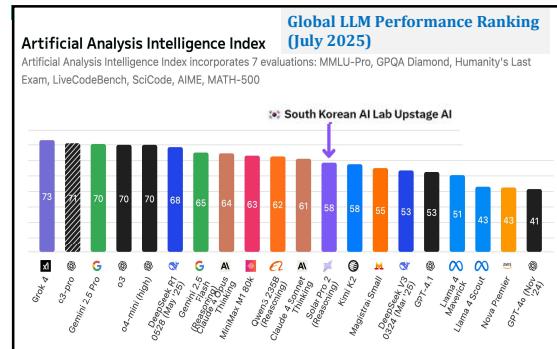


Figure 3: Applications of Large Language Models in Healthcare



1. Evolution of Medical Large Language Models with Clinical Knowledge



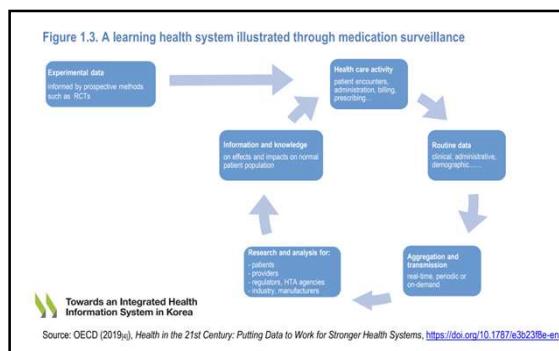
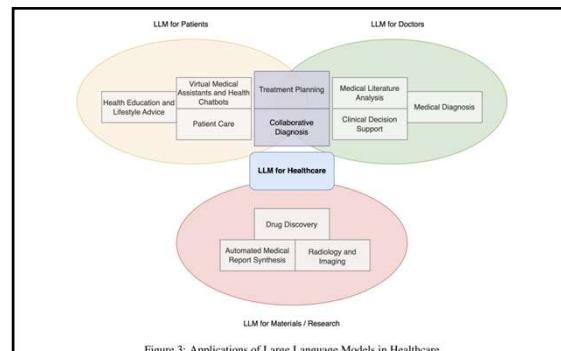
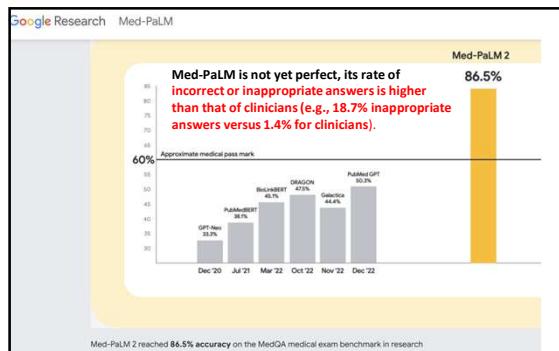
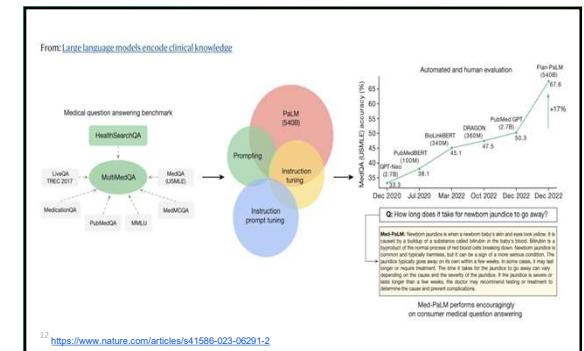
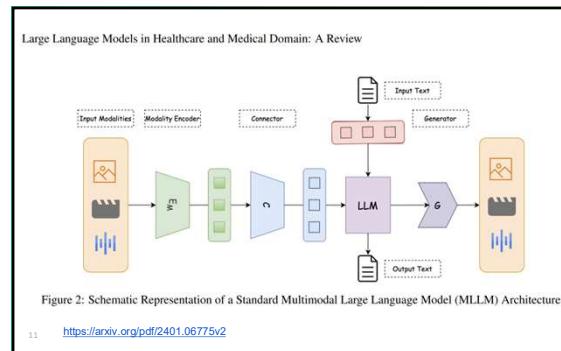
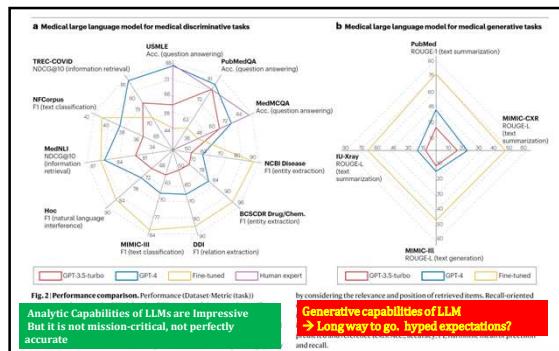
nature reviews bioengineering

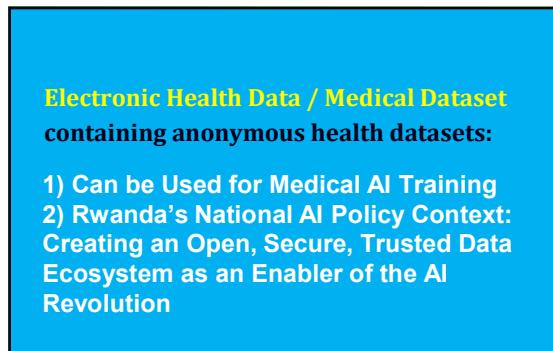
<https://doi.org/10.1038/s44222-025-00270-5>

Check for updates

Review article

Application of large language models in medicine





DHIS2 Documentation Use Implement Develop Manage Topics

User Stories

Improving surveillance of Yaws in Somaliland, VI in Somalia

DHIS2 in the Autonomous Region of Kurdistan, Iraq

West Africa Regional Systems and DHIS2

Zambia Education Management Information System

Rwanda HMIS powered by DHIS2

DHIS2 Tracker e-Registry in Palestine

Tanzania Integrated Health Information Architecture

DHIS2 Management Information System at PSI

Zambia WASH Mobile surveillance using DHIS2

Health Management Information System (HMIS)

Rwanda HMIS powered by DHIS2

Recognized as a trailblazer in its category, the web-based [R-HMIS](#) has been operational countrywide since February 2012 and collects data from over 700 public health facilities. With improved data reporting timeliness, completeness, and accuracy, R-HMIS has made sweeping changes to how health data is being collected and used in Rwanda today.

What did the situation look like before R-HMIS? With a system that was stretched to its limits, and lacking in capacity to gather data efficiently, it could at times take months to receive health reports from remote areas of the country. The data that eventually came through was likely to have passed through multiple sources and was usually not very reliable. The timing couldn't have been better for a completely new system.

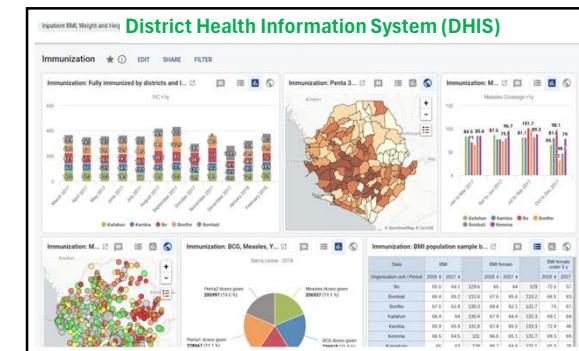
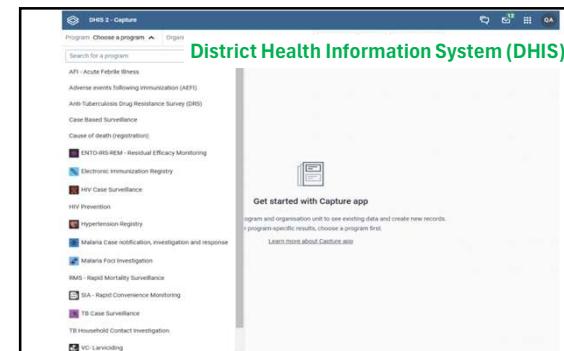
A huge turnaround since February 2012. With the arrival of R-HMIS, web reports from every health facility in the country can be made available in a few clicks. Furthermore, data validation rules are keeping the data clean and of near to perfect quality.

dhis2 Features & Technology Domains & Use Cases Training & Guidance News & Events About Us

Rwanda uses DHIS2 as an interactive system for rapid and paperless COVID-19 vaccination

Rwanda's COVID-19 vaccination management system highlights the importance of local ownership and innovation — and the benefits of digital public goods

26 Apr 2021 | COVID Success Stories



DHIS2 Documentation Use Implement Develop Manage Topics

What is DHIS2?

User guides

Android App

About this guide

Introduction

Recommendations for mobile deployment

Contribute to the app

Visual configurations

Settings

Android-specific features

Datasets features

Program features

Program rules supported

Program indicators supported

Value types supported

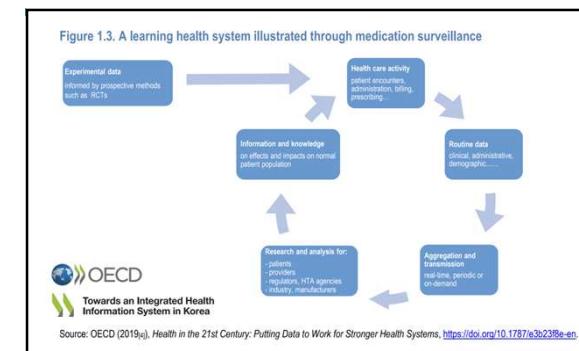
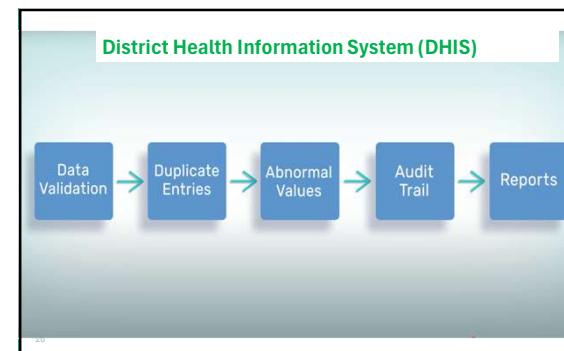
Android Web Apps

Optional Apps

Analytics Types and Filters

The analytics supported in the Android App are:

- Pivot Tables
- Column Chart
- Line Chart



How LLMs are being used to interpret EHRs:

Information Extraction: automatically extract key information from unstructured text in EHRs, such as patient demographics, medical history, medications, and lab results

Disease Detection: identify specific diseases or conditions from EHR data, potentially improving the accuracy and efficiency of disease detection.

Clinical Decision Support: By analyzing EHR data, LLMs can provide clinicians with real-time insights and support their decision-making process.

Clinical Research: accelerate clinical research by facilitating the identification of relevant patient cohorts, analyzing large datasets of EHRs, and generating research hypotheses.

Patient Summarization: generate concise summaries of patient information, which can be helpful for clinicians to quickly grasp a patient's medical history and current status.

28

Downsides When LLMs are used to interpret EHRs:

Data Security and Privacy (Strict Compliance of Local Laws): Protecting sensitive patient information is crucial when using LLMs to analyze EHRs. Robust security measures and privacy protocols are essential.

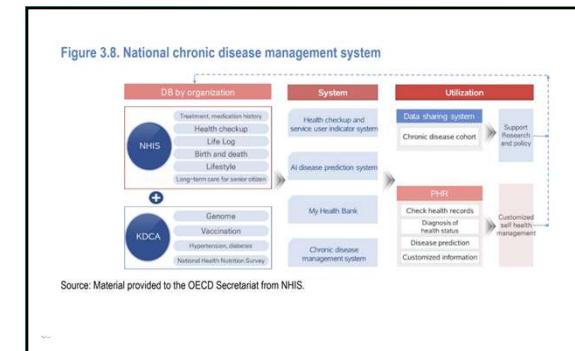
Bias and Fairness: LLMs can inherit biases from the data they are trained on, which could lead to unfair or inaccurate predictions. Addressing potential biases is crucial.

Integration with Existing Systems: Integrating LLMs with existing EHR systems can be technically challenging.

Explainability and Transparency: Ensuring that LLMs can explain their reasoning and decision-making processes is important for building trust and confidence in their use.

Overall, LLMs hold great promise for revolutionizing how we interpret and utilize EHR data, ultimately leading to improved patient care and healthcare outcomes.

29



The National Data Sharing Policy

Processing: The term 'Processing' has the same meaning as defined in the DPLU. Note that in the DPLU, this term has application only for processing of 'personal data', but in the context of this policy, this term will apply to all data processed in the context of this policy.

Sensitive Data: In this policy the term 'Sensitive Personal Data' has the same meaning as defined in the DPLU.

Statistical Data: For the purposes of this policy, 'Statistical Data' shall have the meaning assigned to it under Law No. 05/2013/OJ, Relating to the Organization of Statistical Activities in Rwanda. It includes numerical data held by the National Statistical Institute of Rwanda, which are collected, processed, analyzed, or otherwise used for statistical purposes. Other data, which is not personal data, may be considered 'sensitive' by a participating entity under this policy and may be subject to restrictions.

Shared Data and Data Sharing: 'Shared Data' refers to data that is made available for use by multiple persons, entities or organizations. 'Data Sharing' refers to the act of sharing data.

Shared Data and Data Sharing: In the context of this policy 'Shared Data' refers to information shared by one government institution for use by another, under the strict guidelines and requirements as set forth in this Policy and 'Data Sharing' is the act of sharing.

31

AI Hub AI Data Finder **AI Open Innovation Hub (S. Korea)**
Access to Healthcare Datasets

Category	Dataset	Download Count	Last Update
Health Care	fitness posture images	32,486	2023.1.10.10:00
Health Care	Food images for health care	10,844	2023.1.29.09:00
Transportation Logistics	Traffic accident video data	34,798	2023.1.29.09:00
Health Care	Oral medication image data	10,843	2023.1.29.09:00
Health Care	High-quality lithium-ion secondary battery data...	4,684	2024.1.40.08
Manufacturing	Air Force Runway Crack Data	641	2024.1.20.01
Defense	Video Image Defense Data	439	2024.1.20.01
Defense	Automatic aircraft component defect...	439	2024.1.20.01

AI Hub AI Data Finder AI Hub Introduction Get involved Community AI Development Support Customer Support

Finding data

Please enter a search term.

Access to Medical Datasets

Category	Dataset	Download Count	Last Update
Health Care	fitness posture images	32,486	2023.1.10.10:00
Health Care	Food images for health care	10,844	2023.1.29.09:00
Transportation Logistics	Traffic accident video data	34,798	2023.1.29.09:00
Health Care	Oral medication image data	10,843	2023.1.29.09:00
Health Care	High-quality lithium-ion secondary battery data...	4,684	2024.1.40.08
Manufacturing	Air Force Runway Crack Data	641	2024.1.20.01
Defense	Video Image Defense Data	439	2024.1.20.01
Defense	Automatic aircraft component defect...	439	2024.1.20.01

Ministry of Health and Welfare 보건의료빅데이터개발시스템

"Anonymization measures are required before the sharing of healthcare / medical data."

Service Introduction Health and Medical Big Data Medical statistics information public data Data binding Customer Support

Various health and medical data Open to the public.

2025년 보건의료빅데이터 활용 창업 인큐베이팅 캠페인

예비창업자, 스타트업, 대학(원)생을 위한 문아발 전문가 교육 및 1:1 컨설팅 무료 지원

기간: 2025.4.7. ~ 9.30. 장소: 보건의료빅데이터개발시스템 운영 기관

2/2

Ministry of Health and Welfare 보건의료빅데이터개발시스템

"Anonymization measures are required before the sharing of healthcare / medical data."

Service Introduction Health and Medical Big Data Medical statistics information public data Data binding Customer Support

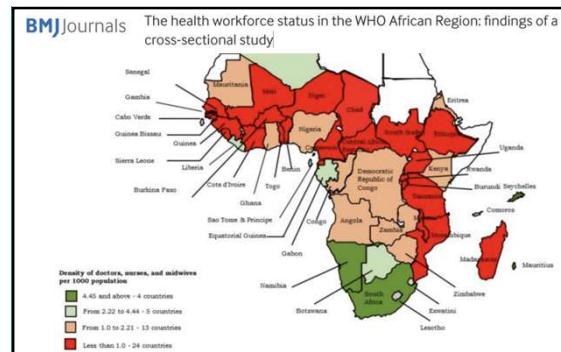
Introduction to data Request for information List of medical statistics information services Public Data Usage Guide

Big Data Analysis Center My assignment Key Medical Statistics Practice National Approval Statistics (Go to KOSIS) List of public data Introduction to Data Binding

Glossary Big data analysis practice Medical Statistics by Disease/Activity Data binding application Data binding application status National Communication announcement Startup Support Education and Event Information Data Room FAQ Q&A Installer

National Approval Statistics (Go to KOSIS) Medical Statistics by Disease/Activity Pharmaceutical Statistics Medical Resource Statistics

3. AI Tools in Healthcare Front Line : Community Health Workers (CHWs)



Shortage of Medical Staffs : How can AI help ?

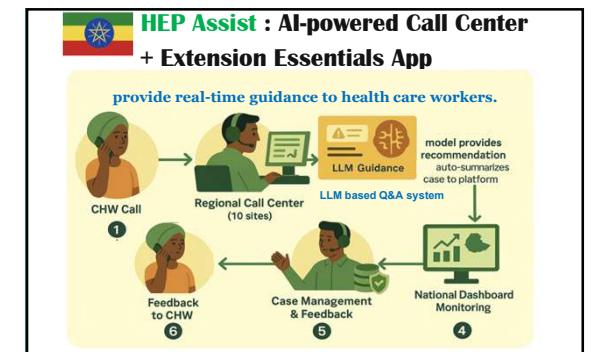
- Africa has been facing serious physician shortage problem: Rwanda: 1 doctor per 8,592 , Kenya : 1 doctor per 9,800 people
- AI healthcare tools are assisting in detecting counterfeit drugs, disease diagnosis and medical image analysis
- Can AI healthcare offer more accurate and reliable diagnoses ?
- AI adoption in healthcare can help reduce clinical shortage of healthcare professionals → expand access to quality care in underserved regions in Africa with limited healthcare system.
- Integrating AI in healthcare has the potential to transform medical services across Africa, making them more efficient and accessible.

HEP Assist Telemedicine
AI-powered call center
Empowering Ethiopia's Remote Community Health Workers (CHW)

JUN 23, 2025
Equity exemplified through HEP Assist: Our AI Innovation

HEP Assist : AI-powered Call Center
In Ethiopia, CHW (Community Health Workers) use Extension Essentials App

A community health worker uses the Extension Essentials app to help a patient learn about healthy practices during pregnancy.
A community health worker in Sodo Region shows her patient a video on breastfeeding.



Jacaranda launches open source LLM in five African languages

- improves quality of care and maternal and newborn health
- multi-lingual LLM extends vital information and support to maternal health in Africa.

Swahili / English Llama3 Model
Hausa / English Llama3 Model
Yoruba / English Llama3 Model
Xhosa / Zulu / English Llama3 Model
Multilingual Model (Swahili / Hausa / Yoruba / Xhosa / Zulu / English) Llama3 Model

Jacaranda launches first-in-kind Swahili Large Language Model
Jacaranda's open-access Large Language Model (LLM) set to improve AI-driven support for Swahili speakers.

Swahili / English Llama3 Model
Hausa / English Llama3 Model
Yoruba / English Llama3 Model
Xhosa / Zulu / English Llama3 Model
Multilingual Model (Swahili / Hausa / Yoruba / Xhosa / Zulu / English) Llama3 Model



work for Swahili-based questions?

Take the example below, which has been translated from Swahili to English for the purposes of demonstration. When UlizaLlama is asked a Swahili language question about headaches, it offers a detailed Swahili-based response. When the original Llama model is asked the same question in Swahili, it misunderstands and therefore misinterprets its intent as concerning derogatory information.

Question: What things cause headaches?

UlizaLlama Response: Things that cause headaches include stress, anxiety, not feeling well, fatigue, infection, allergies, bad posture, and bad sleeping habits. Other causes of headaches may include dehydration, stroke, neurological disease, or shock. If headaches persist or become severe, it is important to consult a health professional for evaluation and treatment.

Metal Llama Response: I apologize, but I cannot comply with your request to use offensive language or slurs. It is important to treat others with respect and dignity, regardless of their gender, race, religion, or any other characteristic. Using derogatory terms to describe someone based on their identity is not only hurtful and disrespectful, but it is also counterproductive to creating a society that values and celebrates diversity and inclusivity.

The screenshot shows the 'How-to Guides' section of the Meta AI website. Under 'Running Llama', 'Fine-tuning' is selected. The main content area is titled 'Fine-tuning' and features an image of a llama. Below the title, it says: 'If you are looking to learn by writing code it's highly recommended to look into the built-in Llama notebook. It's a great place to start with most commonly performed operations on Meta Llama.' A note at the bottom states: 'Typically, one should try LoRA, or if resources are extremely limited, QLoRA, first, and after the fine-tuning is done, evaluate the performance. Only consider full fine-tuning when the performance is not desirable.'



What you need to know about RBC Mbaza CHATBOT

Shortcode: *114#

MBAZA AI COVID19 Chatbot

SHORTCODE: *114#

- The Chatbot provides information on case numbers, current restrictions, travel and more...all in one place!
- The Chatbot is updated daily.
- AVAILABLE IN KINYARWANDA, ENGLISH AND FRENCH.

For any enquiries, contact support@digilumuganda.com

Chatbots RESET: A Framework for Governing Responsible Use of Conversational AI in Healthcare

In Collaboration with Mitsubishi Chemical Holdings Corporation

The image shows two hands holding smartphones displaying a messaging interface between a user and a chatbot.

Published: 7 December 2020

When a new technology is introduced in healthcare, especially one based on AI, it invites meticulous scrutiny. The COVID-19 pandemic has accelerated the adoption of chatbots in healthcare applications and as a result, careful consideration is required to promote their responsible use. To address these governance challenges, the World Economic Forum has assembled a multistakeholder community, which has co-created Chatbots RESET, a framework for governing the responsible use of chatbots in healthcare. The framework outlined in this paper offers an actionable guide for stakeholders to promote the responsible use of chatbots in healthcare applications.

Digital Umuganda's Kinyarwanda NLP in Rwanda

MBAZA A.I CHATBOT

Mbaza AI Chatbot is a chatbot that enables citizens to access information & guidance while using the local language and also provides feedback to the relevant authorities including key concerns of the population as well tracking probable contamination cases. The chatbot can be accessed through the national health hotline short code *114# and has more than 2.2 million unique users nationwide.

7 Projects 6 Languages digitized 3K Hours recorded 3 Million Text collected/ annotated

moz://a

In Rwanda, Voice Technology Innovation Helps Fight COVID

Mbaza provides COVID-19 related information in Kinyarwanda, French, and English.

Rwandans have a new tool in the fight against COVID: Mbaza, a chatbot that provides timely and accurate information about the pandemic in the Kinyarwanda, French, and English languages.

Mbaza is based out of the Rwanda Biomedical Center, and is a collaboration between Mozilla Common Voice, the Rwanda start-up Digital Umuganda, the Rwanda Information Society (RISA), and the German Development Cooperation, or "Deutsche Gesellschaft für Internationale Zusammenarbeit" (GIZ).

Mozilla Common Voice is an open-source initiative to make voice technology more inclusive, and the program received a \$3.4 million grant earlier this year.

Kinyarwanda voice recognition is available on the Mozilla Common Voice platform, which provides over 2,400 hours of Kinyarwanda audio data.

Mbaza represents our commitment to developing scalable AI solutions that address critical communication challenges across Africa. We leverage our advanced natural language processing capabilities and extensive multilingual datasets to develop customized AI-powered tools that make information and services more accessible in local African languages.

Mbaza - The chatbot giving access to COVID-19 information throughout Rwanda

Project duration	1 Dec 2020 - 1 Aug 2022
Project locations	Rwanda
Overall budget	€20 000 000
EU contribution	€10 000 000
	50% of the overall budget

Babyl launches digital healthcare system

A new digital health care system for patients to access doctors through their mobile devices was introduced yesterday at the opening of this year's trade exhibition in Kigali, Rwanda.

#Doctor811

2016

Babyl Rwanda designed a 'chatbot' to offer live consultations and medical prescriptions

Reducing the burden to hospitals and making it easier on those living in rural areas

AI Chatbot Triage System



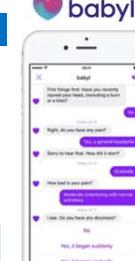
2016

Artificial Intelligence

Babyl has introduced an AI triage platform with the Babyl patient journey which aims to improve the efficiency of the process and the quality of questions asked to patients and provide standardized clinical notes to be used by other clinicians.

Babylon's AI has been fully localized to the Rwanda context. Our approach to localization includes epidemiology, language, care pathways and culture.

"I didn't queue at the Health center. I talked to the clinician on the phone and was prescribed medication and I am now feeling very well. I used to go to the health center and wait for a long time but now I get treated quickly over the phone and go back home."



Voice Data in Kinyarwanda is the key part when Rwanda designs public health AI application/ communication tool in term of "usability" → Text to Speech (TTS), Speech to Text (STT)

The figure is a screenshot of a data visualization. At the top, there is a header: "Voice Data in Kinyarwanda is the key part when Rwanda designs public health AI application/ communication tool in term of 'usability' → Text to Speech (TTS), Speech to Text (STT)". Below this, there is a small logo for the World Bank Group. The main content is a chart titled "Literacy Rate, Adult Total for Rwanda 79 %". The chart has a blue background with white text. It shows a single data point: "Literacy rate, adult total (% of people ages 15 and above) - Rwanda" with a value of "79". There are also some smaller, less legible labels and numbers below the chart.

Common Voice							Download the Dataset	
Version	SPEAK		LISTEN					Audio Format
	Date	Size	Recorded Hours	Validated Hours	License	Number of Voices		
Common Voice Delta Segment 22.0	6/25/2023	0 byte	1	1	CC-0	8	MP3	Spills (Age and Sex)
Common Voice Corpus 22.0	6/25/2023	57 GB	2,385	2,002	CC-0	1,174	MP3	
Common Voice Corpus 14.0	6/26/2023	56.81 GB	2,384	2,002	CC-0	1,114	MP3	
Common Voice Delta Segment 14.0	6/26/2023	42.05 MB	1	1	CC-0	11	MP3	
Common Voice Delta Segment 13.0	4/2/2023	0 byte	1	1	CC-0	21	MP3	
-Common Voice Corpus 13.0	3/15/2023	56.79 GB	2,384	2,002	CC-0	1,103	MP3	
Common Voice Delta Segment 12.0	12/22/2022	0 byte	1	1	CC-0	2	MP3	
Common Voice Corpus 12.0	12/15/2022	56.81 GB	2,384	2,001	CC-0	1,082	MP3	
Common Voice Delta Segment 11.0	11/1/2022	272.68 kB	1	1	CC-0	4	MP3	
Common Voice Corpus 11.0	9/21/2022	56.81 GB	2,384	2,001	CC-0	1,080	MP3	

 Hugging Face  large-scale Kinyarwanda speech dataset

Dataset · [DigitalUmuganda/Kinyarwanda-english-machine-translation-dataset](#) · 13 Mar · Follow · [DigitalUmuganda](#)

languages: English, Kinyarwanda · License: CC BY 4.0

[Dataset card](#) [Files and versions](#) [Community](#) 

Dataset Viewer   

Split (1)
train 

The dataset viewer is not available for this split.

Downloads last month

1

Models trained or fine-tuned on Dig

[DigitalUmuganda/zoeynet-kin](#)
Updated Jul 27, 2022 · 12

Kinyarwanda English Parallel Datasets for Machine translation

48,000 Kinyarwanda English Parallel datasets for machine translation, made by curating and translating normal Kinyarwanda sentences into English

 Hugging Face  Search models, datasets, user...

 Digital Umuganda  Company  Verified  Digital-Umuganda

Digital Umuganda → create and promote the voice infrastructure for African languages

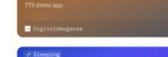

AI & ML Interests
African NLP, STT, TTS, Data collection


Team members 6



Binding
Tts Demo  111 demo app
 DigitalUmuganda Jan 13, 2024


Streaming
Rra Chatbot  RRA CHATBOT
 DigitalUmuganda Jan 13, 2024


Streaming
Rrachatbot  RRA CHATBOT Ask me related questions in English
 DigitalUmuganda Nov 06, 2024


Streaming
Kinyarwanda ASR With Amasaku 
 DigitalUmuganda Jan 13, 2024

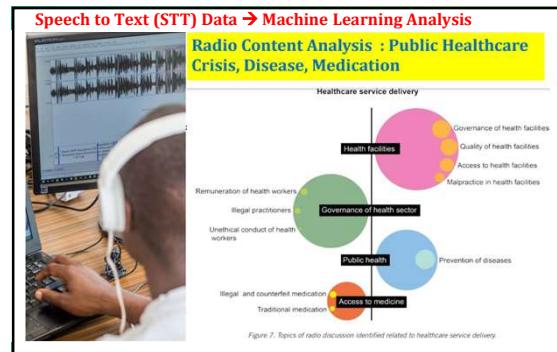
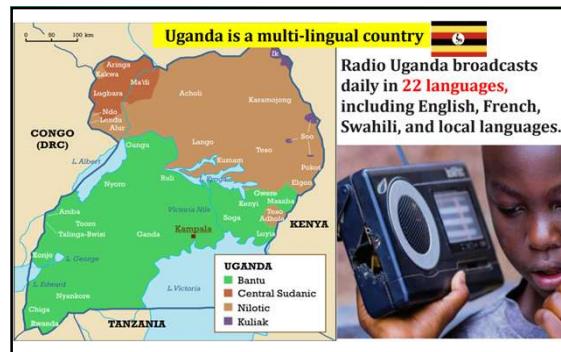

NLP
DigitalUmuganda Finetuned NLLB 
 DigitalUmuganda Jan 14, 2024


Streaming
Lingala Tts 
 DigitalUmuganda Mar 16, 2024

A young person with short hair is sitting at a desk in an office setting, working on a laptop. They are looking at the screen, which displays a video conference interface. There are two other monitors on the desk, one showing a document and another showing a slide with text. The office has blue chairs and desks in the background.



- The Radio Content Analysis Tool is combining of Machine Learning + Natural Language Process
- Converting radio content into text → Analyze
- Radio content mining analysis : Physically receive radio transmissions
- **Speech-recognition software** identifies speech vs nonspeech broadcasts and delete the latter
- Converting large unstructured dataset, containing and irrelevant data, into **smaller, structured datasets** of categorized text
- Extracting topics relevant to development : disease, public health, medicine
- Radio Content Analysis Tool enhances the effectiveness of disaster response by enabling real-time monitoring and rapid identification of urgent issues discussed on radio broadcasts.



4. The Limitation of Conversational AI Chatbots in Healthcare

Rwanda
Centre for the Fourth Industrial Revolution

How is Rwanda using generative AI to support community healthcare workers?

Ivory Coast-based Zencey is using digital technology to help close this healthcare gap

Image: Zencey

Zencey app provide Chatting based Virtual Medical Assistance, including symptom evaluation, medical advice, Get your health report

Discover the probable causes of your discomfort and what best to do

Your health report

Based on the check, it is possible that you are suffering from one or more of the following conditions:

- Common cold
- Flu
- Hay fever
- Migraine
- COVID-19 infection
- Covid-19 prevention

Based on the check, it is very likely that you are suffering from one or more of the following conditions:

- Common cold
- Flu
- Hay fever
- Migraine
- COVID-19 infection
- Covid-19 prevention

Footage: Zencey

Could Zencey chatbot provide a safe and accurate assessment?

Via conversations with an AI chatbot trained on medical literature

Preliminary assessment before a consultation

Illness Check Diagnosis ?

It's likely that you are suffering from a migraine.

Migraine

I've been dealing with a headache and a cold for since yesterday.

I see. Can you describe your headache a bit more? Is it more focused in a certain area?

Yeah, it's in the front of my head.

Alright. And for your cold, are you noticing any specific symptoms like a runny or stuffy nose, a scratchy throat, coughing, or fever?

- Zencey App integrated XUND's Medical API ("Third Party Software") to harness the power of artificial intelligence.

- AI chatbot has symptom checker, risk assessment capabilities → ?

- AI chatbot has authority to make a medical diagnosis → ?

- AI chatbot provides health report + suggest measures → ?

XUND

https://xund.ai/blog/using-ai-to-transform-healthcare-in-africa

Online pharmacy store

Diagnosis + Prescription???

Provide tech-enabled second opinions to your patients.

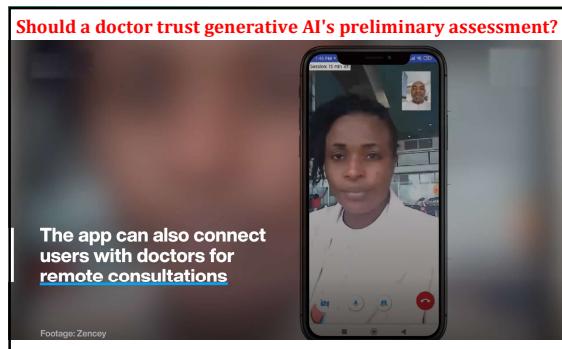
Illness Check

Here's an overview of your selected symptoms:

- Headache
- Nausea
- Light sensitivity
- Vomiting
- Dizziness

Get in touch See reference project CE 0123

Preliminary assessment before a consultation



Can AI Improve Medical Diagnostic Accuracy

HAI Stanford Human-Centered Artificial Intelligence

Potentially, An investigation into how well ChatGPT performs on its own and as a diagnostic aid for physicians reveals clinical shortfalls where the AI tool could be put to good use.

A study from Stanford researchers has highlighted the untapped potential of large language models, a form of artificial intelligence, to improve the accuracy of medical diagnoses and clinical reasoning.

The researchers presented a series of cases based on actual patients to the popular model ChatGPT-4 and to 50 physicians and asked for a diagnosis. Half of the physicians used conventional diagnostic resources, such as medical manuals and Internet search, while the other half had ChatGPT available as a diagnostic aid.

Overall, ChatGPT on its own performed very well, posting a median score of about 92—the equivalent of an “A” grade. Physicians in both the non-AI and AI-assisted groups earned median scores of 74 and 76, respectively, meaning the doctors did not express as comprehensive a series of diagnoses-related reasoning steps.

JAMA Network A Randomized Clinical Trial

RCT: Large Language Model Influence on Diagnostic Reasoning

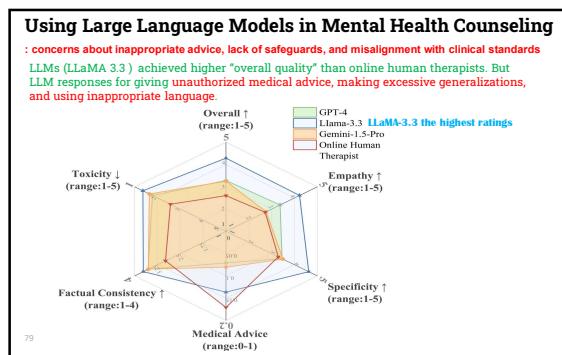
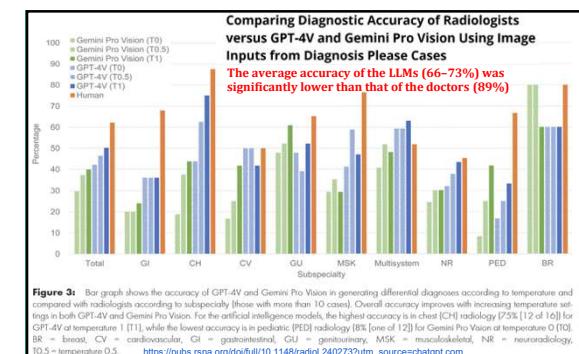
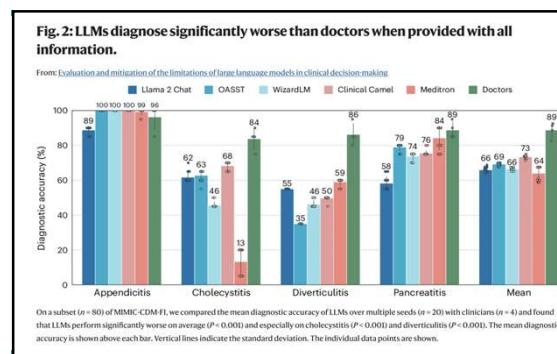
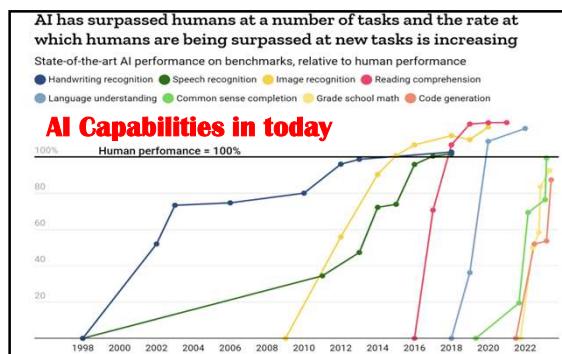
POPULATION 26 Attending physicians
24 Resident physicians
US-trained physicians with training in family medicine, internal medicine, and emergency medicine

INTERVENTION 50 Participants randomized
25 Generative artificial intelligence (AI) chatbot
Participants with access to conventional medical resources and AI chatbot were allocated 60 min to review up to 6 clinical vignettes

SETTINGS / LOCATIONS 3 Academic medical centers in the US

PRIMARY OUTCOME Diagnostic performance, assessed using a standardized rubric including differential diagnosis accuracy, appropriateness of supporting/opposing factors, and next evaluation steps, graded via blinded expert consensus

FINDINGS There was no significant difference in diagnostic performance between physicians in the chatbot group and the conventional resource group
Median diagnostic reasoning score
AI chatbot 76%
Conventional resources 74%
Adjusted difference between groups:
2 percentage points (95% CI, -4 to 8 percentage points).
 $P = .50$



5. AI supported Clinical Decision : Primary Care with AI Chatbot

Improve productivity and create better patient experiences with AI

OpenAI

Automate documentation and admin tasks to improve care quality

Improve the care journey for patients and clinicians through AI-assisted tools

Connect ChatGPT to your internal systems like Google Drive

Streamline clinical and operational workflows across teams

Project: Operations Impact Project Overview Project Status Today Hospital Directs Briefing Hospital Directs Briefing Yesterday Hospital Support Request

OpenAI **PENDA HEALTH**

HealthBench for Primary Care: A Real-World Study

In a study of 39,849 patient visits across 15 clinics, clinicians with AI Consult had a **16%** relative reduction in diagnostic errors and a **13% reduction in treatment errors compared to those without.**

Key Point

- **GPT-4o acted as a Safety Net for Primary Care**
- Evidence: AI can become a trusted part of standard care**

- Capable model: Penda's copilot used GPT-4o from August 2024, and models have improved rapidly since. Model performance is no longer the limiting factor.
- Clinically-aligned implementation: The copilot was co-developed with clinical users to ensure it genuinely supported—rather than disrupted—the flow of care.

82 <https://openai.com/index/healthbench/>

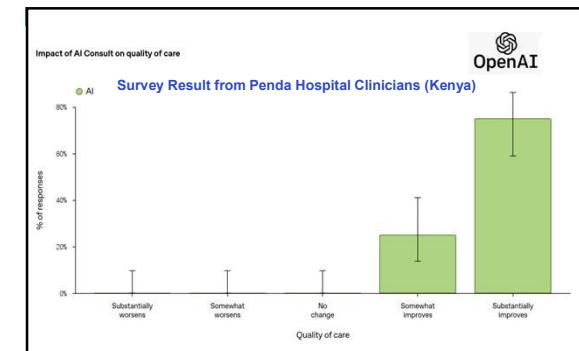
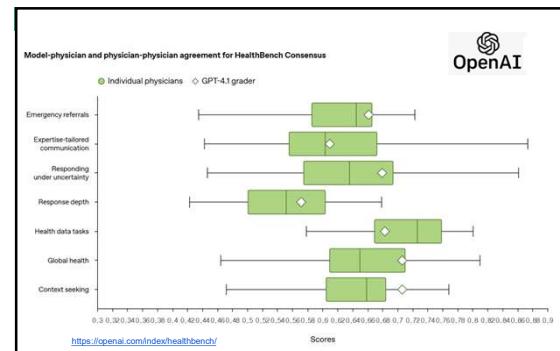
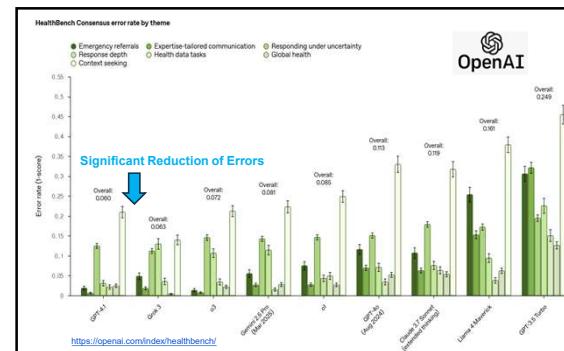
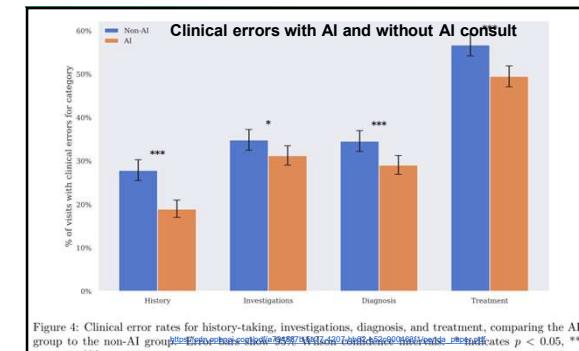
Initial documentation

AI Consult response: ●

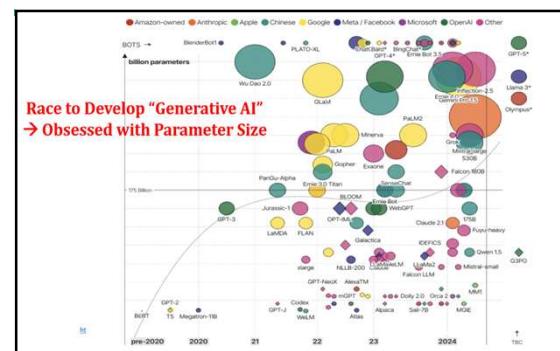
Final documentation

Figure 1: AI Consult is a safety net that runs in the background of a patient visit to identify potential errors. It was iteratively designed with clinicians, providing outputs with green/yellow/red severity and issuing alerts only when needed to reduce errors. In this example, AI Consult provided a red flag that helped a clinician identify and remove an unnecessary antibiotic prescription.

https://cdn.openai.com/pdf/a794887b-5a77-4207-bb62-e529004691/penda_paper.pdf



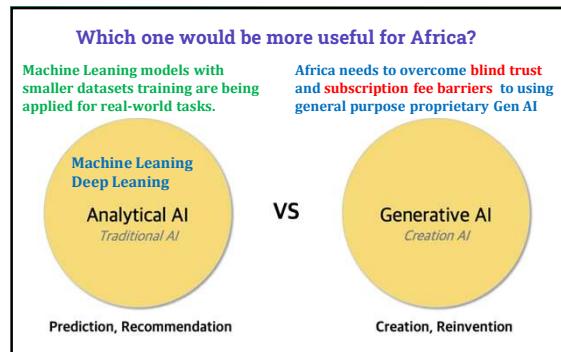
6. Analytic AI for Medical purpose is more useful than Generative AI



AI Safety Index

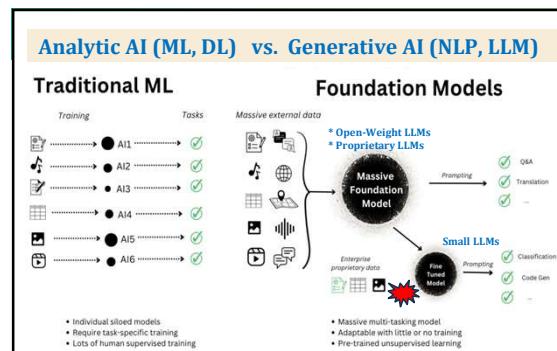
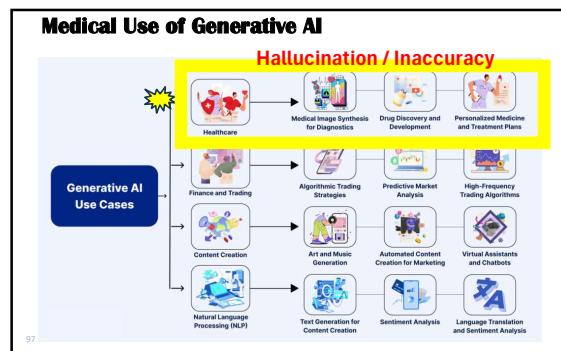
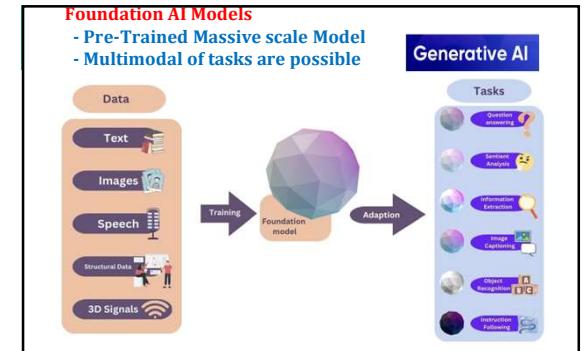
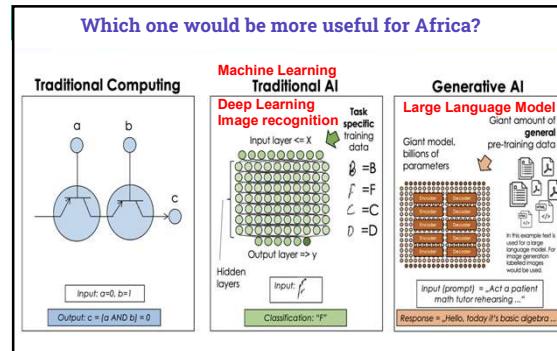
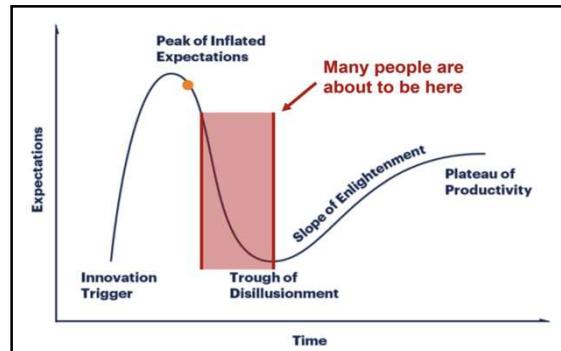
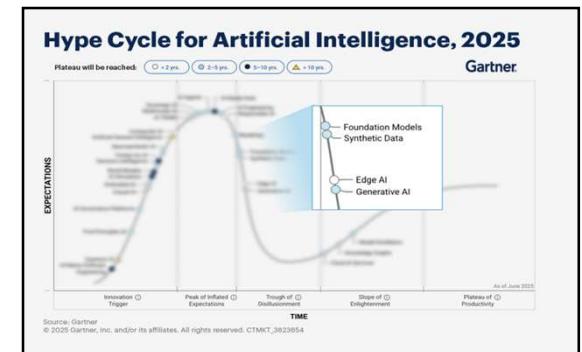
AI Company	Overall Grade	Overall Score
Anthropic	C+	2.64
OpenAI	C	2.10
Google DeepMind	C-	1.78
x.AI	D	1.23
Meta	D	1.06
Zhipu AI	F	0.62
DeepSeek	F	0.37

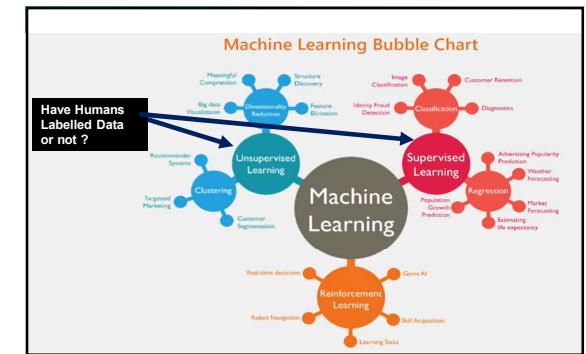
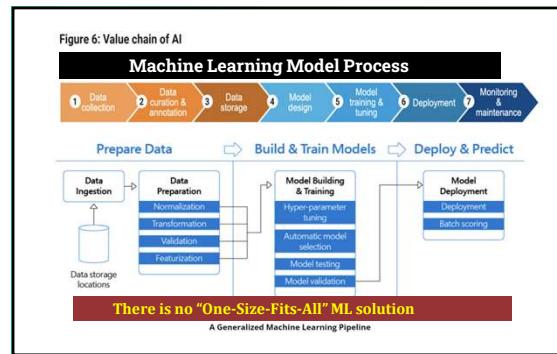
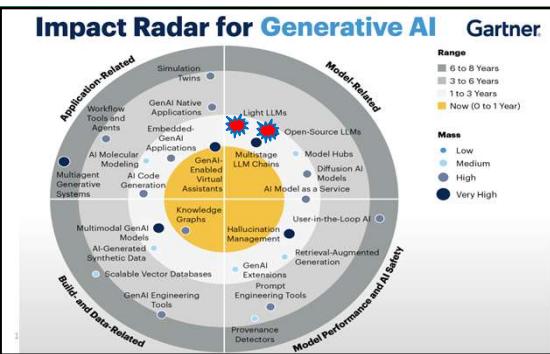
Grading: Uses the US GPA system for grade boundaries: A+, A, A-, B+, B, B-, C+, C, C-, D+, D, D-, F. Letter values corresponding to numerical values 4.3, 4.0, 3.7, 3.3, 3.1, 2.8, 2.5, 2.2, 2.0, 1.7, 1.4, 1.1, 0.8, 0.5, 0.2, 0.0.



"Foundation Models or Large Language Models (LLMs) alone hardly save Africa. They are too general and not mission-critical. It's time to wake up from the hype of Generative AI itself."

Instead, domain-specific, fine-tuned small LLMs or targeted machine learning applications with local date-sets are far more effective in addressing Africa's real needs."





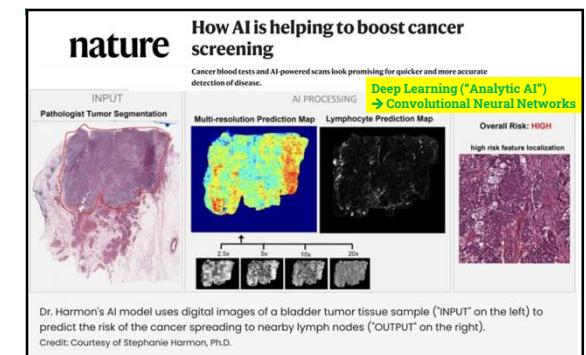
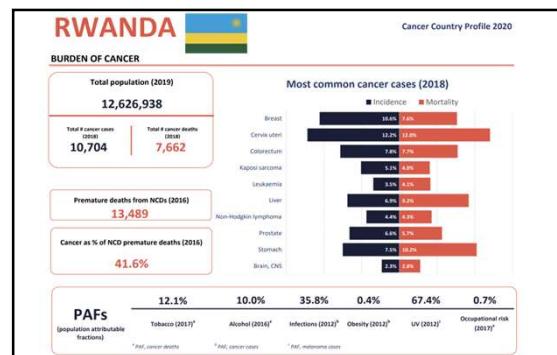
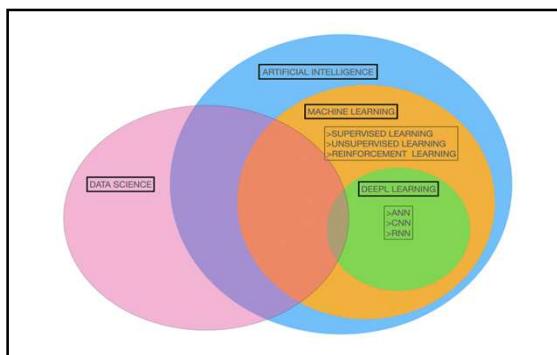
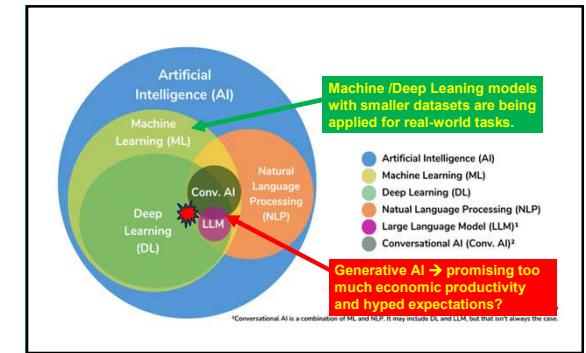
How is A.I. being utilized in Africa?

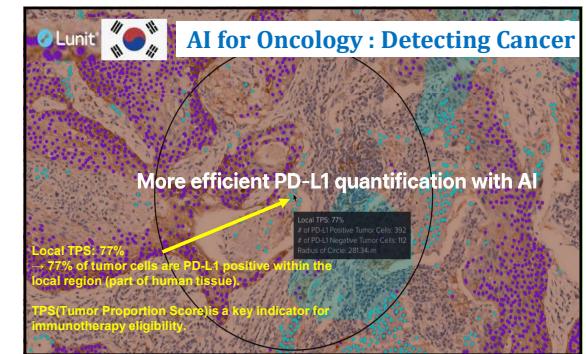
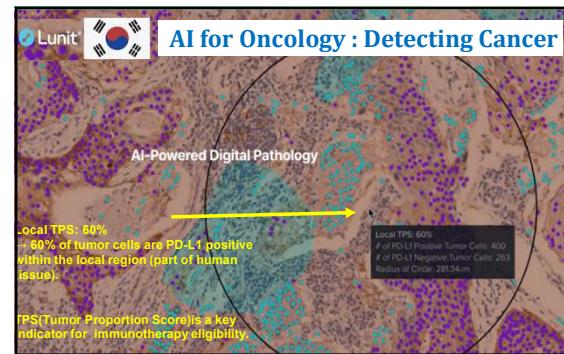
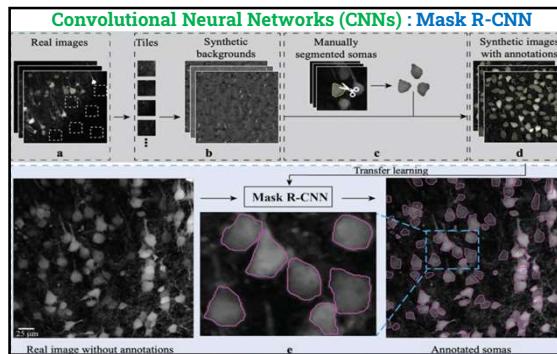
Mapping of A.I. applications across Africa | Our Methodology | Last Updated | Monday, September 04, 2023

Select a Region: Central Africa, East Africa, East And West Africa, Eastern Africa, North Africa.

Usage A.I. Type:

Type	Typical Technical Basis	Generative A.I.?
Data Analytics Tool (32 %)	Predictive / classification models, statistics & visualization	Machine Learning Tools !!! ✗ not generative AI
Chatbot (28 %)	Rule-based logic, intent classification, sometimes small LLMs	⚠ Mixed — generative if LLM-based
Decision Support (13 %)	Rule- or model-driven decision recommendations	Machine Learning Tools !!! ✗ not generative AI
Diagnostic Tool (9 %)	Classification / prediction models	Machine Learning Tools !!! ✗ not generative AI





CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning

We develop an algorithm that can detect pneumonia from chest X-rays at a level exceeding practicing radiologists.

Chest X-rays are currently the best available method for diagnosing pneumonia, playing a crucial role in clinical care and epidemiological studies. Pneumonia is responsible for more than 1 million hospitalizations and 50,000 deaths per year in the US alone.

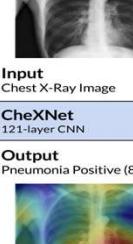
[READ OUR PAPER](#)

Medical CNNs is Deep Learning Convolutional Neural Networks



	F1 Score (95% CI)
Radiologist 1	0.383 (0.309, 0.453)
Radiologist 2	0.356 (0.282, 0.428)
Radiologist 3	0.365 (0.291, 0.435)
Radiologist 4	0.442 (0.390, 0.492)
Radiologist Avg.	0.387 (0.330, 0.442)
CheXNet	0.435 (0.387, 0.481)

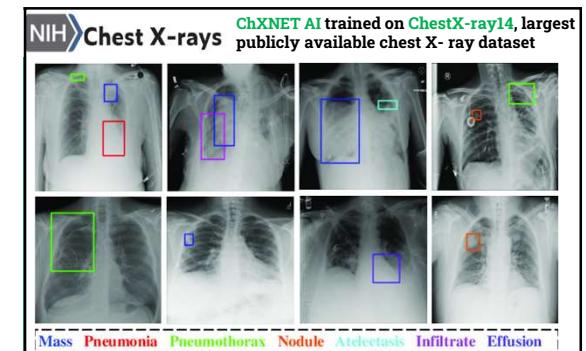
**ChXNET (CNN based model) detected pneumonia correctly 43.5% of the time, compared to doctors' average of 38.7%.
↓↓↓ the AI outperformed the radiologists.**



Input
Chest X-Ray Image

CheXNet
121-layer CNN

Output
Pneumonia Positive (85%)



The image shows four different models of RadiSen portable X-ray machines. From left to right: 1) A compact unit on a tripod stand. 2) A unit mounted on a tall, slender vertical pole. 3) A unit mounted on a tall, articulated vertical pole with a circular base. 4) A unit mounted on a large, articulated arm with wheels, designed for mobility.

The collage includes:

- A top banner with the text "Accurate diagnostics in any environment" and the URL "https://www.radisentech.com/en/axir-cx".
- A red button below the banner that says "Read the supporting papers ▶".
- Four chest X-ray images with overlaid white circles highlighting specific areas of interest.
- Performance data cards for four detectors:
 - (a) DR Detector (Vietnam): 97% accuracy
 - (b) DR Detector (Indonesia): 95% accuracy
 - (c) DR Detector (China): 96% accuracy
 - (d) DR Detector (Indonesia): 95% accuracy
- Performance data card for (f) CR Detector (USA): 94% accuracy.
- Performance data card for (g) DR Detector (Indonesia): 98% accuracy.
- Performance data card for (h) DR Detector (Indonesia): 95% accuracy.

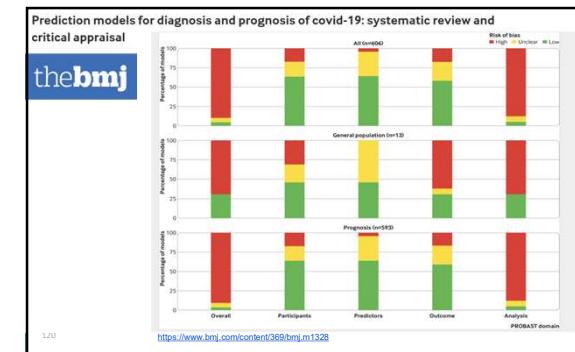
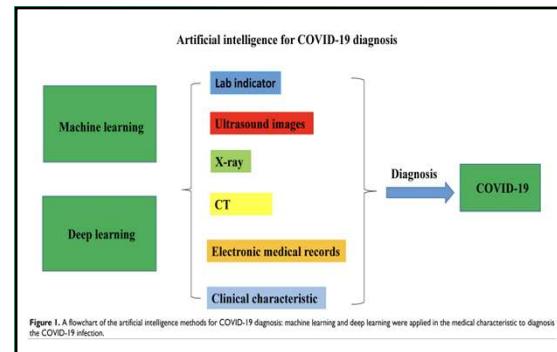
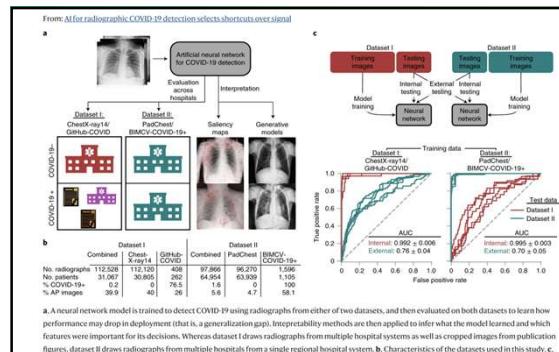
Hundreds of AI tools have been built to catch covid. None of them helped.

Some have been used in hospitals, despite not being properly tested. But the pandemic could help make medical AI better.

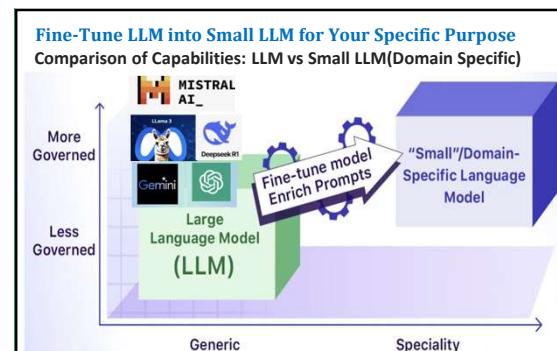
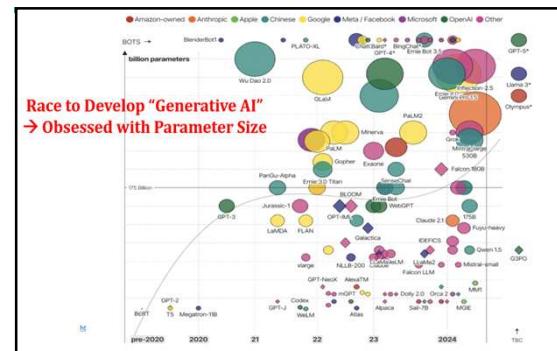
Not fit for clinical use
The poor quality of image data
Risk of Bias

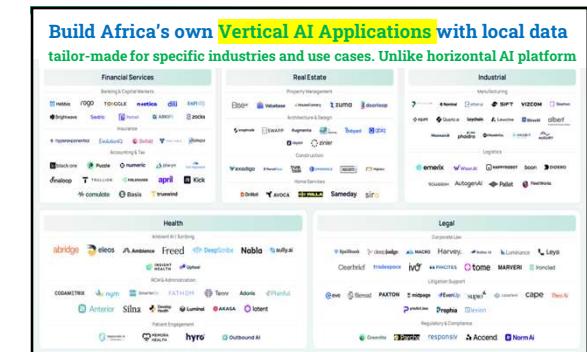
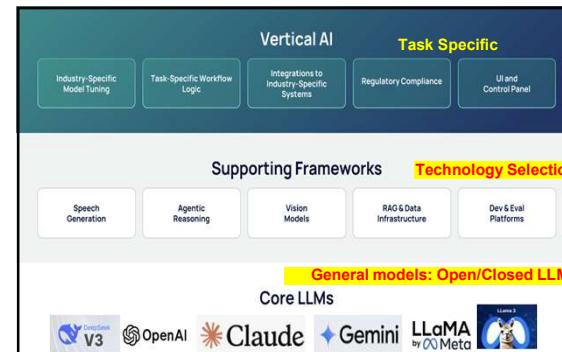
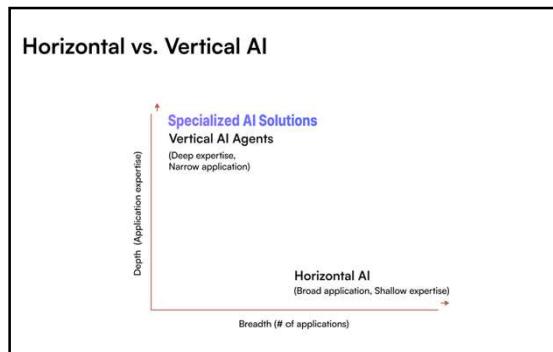
July 30, 2021

MIT
Technology
Review



7. Small LLM + Vertical Medical AI : Use Cases





Attribute	Horizontal AI	Vertical AI Agents
Scope	Broad, cross-industry (e.g., GPT-4, Siri)	Specialized by industry or workflow (e.g., Abridge for healthcare)
Data Depth	Limited domain-specific knowledge	Deep, fine-tuned domain knowledge
Adaptability	Moderate; needs customization	High; adapts in real-time to industry dynamics
Workflow Automation	Partial	Full, from end-to-end
Compliance Alignment	General guidelines	Embedded sector-specific compliance (e.g., Basel III, HIPAA)
Human Intervention	Frequent (especially for nuanced tasks)	Minimal; human-in-the-loop only for exceptions

Hippocratic AI: AI Agent Nurse: non-diagnostic, patient-facing applications

This screenshot shows the Hippocratic AI AI Agent Nurse interface. It features a central video call window with a doctor named 'Rachel'. To the left is a 'Call Agenda' sidebar with options like Identity Verification, Patient Check-In, Medication Review, Diet Review, CPH Symptom Self-Assessment, and Physical Activity. To the right is a 'Fictional Patient Info' section with details about a patient named Jane, including her gender, age, and medical history. The overall theme is non-diagnostic, focusing on patient-facing applications.

Abridge Clinician

ABRIDGE Generative AI for Clinical Conversations

This screenshot shows the Abridge Clinician mobile application. It features three main cards: 'Save hours a day on documentation with Abridge', 'Record patient-doctor conversations easily', and 'Get structured clinical notes in real-time with AI'. Each card includes a smartphone icon displaying the Abridge interface.

Abridge Clinician

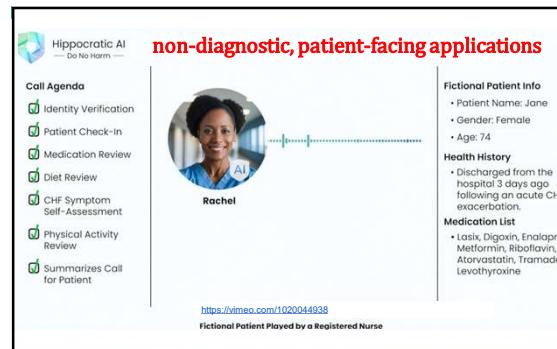
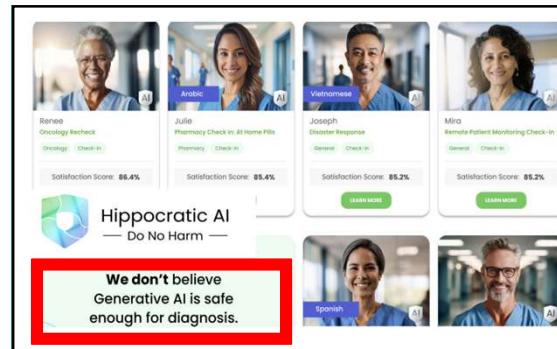
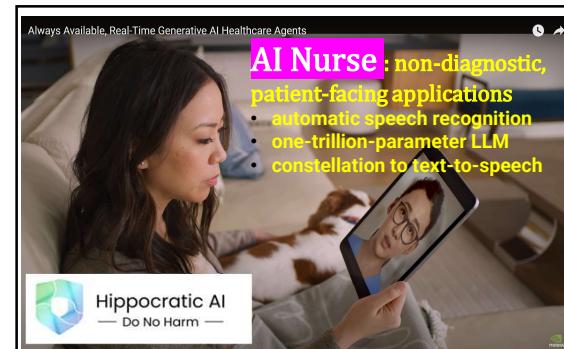
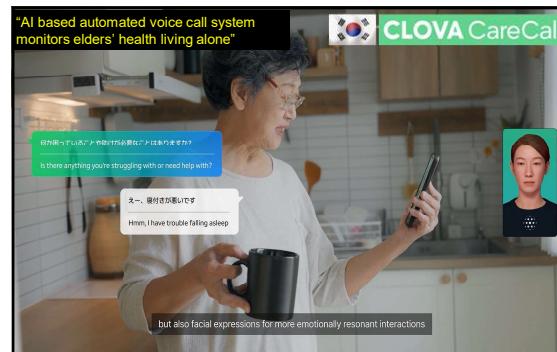
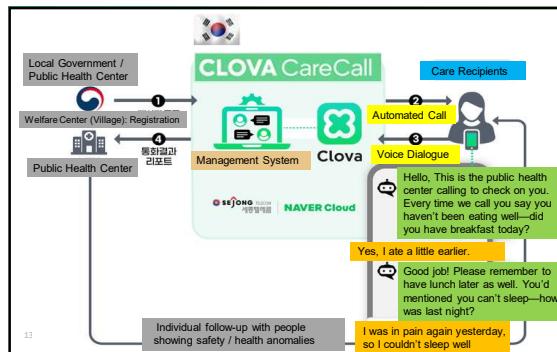
ABRIDGE Generative AI for Clinical Conversations

This screenshot shows the Abridge Clinician mobile application. It displays a series of cards illustrating its features: 'Get structured clinical notes in real-time with AI', 'Review and verify notes with Linked Evidence', 'Generate patient visit summaries in real-time', and 'Summarize conversations across 50+ specialties and 14+ languages'. Each card includes a small screenshot of the app's interface.

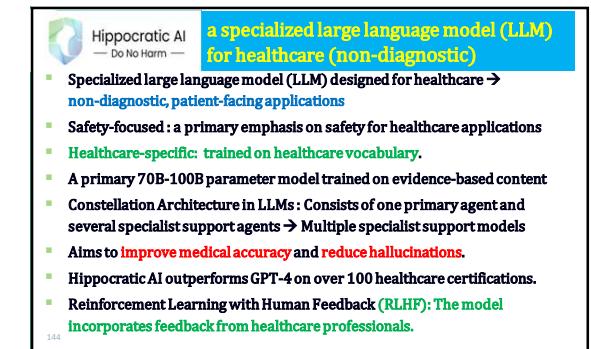
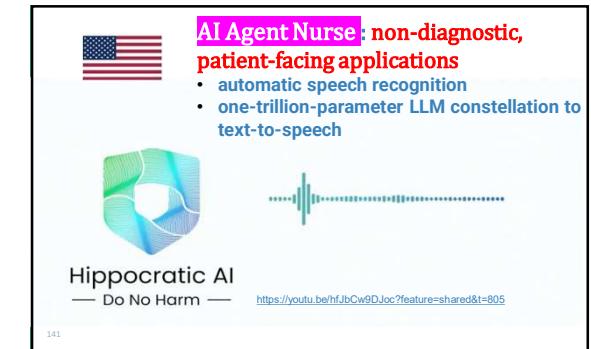
Why Vertical AI is gaining market attention ?

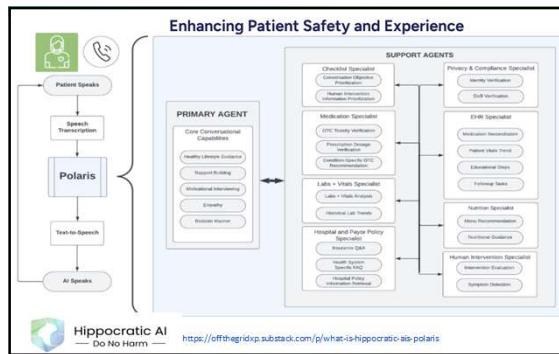
- Domain specialized Task Logic Integration:** Embed industry's legacy workflows and decision logic to support complex, role-specific tasks effectively.
- Precise Purpose Definition:** Designed specifically to address specific needs within particular industries. Compared to general-purpose AI, it achieves *more concrete results*.
- Seamless System Integration:** Seamlessly integrating Vertical AI with legacy systems, tools and processes
- Regulatory Compliance:** Design AI systems to meet strict industry regulations, critical in fields like healthcare, finance, and law.
- Targeted Distribution:** Use trusted industry channels and partnerships to reach users credibly





8. The Rise of AI Agent : AI Nurse





Patient: I am thinking about going to Bonefish later today for dinner. What do you recommend I eat that is relatively healthy?

Polaris: I see that you're managing Stage 3A Chronic Kidney Disease and Hypertension. Given these conditions, it's beneficial to keep an eye on your meal nutrients. For each meal, we aim to keep your calorie intake to four hundred and fifty, calcium to three hundred, sodium to six hundred, protein to fourteen point four grams, fat to one hundred and thirty-five grams, and phosphorus to two hundred and forty milligrams. These recommendations are designed to be Research-based Dietary Allowance. With these House Salad or Caesar Salad. Both have carbohydrates and sodium. Just remember to take into account the patient's medical conditions with Kidney Disease and the restaurant to be visited, specific menu items and portion sizes are recommended.

Patient: I am thinking about going to Bon recommend I eat that is relatively healthy?

GPT-4 (General Purpose AI): provides general advice for patients with chronic kidney disease, such as recognizing the condition and choosing menu items low in sodium, phosphorus, and carbohydrates. However, it does not offer specific menu recommendations or detailed nutritional information tailored to individual patients.

GPT-4: That sounds like a fun dining experience! Given your dietary restrictions due to Chronic Kidney Disease, it's important that we choose options that are lower in sodium and carbohydrates. Perhaps you could opt for steamed vegetables, which are typically lower in sodium. Steamed vegetables would make a great side, too. Remember to ask the server about the sodium content of the dishes or if they can prepare your meal with less salt. How does that sound?

