Transforming cancer surveillance

Emerging directions in multimodal health data analytics

Dr. Pete Clardy, MD

Lead, Clinical Enterprise Team

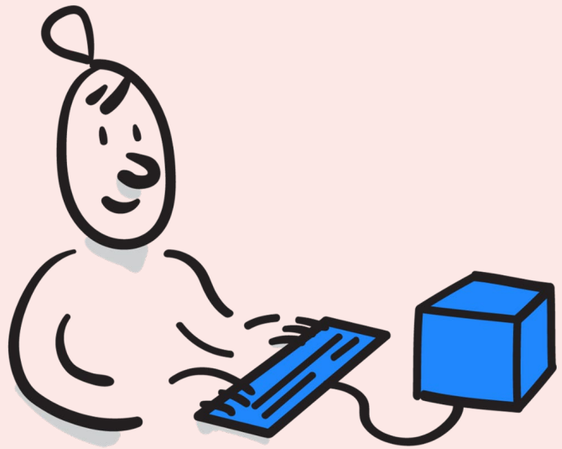
Google Health

Google Health

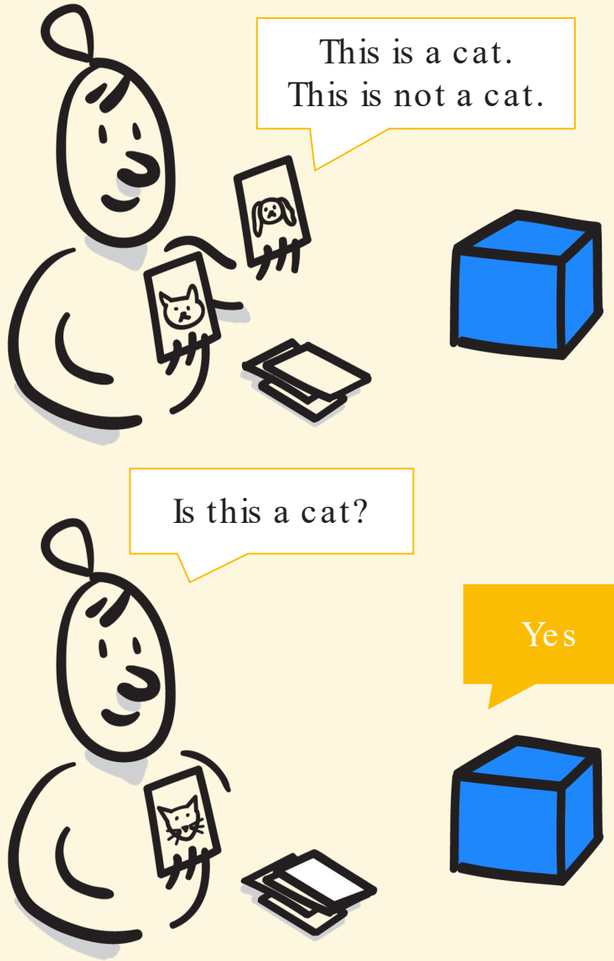
We are in a new era of generative AI…

Cat:
type: animal
legs: 4
ears: 2
fur: yes
likes: yarn, catnip

(etc …)

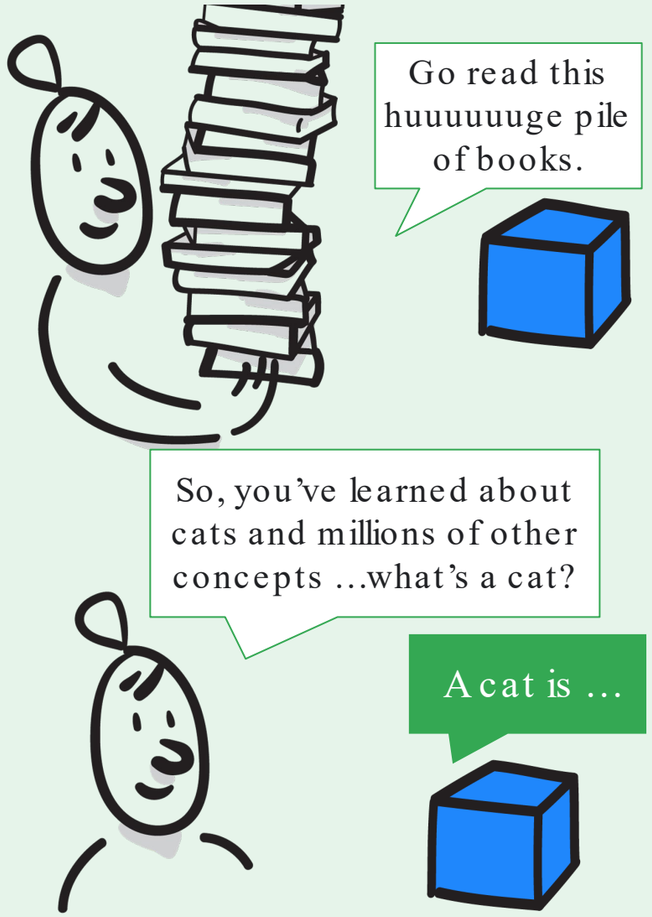


Traditional programming



Wave of neural networks

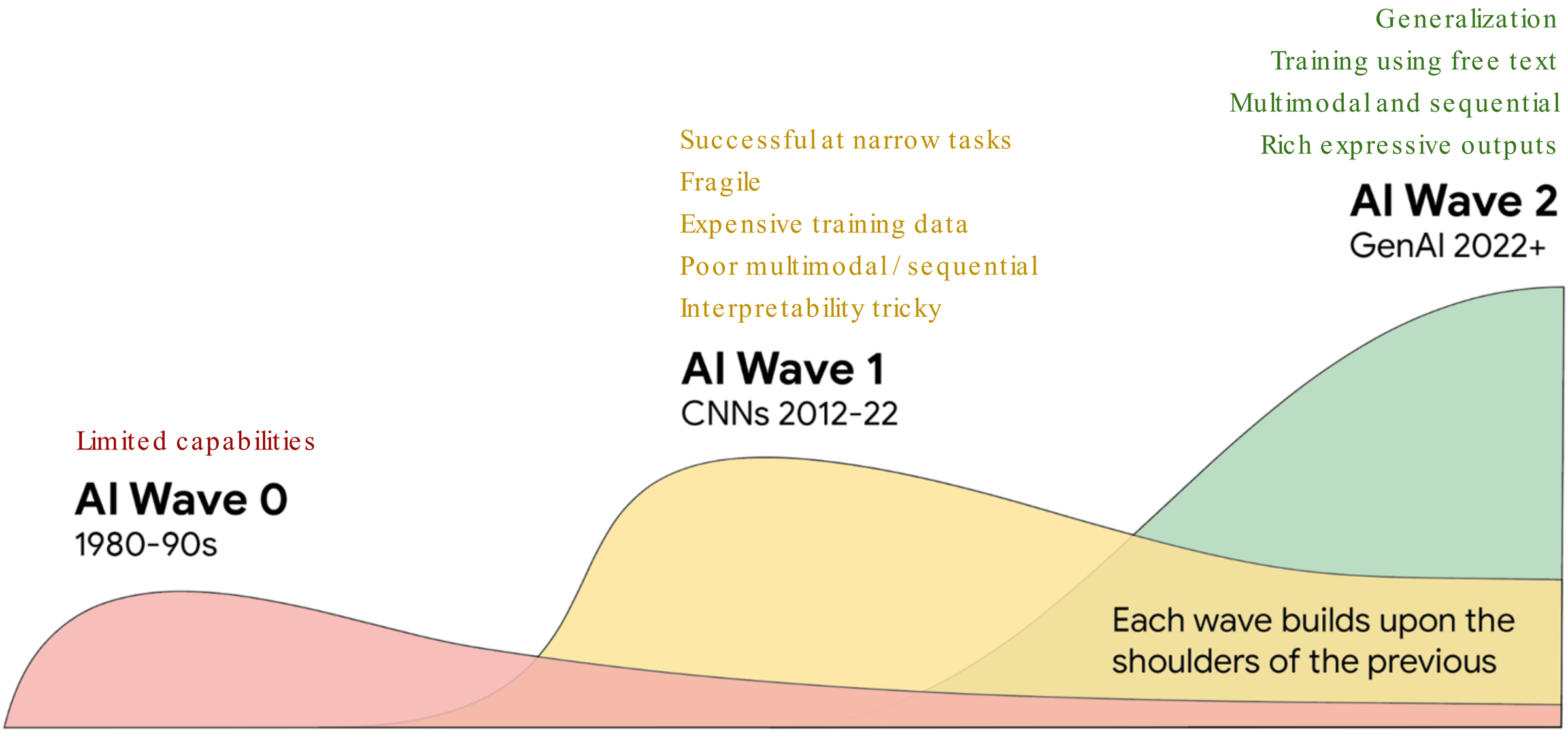
~2012



Generative language models

GPT, Gemini, etc

We are in a new era of generative AI...



JAMA 2024; 331:242

Initial efforts focused on narrow applications:

- Highly constrained inputs
- Simple task
- High volume / lots of data
- Constrained outputs

Lung cancer detection

Diabetic retinopathy detection

Genomics

Pathology classifiers / predictors

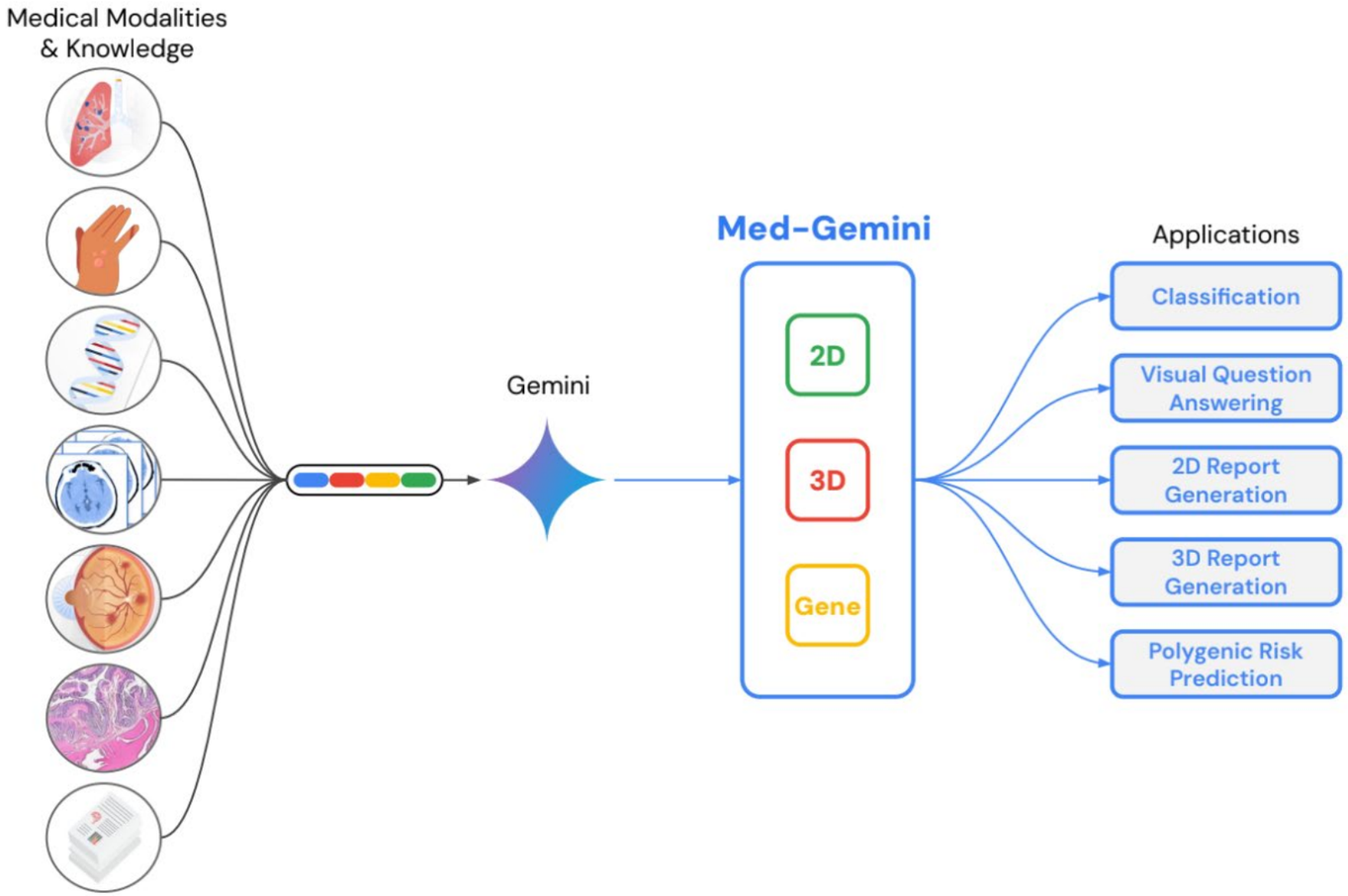
Breast cancer detection

Dermatology

AKI prediction

...and many others

Attention has shifted to multimodal models:



Google Health

Google DeepMind and Google Research 2024-5-7

Advancing Multimodal Medical Capabilities of Gemini

Google Research and Google DeepMind †

Many clinical tasks require an understanding of specialized data, such as medical images and genomics, which is not typically found in general-purpose large multimodal models. Building upon Gemini's multimodal models, we develop several models within the new *Med-Gemini* family that inherit core capabilities of Gemini and are optimized for medical use via fine-tuning with 2D and 3D radiology, histopathology, ophthalmology, dermatology and genomic data. Med-Gemini-2D sets a new standard for AI-based chest X-ray (CXR) report generation based on expert evaluation, exceeding previous best results across two separate datasets by an absolute margin of 1% and 12%, where 57% and 96% of AI reports on normal cases, and 43% and 65% on abnormal cases, are evaluated as "equivalent or better" than the original radiologists' reports. We demonstrate the first ever large multimodal model-based report generation for 3D computed tomography (CT) volumes using Med-Gemini-3D, with 53% of AI reports considered clinically acceptable, although additional research is needed to meet expert radiologist reporting quality. Beyond report generation, Med-Gemini-2D surpasses the previous best performance in CXR visual question answering (VQA) and performs well in CXR classification and radiology VQA, exceeding SoTA or baselines on 17 of 20 tasks. In histopathology, ophthalmology, and dermatology image classification, Med-Gemini-2D surpasses baselines across 18 out of 20 tasks and approaches task-specific model performance. Beyond imaging, Med-Gemini-Polygenic outperforms the standard linear polygenic risk score-based approach for disease risk prediction and generalizes to genetically correlated diseases for which it has never been trained. Although further development and evaluation are necessary in the safety-critical medical domain, our results highlight the potential of Med-Gemini across a wide range of medical tasks.

† See Contributions and Acknowledgments section for full author list.
Corresponding authors: {dangolden, shekazizi, kellych, roryp}@google.com.

Problem statement:

- Longitudinal and comprehensive patient records are inherently multimodal and heterogeneous
- Questions we ask of healthcare data are often complex and require nuance
- Pre-processing data and questions increases cost, rigidity, and signal loss
- Interop and FHIR have significantly improved this dynamic, yet it still remains a challenge

Implications for future clinical information systems design:

Clear Box → Opaque Box

	Deterministic Search/SQL on FHIR Record	Hybrid RAG on FHIR+ Record	Generative Long Context LLM on Raw Record
Pre-Process Data	- High storage cost - Higher map/merge burden	- Higher storage cost + Lower map/merge burden	+ No map/merge burden + Multimodal ingest
Process Query	+ Lower latency & cost - Requires explicit logic	- Requires concept expansion & knowledge graphs	- Higher latency & cost + High flexibility w/ prompting
Post-Process Response	+ Accurate & consistent - Requires parsing	+ Grounding - Hallucination risk	+ Human ready - Hallucination risk
Use Cases	End Users / System to system Production	Human-in-the-loop Pilot	Professional-in-the-loop Proof of Concept