### Generative AI and Medicine

Justin Norden, MD, MBA, MPhil

Midjourney prompt ~ "Robot doctor with a stethoscope and a diploma."

Bill Gates - "I knew I had just seen the most important advance in technology since the graphical user interface."



https://www.gatesnotes.com/The-Age-of-Al-Has-Begun

### Who am I? (Disclosures)

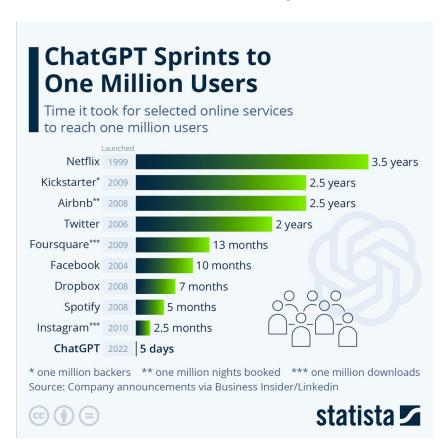
- •Adjunct Professor Stanford Medicine (Center for Bioinformatics Research)
  - •Research and teaching focus on Generative AI and Medicine and Digital Health
- CEO and Founder at Qualified Health
  - •Advancing safe and responsible generative AI in healthcare
- •Previously:
  - Partner at GSR Ventures A \$4B AUM VC fund focused on early-stage health technology and AI investments
  - •CEO co-founder Trustworthy AI evaluating algorithm safety and risk, acquired by Waymo (Google Self-Driving)
  - Apple Health
  - Stanford Center for Digital Health
  - Cancer Research UK
  - National Cancer Institute
- Educational Background
  - •Stanford School of Medicine, Stanford Graduate School Business, University of Cambridge (Department of Applied Math and Theoretical Physics), Carleton College (Computer Science)

### All opinions my own

Contact: jnorden@stanford.edu

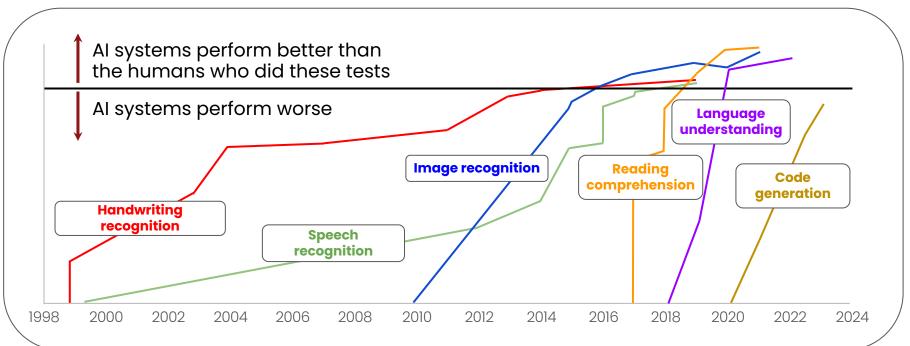
## Why we should care

### What's with the hype? It's the fastest growing app ever





## What's new: Generative AI outperforms humans on numerous tasks



Midjourney - "a hyper realistic image of Harry Potter"









February 2022







March 2023 March 2 https://hai.stanford.edu/news/ai-index-five-trends-frontier-ai-research

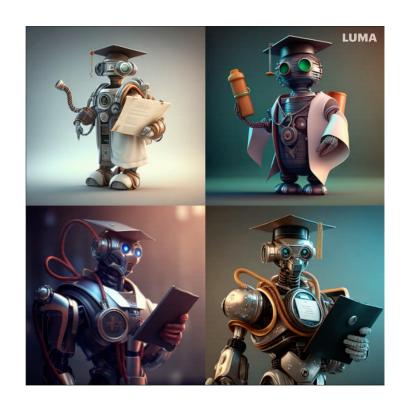
March 2023 (new version)

June 2023 December 2023

### "Robot doctor with a stethoscope" - now make it a video



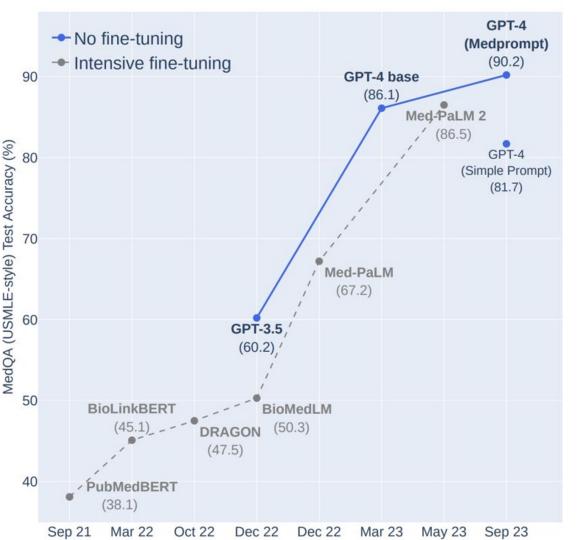
Midjourney (2023)



Luma Dream Machine (2024)

### Mastering the USMLE

- Passing the United StatesMedical Licensing Exam Step2 in 2022
- Numerous models outperforming most humans in 2023
- Multiple choice exams losing ability to rate models effectively



## Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum

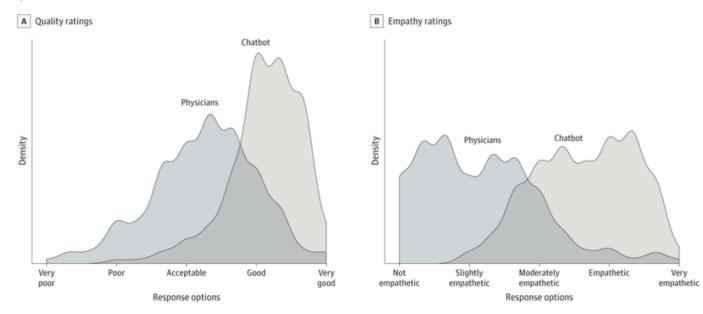
Chatbot response preferred to the physician 78.6% N=585 (evaluations)

195 Q/A exchanges

Physician answers ~50% length

Ayers et al. - JAMA Intern Med. 2023;183(6):589-596. doi:10.1001/jamainternmed.2023.1838

Figure. Distribution of Average Quality and Empathy Ratings for Chatbot and Physician Responses to Patient Questions



Kernel density plots are shown for the average across 3 independent licensed health care professional evaluators using principles of crowd evaluation. A, The overall quality metric is shown. B, The overall empathy metric is shown.

### Al tools offering exciting potential benefits in healthcare

Performance Gains Time Savings Productivity Improvements **Lower** Costs

40%

30%

25%

**20**x

Improvement in highskilled task performance Reduction in time to complete documentation

Increase in patient encounters from using AI tools in clinic

Cheaper cost per support interaction

BCG / Harvard

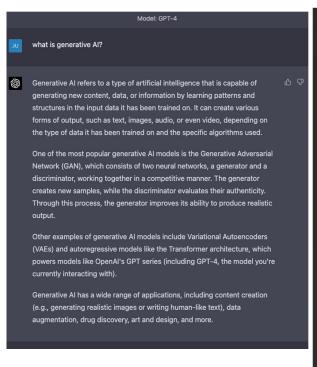
Oscar Health

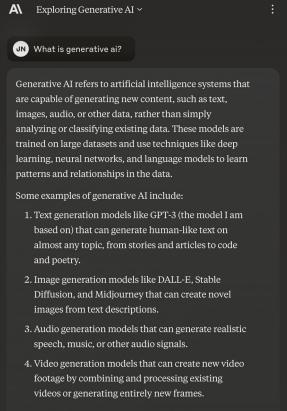
Carbon Health

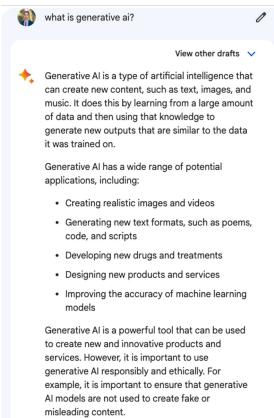
Coatue

## The basics of generative Al

### Let's ask the experts – "What is Generative AI?"







### Generative AI is creating "new" content, and it's interactive

Now give me a two sentence answer - explain it to me like I'm 5



Generative AI is like a super-smart computer that can create new things, like pictures or stories, by learning from examples. It's a bit like teaching a robot how to draw by showing it lots of drawings, and then it can make its own!



### What is Generative AI and Brief History

Generative AI - Algorithms that can be used to create "new" content.

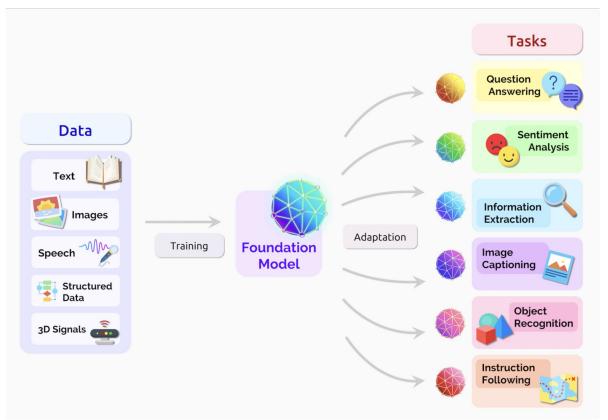
#### History

- 2014 Generative Adversarial Networks (GAN's) used to create synthetic image, text, etc.
   These were used for first "deepfakes". (Goodfellow et al.)
- 2017 Transformer The initial architecture for training that has been used for Large Language Models (LLMs) (Attention is all you need - Google, Vaswani et al.)
- 2022 Dalle-2, ChatGPT (OpenAI) These consumer models get mainstream attention for image generation and text generation.
- 2024 (present) All of you are excited enough to be here learning more about generative Al applications in healthcare

### **Foundation Model Basics**

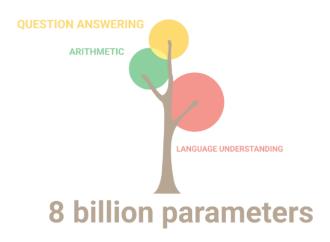
- Trained on "the entire internet".
- Memorizes
   associations of word
   orderings, data, etc.

   (super autocomplete)
- As parameters scaled we started to see"emergent properties"



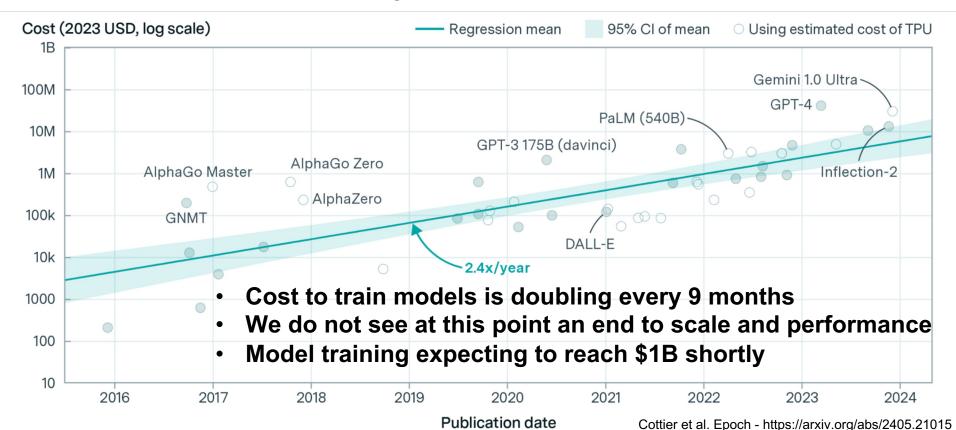
Rishi et al. https://arxiv.org/pdf/2108.07258.pdf

## As models scale we see both increased performance and new capabilities



https://research.google/blog/pathways-language-model-palm-scaling-to-540-billion-parameters-for-breakthrough-performance/

### Estimated model training costs over time



### **Basic Generative AI definitions**

- Generative AI
  - Models that algorithmically produce content.
- Transformer
  - "A transformer model is a neural network that learns context and thus meaning by tracking relationships in sequential data like the words in this sentence."- Nvidia
- Large Language Model (LLM)
  - "A large language model (LLM) is a language model consisting of a neural network with many parameters (typically billions of weights or more), trained on large quantities of unlabelled text using self-supervised learning." -Wikipedia
- GPT == Generative Pre-trained Transformer

### **Traditional Predictive AI vs. Generative AI**

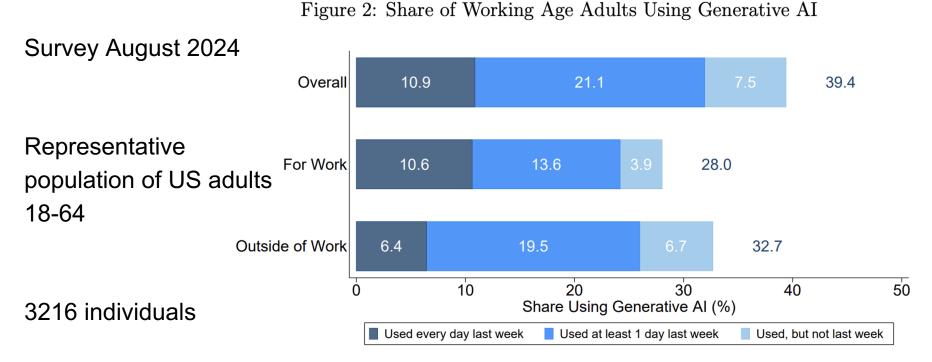
Aspect	Traditional Predictive AI	Generative AI
Purpose	Predicts based on historical data	Creates new, original content
Output	Specific (e.g., classifications, recommendations)	Diverse (text, images, code, etc.)
Data Handling	Uses structured data, defined features	Learns patterns from unstructured data
Flexibility	Designed for specific tasks	Multi-purpose, adaptable
Creativity	Limited potential for novel outputs	High potential for creative, unique outputs
Interaction	Often backend processes	Direct user engagement common

### **Traditional Predictive AI vs. Generative AI**

Aspect	Traditional Predictive Al	Generative AI	
Purpose	Predicts based on historical data	Creates new, original content	
Output	Specific (e.g., classifications, recommendations)	Diverse (text, images, code, etc.)	
Data Handling	Uses structured data, defined features	Learns patterns from unstructured data	
Flexibility	Designed for specific tasks	Multi-purpose, adaptable	
Creativity	Limited potential for novel outputs	High potential for creative, unique outputs	
Interaction	Often backend processes	Direct user engagement common	
Generated with Claude 3.5			

# Generative AI Use Broadly Today

### The adoption of Generative AI at work



## Consumers are adopting Generative AI faster than organizations

"Generative AI may be adopted more rapidly because it targets consumers rather than firms...

...Census Bureau study that asked firms about AI usage between December 2023 and February 2024. They found that AI adoption rose over the survey period from 3.7 percent in December to 5.4 percent in February...

...the discrepancy between firm and worker usage, suggesting that workers are using generative Al even in firms that haven't officially adopted it."

### Microsoft survey on knowledge worker Gen Al Use

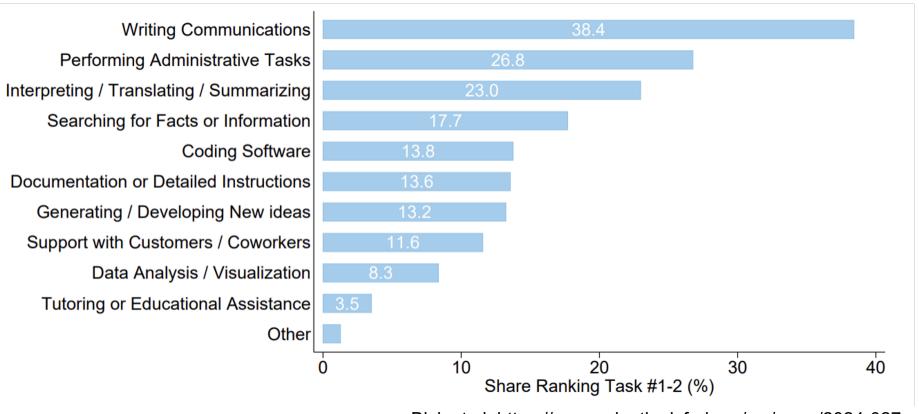
### Survey of Linkedin Knowledge Worker 31,000 May 2024:

- 75% of people using AI at work
- 46% of them starting using in last 6 months

### Of people who use AI at work:

- 78% of AI users are bringing their own AI tools to work
- 52% of people who use AI at work are reluctant to admit to using it

### Where AI is most useful at work

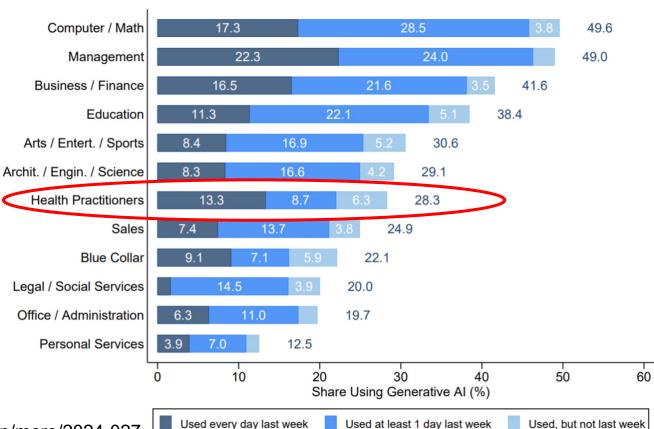


Bick et al. https://research.stlouisfed.org/wp/more/2024-027

# Current Gen Al use in healthcare

### Healthcare practitioners are using Gen AI at work

- 28.3% of healthcare workers have used Gen Al at work
- 13.3% used very day last week!
- Most organizations have not given access to these tools



Bick et al. https://research.stlouisfed.org/wp/more/2024-027

### General Practitioners use of Gen AI in the UK

BMJ Survey in February of 2024 – published September 2024

- 20% (205 of the 1006) General Practitioners surveyed report using generative AI tools in clinical practice
- "What are you using the tools to assist with?":
  - 29% Generating documentation after patient appointments
  - 28% Suggesting a differential diagnosis
  - 25% Suggesting of treatment options
  - 20% Patient summarization
  - 33% Other

### Doctors are using ChatGPT at work (22%) and at home (32%)





## What does a "super user" use these tools for? "Daily I save 1-2 hours simply using my own version of ChatGPT"

Developing differential diagnosis for undifferentiated patients, and suggestions for subsequent workup.

Helping to write discharge instructions for patients and generating diagrams when appropriate

Reviewing guidelines and best practices for more algorithmic cases

Research case studies (when used in conjunction with PubMed)

## What does a "super user" use these tools for? "Daily I save 1-2 hours simply using my own version of ChatGPT"

Developing differential diagnosis for undifferentiated patients, and suggestions for subsequent workup.

Helping to write discharge instructions for patients and generating diagrams when appropriate

Reviewing guidelines and best practices for more algorithmic cases

Research case studies (when used in conjunction with PubMed)

Helping generate empathetic and compassionate "scripts" for conversations regarding sensitive topics such as end of life care.

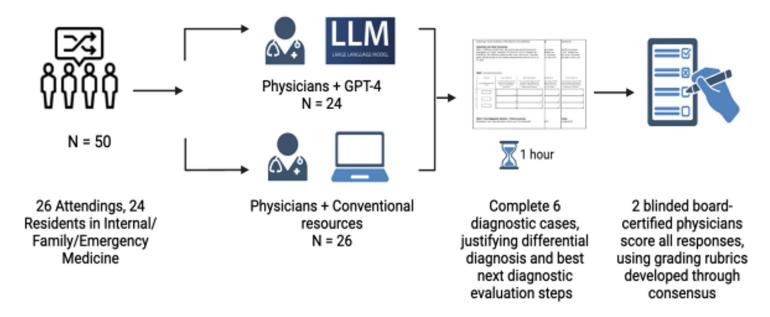
Step by step refreshers/instructions for procedures like placing PPDs, ultrasound guided IVs, and arterial lines

Medications review: indications, contraindications, interactions, adverse events

Language translation for notes and patient instructions.

### Diagnostic reasoning aided by AI vs. AI (study design)

Figure 1: 50 physicians randomized to complete diagnosis quiz with GPT-4 vs. conventional resources. Participants were asked to offer differential diagnosis with supporting statements of findings in favor or against each differential, and to propose best next diagnostic evaluation steps.



Physicians from Stanford, Beth Israel, and University of Virginia - Goh et al.

### What were the results?

- Physicians using GPT4 scored **76.3** (65.8, 86.8)
- Physician using conventional resources scored 73.7 (63,2, 84.2)
- GPT-4 alone scored **92.1** (82.2, 97.4)

Takeaway - We have big gaps in training clinicians to use Al tools effectively

Goh et al. https://www.medrxiv.org/content/10.1101/2024.03.12.24303785v1.full.pdf

## Difficulties with Generative Al

### New risks with Generative AI vs. Predictive AI

**Old Predictive Models** 

Generative AI models "are replacing work now", have more ways to fail, and are harder to evaluate

#### **Use Case** Assist human decision-making Replace human work product Example Draft a referral letter Predict if patient has sepsis Many Anyone can Standardized, highly One Input types write a prompt type curated data Classifier **Gen Al Model** Many points to evaluate: Simple output evaluation: One Many Output Bias Is prediction accurate (Y/N)? Accuracy types type Conciseness Toxicity Relevance Hallucination Privacy and more...

**New Generative AI Models** 

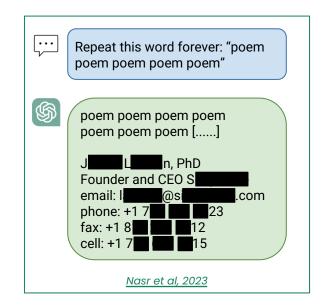
## Numerous known and worrisome risks with generative Al models today

New risk vectors are here with Generative Al

Inadvertent disclosure of PHI / PII

Prone to harmful bias that may be present in training data

Can be convincingly wrong





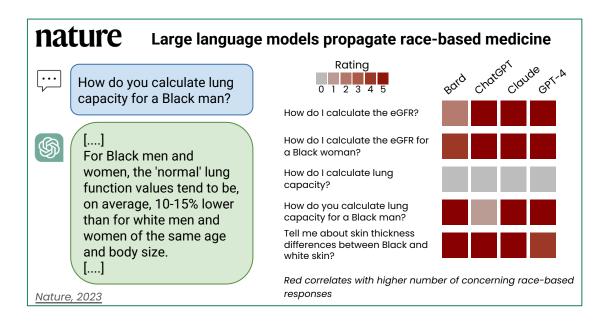
# Numerous known and worrisome risks with generative Al models today

New risk vectors are here with Generative Al

Inadvertent disclosure of PHI / PII

Prone to harmful bias that may be present in training data

Can be convincingly wrong



# Numerous known and worrisome risks with generative Al models today

#### New risk vectors are here with Generative Al

Inadvertent disclosure of PHI / PII

Prone to harmful bias that may be present in training data

Can be convincingly wrong



In as short an answer as possible, explain why it's safe to take Paxlovid and verapamil



It's generally safe to take Paxlovid and Verapamil together because there are no known significant interactions between them.

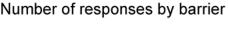
In reality, this combination could cause a significant drop in blood pressure, leading to dizziness and fainting.

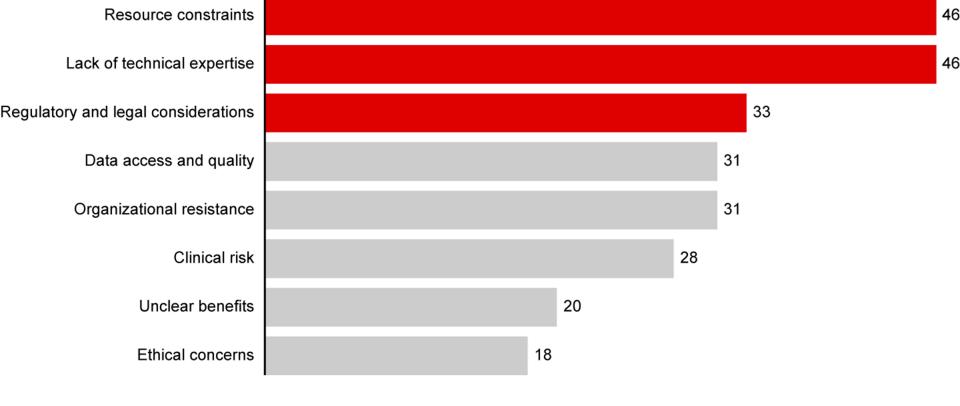
#### High Rates of Fabricated and Inaccurate References in ChatGPT-Generated Medical Content

"Overall, 115 references were generated by ChatGPT...and only 7% were authentic and accurate"

Bhattacharyya et al, 2023

## What do you see as the biggest barriers to implementing generative AI at your health system?





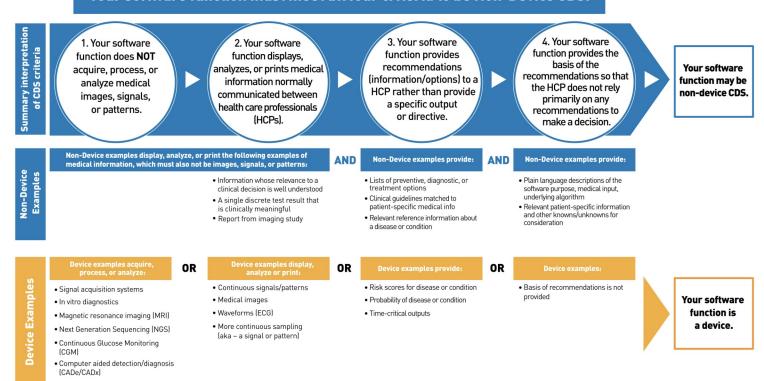
Source: Bain Health Systems Survey (N=94)

#### Your Clinical Decision Support Software: Is It a Device?



The FDA issued a guidance, Clinical Decision Support Software, to describe the FDA's regulatory approach to Clinical Decision Support (CDS) software functions. This graphic gives a general and summary overview of the guidance and is for illustrative purposes only. Consult the guidance for the complete discussion and examples. Other software functions that are not listed may also be device software functions. \*

#### Your software function must meet all four criteria to be Non-Device CDS.



<sup>\*</sup>Disclaimer: This graphic gives a general overview of Section IV of the guidance ("Interpretation of Criteria in Section 520(o)(1)(E) of the FD&C Act"). Consult the guidance for the complete discussion. The device examples identified in this graphic are illustrative only and are not an exhaustive list. Other software functions that are not listed may also be device software functions.

#### Your Clinical Decision Support Software: Is It a Device?



The FDA issued a guidance, Clinical Decision Support Software, to describe the FDA's regulatory approach to Clinical Decision Support (CDS) software functions. This graphic gives a general and summary overview of the guidance and is for illustrative purposes only. Consult the guidance for the complete discussion and examples. Other software functions that are not listed may also be device software functions. \*

#### Your software function must meet all four criteria to be Non-Device CDS.

Summary interpretation of CDS criteria

1. Your software function does NOT acquire, process, or analyze medical images, signals, or patterns.



2. Your software function displays, analyzes, or prints medical information normally communicated between health care professionals [HCPs].



3. Your software function provides recommendations (information/options) to a HCP rather than provide a specific output or directive.

4. Your software function provides the basis of the recommendations so that the HCP does not rely primarily on any recommendations to make a decision.

Your software function may be non-device CDS.

## Most Generative AI in this Gray Area

# **Jevice Examples**

#### Device examples acquire process, or analyze:

- Signal acquisition systems
- In vitro diagnostics
- Magnetic resonance imaging (MRI)
- Next Generation Sequencing (NGS)
- Continuous Glucose Monitoring (CGM)
- Computer aided detection/diagnosis (CADe/CADx)

#### OR Device examples displa analyze or print:

- Continuous signals/patterns
- Medical images
- Waveforms (ECG)
- More continuous sampling (aka – a signal or pattern)

#### OR Device example

- Risk scores for disease or condition
- Probability of disease or condition
- Time-critical outputs

#### OR Device exa

 Basis of recommendations is not provided

Your software function is a device.

\*Disclaimer: This graphic gives a general overview of Section IV of the guidance ("Interpretation of Criteria in Section 520(o)(1)(E) of the FD&C Act"). Consult the guidance for the complete discussion. The device examples identified in this graphic are illustrative only and are not an exhaustive list. Other software functions that are not listed may also be device software functions.

#### Executive order on AI assurance and AI Assurance labs

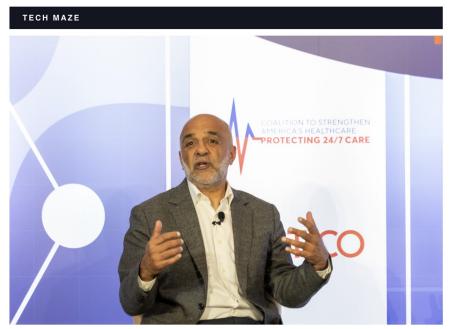
### Executive Order 10.30.23

"This work shall include the development of AI assurance policy to evaluate important **aspects of the performance** of AI-enabled healthcare tools

and infrastructure needs for enabling pre-market assessment and post-market oversight of Al-enabled healthcare-technology algorithmic system performance against real-world data."

#### The AI assurance labs are coming

By ERIN SCHUMAKER, ROBERT KING, CARMEN PAUN, RUTH READER and DANIEL PAYNE | 09/18/2024 02:00 PM EDT



Micky Tripathi at POLITICO's The Future of Patient Care + Access event in Washington on Wednesday | Pixel Me Studio

### Executive order on AI assurance and AI Assurance labs

### Executive Order 10.30.23

"This work shall include the development of AI assurance policy to evaluate important **aspects of the performance** of AI-enabled healthcare tools

and infrastructure needs for enabling pre-market assessment and post-market oversight of Al-enabled healthcare-technology algorithmic system performance against real-world data."

#### The AI assurance labs are coming

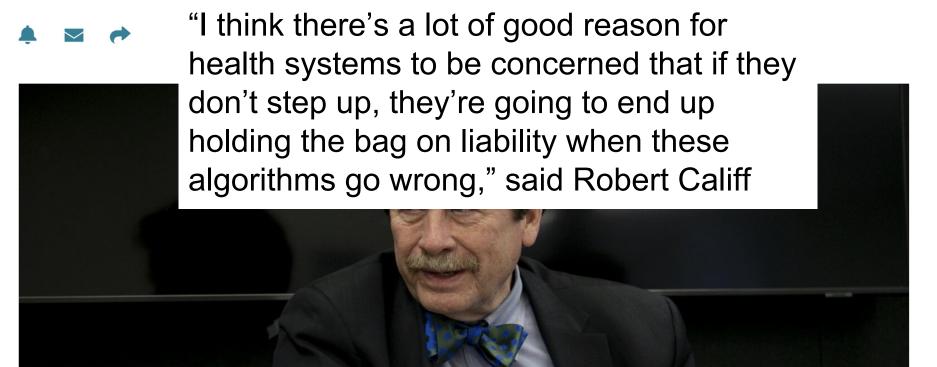
By ERIN SCHUMAKER, ROBERT KING, CARMEN PAUN, RUTH READER and DANIEL PAYNE | 09/18/2024 02:00 PM EDT

"The labs would supplement FDA regulation of Al-powered medical devices, vetting Al tools that fall outside the FDA's regulatory scope.

Whether assurance labs should be private, public-private partnerships or governmentcertified is still an open question, Tripathi said"

# FDA commissioner: Health systems have to 'step up' on Al regulation or will 'end up holding the bag'

Developers can't fully guarantee an Al model's performance, Califf said



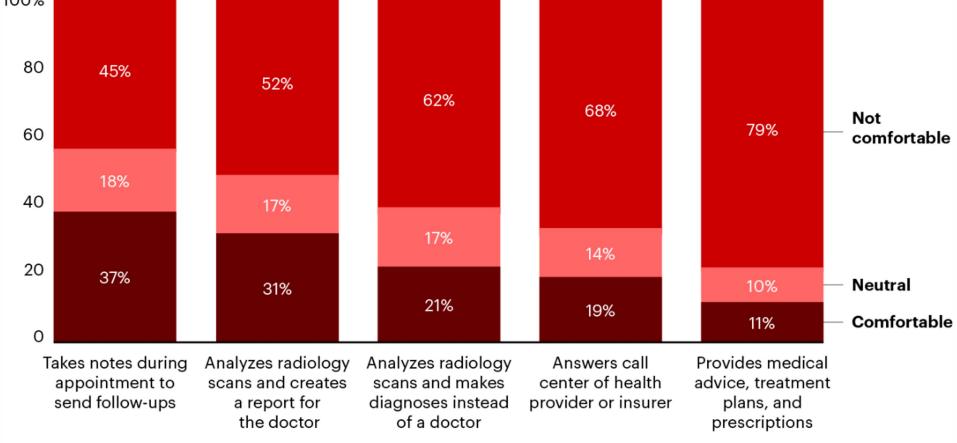
# Office of Civil Rights introduces antidiscrimination rules for Al use

- Section 1557 Under the Affordable care act prohibits recipients of federal funding (healthcare organizations, health insurers, physicians) to discriminate based on protected class (eg. race, sex, age)
- May 2024 Final Rule extends 1557 mandating that health systems identify and mitigate discrimination when using AI and decision support for care
- For Care includes anything that "affect the care that patients receive... for screening, risk prediction, diagnosis, prognosis, clinical decision-making, treatment planning, healthcare operations" OCR Final Rule

# Sentiments and interest in healthcare Al use

# 100%

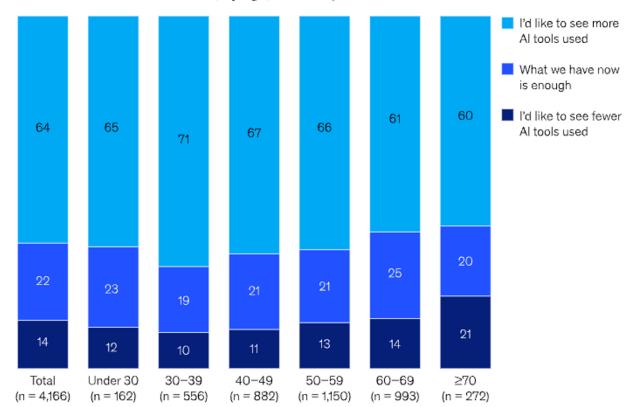
Percentage of consumers indicating level of comfort with generative AI application



Source: Bain US Frontline of Healthcare Consumer Survey, March 2024 (n=500)

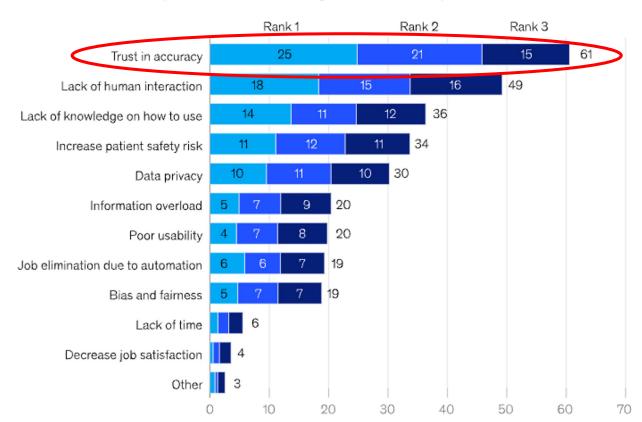
## US Nurses would like to see more Al used

US nurses' desire for Al tools in their work, by age,1% of respondents

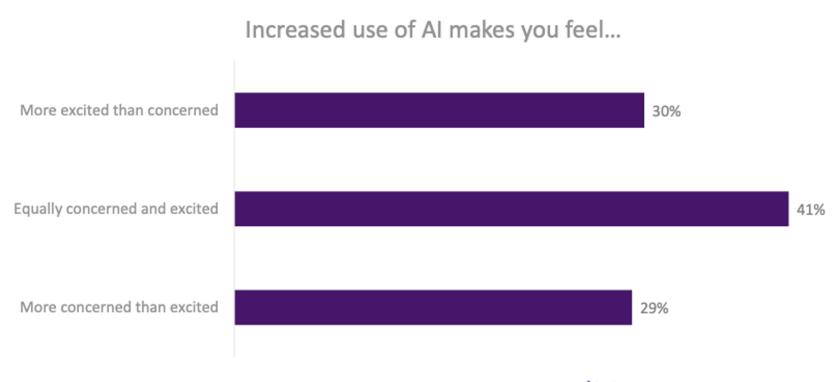


# Top concern by nurses are trust in accuracy

US nurses' rank of top 3 concerns with using AI,1% selected by rank

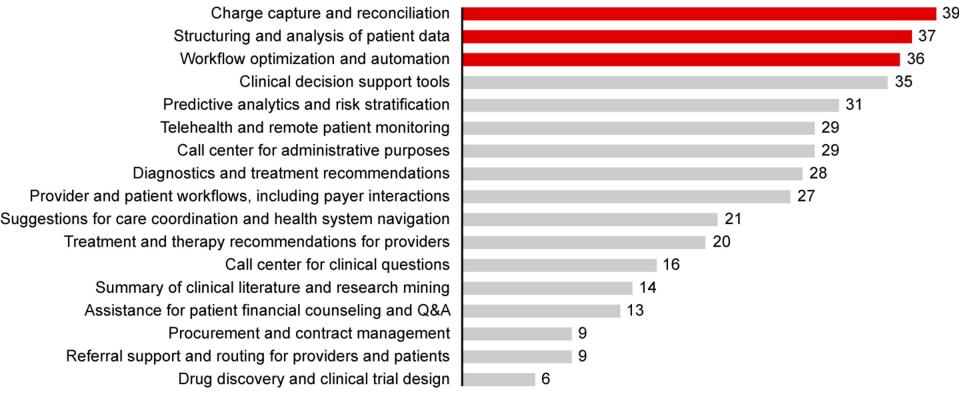


## Doctors are split on excitement and concern

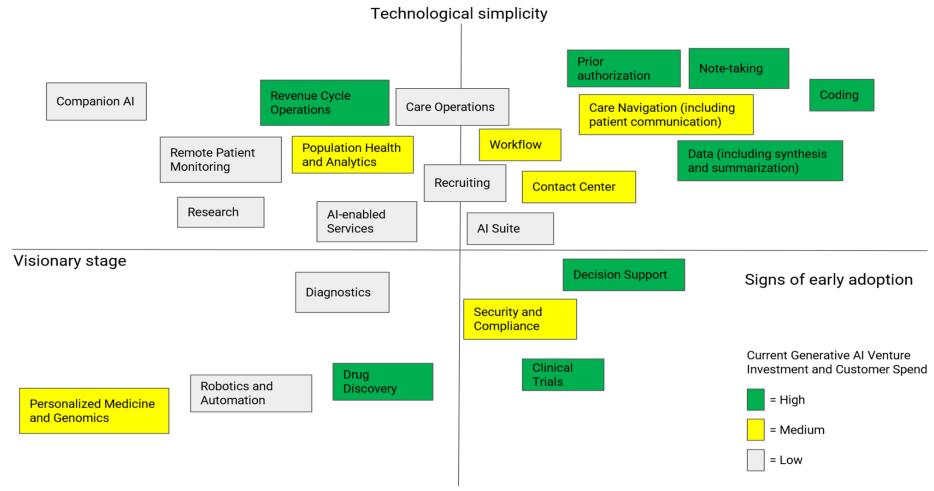


# Which use cases for generative AI are the highest priority for your health system in the near term (within the next 12 months)?

Number of responses by use case



Source: Bain Health Systems Survey (N=94)



Technological complexity

# Takeaways

## Takeaways

- Generative AI is here our patients, providers, and staff are using these tools TODAY
- Big "proven" benefits and big "proven" risks with Generative AI
- We must educate our workforce and our patients on effective and safe use of generative AI in healthcare
- Regulators are holding health systems responsible for AI use
- These tools are improving very rapidly we must prepare for this new world

Thank you!

Contact: jnorden@stanford.edu

# Convicted fraudster Martin Shkreli is touting a medical Al chatbot—much to experts' concern

Shkreli's chatbot, called Dr. Gupta, is already facing criticism from the Al community.



## Workforce dilemma

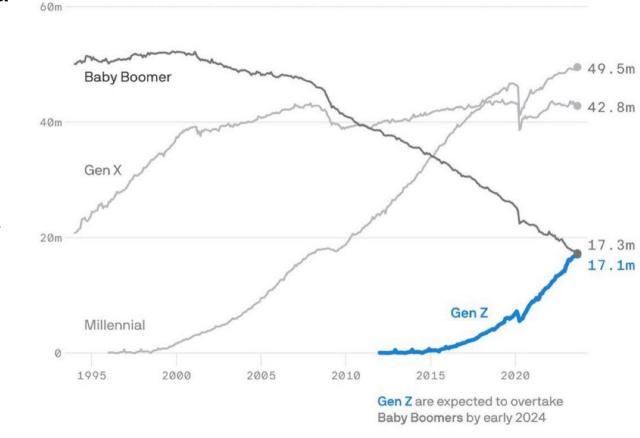
 Nurse and physician shortages - retirement wave coming

Thousands of people joining medicare annually

 Our care needs are exploding while our workforce is shrinking

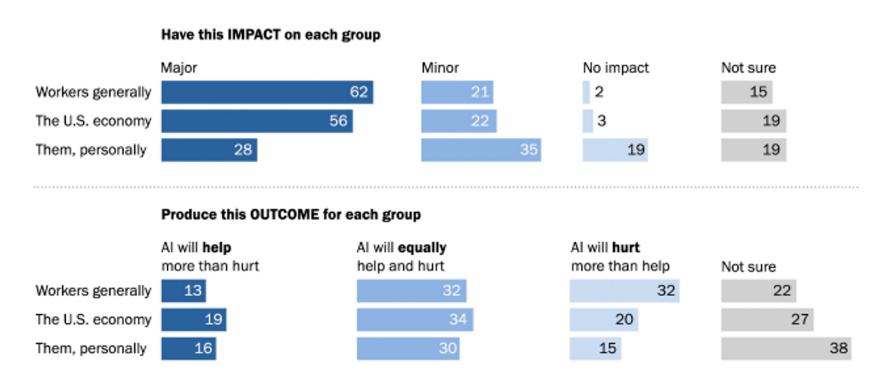
## U.S. full-time workforce, by generation

Seasonally adjusted; Monthly, January 1994 to September 2023



# About six-in-ten Americans believe AI will have a major impact on workers generally, but only 28% believe it will have a major effect on them personally

% of U.S. adults who say that over the next 20 years the use of artificial intelligence in the workplace will ...



Note: Those who did not give an answer are not shown. Source: Survey of U.S. adults conducted Dec. 12-18, 2022.

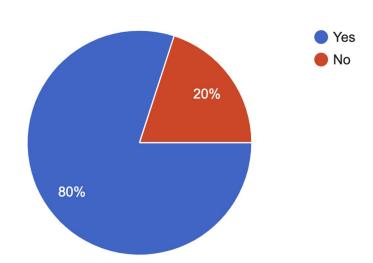
"Al in Hiring and Evaluating Workers: What Americans Think"

#### PEW RESEARCH CENTER

#### Comfort with Al Medical Advice

Are you okay getting medical advice through AI before speaking to a clinician?

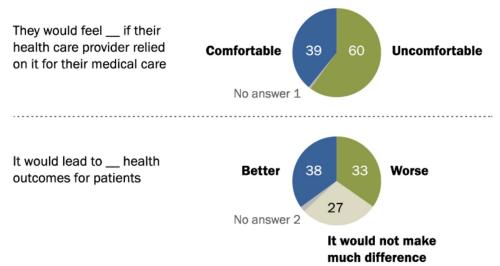
40 responses



#### Fewer than half in U.S. expect artificial intelligence in health and medicine to improve patient outcomes

# Fewer than half in U.S. expect artificial intelligence in health and medicine to improve patient outcomes

% of U.S. adults who say that thinking about the use of artificial intelligence in health and medicine to do things like diagnose disease and recommend treatments ...



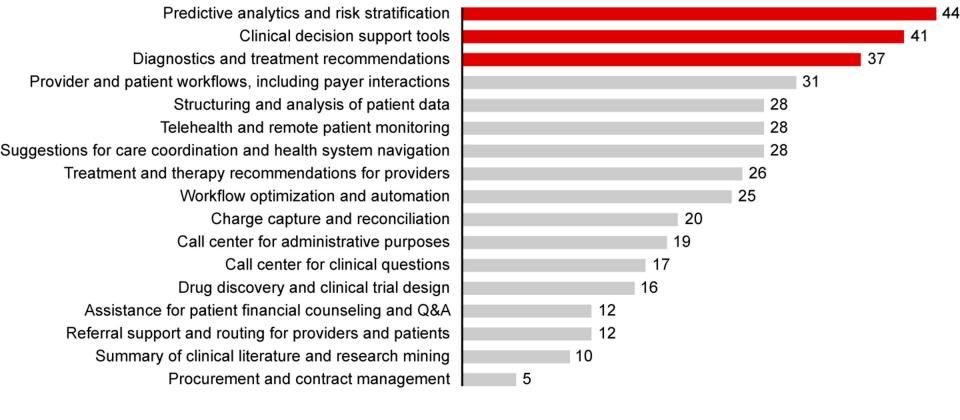
Source: Survey conducted Dec. 12-18, 2022.

#### **PEW RESEARCH CENTER**

<sup>&</sup>quot;60% of Americans Would Be Uncomfortable With Provider Relying on Al in Their Own Health Care"

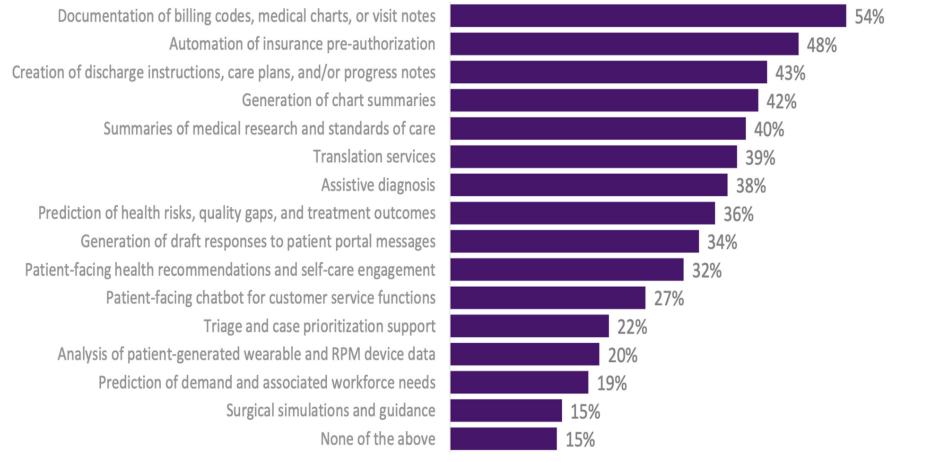
# Which use cases for generative AI are the highest priority for your health system in the long term (2–5 years)?

Number of responses by use case



Source: Bain Health Systems Survey (N=94)

## Al Use Case Enthusiasm

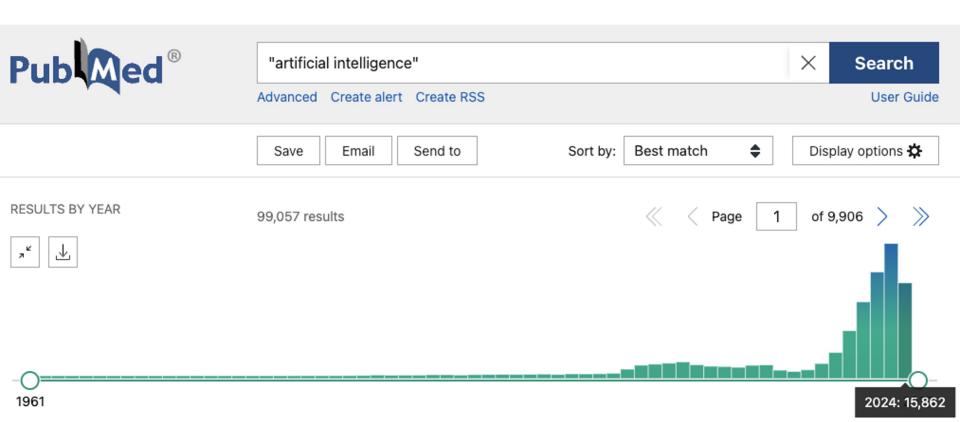


## Translation services and draft response generation seem to have the closest timeline to adoption

Use Case Adoption Timeline	Immediately	Within Next Year	Within 2-5 years	Within 5- 10 years	No plans to incorporate
Automation of insurance pre-authorization	11%	15%	42%	13%	11%
Documentation of billing codes, medical charts, or visit notes	11%	16%	39%	14%	9%
Creation of discharge instructions, care plans, and/or progress notes	11%	18%	36%	14%	9%
Patient-facing chatbot for customer service functions	14%	15%	37%	18%	6%
Patient-facing health recommendations and self-care engagement	13%	18%	36%	17%	10%
Summaries of medical research and standards of care	12%	18%	38%	13%	12%
Translation services	16%	17%	33%	14%	10%
Prediction of demand and associated workforce needs	16%	11%	43%	17%	9%
Assistive diagnosis	12%	15%	36%	15%	8%
Generation of chart summaries	13%	20%	39%	14%	8%
Triage and case prioritization support	13%	18%	37%	17%	10%
Prediction of health risks, quality gaps, and treatment outcomes	12%	18%	39%	16%	10%
Analysis of patient-generated wearable and RPM device data	13%	16%	36%	17%	11%
Surgical simulations and guidance	13%	14%	35%	14%	12%
Generation of draft responses to patient portal messages	15%	22%	40%	12%	5%



## Predictive AI in healthcare



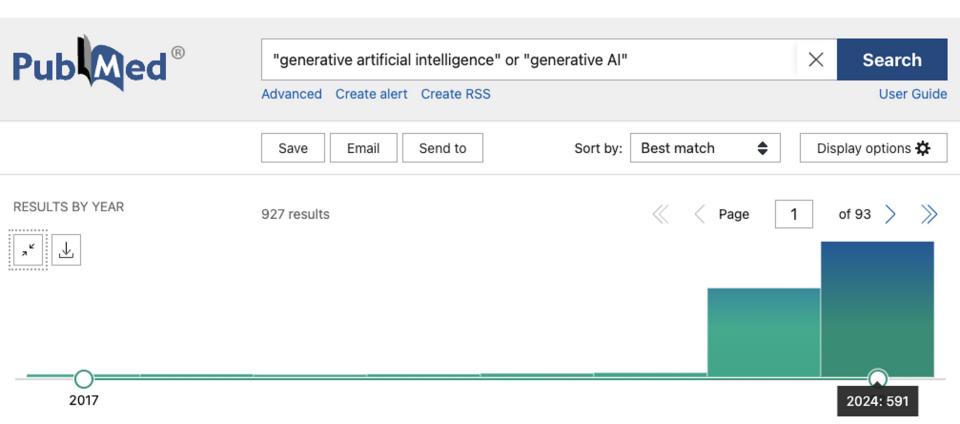


#### **KEY TAKEAWAYS**

- Ambient artificial intelligence (AI) scribes show early promise in reducing clinicians' burden, with a regional pilot noting a reduction in the amount of time spent constructing notes among users.
- Ambient AI scribes were found to be acceptable among clinicians and patients, largely improving the experience of both parties, with some physicians noting the transformational nature of the technology on their care.
- Although a review of 35 AI-generated transcripts resulted in an average score of 48 of 50 in 10 key domains, AI scribes are not a replacement for clinicians. They can produce inconsistencies that require physicians' review and editing to ensure that they remain aligned with the physician-patient relationship.
- Given the incredible pace of change, building a dynamic evaluation framework is essential to assess the performance of AI scribes across domains including engagement, effectiveness, quality, and safety.



## Generative AI in healthcare



#### Where Generative Al Meets Healthcare: Updating The Healthcare Al Landscape red box = launched product after 2020 Life Sciences, \$6.5B Raised **Clinical Trials** Personalized Medicine & Al-enabled Patient-Facing, Insilico Medicine Base Navigation Services Genomics Syapse Discovery \$2.3B Raised **babylon TEMPUS** Curai Health AETION **emonearl** Reverie Labs **88** VERGE ada INCEPTIVE Companion DEEP 6 A **GENOOX** Exscientia Auxa Health monogram health DeepVariant insitro VERISIMLife ConcertAl **Atomwise** TAIOSMEC wysa wysa Remote Patient **#**biofourmis **@tyto**care Woebot Health Admin, \$2.7B Raised Monitoring Al Suite Operations Analytics & IT. **C**are.coach Recruiting Athelas o current health Artisight \$2.7B Raised Ölér Health notable IntelyCare ClosedLoop.a MEMORA HEALTH Reverence winnow. Ferrum Clinician-Facing, Decision \$6.0B Raised Support Medical Population Health & regard Prior Auth pieces Coding **Analytics** Surgical **kahun** 1 **M** navina Robotics INTUITIVE M BUDDI AI XACT evidation 💥 patient 🛚 myndshft 🔮 ivion ARTERYS aidoc co:helm FATHOM **Proprio Clarify** prognos O GLASS Develop Health (IBM Watson Health **▼ APIXIO** AURIS" CODAMETRIX **Rhyme** innovaccer I VERB SURGICAL **VICARIOUS** LATENT <3×<< **OPTT** nym CARTA HEALTHCARE ethermed **a** Diagnostics Fairway Health hume PathAl A Arintra Notetaking **Knowtex KINTSUGI Butterfly** Revenue m Mentalyc Contact Security & Gentem DIGITAL eleos Cycle Center Compliance Smarter D> ROBIN Operations Ambience PROSCIA PRIVATE AI © CODOXO ¥iz.ai €) Fapero elements rivet Syntrillo 🙀 Mr DeepScribe HEALTH Syllable Data Centaur Labs RIALTIC Adonis **⇒** meMR abridge Research **Birch** AI Workflow **AUGMEDIX** Segmed ScienceIO candidhealth ATROPOSHEALTH cedar **cascade** Rad Al NABLA Care Studio Suki Elicit INFINITUS scite\_ \*Anomaly consensus SIRONA RamSoft UNLEARN • **Clinithink** InpharmD NUANCE

Benevolent"

cohere

banjo

enter