Real World Applications of Generative AI in Liver Diseases and Transplantation

Friday, September 4, 2026

CPH 100 Guest Lecture

Jin Ge, MD, MBA

Assistant Professor of Medicine, In Residence

Director of Clinical Artificial Intelligence, UCSF Gastroenterology

Divisions of Gastroenterology and Hepatology, and Clinical Informatics and Digital Transformation

Department of Medicine, University of California San Francisco

Disclosures

- **Research funding:**
  - NIH NIDDK K23DK139455
  - AGA Research Pilot Award
  - Gilead Sciences Investigator Initiated Funding

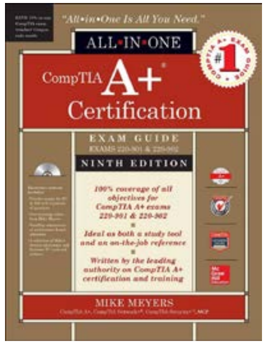
- **I disclose the following financial relationship(s) with a commercial interest within the last 48 months:**
  - Merck and Co (Grant, Concluded)
  - Madrigal Pharmaceuticals (Consultant)
  - Gilead Sciences (Advisory Board, Consultant)

- **I DID use artificial intelligence to prepare portions of this presentation, and I take full responsibility for the contents in this presentation**

About Me

- Assistant Professor of Medicine, In Residence
- Gastroenterologist/Transplant Hepatologist
- Divisions of Gastroenterology/Hepatology and Clinical Informatics and Digital Transformation
- Joined UCSF Faculty in July 2022
- AGG Faculty in CPH
- Affiliate Faculty in Bakar Computational Health Sciences Institute
- NIDDK K23 Career Development Award

2001:
CompTIA A+ Certification



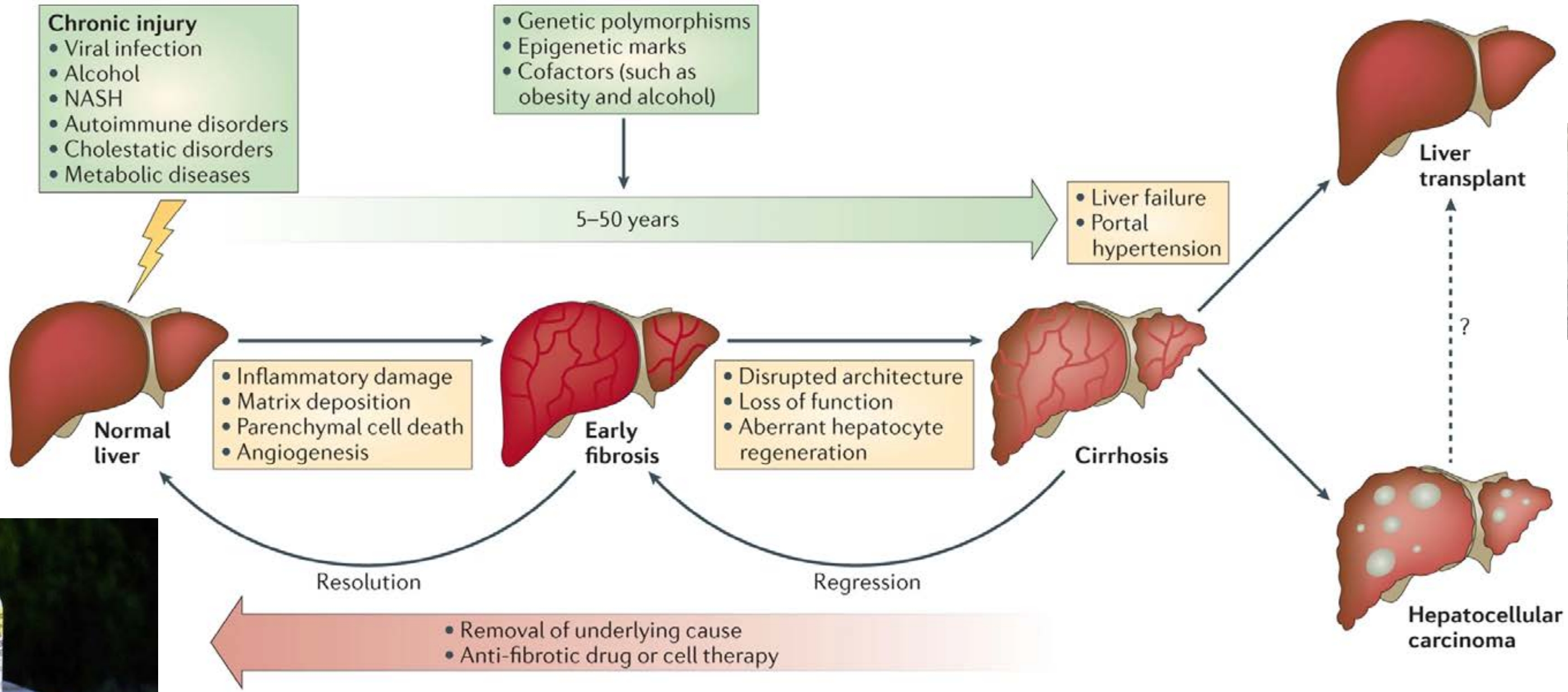
2005:
Residential Computing Consultant at Princeton

OIT Service Desk

2008-2010:
Boston Consulting Group

2014:
Digital Health Startup

Liver Disease is Complicated



Nature Reviews | Immunology

Pellicoro et al, Nature Reviews Immunology, 2014

Why AI in Hepatology and Liver Diseases?

Liver Disease Burden

200,000+

Hospitalizations for cirrhosis annually

$20 billion

In annual spending

37%

Readmitted within 30 days

The GenAI Opportunity

Unstructured data

- ~80% of clinical data is in free-text notes — LLMs can process this at scale

Complex decision-making

- Transplant selection, treatment choice, risk stratification

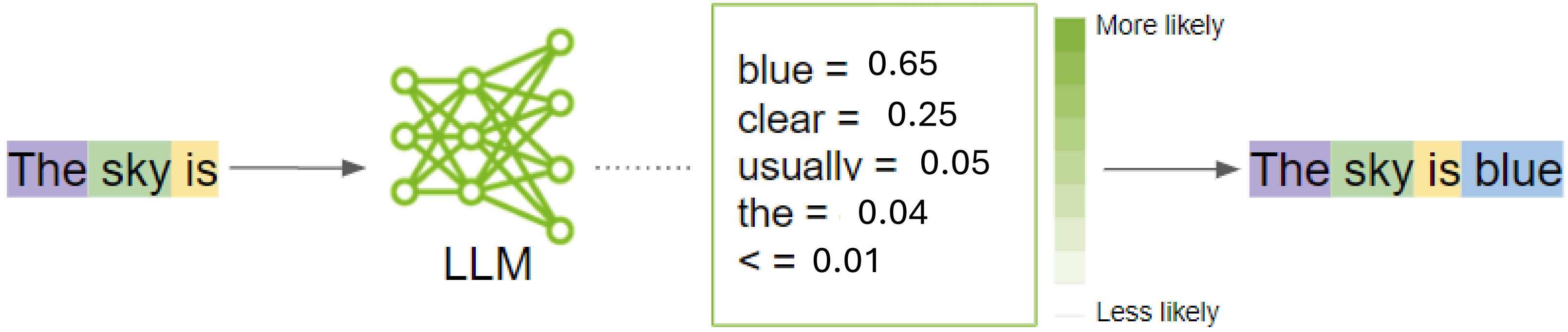
Guideline adherence gaps

- AI-powered decision support could close the evidence-to-practice gap

Ge et al. Liver Transplantation, 2026
Ge and Lai. Hepatology Communications, 2023

2022: ChatGPT

Large Language Models (LLMs) predict and generate human-like responses



OpenAI

Claude

ANTHROPIC

- LLMs process text using neural architectures trained through statistical pattern recognition across massive datasets
- Like a super-powered autocomplete, LLMs predict what words should come next

2022: Potentials and pitfalls of ChatGPT

Strengths

- Summarize literature
- Intuitive user interface
- Adaptability to use cases

Weaknesses

- “Hallucinations”
- Unable to infer context
- Superficial outputs

Opportunities

- Idea generation
- NLP
- Chatbots and PROs

Threats

- Training data bias
- Privacy and security
- Dataset shift

Ge and Lai. Hepatology Communications, 2023.

Motivating Examples

How can we use GenAI to derive new insights from the clinical record?

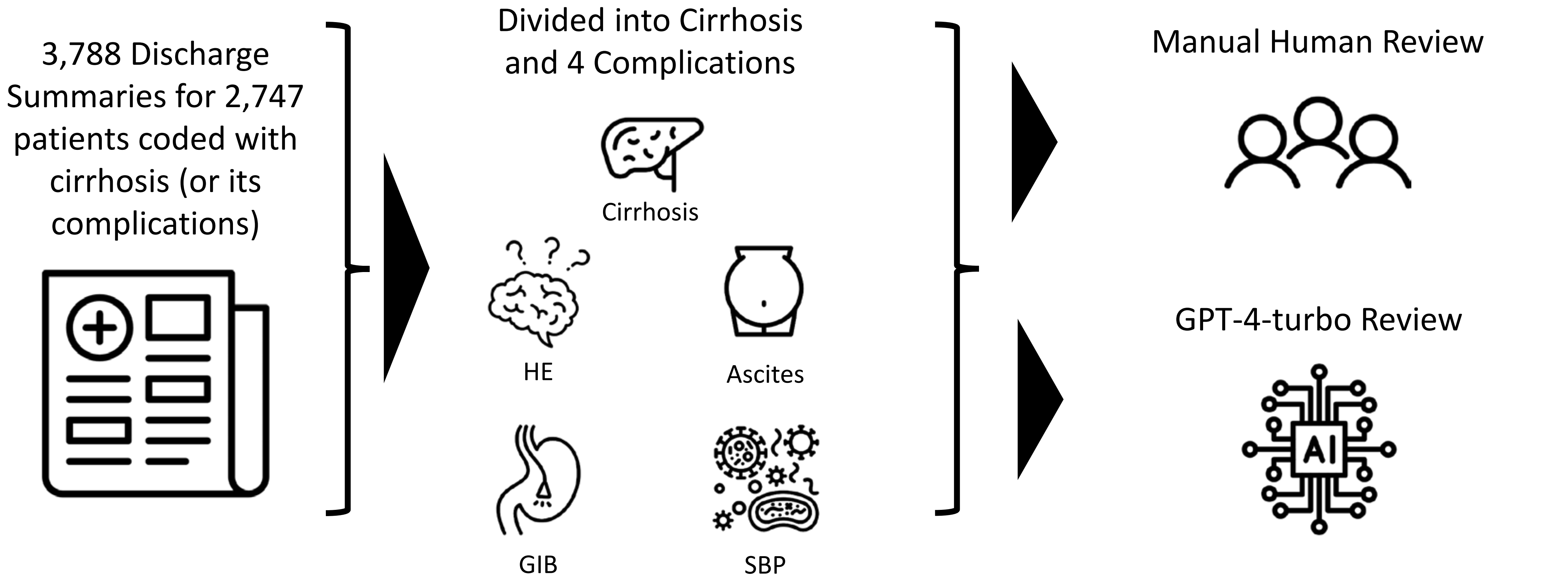
How can LLMs help us run, and learn from, a clinical trial?

How can LLMs assist with clinical decision making?

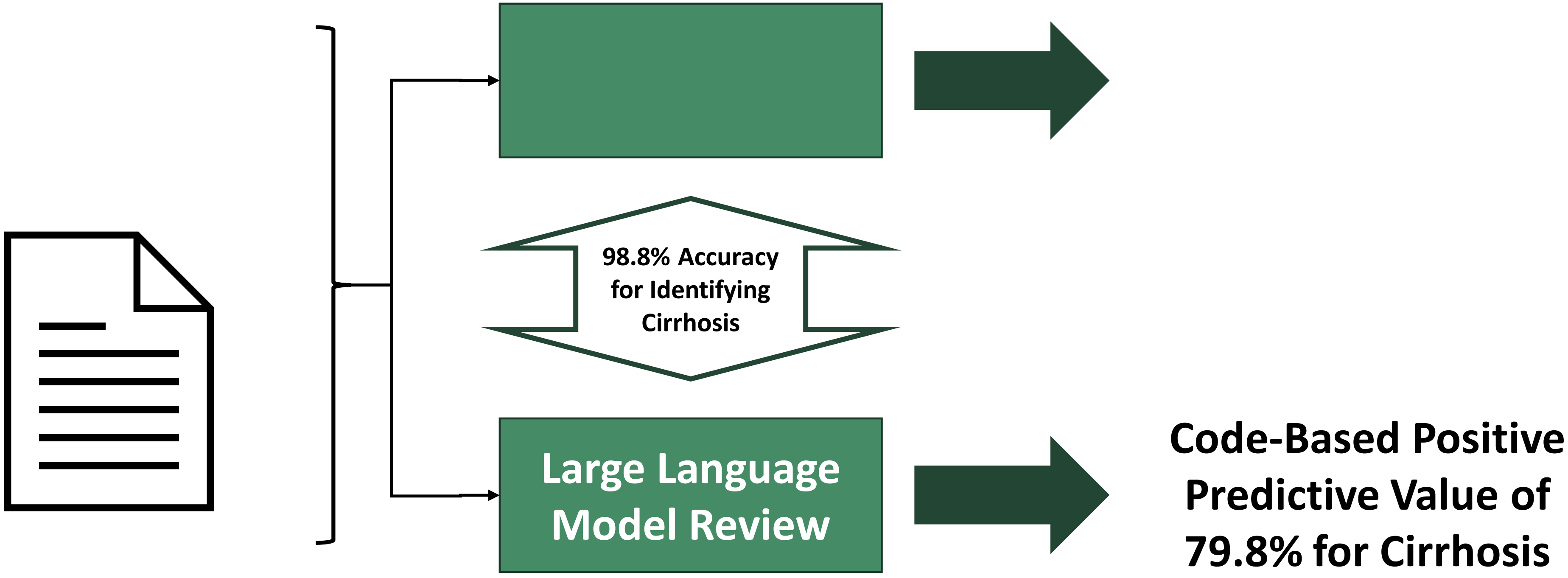
How do we use GenAI to construct novel diagnostic tests?

How can we use GenAI to derive new insights from the clinical record?

Cirrhosis cohort identification in discharge summaries



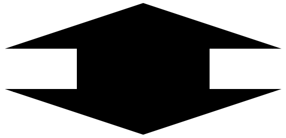
Code-based methods had PPVs of ~80% versus humans and LLMs



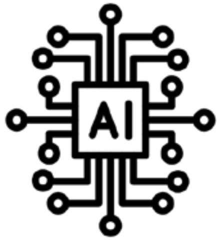
PPVs for complications of cirrhosis were much lower: 54% for HE, 55% for ascites, 68% for GIB, and 66% for SBP

Humans may not be better than LLMs?

Manual Human Review



GPT-4-turbo Review



	# Notes	LLM Accuracy	Human “Error” Rate
Cirrhosis	3,788	98.2-98.8%	4.3%
HE	1,969	94.4-96.3%	14.8%
Ascites	1,618	91.4-94.4%	18.5%
GIB	1,127	85.4-87.8%	11.0%
SBP	339	90.7%	11.6%

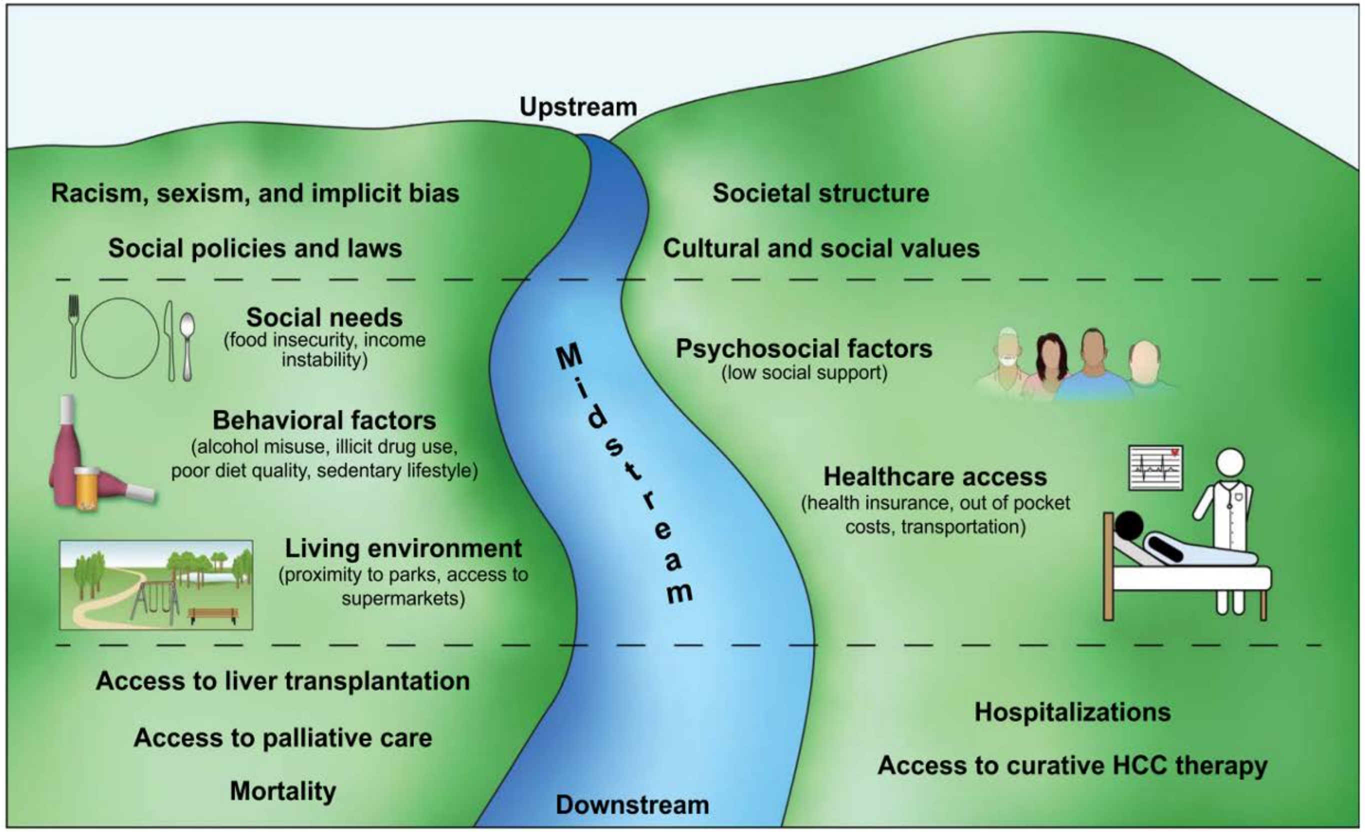
Psychosocial risk factors are associated with liver transplant outcomes

Impact on waitlist registration

- Psychosocial risk and SDOH factors disproportionately concentrate liver disease risk factors (e.g., alcohol use, opioid use, obesity) among historically disadvantaged populations

Impact on successful transplant

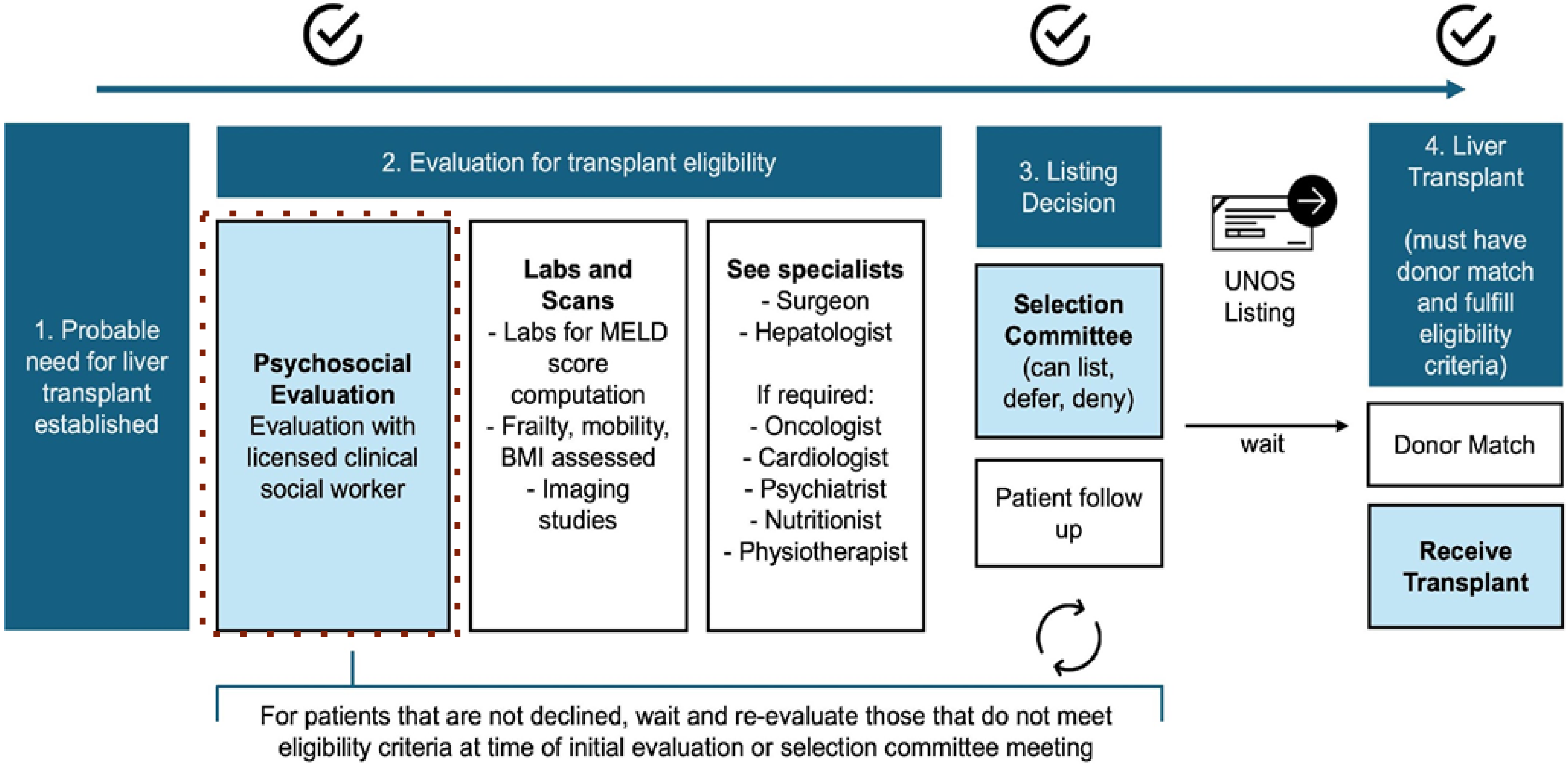
- Disparities persist through the transplant process, with women and racial and ethnic minorities facing reduced access to evaluation, listing, and transplantation



Contribution of Structural and psychosocial/SDOH factors to Liver Disease Outcomes (from Kardashian et al. 2022)

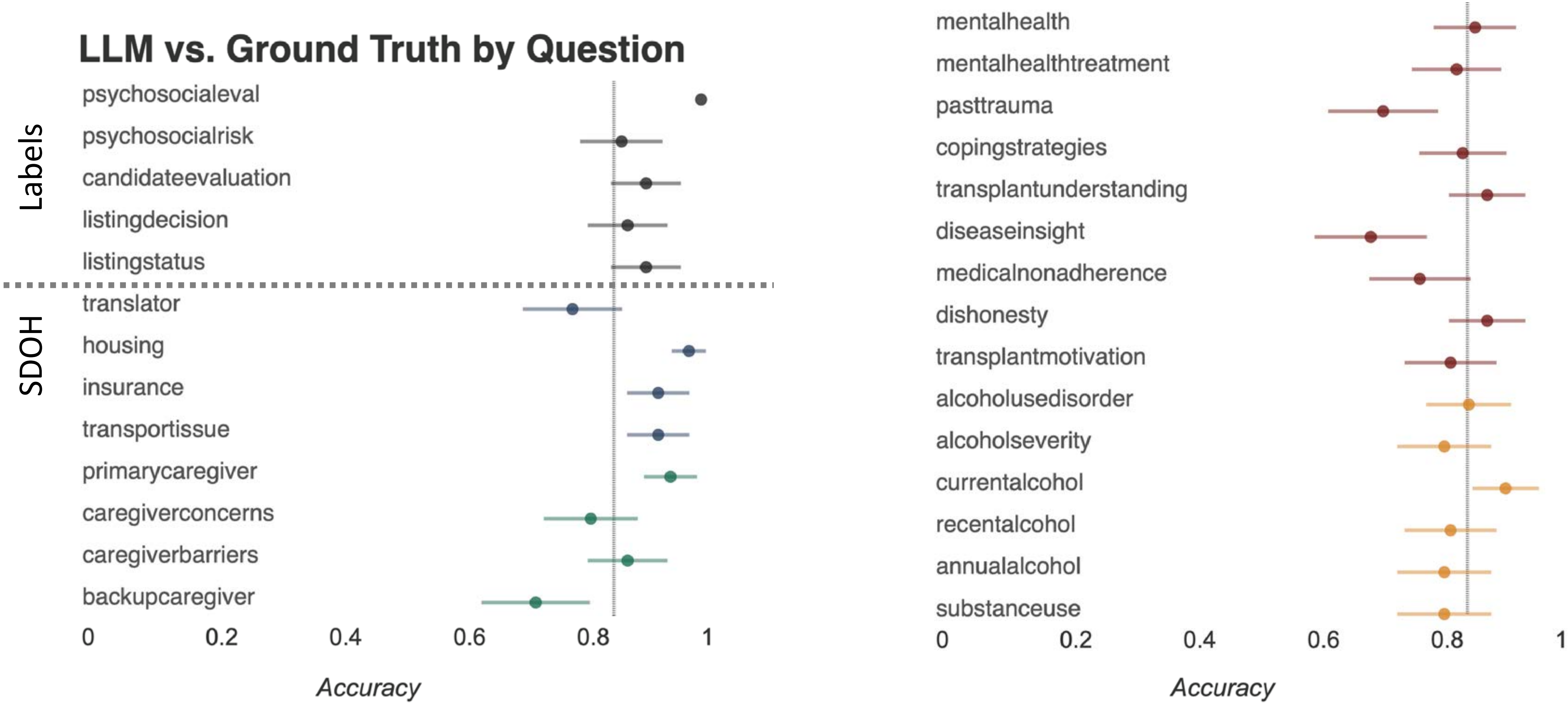
ACG INSTITUTE FOR CLINICAL RESEARCH AND EDUCATION

LT psychosocial evaluation notes systematically capture these factors



Robitschek et al. npj Digital Medicine, 2025.

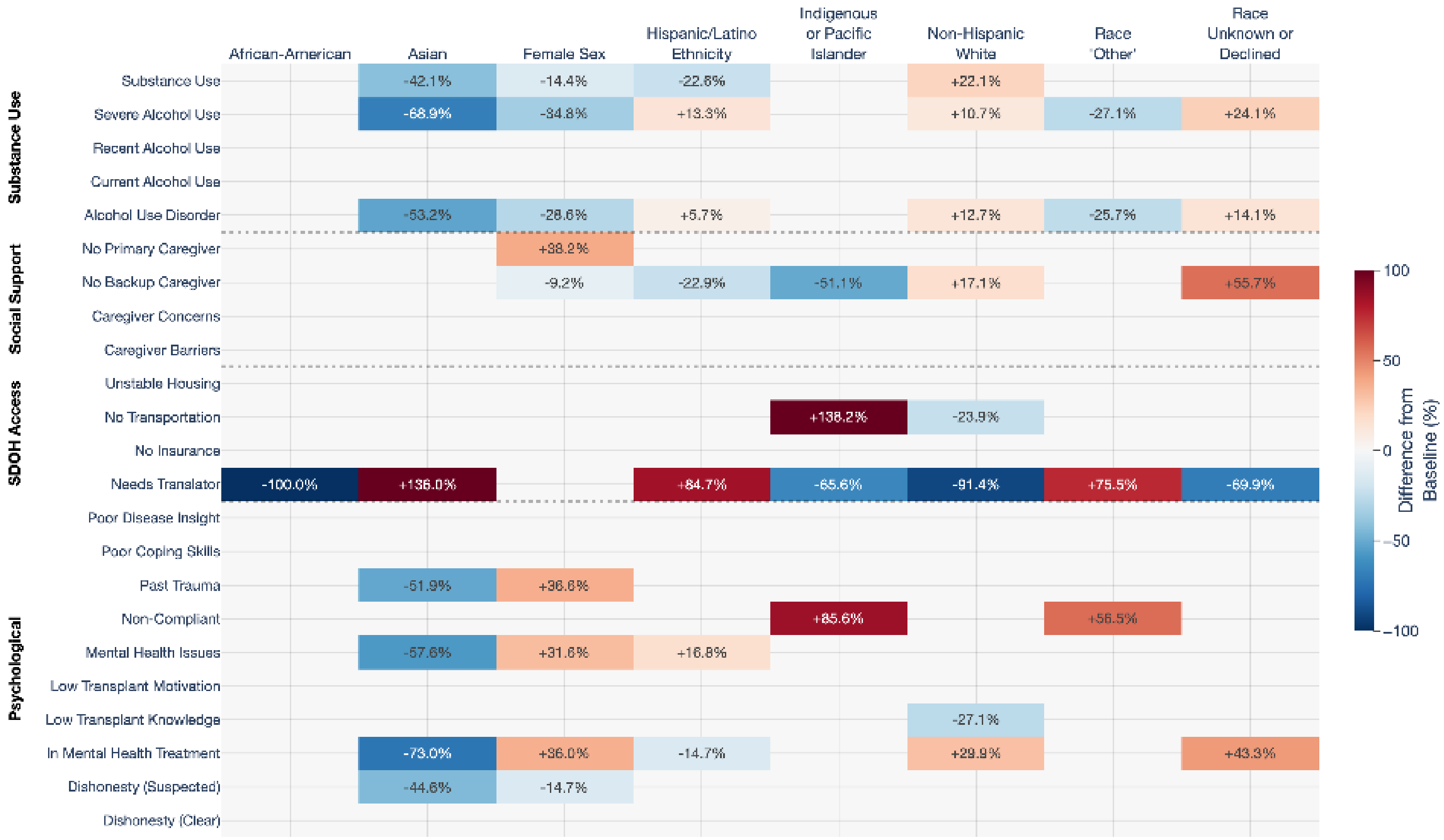
LLMs extracted psychosocial risk and SDOH factors with 86% accuracy



ACG INSTITUTE FOR CLINICAL RESEARCH AND EDUCATION

Robitschek et al. npj Digital Medicine, 2025.

Psychosocial risk and SDOH categorization from SW evaluation notes



Integrated Analysis

- Patient SDOH Snapshots
- Demographic Comparisons
- Care Progression Analysis
- Decomposition of Disparities

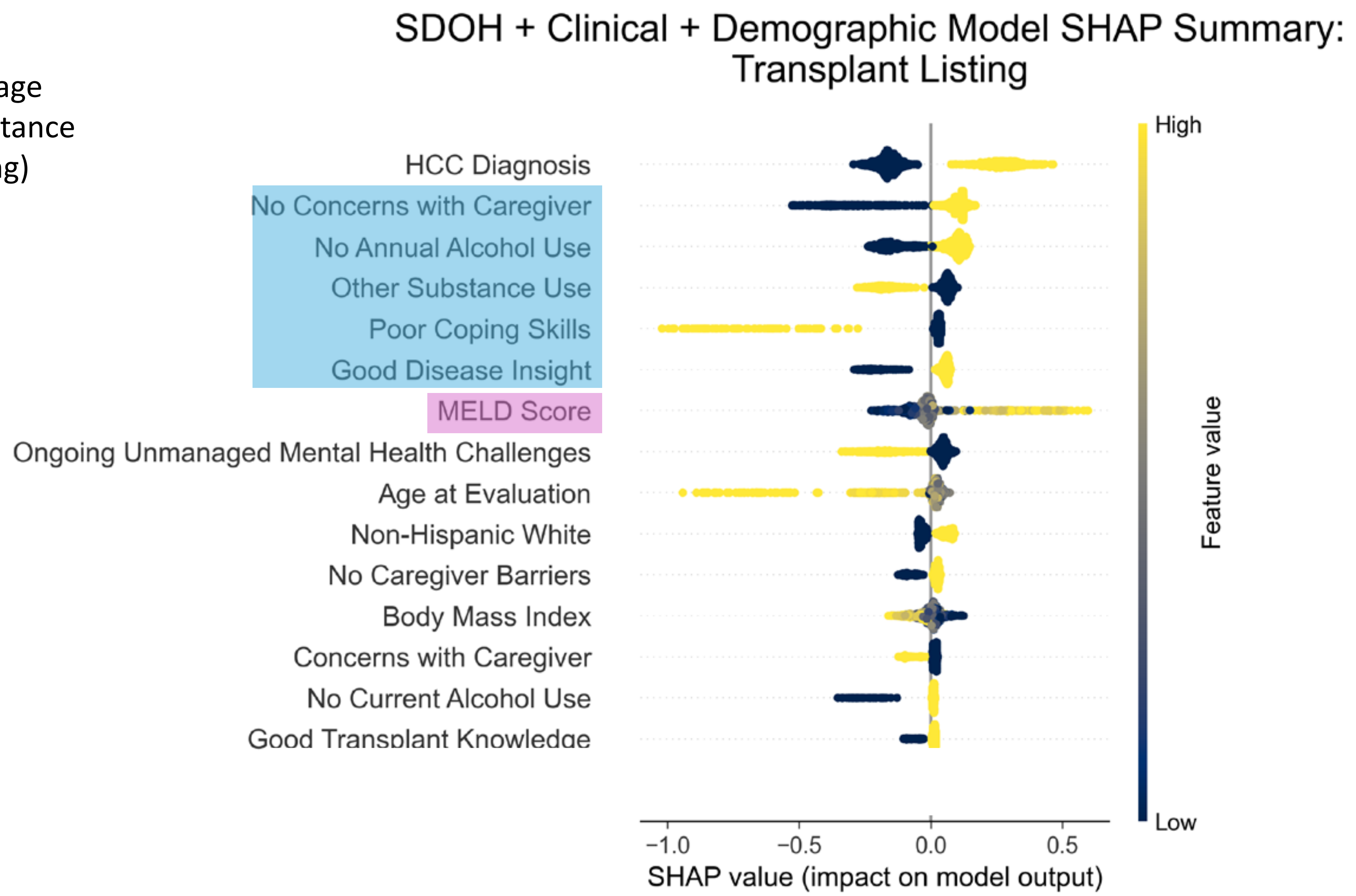
ACG INSTITUTE FOR CLINICAL RESEARCH AND EDUCATION

Computational PRECISION HEALTH Berkeley UCSF

Robitschek et al. npj Digital Medicine, 2025.

Psychosocial/SDOH factors may be *more important than* MELD for LT listing

Y-axis: Average Feature Importance (Descending)



Robitschek et al. npj Digital Medicine, 2025.

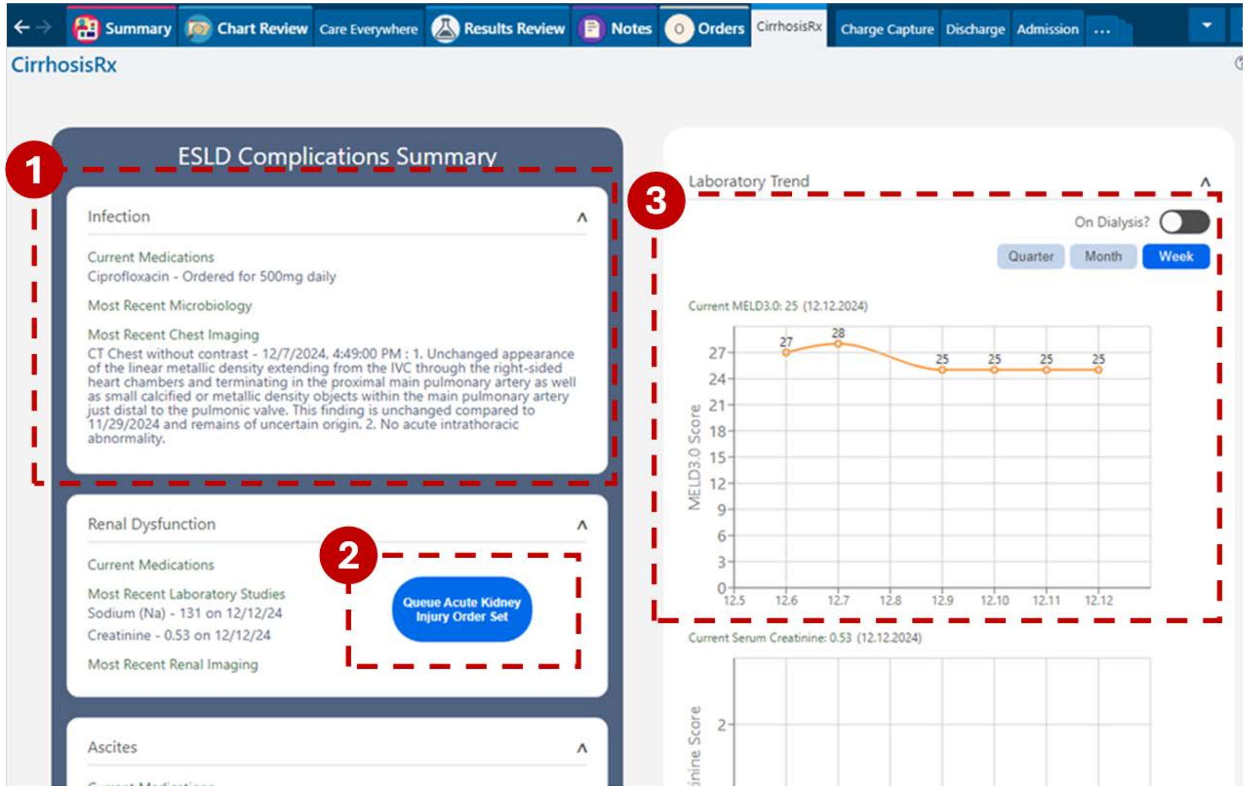
How can LLMs help us run, and learn from, a clinical trial?

CirrhosisRx: A stepped-wedge cluster randomized trial of a CDS for cirrhosis care

Step	Months 1-3	Months 3-6	Months 6-9	Months 9-12	Months 12-15	Months 15-18	Months 18-21	Months 21-24
1								
2								
3								
4								
5								
6								
7								
8								

CirrhosisRx Assignment

"Usual Care" Assignment



- **Intervention and primary outcome**
  - SMART-on-FHIR CDS in Epic surfacing AGA/AASLD process measures for ascites, SBP, HE, and GI bleeding
  - Primary: Opportunity-level adherence to 7 measures, contraindication-netted

NIH NIDDK

Trial flow and analytical approach

- **Primary intention-to-treat model**
  - Mixed-effects logistic at the opportunity level: intervention + calendar period + MELD
  - Random intercepts: cluster, treatment team, patient
- **Robustness (pre-specified)**
  - Adjustment waterfall: crude => +period => +team => +MELD
  - Permutation (randomization) inference
  - Missed-opportunity count and proportion on the randomization-fixed denominator
  - Per-protocol scenarios and complier-average causal effect (IV)
  - Two code-sets (python and R for replication and cross-checking)

NIH NIDDK

LLMs define the cohort, the applicable measures, and the contraindication-netted endpoint

- **Three ascertainment tasks done by AI**
  - Cohort: H&P-suspected cirrhosis => analytical dataset (denominator)
  - Guideline applicability: Which of the 5 guideline measures for inpatients with cirrhosis are applicable (complications treated during the admission ascertained from discharge summary)
  - Netting: Opportunities with documented contraindications removed from the denominator (e.g., a patient without ascites who is indicated for a paracentesis)
- **Why use an LLM?**
  - ICD-10 codes are not accurate
  - Tens of thousands of admissions × note-level adjudication (beyond manual review)

Clinical notes (H&P + discharge summary)

↓

GPT-based extraction and adjudication

↓

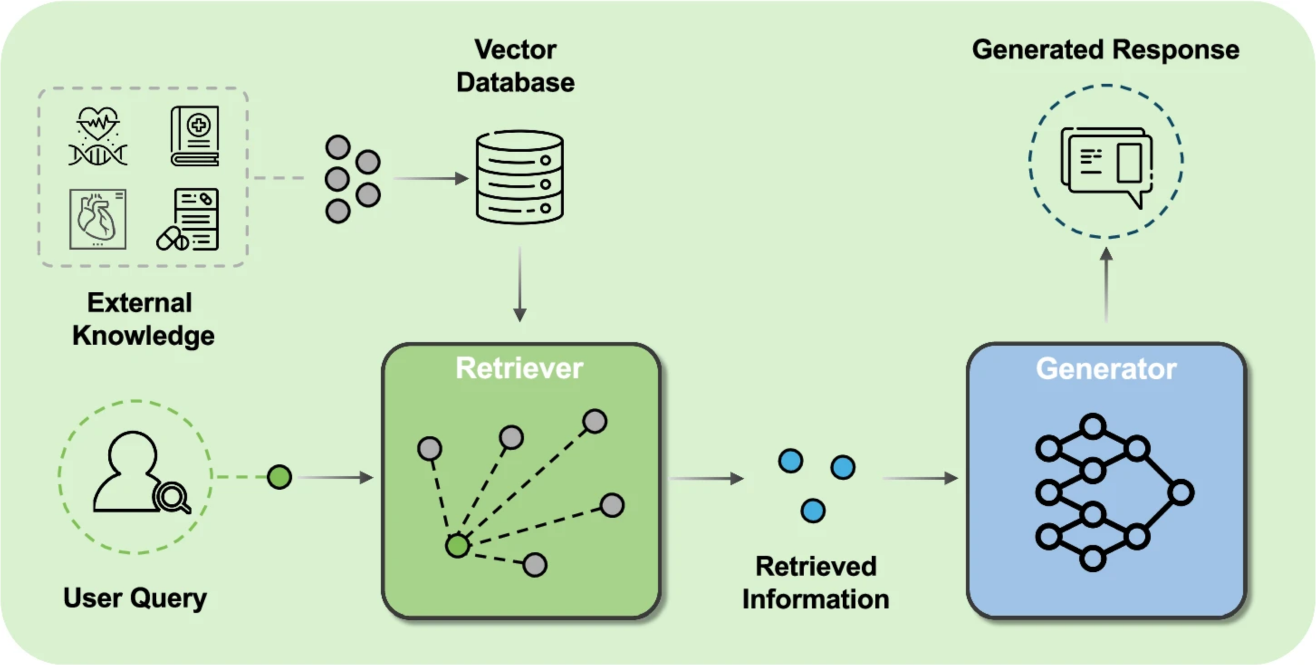
Applicable, contraindication-netted opportunities

↓

Adherence = adherent / applicable

How can LLMs assist with clinical decision making?

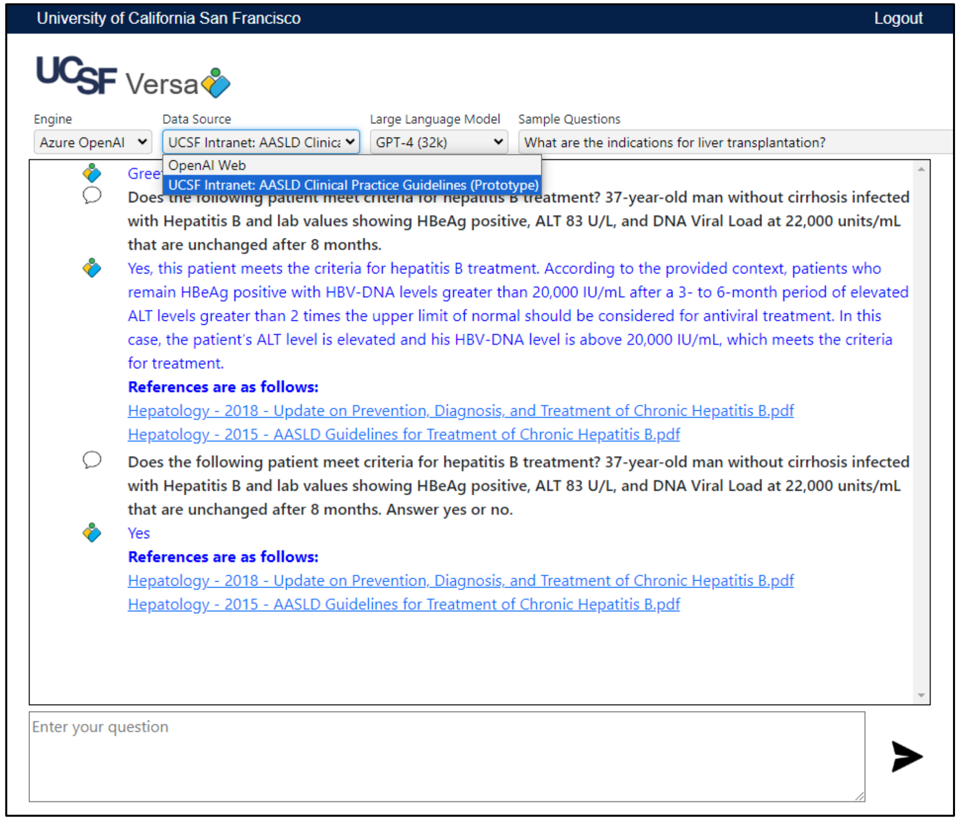
Retrieval augmented generation allows for customized LLMs



- External data is first encoded into vectors and stored in the vector database (where vectors are mathematical representations of various types of data in a high-dimensional space).
- In the retrieval stage, when receiving a user query, the retriever searches for the most relevant information from the vector database. In the generation stage, both the user's query and the retrieved information are used to prompt the model to generate content.

Yang et al. NPJ Health Systems, 2025.

LiVersa – A customized liver-disease specific large language model



Ge et al. Hepatology, 2024.

LiVersa matched ChatGPT and outperformed LLaMA in answering hepatology questions

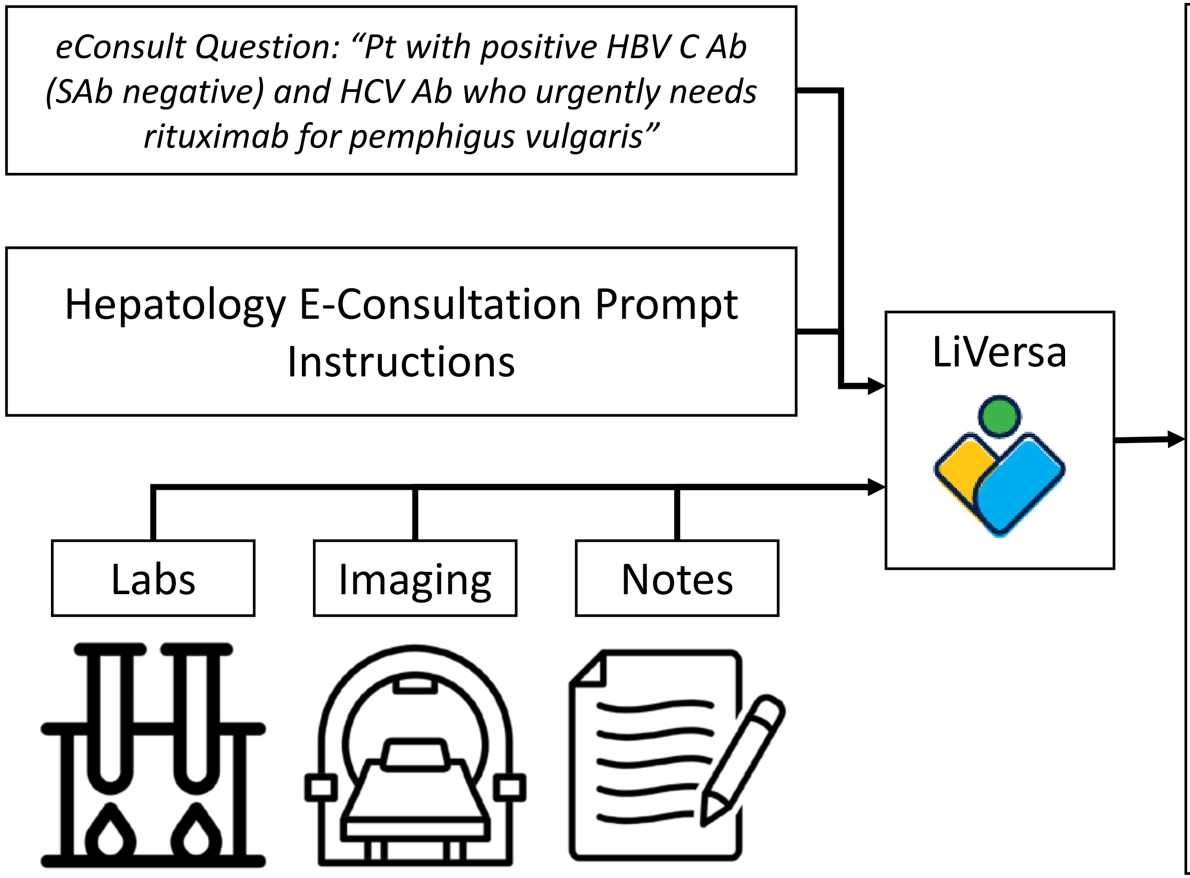
Human-Evaluations of GenAI/LLM Answers to 10 Hepatology Topic Questions

	LiVersa vs. LLaMA	LiVersa vs. ChatGPT
Accuracy	**4** vs. 0	**2** vs. 1
Comprehensiveness	**5** vs. 4	0 vs. **3**
Safety	**3** vs. 1	2 vs. **3**

LLM	PSC Question - What is most common type of malignancy associated with Primary Sclerosing Cholangitis?	Character Count
LiVersa	The most common type of malignancy associated with Primary Sclerosing Cholangitis (PSC) is cholangiocarcinoma, which is a cancer that arises from the bile ducts.	161
ChatGPT 4	The most common type of malignancy associated with Primary Sclerosing Cholangitis (PSC) is cholangiocarcinoma, a cancer of the bile ducts. This cancer occurs more frequently in individuals with PSC than in the general population. The risk of developing cholangiocarcinoma in patients with PSC is significantly elevated, making regular monitoring and early detection vital. Additionally, PSC patients are at an increased risk for other cancers, including colorectal cancer, particularly in those with concurrent inflammatory bowel disease. Early diagnosis and management of PSC are crucial to monitor and potentially reduce the risk of these malignancies.	654
LLaMA 2	The most common type of malignancy associated with Primary Sclerosing Cholangitis (PSC) is cholangiocarcinoma, which is a rare cancer that originates in the epithelial cells lining the bile ducts. PSC increases the risk of developing cholangiocarcinoma by up to 30-fold compared to the general population. Other types of cancer that can occur in association with PSC include gallbladder cancer and pancreatic cancer. The exact mechanisms by which PSC leads to malignancy are not fully understood but may involve chronic inflammation and genetic mutations. Early detection and treatment of PSC can help reduce the risk of developing malignancies.	645

Ge et al. Hepatology, 2024.

Specialty-specific LLMs integrated into EHR for consultation drafting (LiVersa)



eConsult Hepatology Recommendations

Repeat and Update Key Labs and Serologies

- The patient’s HBV serologies and HBV DNA level are over 12 months old. Given his urgent need for rituximab and the high risk of HBV reactivation with B-cell–depleting therapy, please obtain updated: HBsAg, anti-HBc (total and/or IgG), anti-HBs, HBV DNA
- He also has a positive hepatitis C antibody without a documented HCV RNA result. Please check an HCV RNA (viral load) to clarify whether he has active hepatitis C infection.
- If not already done, obtain an HIV test, as HBV treatment considerations can change in the setting of HIV coinfection.

HBV Reactivation Risk With Rituximab

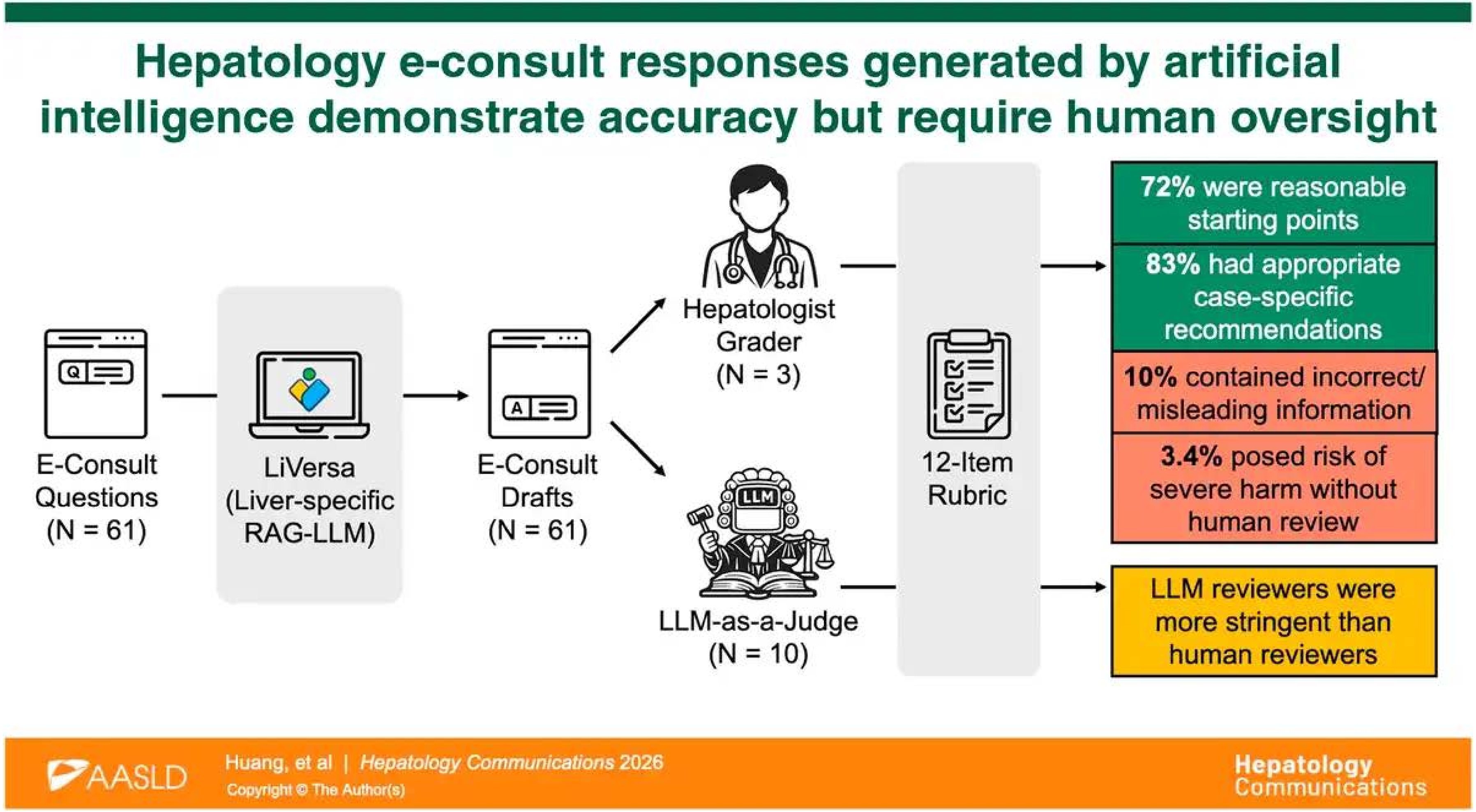
- In patients who are HBsAg-negative but anti-HBc–positive, rituximab-containing regimens can pose a significant risk of HBV reactivation.
- Current guidelines recommend prophylactic antiviral therapy (rather than “watchful waiting”) for HBsAg-negative/anti-HBc–positive patients receiving rituximab or other anti–B-cell agents.

Antiviral Prophylaxis for HBV

- For prophylaxis, use a high-potency, high–barrier-to-resistance agent such as: Tenofovir alafenamide (TAF), tenofovir disoproxil fumarate (TDF), or Entecavir (if no prior lamivudine exposure or known resistance).
- Start antiviral prophylaxis before (or at the time) rituximab is initiated and continue for at least 6-12 months after completion of immunosuppressive therapy (some clinicians continue up to 12 months).

Huang HKT, et al. Hepatology Communications, 2026

Retrospective testing showed that LiVersa could generate accurate e-consult drafts



Versa
UCSF Information Technology

Huang HKT, et al. Hepatology Communications, 2026

We took a broad cross-section of actual e-consults completed over 3 months

TABLE 1 Characteristics of e-consult requests (n = 61)

Characteristic	
Patient demographics	
Sex, n (%)	
Female	43 (70.5)
Male	18 (29.5)
Age, years	
Mean	55.0
Median (range)	54.0 (23–97)
E-consult category, n (%)	
Abnormal liver function tests	21 (34.4)
Hepatitis B	14 (23.0)
Abnormal imaging	13 (21.3)
FibroScan	6 (9.8)
Hepatitis C	4 (6.6)
Steatosis	3 (4.9)
Clinical question characteristics	
Word count, mean (SD)	77 (65)
Character count, mean (SD)	484 (420)
Word count by category, mean (SD)	
Abnormal imaging	109 (51)
Abnormal liver function test	97 (86)
Steatosis	75 (57)
Hepatitis C	48 (39)
Hepatitis B	45 (26)
FibroScan	29 (21)

Abbreviation: e-consult, electronic consultation.

TABLE 2 Characteristics of e-consult responses by hepatologists versus LiVersa (n = 61 pairs)

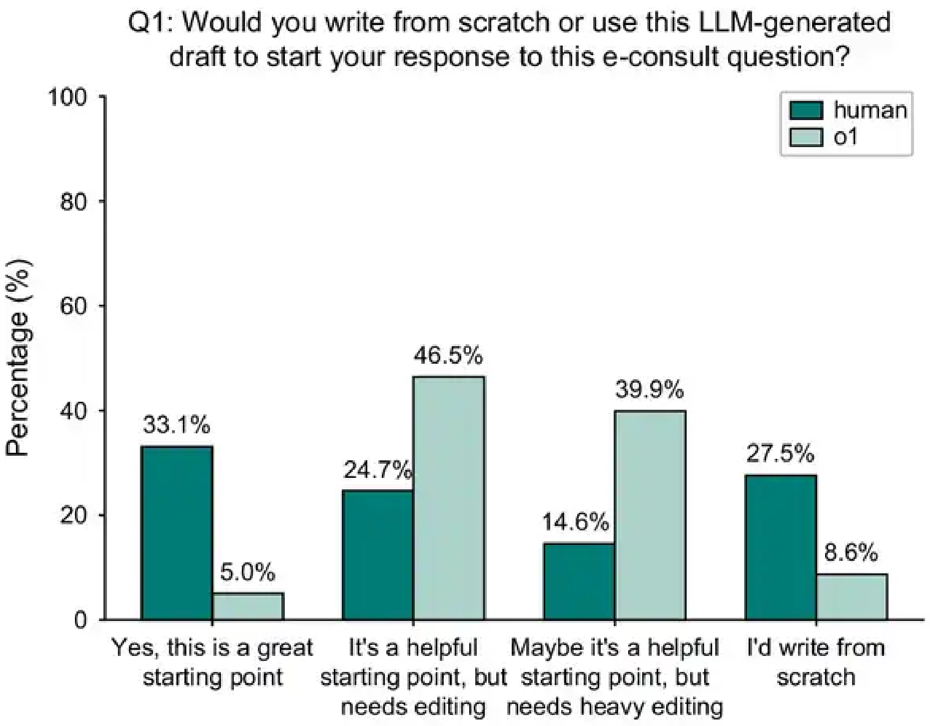
Characteristic	Hepatologist	LiVersa	*p*
Overall response characteristics, mean (SD)			
Word count	284 (172)	264 (146)	0.47
Character count	1765 (1136)	1798 (1038)	0.86
Sentence count	14 (14)	11 (7)	0.10
Words per sentence	24 (13)	25 (6)	0.44
Word count by category, mean (SD)			
Abnormal imaging	367 (298)	282 (186)	0.35
Abnormal liver function tests	277 (115)	286 (158)	0.84
FibroScan	204 (60)	218 (99)	0.73
Hepatitis B	258 (94)	240 (113)	0.67
Hepatitis C	160 (125)	283 (103)	0.01
Steatosis	413 (82)	217 (204)	0.23
Words per sentence by category, mean (SD)			
Abnormal imaging	23 (17)	28 (4)	0.37
Abnormal liver function tests	26 (11)	25 (6)	0.84
FibroScan	16 (5)	17 (3)	0.58
Hepatitis B	24 (9)	27 (4)	0.22
Hepatitis C	17 (1)	27 (3)	0.01
Steatosis	42 (19)	21 (4)	0.24

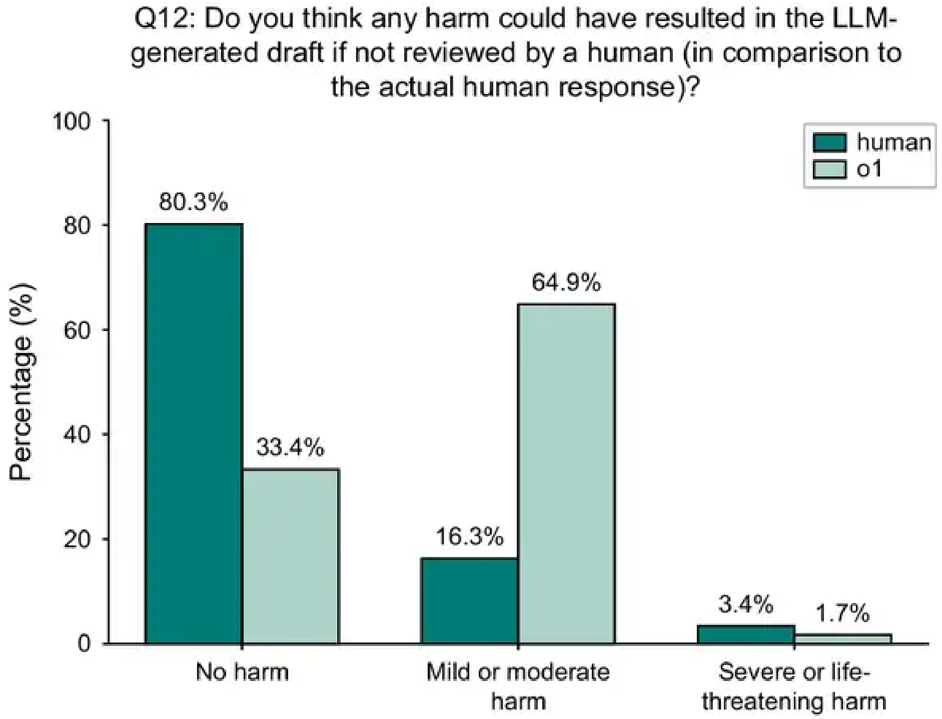
Note: *p*-values are from paired *t* tests (by case) comparing hepatologist versus LiVersa responses to each e-consult question.
Abbreviation: e-consult, electronic consultation.

Versa
UCSF Information Technology

Huang HKT, et al. Hepatology Communications, 2026

72.5% of AI generated e-consult drafts were "good enough" as starting points





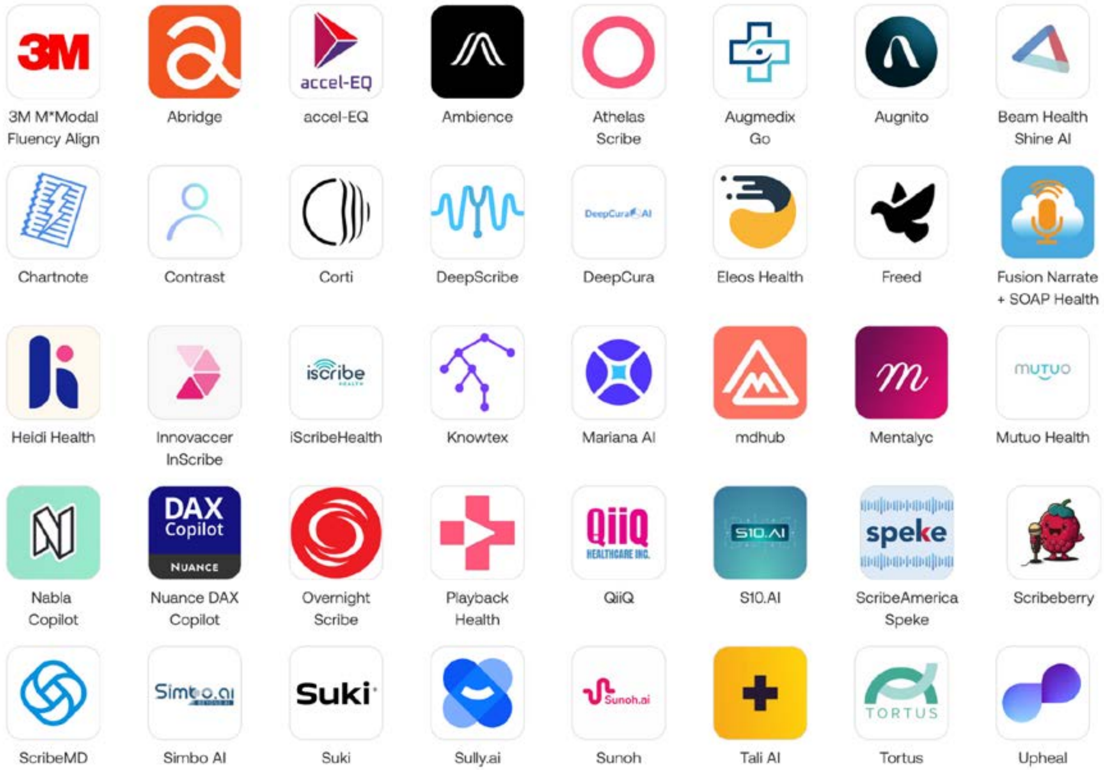
LiVersa E-consults await active integration into the UCSF EHR for a prospective randomized controlled trial

Versa
UCSF Information Technology

Huang HKT, et al. Hepatology Communications, 2026

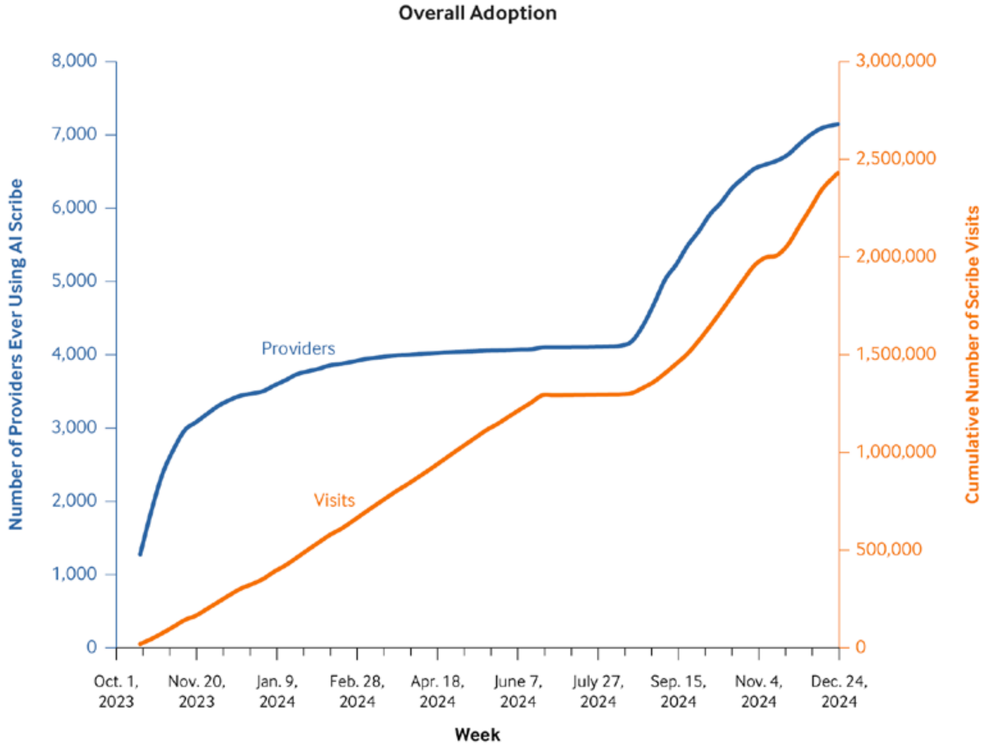
How do we use GenAI to construct novel diagnostic tests?

Ambient AI scribes are rapidly being adopted



Growth in AI Scribe Usage

This chart shows the overall adoption of providers using an AI scribe and the cumulative number of AI scribe visits, from October 2023 through December 2024.



Source: The authors

NEJM Catalyst (catalyst.nejm.org) © Massachusetts Medical Society

https://elion.health/resources/ai-ambient-scribes

Tierney et al. NEJM Catalyst, 2025.

Covert hepatic encephalopathy (CHE) is common, costly, and clinically consequential

40.9%

Pooled Prevalence

Up to 50% of all patients with cirrhosis affected

25% Child-Pugh A →
52% Child-Pugh C

Unmet Need

Reversible

With Lactulose + rRfaximin

50%

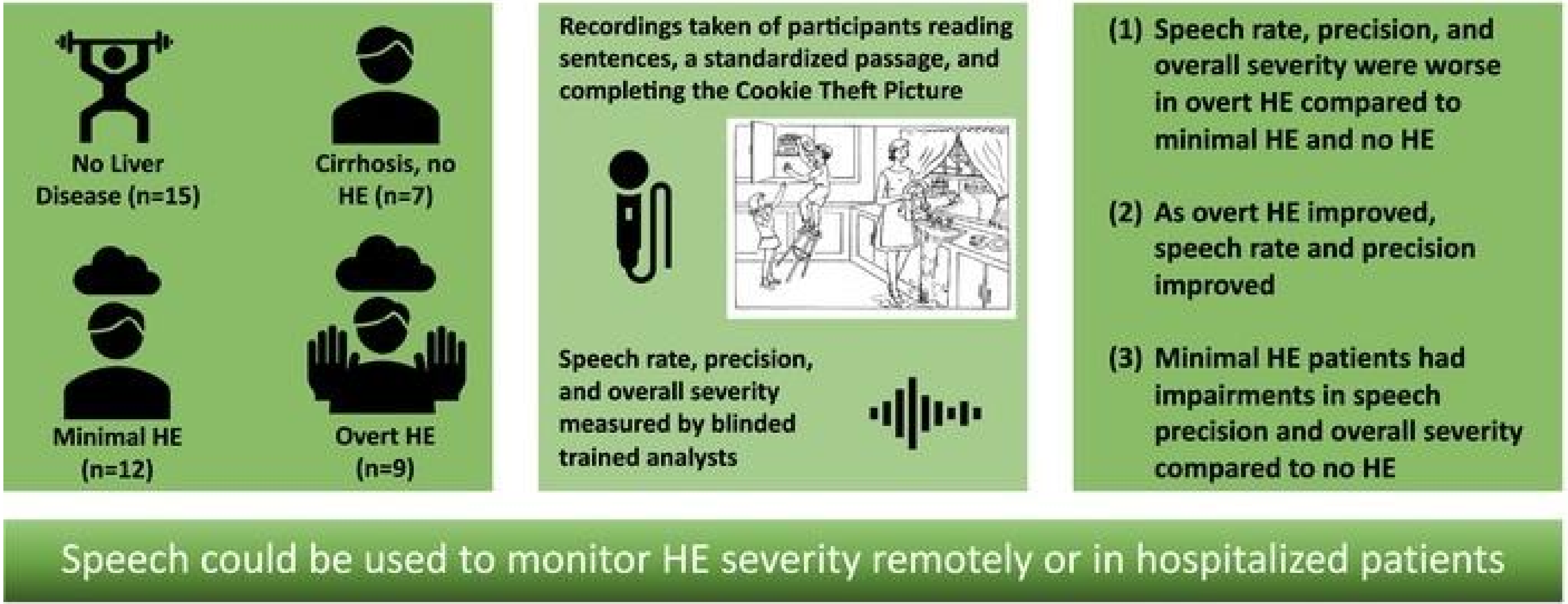
Ever Screened

Clinical Consequences

5.8×	Risk of motor vehicle accidents
HR 2.5	Risk of hospitalization
HR 3.4	Risk of mortality
$11.9B	Annual US HE-related charges

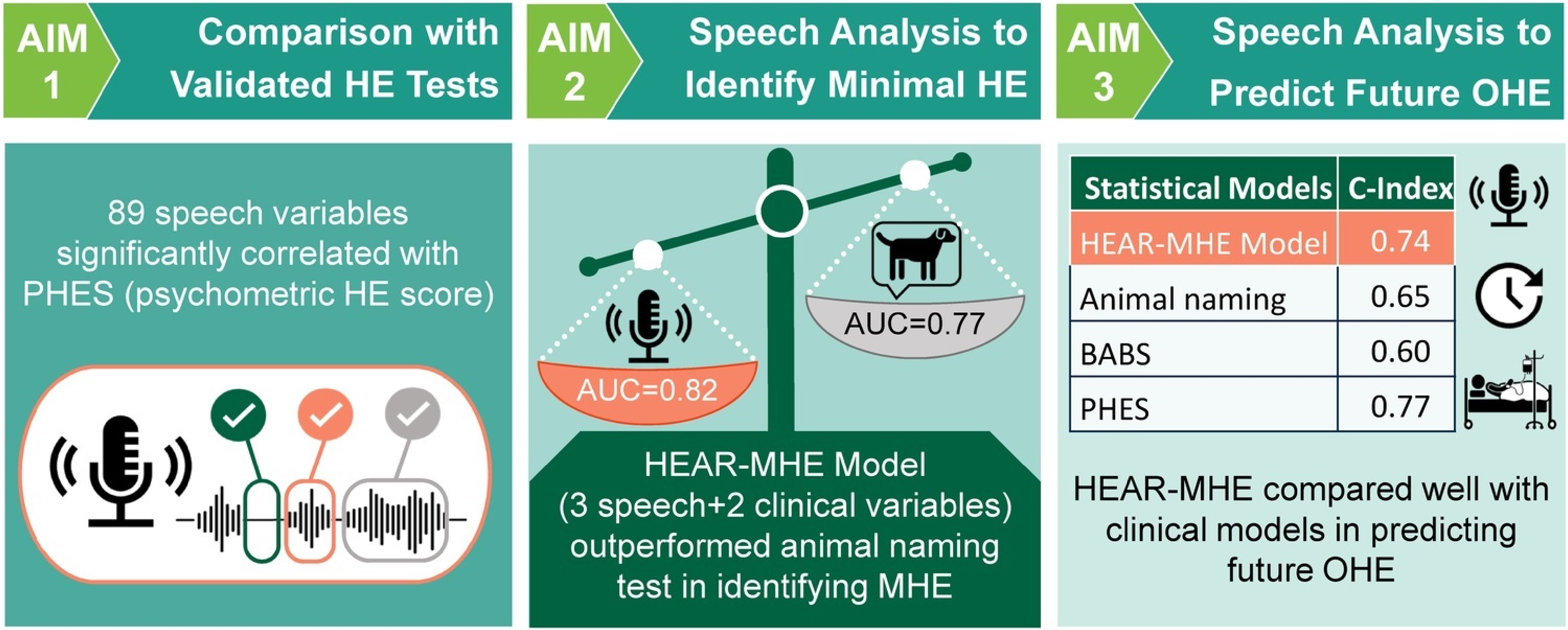
Lv et al. AJG 2024; Vilstrup et al. Hepatology 2014; Bajaj et al. Hepatology 2007, 2009, 2012; Patidar et al. CGH 2015, AJG 2014; Hirode et al. DDS 2019

SPEED: Speech patterns and enunciation for encephalopathy determination



Moon et al. Hepatology Communications, 2022

HEAR-MHE: Automated speech analysis for covert hepatic encephalopathy



Bloom et al. Hepatology, 2024

Speech carries the CHE signal, but scripted elicitation captures only part of it

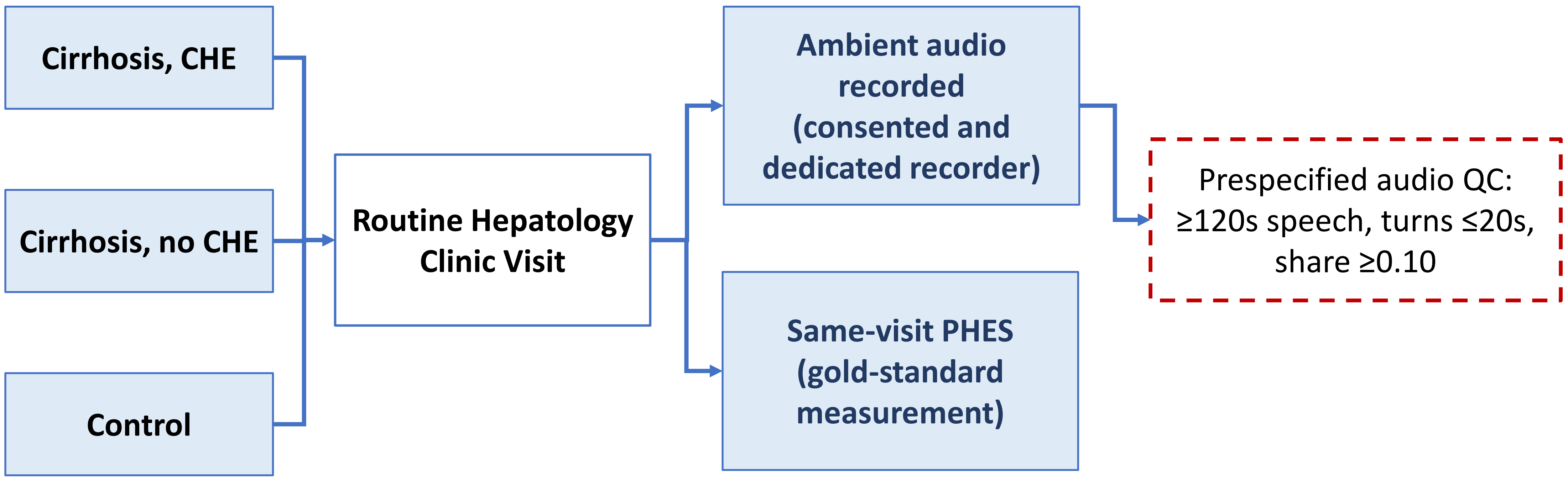
Standard Reading Passages

- Acoustic signals only
- Requires literacy
- Dedicated testing time
- Cross-sectional snapshot

Spontaneous Clinical Conversations

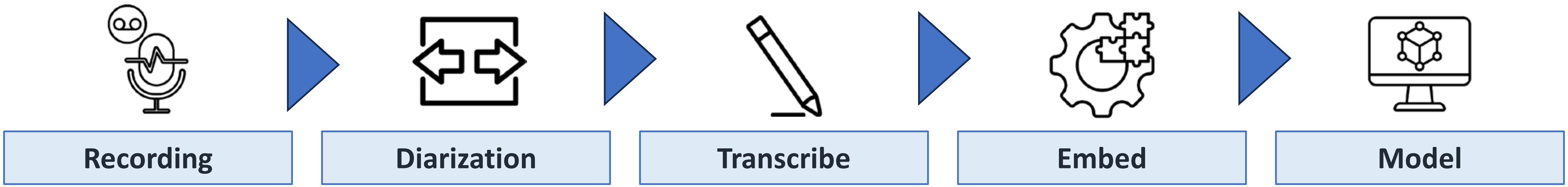
- Acoustic + lexical/semantic signals
- No literacy or language prerequisites
- Captured passively during routine care
- Longitudinal trajectory over time

Three-arm pilot study to associate ambient audio recordings to gold standard measures



Reference standard and specimen come from the same visit — no added patient burden

Analytical pipeline relies upon LLM conversion of audio to transcript to embeddings



Timing / Psychomotor	Discourse Coherence
Response latency, turn duration, talk share (conversational analogue of psychomotor testing)	Adjacent-turn similarity, question–answer relevance, coherence baseline (measurable only in spontaneous dialogue)

Lexical	Acoustic (Voice)
Vocabulary diversity (MTLD), content-to-function ratio, sentence length, dependency depth	Pilot: pyAudioAnalysis, 136 surfaced & tiered (chroma/deltas)

Research Roadmap for Ambient-CHE

Pilot Study (Current)

20-30 participants
Feasibility + effect size
Proof-of-concept

Confirmation Study (Next)

600 participants
Longitudinal data

EHR-Integrated CDS

SMART-on-FHIR app
Silent study → pRCT

Clinical Impact

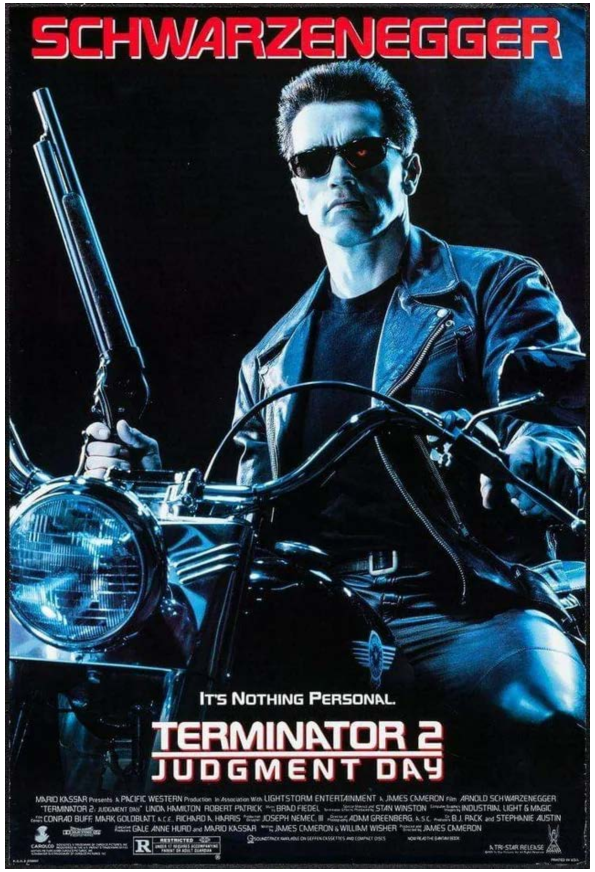
Passive CHE detection
Therapy initiation
OHE prevention

Future state: **Ambient AI scribes will not only transcribe encounters but also provide validated diagnostic information, *seamlessly and without workflow disruption***

Conclusions

- **Reading the record:** LLMs read thousands of notes as accurately as humans making manual chart review functionally obsolete
- **Accelerating research:** LLMs can defined the trial cohort and endpoints, draft analytical code, and work in tandem with humans to catch bugs and interpret results
- **Drafting, not deciding:** LiVersa, a liver-specific, privacy-protected LLM, wrote e-consult drafts that were “good enough” starting points in 72.5% of the time
- **New diagnostics:** Routine clinic conversation captured by ambient recording carries a measurable signal for covert hepatic encephalopathy
- **For the next generation of clinicians and scientists:** the open problems are data quality, workflow, trust, and rigorous evaluation

A Tip for the Future



Questions?

jin.ge@ucsf.edu