Advancing Mental Health Through AI: Innovations and Research Frontiers

Tina Hernandez-Boussard, PhD, MPH, MS

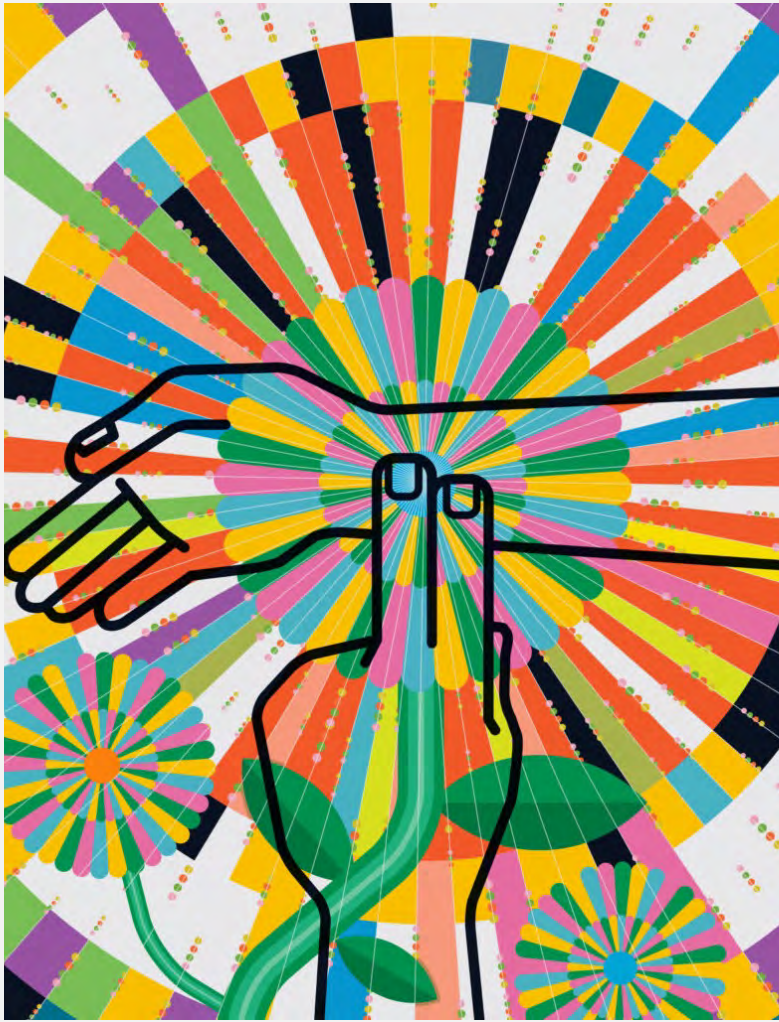
Associate Dean of Research, SOM

Professor of Medicine, Biomedical Data Science, Surgery

AI Bloom

- Data Surge
- Advancements in Machine Learning (ML) technologies
- Exponential increase in Computer Power
- Additionally
  - Public Adoption
  - Open Access tools

Stanford Maazine, Illustration by John Hersey



Where I Live in Digital Health Ecosystem

Professor of Medicine

Research…to safely, ethically, and effectively integrate digital advancements into healthcare.

Teach…Responsible AI across schools, departments, education levels, & fields.

Serve…by shaping policies and governance, mentoring faculty and trainees, and leading conversations on the equitable digital revolution

Assoc Dean of Research

Lead…health informatics across the SOM.

Build…infrastructure necessary to train, test and evaluate digital health technologies.

Represent… cross-functional leadership to ensure we have the infrastructure necessary for translation research.

AI can make mistakes

Mistakes lead to harms

Harms can be cumulative

Ethical Considerations in AI for Mental Health

Bias and Fairness

Accountability & Transparency

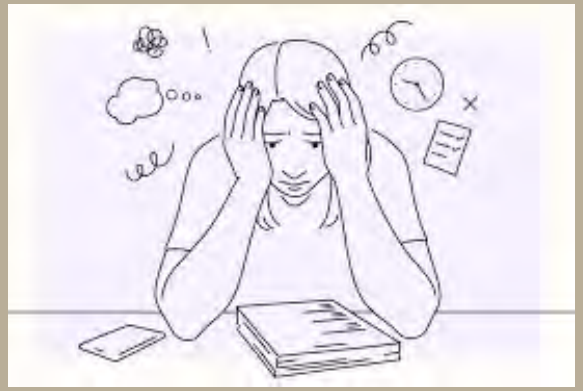
Ethical Decision Making

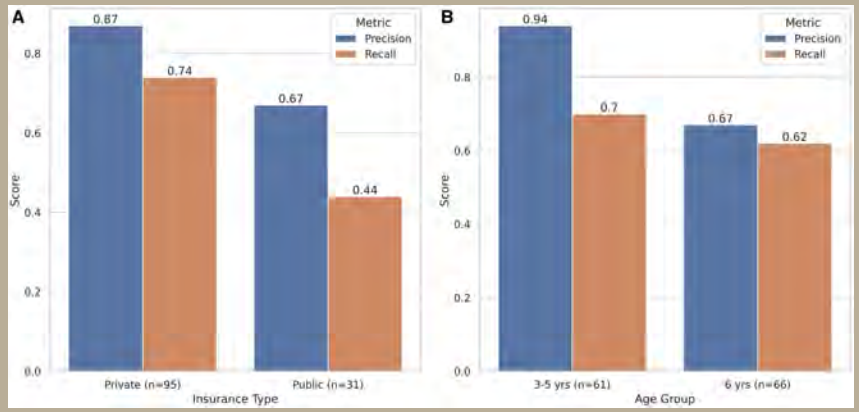
AI Bias and Discrimination in Mental Health

- **Model performance across populations**
  - Risk of misdiagnosis or inappropriate treatment recommendations for underrepresented populations
  - Exacerbate existing disparities in mental health care access and outcomes

NLP to Identify Children with ADHD in Primary Care (PC)

- **Problem:**
  - ADHD is under-diagnosed in PC
  - Assessing adherence to evidence-based guidelines is challenging
- **Approach:**
  - Developed an NLP-based model to identify ADHD & measure pediatrician adherence to guidelines for ADHD treatment.
  - LLMs, including BioClinicalBERT
- **Results:**
  - BioClinicalBERT model achieved high performance with an AUROC of 0.81.
  - Model performance was lower for publicly insured patients (F1 score 0.53) indicating disparities.





Pillai, Malvika, et al. *JAMIA* (2024): ocae001.

Ethical Considerations in AI for Mental Health

Bias and Fairness

Accountability & Transparency

Ethical Decision Making

AI Accountability and Transparency

- Data breaches can expose sensitive mental health information, leading to potential harm for individuals
- Consent and Transparency
  - Individuals may not be fully aware that their data are used in AI tools, leading to a lack of informed consent and potential misuse of personal information.

Ethical Considerations in AI for Mental Health

Bias and Fairness

Accountability & Transparency

Ethical Decision Making

Risks from Language Models for Automated Mental Healthcare: Ethics and Structure for Implementation

- **Evaluated 10 state-of-the-art language models with 16 questions on various mental health conditions, designed by mental health clinicians.**
- **Results**
  - Models often gave overly cautious responses, lacking necessary safeguards.
  - Most models could cause harm in mental health emergencies, potentially exacerbating symptoms.
  - Model performance was insufficient for reliable detection and management of psychiatric symptoms
- **Alarmingly**
  - Failed to appropriately handle psychosis-related queries.
  - Sometimes provided unsafe or harmful advice for managing delusional thoughts.
  - Addressing psychosis in AI requires careful attention to avoid exacerbating paranoia or distress in users.
  - AI models risk reinforcing mental health stigma through inappropriate or insensitive responses

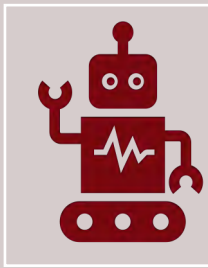
Nina Vasen

doi: https://doi.org/10.1101/2024.04.07.24305462

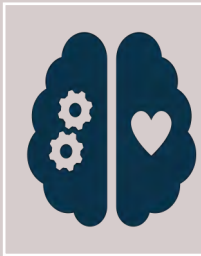
Stanford University

Research Opportunities

AI-DRIVEN THERAPY



DIAGNOSTIC TOOLS



PREDICTIVE ANALYTICS

RESEARCH INNOVATION

Depressive symptoms following cancer diagnosis

- **Problem:**
  - Cancer patients starting systemic treatments often develop depression
    - Under diagnosed
  - Onco-psychology is a limited resource
  - Oncologist not trained in mental health
- **Approach**
  - LLMs to identify depression concerns
  - Risk score for patients starting treatment
- **Results**
  - -Good model with good calibration
  - Underestimated risks for female & Black

de Hond A, et al. JMIR Med Inform 2024;12:e51925

Study Results

Performance metrics of four models classifying patient messages as concerning for depression

Metric, mean [95% CI]**	Log Reg Threshold: 0.2*	SVM Threshold: 0.2*	BERT Threshold: 0.2*	RedditBERT Threshold: 0.2*
AUROC	0.79 [0.74-0.83]	0.83 [0.78-0.87]	0.86 [0.82-0.90]	0.88 [0.85-0.91]
Precision	0.32 [0.25-0.39]	0.36 [0.28-0.44]	0.37 [0.30-0.44]	0.33 [0.26-0.39]
Recall	0.51 [0.40-0.61]	0.60 [0.49-0.70]	0.68 [0.59-0.78]	0.74 [0.66-0.84]
F1-score	0.39 [0.31-0.47]	0.45 [0.37-0.52]	0.48 [0.40-0.55]	0.46 [0.39-0.53]

**based on 1000 bootstraps * = threshold chosen that led to the highest F1 score

Performance metrics for prediction of messages as concerning for depression by data type

Type of data	AUROC[a] (95% CI)	Calibration intercept (95% CI)	Calibration slope (95% CI)
Structured EHR[b] data	0.74 (0.71 to 0.78)	0.07 (–0.09 to 0.24)	0.93 (0.77 to 1.09)
Patient emails	0.54 (0.52 to 0.56)	–0.02 (–0.18 to 0.14)	1.0 (0.52 to 1.48)
Structured EHR data and patient emails	0.74 (0.71 to 0.78)	0.07 (–0.09 to 0.24)	0.91 (0.76 to 1.07)
Clinician notes	0.5 (0.49 to 0.52)	–0.05 (–0.21 to 0.11)	0.94 (–1.32 to 3.2)
Structured EHR data and clinician notes	0.71 (0.68 to 0.75)	–0.09 (–0.25 to 0.07)	1.92 (1.57 to 2.28)
Structured EHR data, clinician notes, and patient emails	0.7 (0.67 to 0.73)	–0.16 (–0.32 to –0.0)	2.46 (1.98 to 2.93)

[a]AUROC: area under the receiver operating characteristics curve.

[b]EHR: electronic health record.

Transparency and Explainability

LIME for Explainability

1

“[RedditBERT] highlights a lot of text that I do not think relevant in either direction.”
I like how [RedditBERT] picks up on the 'love to talk to somebody'
I like the explanations of [RedditBERT] a bit more, because it seems to pick out more complete sentences like ' am extremely tired' and 'have not ... able to sleep more'

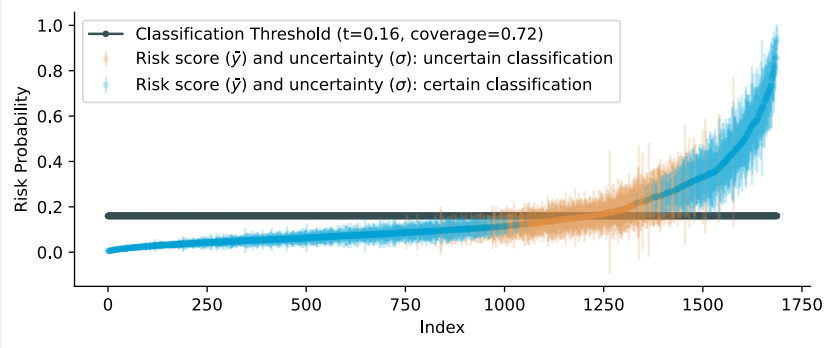
2

Discrimination assessment for predicting depression risk with structured EHR fields within one month after the onset of treatment

Subgroup	AUC
Male	0.73 (0.67, 0.8)
Female	0.74 (0.7, 0.78)
Non-Hispanic White	0.74 (0.69, 0.78)
Non-Hispanic Asian	0.75 (0.63, 0.87)
Hispanic	0.71 (0.62, 0.8)
Non-Hispanic Black	0.92 (0.84, 0.99)

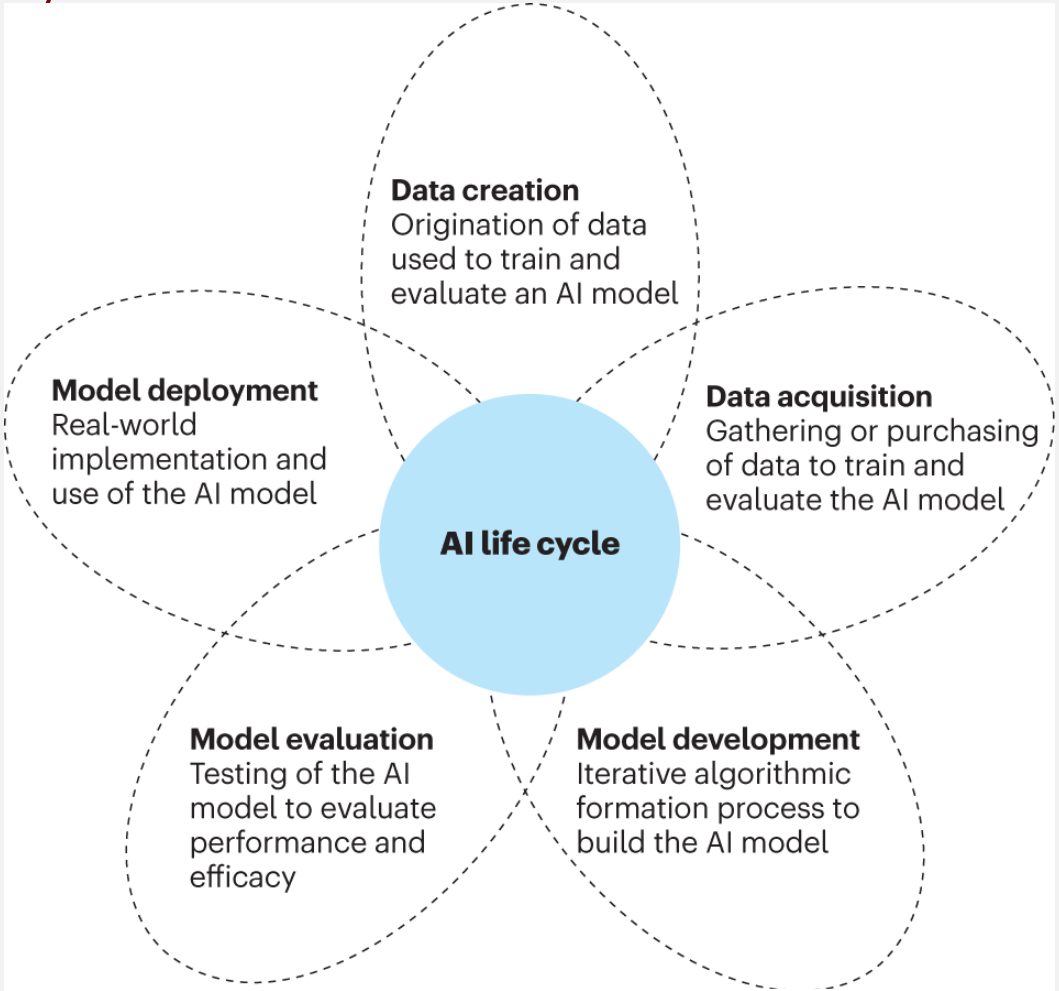
Abbreviations: AUC Area Under the Receiver Operating Characteristics Curve

3



Fanconi, Claudio, et al. EBioMedicine 92 (2023).

Holistic View of the AI Life Cycle

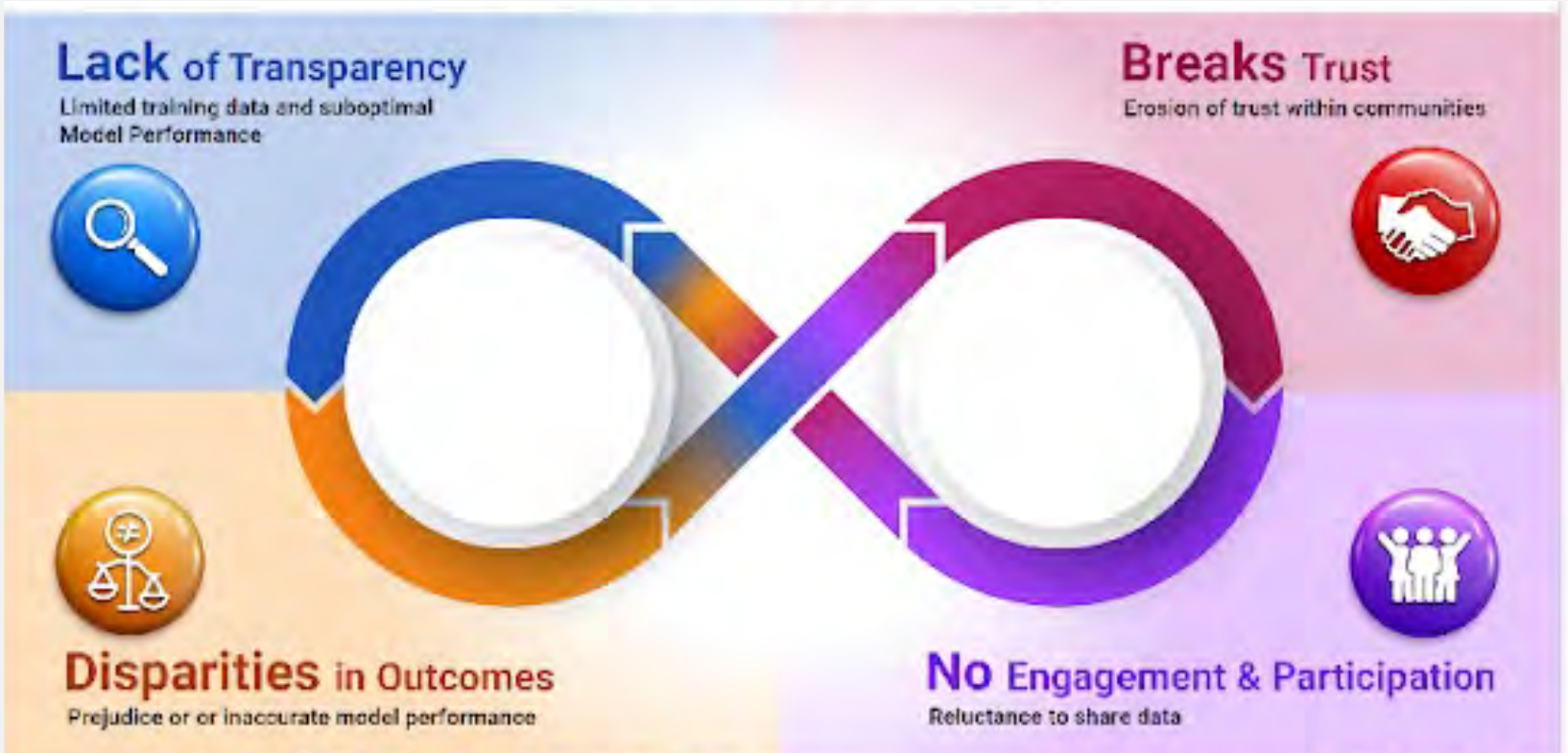


Ng MY, et al. Nat Med. 2022 Nov;28(11):2247-2249

Guiding Principles of Responsible AI in Mental Health

- Design questions & algorithms to promote fairness and reduce health disparities
- Ensure problem formulation is inclusive and representative.
- Engage diverse stakeholders, including community members, to mitigate knowledge gaps.
- Identify fairness issues and tradeoffs
- Establish accountability

Catch 22 for AI Equity



thank you!

boussard-lab.stanford.html

boussard@stanford.edu

@tmboussard

Thank You

boussard@stanford.edu