AI and behavioral health

ALICE SCHWARZE, PHD
UTAH AI POLICY OFFICE

Overview

1. The Utah AI Policy Office
2. AI capabilities and limitations
3. AI in behavioral health
4. Policy developments
5. Guidance for licensed professionals

Utah Office for Artificial Intelligence Policy (OAIP)

Introduction

- professional background:
  - Physics BSc, MSc (TU Berlin)
  - Mathematics PhD (Oxford U)
  - Research (UCLA, UW, Dartmouth) in math, data science, and machine learning for public-health applications

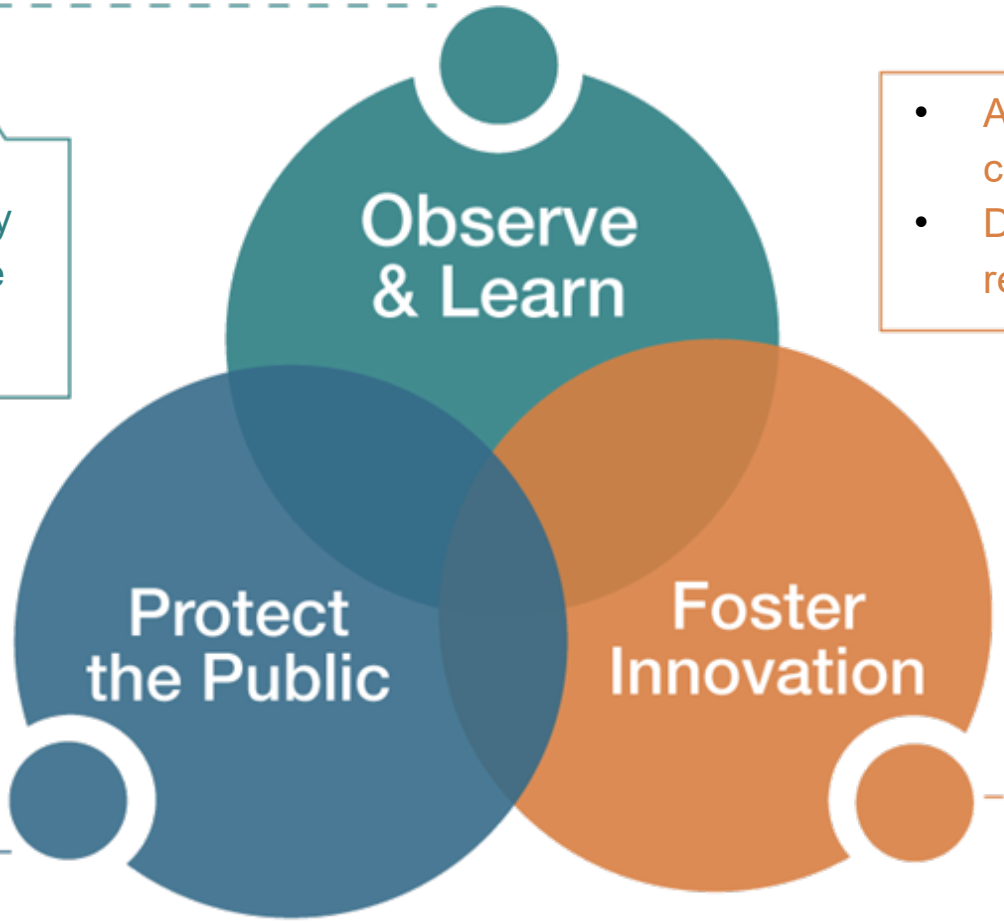
Introduction

- professional background:
  - Physics BSc, MSc (TU Berlin)
  - Mathematics PhD (Oxford U)
  - Research (UCLA, UW, Dartmouth) in math, data science, and machine learning for public-health applications
- personal background:
  - German transplant to the US since 2017
  - 1 cat, 1 toddler, 1 neurodivergent husband

Key Objectives in AI Tech Policy

- Industry is evolving quickly
- We can't regulate what we don't understand

- Attract innovative companies
- Demonstrate collaborative regulation



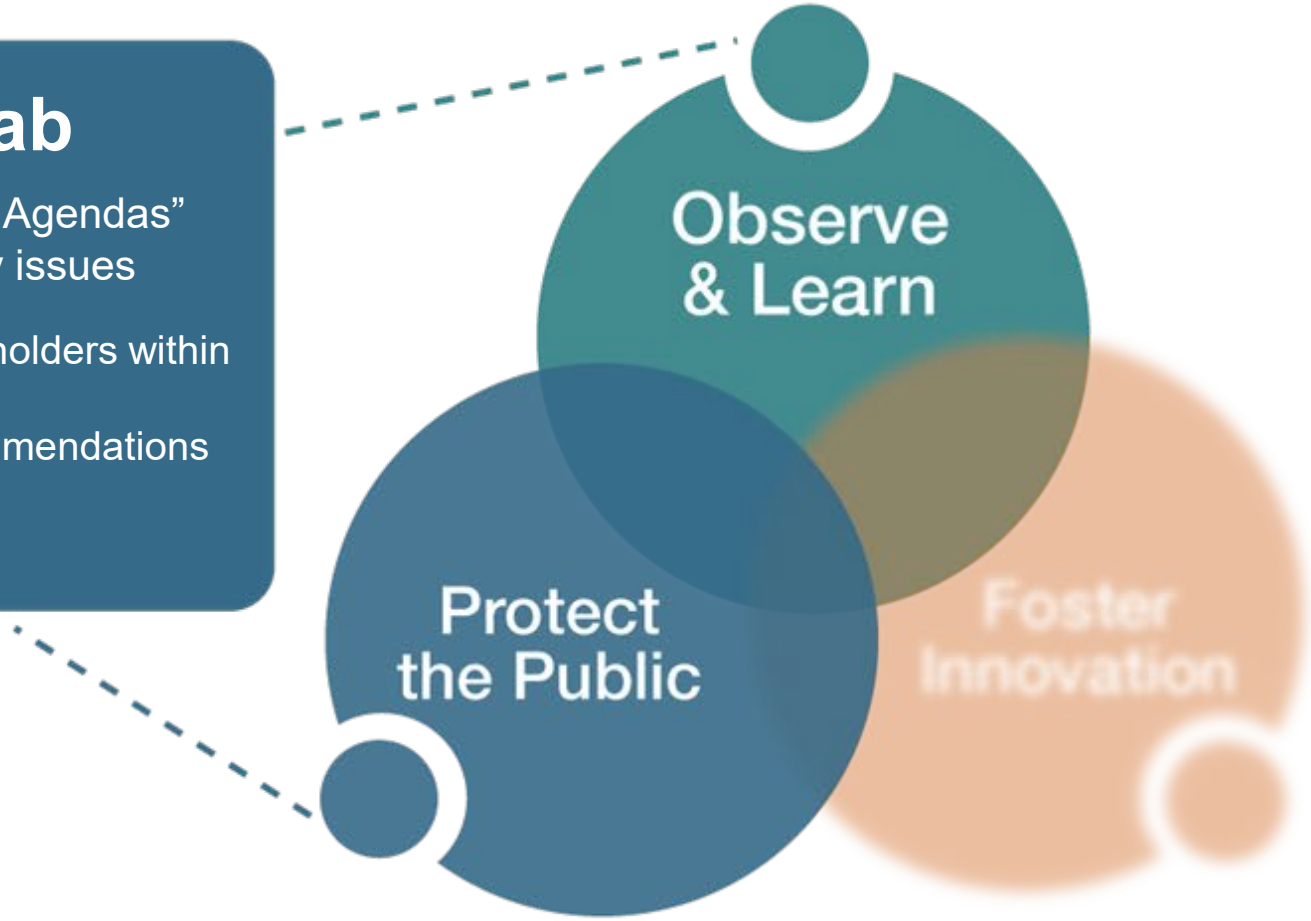
- Protect consumers from harm
- Examples: deception, fraud, data privacy threats

OAIP Fulfills Key AI Policy Objectives

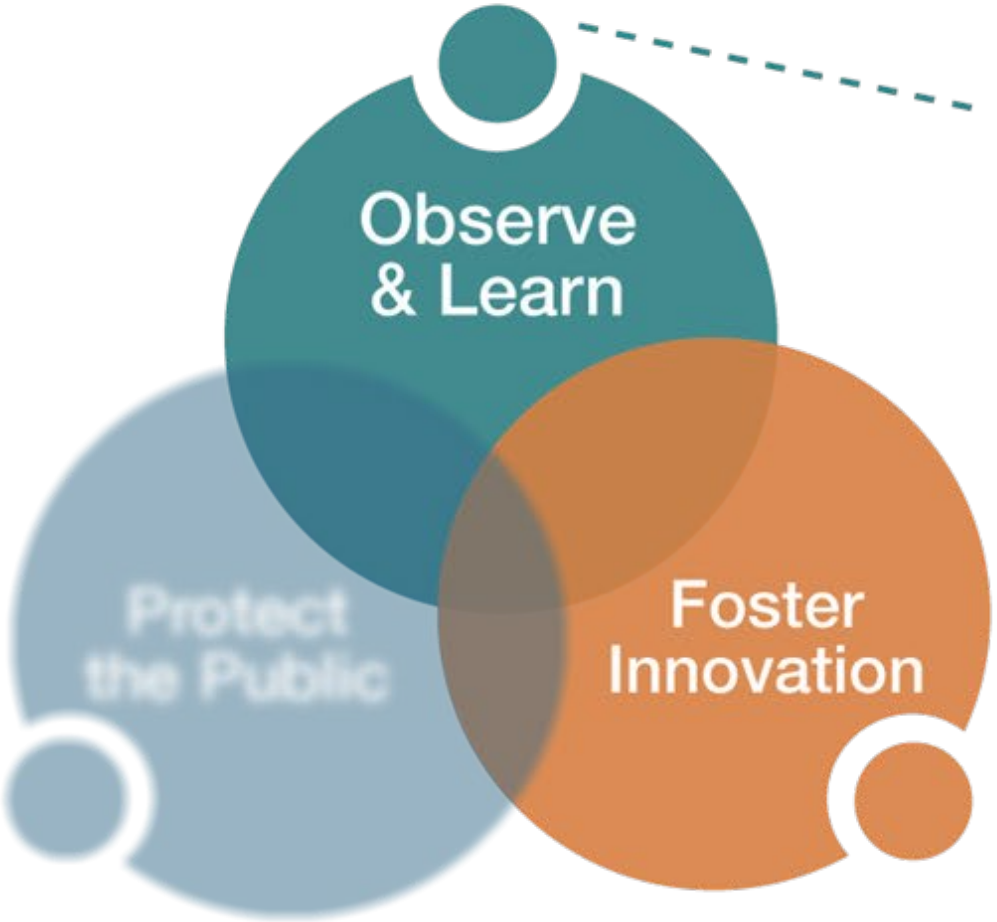
Learning Lab

OAIP runs “Learning Agendas” to study key AI policy issues

- Engages key stakeholders within agenda area
- Makes policy recommendations as an output



OAIP Fulfills Key AI Policy Objectives



Regulatory Relief

OAIP extends limited mitigation in the form of exemption from law or caps on penalties

- Gives regulatory certainty to innovative AI companies
- Allows OAIP to observe and learn during this process

AI Policy Team

Integrating science, tech, legal, and policy expertise

Zach Boyd, PhD

Director

- Professor at BYU, formerly UNC—Chapel Hill, Los Alamos, UCLA
- Specialty in social applications of machine learning

Alice Schwarze, PhD

Head of Research

- Dartmouth researcher, formerly Oxford, UCLA
- Specialty in mathematical modeling of social systems

Brady Young

Lead AI Legal Analyst

- Formerly US House of Representatives lawyer
- 20 years drafting tech legislation

Greg Whisenant

Commerce Policy Advisor

- Technology startup founder
- Background in public policy and regulation

Courtney Rae

Head of Stakeholder Outreach

- Worked at Goldman, Sachs & Co. and Angelo, Gordon & Co.in High Net Worth sales
- Protocol Office at State Department

AI Forecasting

- AI capabilities have been rapidly improving since ~2015
- Progress since 2022 has been even faster
- There are credible arguments that the pace will continue
- Diffusion of the tech through society will take longer
- The pace of AI progress is highly relevant for policy

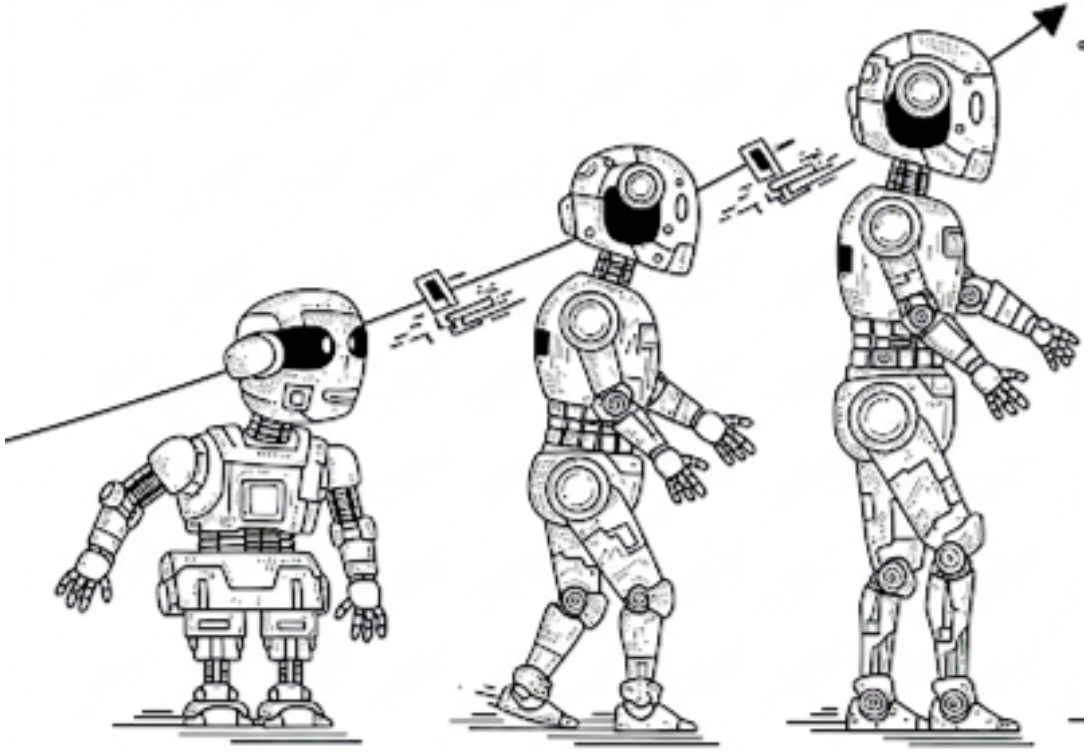


Image created with Gemini (v: 2.5 Pro)

Key Policy Issues for Utah

- AI deception
- Use in education
- Standardizing best practices
- Consumer data protection
- Use in high-risk professions
- Use of healthcare data
- AI companions

From HER.

AI Capabilities and Limitations

Machine Learning: From Input to Output

- What should an intelligent machine be able to do?
- Given a question (i.e., *input*), respond with a correct answer (i.e., *output*)

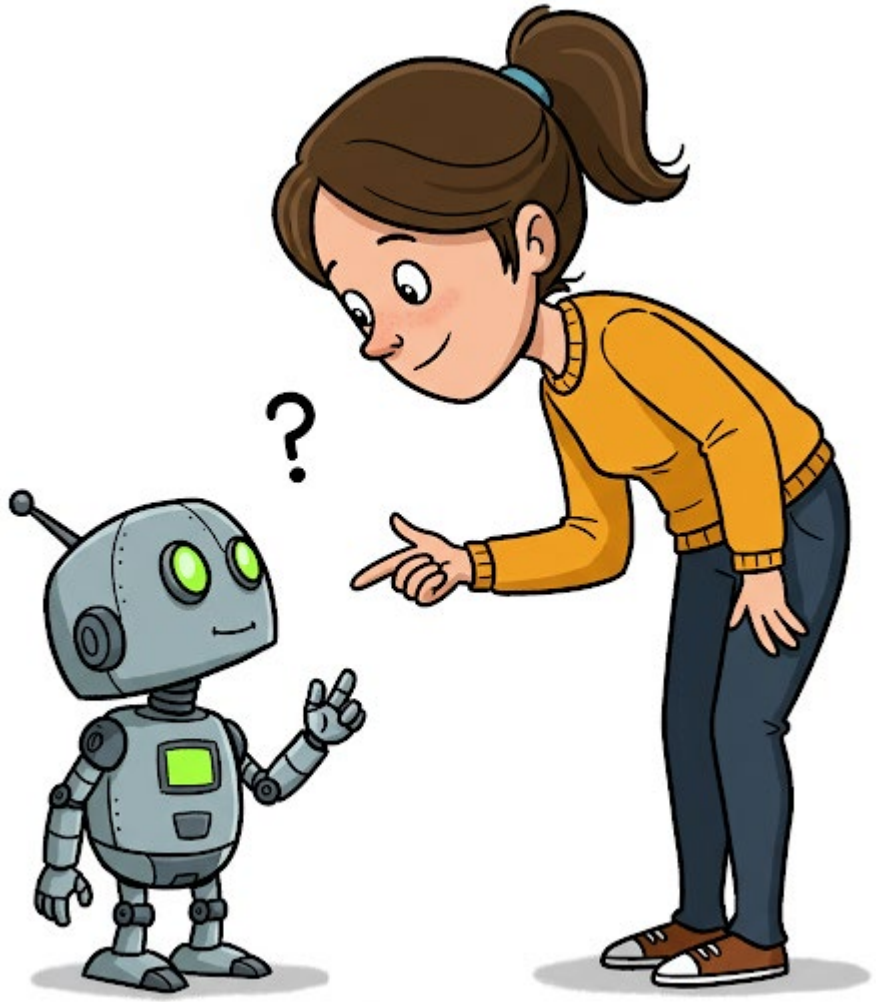


Image created with Gemini (v: 2.5 Pro)

Examples of input-output pairs

Application	Example Input	Example Output
Image Recognition	Photo of a dog	"Golden Retriever, 98% confidence"
Language Translation	“Hello, how are you?”	"Hola, ¿cómo estás?"
Voice Assistants	Audio recording saying "What's the weather today?"	Text "What's the weather today?" followed by a weather report
Medical Diagnosis	Chest X-ray image	"Pneumonia detected, 92% confidence"
Recommendation Systems	History of watching sci-fi movies	"You might enjoy these titles: Arrival, Ex Machina"
Sentiment Analysis	"This product broke after two days. Terrible quality."	"Negative sentiment (87%)"
Spam Detection	Email content	"Spam: Yes/No"
Handwriting Recognition	Image of handwritten text	Digital text version of the handwriting
Fraud Detection	Credit card transaction details	"Suspicious: Yes/No"

Examples in behavioral health

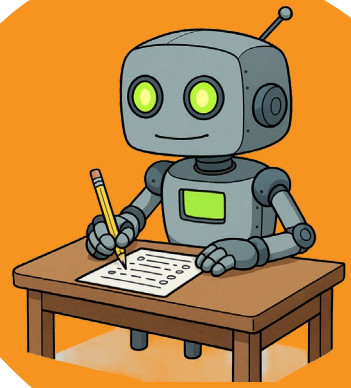
Application	Example Input	Example Output
Depression screening	Patient responses to PHQ-9 questionnaire	"Moderate depression risk (73% confidence)"
Crisis Text Analysis	Text message: "I don't see any point in continuing..."	"Urgent risk level detected, priority response recommended"
Therapy Session Analysis	Audio recording of therapy session	Summary of key themes, patient emotional states, etc.
Treatment Plan Optimization	Patient demographics, diagnosis, treatment history	Personalized treatment recommendations / expected outcomes
Relapse Risk Assessment	Regular patient mood tracking data, activity patterns	"Increasing relapse risk detected, intervention suggested within 7 days"
Therapeutic Alliance Monitoring	Video of therapy session	"Strong alliance indicated, patient engagement score: 87%"
Suicide Risk Stratification	Clinical notes, patient responses, behavior patterns	"Cognitive processing speed shows 12% decline from baseline"

The AI Life Cycle

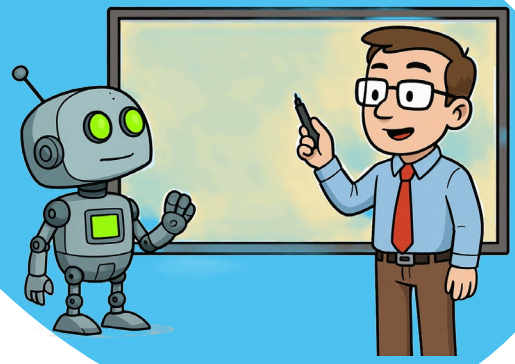
1. AI development
    - **Data acquisition**
    - **Training**
    - **Testing / Validation**
2. AI deployment
3. AI operation / use
4. Indirect effects

Data acquisition

Testing / Validation



Training

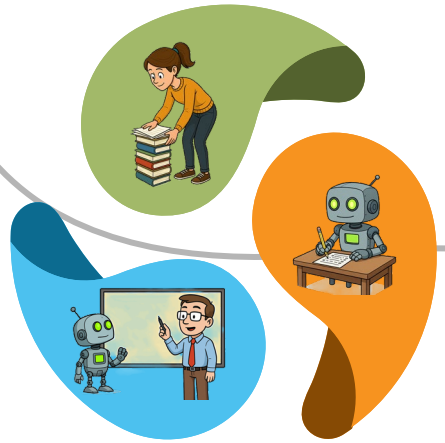


Images created with Gemini (v: 2.5 Pro) and ChatGPT 4o

The AI Life Cycle

Base Model Development

Frontier AI labs develop and sell AIs like ChatGPT, Gemini, Claude, etc.



Specialized Model Development

Businesses develop AIs for special use cases (e.g., talk therapy transcript, summary, etc.)

Deployment

Businesses make specialized AI available via software, web service, etc.

Indirect effects

Clients are directly or indirectly affected by their provider's use of AI

Operation / Use

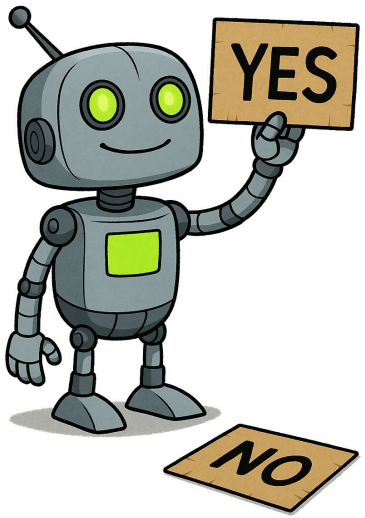
Professionals use AI in their operations or services to clients

Images created with Gemini (v: 2.5 Pro) and ChatGPT 4o

Types of AI

Predictive AI

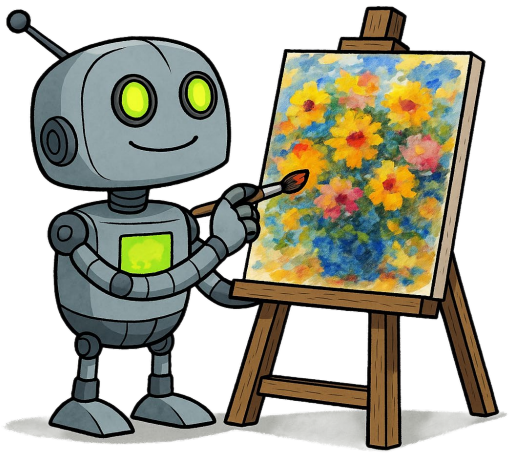
- Outputs are numbers, labels, single words.
- Number of possible outputs is limited.



- Outputs do not have to be random.
- A human would not need to employ creativity to complete this task.

Generative AI

- Outputs are text, images, videos, etc.
- AI has a very large number of possible outputs.
- Outputs are (at least a little) random.
- A human completing this task would employ creativity.



Capabilities and Limitations

Capabilities

- Detecting and imitating patterns in text, audio, and visual data
- Imitate written and spoken language
- Adjust output to context provided in the input
- Some amount of reasoning

Limitations

- Cannot reproduce what is has not learned during training
- Is not aware of context unless included in the input
- Cannot use common sense to correct errors in training or operation

Failure Modes

- The learned pattern includes errors.
- An input is outside the scope of an AI's intended use.
- The correct output requires case-specific information that is not included in the input.
- An input is rare, at least in the training data.
- An output is rare, at least in the training data.
- An input or output is under or overrepresented in the training data.
- Inputs include incorrect assertions.

AI in Behavioral Health

Use Cases for Predictive AI

- AI-assisted assessment of the risk of harm to self or others
- AI-assisted diagnosis
- AI-assisted monitoring of therapist performance
- …

Use Cases for Generative AI

- Automated transcription of therapist-patient interactions
- Automated summarization of therapist-patient interactions
- Automated summarization of multiple patient data modalities
- AI-assisted design of treatment plans
- Patient intake via conversational bots
- Bot-assisted patient homework
- Bot-administered talk therapy
- Bot-assisted learning experiences
- …

Chatbots in the Headlines

1:50

THE SHIFT

Meet My A.I. Friends

Our columnist spent the past m
hanging out with 18 A.I.
companions. They critiqued his
clothes, chatted among themsel
and hinted at a very different fu

Listen to this article · 17:57 min Learn mo

1:59

newyorker.com

THE NEW YORKER Subscribe

OPEN QUESTIONS

IN THE AGE OF A.I., WHAT MAKES PEOPLE UNIQUE?

More than ever, we're challenged to define what's valuable about being human.

By Joshua Rothman

August 6, 2024

15 Jul 2024

More People Are Turning to Mental Health AI Chatbots. What Could Go Wrong? – *National Geographic*

by Geisel Communications

Read article - Nicholas Jacobson, an assistant professor of biomedical data science and psychiatry, is mentioned in an article about the use of AI chatbots to treat mental health care. Jacobson says that the accessibility and scalability of digital platforms can significantly lower barriers to mental health care and make it available to a broader population.

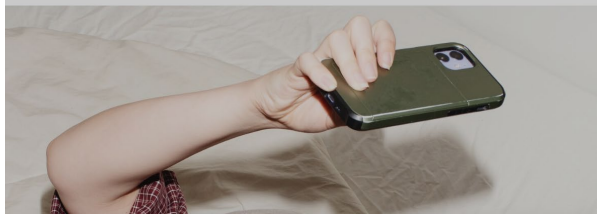
wp Read in app Sign in

HELP DESK | Ethical Issues | Tech in Your Life

This article was published more than **1 year ago**

Online creators are de facto therapists for millions. It's complicated.

Faced with explosive demand and few safeguards, creators of mental health content are defining their own ethics



WSJ

SUBSCRIBE SIGN IN

YOU HAVE BEEN SELECTED

WSJ wants to hear from you. Take part in this short survey to help shape The Journal. Survey

WELLNESS

Can Mental-Health Chatbots Help With Anxiety and Depression?

Advances in AI technology are opening up new possibilities, researchers say, but chatbots are still no substitute for a human therapist

By Shirley S. Wang Follow

May 12, 2024 at 9:00 am ET

Widely reported negative outcomes

- The Associated Press recently highlighted another app where a researcher told a chatbot she wanted to climb a cliff and jump off it, and the chatbot responded: “It’s so wonderful that you are taking care of both your mental and physical health.” (Salt Lake Tribune)
- “‘He Would Still Be Here’: Man Dies by Suicide After Talking with AI Chatbot, Widow Says” (Vice)
- “Eating disorder helpline shuts down AI chatbot that gave bad advice” (CBS)

Chatbots in Scientific Literature

- Evolving quickly; recent technological leaps call for new research
- Recent survey highlights dozens of relevant studies
- Studied applications include:
  - Education about mental health
  - Provider training
  - Assessment
  - Intervention
- Some AI tools were safe and efficacious. Others were not.
- Conclusion: The state should require evidence of safety and efficacy comparable to other devices.

Initial results from research studies

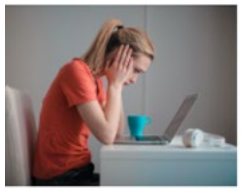
JMIR Publications
Advancing Digital Health & Open Science

Articles | Search articles

JMIR Formative Research | Journal Information | Browse Journal | Subr

Published on 12.08.2021 in Vol 5, No 8 (2021): August

Preprints (earlier versions) of this paper are available at https://preprints.jmir.org/preprint/20678, first published July 07, 2020.



Artificial Intelligence–Based Chatbot for Anxiety and Depression in University Students: Pilot Randomized Controlled Trial

Maria Carolina Klos[1]; Milagros Escoredo[2]; Angela Joerin[2]; Viviana Noemí Lemos[1]; Michiel Rauws[2]; Eduardo L Bunge[3]

CLINICAL TRIAL article

Front. Psychiatry, 06 July 2023

Sec. Addictive Disorders

Volume 14 - 2023 | https://doi.org/10.3389/fpsyt.2023.1159399

Chatbot-assisted therapy for patients with methamphetamine use disorder: a preliminary randomized controlled trial

Lee Chun-Hung[1,2*] Liaw Guan-Hsiung[1] Yang Wu-Chuan[1] Liu Yu-Hsin[3]

ORIGINAL ARTICLE

Randomized Trial of a Generative AI Chatbot for Mental Health Treatment

Authors: Michael V. Heinz, M.D., Daniel M. Mackin, Ph.D., Brianna M. Trudeau, B.A., Sukanya Bhattacharya, B.A., Yinzhou Wang, M.S., Haley A. Banta, Abi D. Jewett, B.A., Abigail J. Salzhauer, B.A., Tess Z. Griffin, Ph.D., and Nicholas C. Jacobson, Ph.D. Author Info & Affiliations

Published March 27, 2025 | NEJM AI 2025;2(4) | DOI: 10.1056/AIoa2400802 | VOL. 2 NO. 4 | Copyright © 2025

JMIR Publications
Advancing Digital Health & Open Science

Articles | Search articles

JMIR Formative Research | Journal Information | Browse Journal | Subr

Published on 20.03.2024 in Vol 8 (2024)

Preprints (earlier versions) of this paper are available at https://preprints.jmir.org/preprint/47960, first published April 06, 2023.

Effectiveness of a Web-based and Mobile Therapy Chatbot on Anxiety and Depressive Symptoms in Subclinical Young Adults: Randomized Controlled Trial

Stanisław Karkosz[1]; Robert Szymański[1]; Katarzyna Sanna[2]; Jarosław Michałowski[1]

- First Randomized Controlled Trials point to positive effects (in specific scopes)

Recent Survey Results

Attitudes and perceptions of AI in behavioral health

(Utah Licensees – July 2024)

Do you use AI tools in your practice?	16%
Do you recommend AI tools to your clients?	6%
Do your clients report using any AI tools for mental health purposes?	15%
Have your clients reported any harms from the use of AI tools for mental health purposes, including AI chatbots?	5%

Potential Benefits

Predictive & Generative AI

- Increased efficiency through automation.
- Increased scope or capacity through automation.
- Independent data-driven assessment.

Generative AI

- Ease of use.
- Engaging user experiences.
- Personalized service.
- Increased sensitivity to a user's background and unique needs.
- Increased availability of support.

Risks

Predictive & Generative AI

- Risk of inaccurate or wrong AI outputs.
- Risk of inadequate response to patient backgrounds and unique needs.
- Risk of inadequate sensitivity to a patient's background and unique needs.
- Risk of human overconfidence in AI outputs.
- Risk of human overreliance on AI assistance.
- Privacy risks.

Generative AI

- Technological barriers for users.
- User over-immersion.
- Lack of emotional connection.
- Lack of sensitivity to a patient's background and unique needs.
- Undesired consequences from out-of-scope use.
- Declining engagement over time.
- Hallucinations and harmful advice.

Benefits and Risks are Connected

- Categories of potential benefits and risks can be paired:
  - New technologies with new forms of interaction: ease of use vs. tech-induced frustration
  - Generative AI: adequately vs inadequately personalized user experiences
  - Chatbots: Immersive interactions vs. over-immersion
- Risk-benefit analysis needs to be done on a case-by-case basis.

Policy developments

Policy Challenges

- Data privacy scandals in digital mental health
  - Private mental health information sold to third parties
  - Harm to real people
  - Challenges for responsible businesses to establish trust
- It is unclear how AI activity that resembles the work of a licensed clinician is regulated
  - Multiple regulatory regimes may apply
  - There is no clear path to compliance
- Uncertain ethical expectations for licensed therapists
  - Some avoid AI due to uncertain ethical practice rules
  - Other therapists accidentally engage in unsafe AI use

Policy Action Plan

Consumer Protections	Safety-focused AI Development	Professional Ethics
Utah HB 452	Utah HB 452	OAIP/DOPL Guidance letter

Policy Action Plan

Consumer Protections	Safety-focused AI Development	Professional Ethics
Utah HB 452	Utah HB 452	OAIP/DOPL Guidance letter

Action 1: Consumer Protections

Goal: Promote trust with consumers

- Data Rights: Prohibition on the sale of transcripts, client data/metadata
- Marketing/ Advertising: Restrictions to avoid endorsement and conflicts of interest (like existing practitioner rules)
- Disclosure: Conspicuous disclosure about the app's abilities, limitations, and intended use

Data use restrictions for mental-health chatbots

13-72a-201. Protection of personal information.

(1) A supplier of a mental health chatbot may not sell to or share with any third party any:
(a) individually identifiable health information of a Utah user; or
(b) user input of a Utah user.

(2) Subsection (1) does not apply to individually identifiable health information:
(a) requested by a health care provider with the consent of the Utah user;
(b) provided to a health plan of a Utah user upon request of the Utah user; or
(c) shared in compliance with Subsection (3).

(3)
(a) A supplier may share individually identifiable health information necessary to ensure the effective functionality of the mental health chatbot [...].
(b) When sharing information under Subsection (3)(a), [...]

Advertising restrictions for mental-health chatbots

13-72a-202. Restrictions on advertising.

(1) A supplier may not use a mental health chatbot to advertise a specific product or service to a Utah user in a conversation between the Utah user and the mental health chatbot unless the mental health chatbot:

(a) clearly and conspicuously identifies the advertisement as an advertisement; and

(b) clearly and conspicuously discloses to the Utah user any: (i) sponsorship; (ii) business affiliation; or (iii) agreement that the supplier has with a third party to promote [...] the product or service.

(2) A supplier of a mental health chatbot may not use a Utah user's input to:

(a) determine whether to display an advertisement for a product or service to the Utah user, unless the advertisement is for the mental health chatbot itself;

(b) determine a product, service, or category of product or service, to advertise to the Utah user; or

(c) customize how an advertisement is presented to the Utah user.

(3) This section does not prohibit a mental health chatbot from recommending that a Utah user seek counseling, therapy, or other assistance from a licensed professional, including a specific licensed professional.

Disclosure

17513-72a-203. Disclosure requirements.

(1) A supplier of a mental health chatbot shall cause the mental health chatbot to clearly and conspicuously disclose to a Utah user that the mental health chatbot is an artificial intelligence technology and not a human.

(2) The disclosure described in Subsection (1) shall be made:

(a) before the Utah user may access the features of the mental health chatbot;

(b) at the beginning of any interaction with the Utah user if the Utah user has not accessed the mental health chatbot within the previous seven days; and

(c) any time a Utah user asks or otherwise prompts the mental health chatbot about whether artificial intelligence is being used.

Policy Action Plan

Consumer Protections

Utah HB 452

Safety-focused AI Development

Utah HB 452

Professional Ethics

OAIP/DOPL Guidance letter

Action 2: AI Safety Standards

Goal: Maximize safety and efficacy

- We want to
  - Give responsible developers a clear advantage in the marketplace
  - Ensure adequate liability, especially for irresponsible actors
- Clarify in law what responsible development entails
  - Informed by clinical best practices
  - Safety established through rigorous testing
  - Continuous negative outcome monitoring
- Create limited legal protections responsible developers
  - Limited liability protections
  - Limited protections from DOPL enforcement
  - Balance positive incentives with consequences for irresponsibility

Mental-health chatbot development

58-60-118. Mental health chatbots -- Affirmative defense.

- AI supplier receives affirmative defense to liability by meeting transparency requirements and filing a written policy
- Policy should describe:
  - involvement of licensed professional in AI development and review
  - consistency of AI development with clinical best practices;
  - regular product testing to ensure that the output of the mental health chatbot poses no greater risk than a licensed mental health therapist
  - reporting mechanism for harmful interactions
  - assessment of risk of harm to user and risk mitigation plan
  - protocols to respond in real time to acute risk of physical harm;
  - regular, objective reviews of safety, accuracy, and efficacy
  - adequate user education
  - prioritization of user mental health and safety over engagement metrics or profit
  - prevention of discriminatory treatment
  - compliance with relevant privacy laws

Mental-health chatbot development

58-60-118. Mental health chatbots -- Affirmative defense.

- AI supplier receives affirmative defense to liability by meeting transparency requirements and filing a written policy
- Policy should describe:
  - involvement of licensed professional in AI development and review
  - consistency of AI development with clinical best practices;
  - regular product testing to ensure that the output of the mental health chatbot poses no greater risk than a licensed mental health therapist
  - reporting mechanism for harmful interactions
  - assessment of risk of harm to user and risk mitigation plan
  - protocols to respond in real time to acute risk of physical harm;
  - regular, objective reviews of safety, accuracy, and efficacy
  - adequate user education
  - prioritization of user mental health and safety over engagement metrics or profit
  - prevention of discriminatory treatment
  - compliance with relevant privacy laws

Mental-health chatbot development

58-60-118. Mental health chatbots -- Affirmative defense.

- AI supplier receives affirmative defense to liability by meeting transparency requirements and filing a written policy
- Policy should describe:
  - involvement of licensed professional in AI development and review
  - consistency of AI development with clinical best practices;
  - regular product testing to ensure that the output of the mental health chatbot poses no greater risk than a licensed mental health therapist
  - reporting mechanism for harmful interactions
  - assessment of risk of harm to user and risk mitigation plan
  - protocols to respond in real time to acute risk of physical harm;
  - regular, objective reviews of safety, accuracy, and efficacy
  - adequate user education
  - prioritization of user mental health and safety over engagement metrics or profit
  - prevention of discriminatory treatment
  - compliance with relevant privacy laws

Policy Action Plan

Consumer Protections

Utah HB 452

Safety-focused AI Development

Utah HB 452

Professional Ethics

OAIP/DOPL Guidance letter

Guidance for licensed professionals

Why is responsible use important?

- **Use of AI tools can lead to undesired outcomes**
  - Risk of patient data being exposed
  - Risk of incorrect AI outputs
  - Exposure of patients to Gen AI can cause tech-induced frustration or over-immersion
- **Clinician's responsible use can prevent and mitigate many undesired outcomes**
  - Assess risks posed by individual AI tools, review of AI outputs critically, avoid overconfidence / overreliance on AI tools, …

Encouraging Responsible Use

- **The State aims clarify that clinicians may choose to use AI, within certain boundaries:**
  - Must ensure data privacy, extending to intake
  - Informed consent still applies, including to case management and transcription
- **Clinician must observe existing professional standards and monitor client progress**
- **Clinician still bears ultimate responsibility for care**

General use of AI technologies

- Informed consent
- Disclosure
- Data privacy and data safety
- Mental health therapist competence with AI technology
- Patient safety and competence with AI technology
- Unique patient needs
- Continuous monitoring and reassessment

Therapists' Interactions with Gen AI

- Authorship transparency
- Review of AI-generated text
- Reasonable doubt about AI-generated diagnosis and treatment proposals

Patients' Interactions with Gen AI

- Patients' relationship with AI technology
- Primary commitment to patients
- Service interruptions and malfunctions
- Conduct of generative AI
- Responsibility in emergency situations
- Continuous monitoring of patient-AI interactions
- Mandated reporting of patient-AI interactions

Gen AI in Behavioral Health Ed

Supervisors <> Generative AI

- Review of AI-generated text
- Reasonable doubt about AI-generated student assessment

Trainees <> Generative AI

- Student's relationship with AI technology
- Primary commitment to students
- Service interruptions and malfunctions
- Continuous monitoring of student-AI interactions

Office of Artificial Intelligence Policy
UTAH DEPARTMENT OF COMMERCE

Division of Professional Licensing
UTAH DEPARTMENT OF COMMERCE

Guidance Letter

Best Practices for the Use of Artificial Intelligence by Mental Health Therapists

April 2025

Prepared for the Utah Office of Artificial Intelligence Policy and the Utah Division of Professional Licensing

Concluding remarks

How to engage in the AI policy space?

- Cautious and curious: Try new things while following best practices
- Consider asking clients if/how they are using AI for mental health purposes
- Reach out to my office with any comments: ai@utah.gov

Thank you.

Dr. Alice Schwarze
Head of Research, Office of AI Policy
aschwarze@utah.gov