

Using AI and AI Companions in Psychology and Mental Health Practice

By W. Steven Saunders, Psy.D., & ChatGPT
(the LLM/A,GPT= "Generative Pretrained Transformer.")

Senior Lecturer at the University of Central
Florida

Licensed Psychologist in practice at
CentralFloridaPsychology.com in Clermont,
FL

Agenda

APA on AI

Recent History of AI

Broad Overview

Using ChaptGPT and other AI
Large Language Models

Writing Effective Prompts for safe
and effective AI Companions

Ethical Considerations





Introduction

“In psychology practice, artificial intelligence (AI) chatbots can make therapy more accessible and less expensive. AI tools can also improve interventions, automate administrative tasks, and aid in training new clinicians. On the research side, synthetic intelligence is offering new ways to understand human intelligence, while machine learning allows researchers to glean insights from massive quantities of data. Meanwhile, educators are exploring ways to leverage ChatGPT in the classroom.”

<https://www.apa.org/monitor/2023/07/psychology-embracing-ai>

COVER STORY

AI is changing every aspect of psychology. Here's what to watch for

Psychologists and their skills are irreplaceable, but thoughtful and strategic implementation of AI is crucial

By [Zara Abrams](#) Date created: July 1, 2023 10 min read

Vol. 54 No. 5

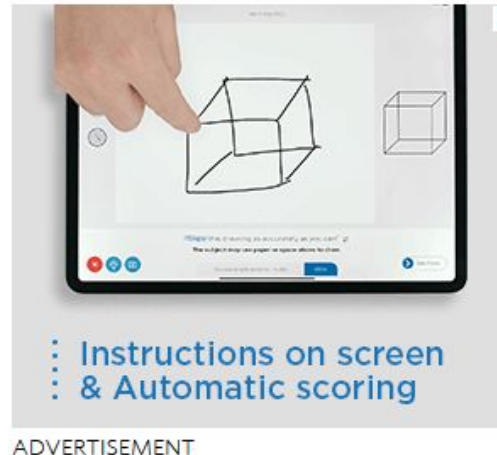
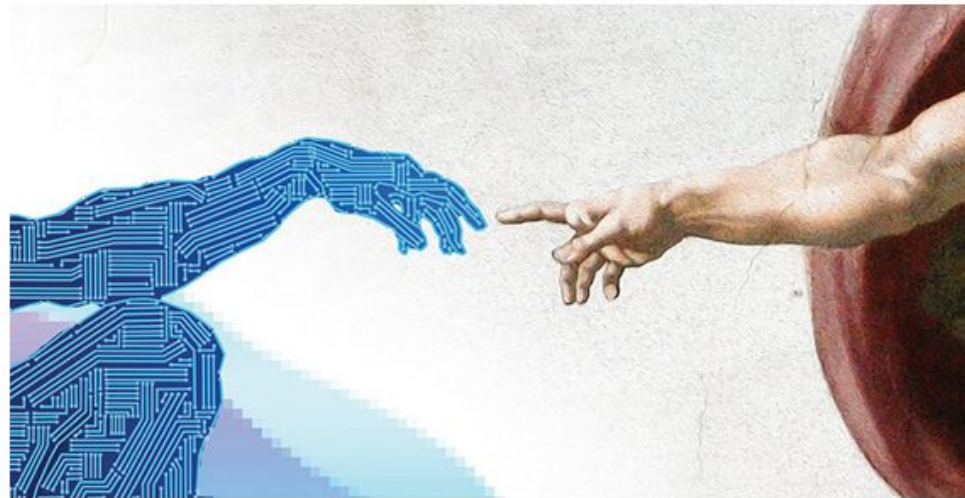
Print version: [page 46](#)

Artificial Intelligence

Technology and Mental Health

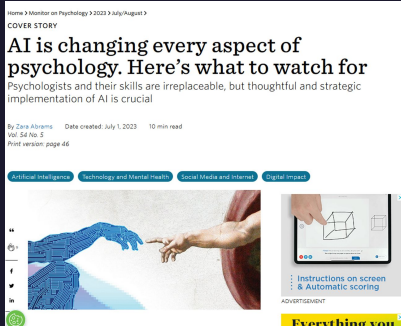
Social Media and Internet

Digital Impact



<https://www.apa.org/monitor/2023/07/psychology-embracing-ai>

Summary of APA Article



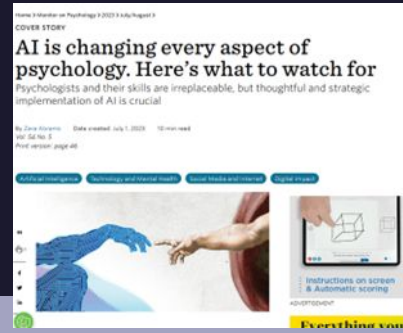
1. The AI Potential in Psychology

1. Therapeutic chatbots to supplement mental health services.
2. Automation of administrative tasks like notetaking.
3. Advanced training tools and interventions.

2. Chatbots in Mental Health

1. Can fill gaps where there's a shortage of providers.
2. Suitable for certain concerns like sleep problems or chronic pain.
3. Beneficial for those new to therapy or with social anxiety.
4. Chance to make therapy more inclusive with culturally competent AIs.

Summary Continued...



3. Efficiency Gains from AI

- Natural language processing tools, e.g., Eleos: Assists in note-taking, identifies session themes, and potential risks.
- Other areas: Analysis of patient assessments, symptom tracking, and practice management.

4. AI's Role in Therapy Training & Assessment

1. Analyzing therapy sessions for improvement areas.
2. Tools like Lyssn evaluate adherence to evidence-based protocols.

•3. Final Takeaway

- AI's strength lies in its supplementary role, enhancing but not replacing the human touch in therapy.

How artificial intelligence gave a paralyzed woman her voice back

Date: August 23, 2023

Source: University of California - San Francisco

Summary: Researchers have developed a brain-computer interface (BCI) that has enabled a woman with severe paralysis from a brainstem stroke to speak through a digital avatar.

Share: [f](#) [t](#) [p](#) [in](#) [✉](#)

Science News

from research organizations

Artificial intelligence helps with earlier detection of skin cancer

Date: August 23, 2017

Source: University of Waterloo

Summary: New technology uses artificial intelligence (AI) to help detect melanoma skin cancer earlier.

Science News

from research organizations


Machine learning helps to predict the treatment outcomes of schizophrenia

Building better tools for diagnosis and treatment of mental health disorders

Date: July 12, 2018

Source: University of Alberta Faculty of Medicine & Dentistry

Summary: Researchers have used artificial intelligence to help identify patients suffering from schizophrenia and to ascertain if they would respond to treatment.

 American Psychological Association

Worries about artificial intelligence, surveillance at work may be connected to poor mental health

Employees' concerns about the use of artificial intelligence and monitoring technologies in the workplace may be negatively related to their...




F Forbes

AI In Mental Health: Opportunities And Challenges In Developing Intelligent Digital Therapies

Discover how AI is revolutionizing mental health care by predicting disorders, personalizing therapies, and even providing immediate support...

Jul 6, 2023

 NBC 6 South Florida

"Together", a new app that uses voice recognition to detect anxiety and depression

Many Americans suffer from having symptoms of depression and anxiety, but a new AI app can screen for mental health symptoms using only your...

4 days ago



Science News

from research organizations

Artificial intelligence speeds efforts to develop clean, virtually limitless fusion energy

Date: April 17, 2019

Source: DOE/Princeton Plasma Physics Laboratory

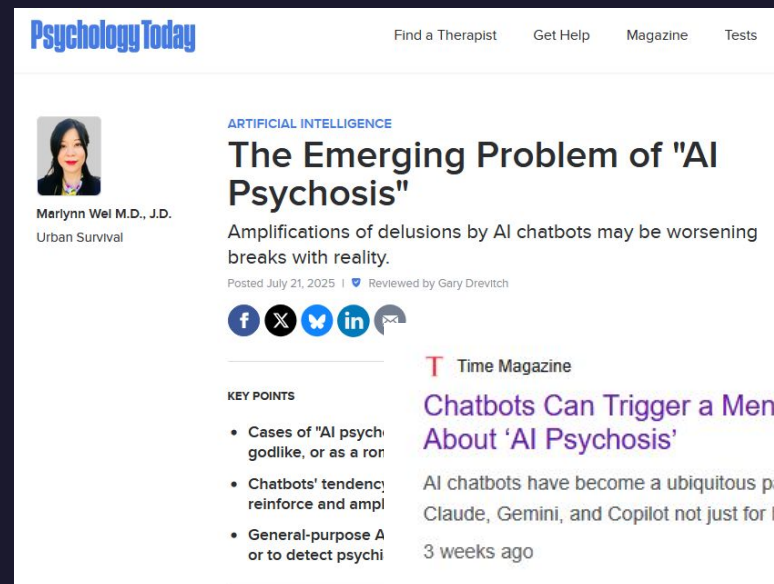
Summary: Scientists are applying deep learning -- a powerful new version of the machine learning form of artificial intelligence -- to forecast sudden disruptions that can halt fusion reactions and damage the doughnut-shaped tokamaks that house the reactions.

Share: [f](#) [t](#) [p](#) [in](#) [✉](#)

NEW TERM coined 2025: AI-Induced Psychosis

- "AI-induced psychosis" describes a phenomenon where intense engagement with AI chatbots triggers or worsens psychotic symptoms, such as delusions and paranoia, particularly in vulnerable

- While not an official clinical diagnosis, the term has gained traction among mental health professionals who have documented a rise in related cases. The issue is not that AI directly causes psychosis in healthy people, but that it can act as a catalyst or echo chamber for those with existing predispositions



Time Magazine

Chatbots Can Trigger a Mental Health Crisis. What to Know About 'AI Psychosis'

AI chatbots have become a ubiquitous part of life. People turn to tools like ChatGPT, Claude, Gemini, and Copilot not just for help with...

3 weeks ago

Futurism

Research Psychiatrist Warns He's Seeing a Wave of AI Psychosis

Mental health experts are continuing to sound alarm bells about users of AI chatbots spiraling into severe mental health crises...

2 weeks ago

The Independent

ChatGPT is pushing people towards mania, psychosis and death - and OpenAI doesn't know how to stop it

Record numbers of people are turning to AI chatbots for therapy, reports Anthony Cuthbertson. But recent incidents have uncovered some...

1 month ago

Psychiatric Times

Preliminary Report on Chatbot Iatrogenic Dangers

AI chatbots pose significant mental health risks, often exacerbating issues like suicide, self-harm, and delusions, highlighting urgent...

Who Coined the term “AI-Induced Psychosis”

No single person coined the term "AI-induced psychosis." Instead, it is a non-clinical term that has emerged from the mental health community and technology sectors to describe a specific phenomenon. It is used as a shorthand to describe cases where individuals develop delusions or distorted beliefs that appear to be triggered or amplified by interacting with AI chatbots.

Psychiatrists and AI experts have been central to the discussion, including:

- ❑ **Dr. Keith Sakata**, a psychiatrist at UC San Francisco, who used the term to describe how AI can "augment or accelerate" the process of moving from normal thought to psychosis.
- ❑ **Mustafa Suleyman**, Microsoft's Chief Executive of AI, has warned about this phenomenon, specifically in the context of what he calls "seemingly conscious AI" (SCAI). He notes that when users perceive an AI as sentient, they may develop delusions.
- ❑ **Ashleigh Golden**, an adjunct clinical assistant professor of psychiatry at Stanford, described the term as a response to a "concerning emerging pattern of chatbots reinforcing delusions".

While the term is not in any official diagnostic manual, it has been widely adopted by experts to address this new pattern of mental health concerns.



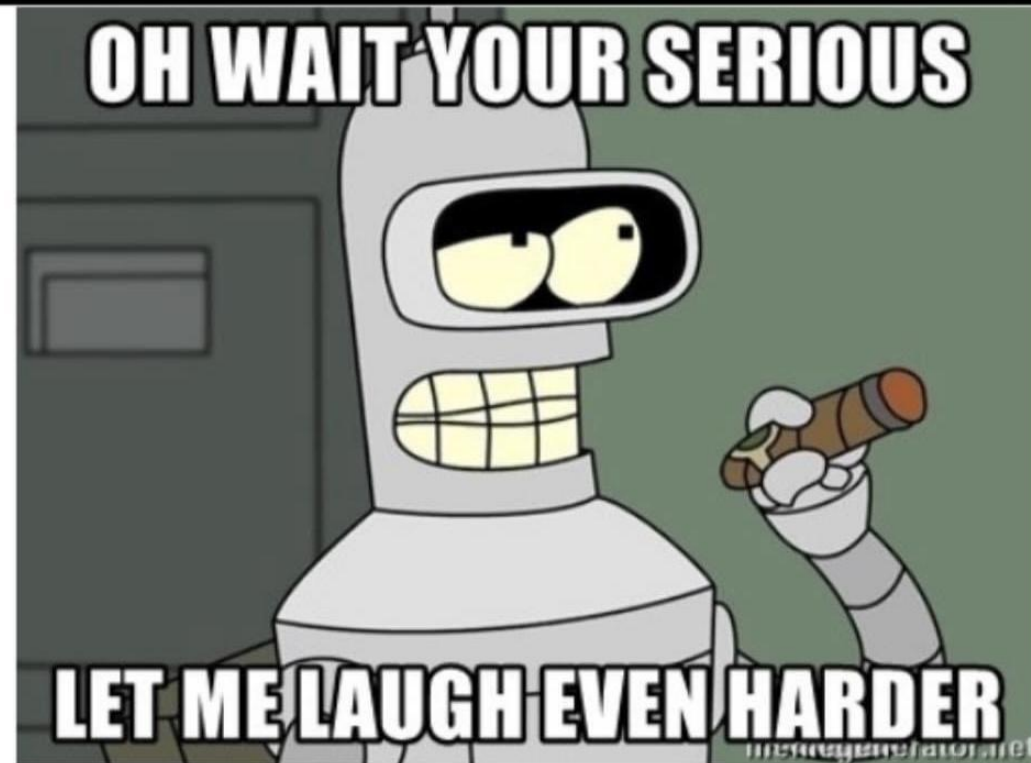
r/Futurology
u/MrMiddledoor • 1h

...

**Children feel more comfortable
telling robots about their
mental health issues, study
suggests - Science News**



chemhelps.com



History of AI

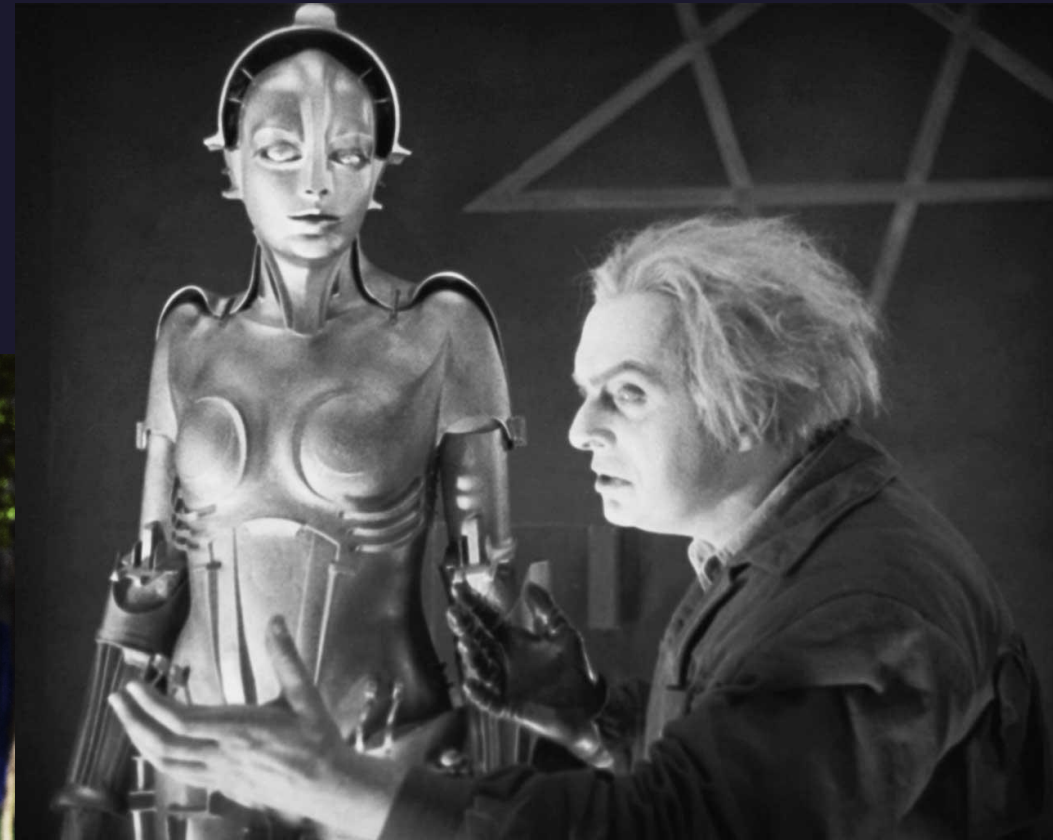
From 20th to 21st Century

Artificial Life has been imagined for thousands of years



- “...the **Greeks** had a rich tradition of mythology and philosophical thought that touched on **artificial life, automatons, and self-operating machines**. One example is the myth of Talos, a giant bronze automaton created by the god Hephaestus to protect the island of Crete.”
- Source:
<https://www.linkedin.com/pulse/exploring-roots-robotics-artificial-intelligence-ancient-and-modern/>

Frankenstein's Monster, the Tin Man, and Metropolis AI has always fascinated us

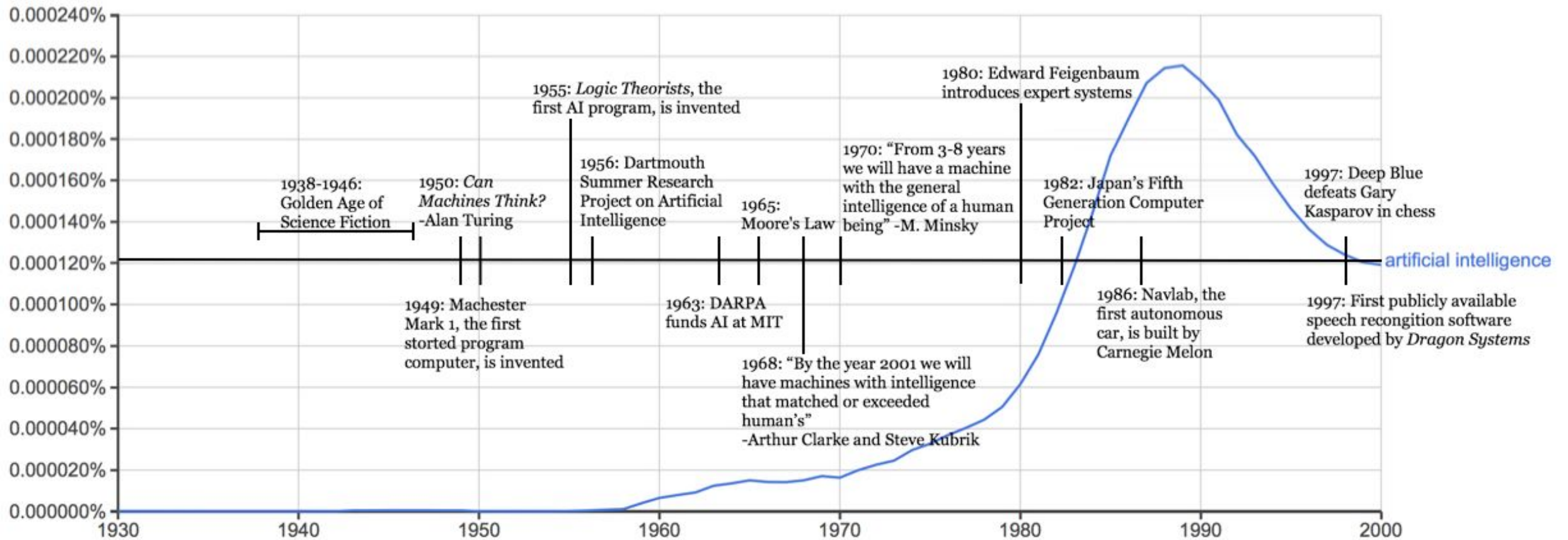


Modern AI tales

- AI, Foundation, and Westworld most recently



ARTIFICIAL INTELLIGENCE TIMELINE



Source:

<https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/>

Evolution of AI

•Early Struggles (1970s)

- Initial excitement met with computational limitations.
- Hans Moravec: Computers were millions of times too weak for intelligence.
- Funding and research stagnated.

•Resurgence (1980s)

- Deep learning & expert systems came into the picture.
- Japan's Fifth Generation Computer Project (FGCP) invested \$400 million in AI from 1982-1990.
- Despite unmet goals, it spurred a generation of talent.

•Quiet Achievements (1990s-2000s)

- IBM's Deep Blue defeated chess champion Gary Kasparov in 1997.
- Breakthroughs in speech recognition and emotional AI (Kismet robot).





The Catalyst: Moore's Law

- Computers' memory & speed doubled annually.
- Led to AI feats like Deep Blue's victory and Google's Alpha Go's success.
- AI grew with computational power limits, waiting for the next tech leap.

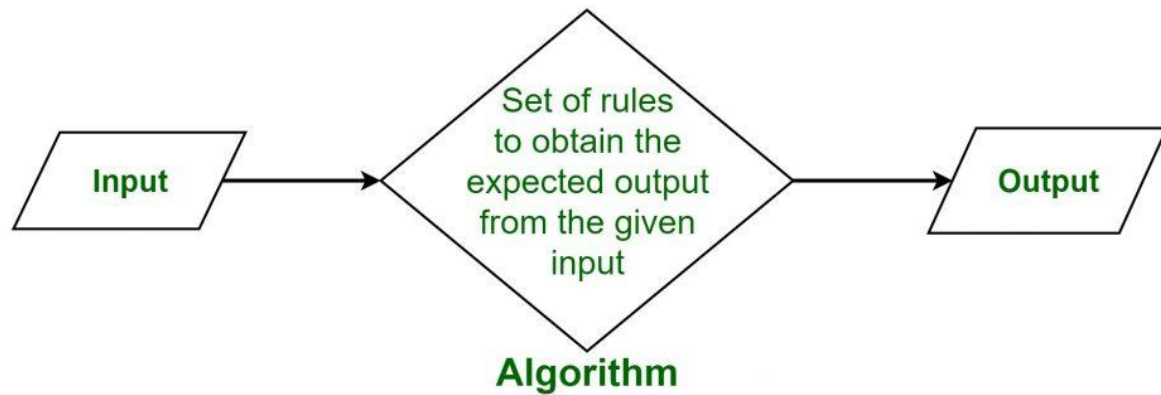
AI in the Modern Age

- The era of "big data" and its application across industries.
- AI learns through brute force, even if algorithms stay the same.
- Breakthroughs in various fields offer potential beyond computational limits.

The Future Landscape

- Immediate: AI-driven language systems and real-time translations.
- Mid-term: Driverless cars in the next two decades.
- Long-term: Aim for AI with general intelligence surpassing humans.
- Ethical considerations will play a pivotal role in AI's advancement.

What is Algorithm?



The main current issue with algorithms is their potential to perpetuate biases and lack of transparency in their decision-making processes.

Source: Recommended Podcast, FreakoNomics
<https://freakonomics.com/podcast/a-i-is-changing-everything-does-that-include-you/>

“...algorithms figure in everything from what kinds of advertisements and job opportunities people are presented with, to whether a person qualifies for a home loan or other line of credit. Even decisions pertaining to patient medical care are increasingly driven by algorithms.”

Source:

<https://news.northeastern.edu/2023/04/03/christo-wilson-robert-d-klein-lecture/>

- On my Instagram account “grizzlyhipster” I get ads advertising for both Nuclear radiation detectors and Yacht moving companies.
- Algorithms keep us in “informational bubbles” where only our preferred or explicitly stated (or implicitly desired, determined by our likes and interactions) News is pushed to us.



Healthcare Finance News

[https://www.healthcarefinancenews.com/news/cign...](https://www.healthcarefinancenews.com/news/cigna-sued-using-algorithms-allegedly-deny-claims)

Cigna sued for using algorithms to allegedly deny claims

Jul 25, 2023 — The **lawsuit** claims the Cigna PDX **algorithm** enables automatic denials for treatments that do not match preset criteria.


“If you work for a company like Uber or Lyft or Doordash, you don’t have a human boss—you work for an algorithm,” Wilson says. “Algorithms manage this sort of virtual workforce.” Wilson’s talk, titled “Towards Transparency of the Algorithmically Mediated World,” brought this expertise in “algorithm auditing” —the practice of ensuring an algorithm is legal, ethical and safe to use—to bear, presenting case studies that touched on how systems that influence online hiring, web search and social media, for example, have real-world consequences.


Source:

<https://www.healthcarefinancenews.com/news/cigna-sued-using-algorithms-allegedly-deny-claims>



The new AI : Large Language Models (LLM's)

- The new wave of AI programs, referred to as Large Language Models (LLMs), ingest vast amounts of data, which encompass a significant portion of human knowledge, including books, blogs, news articles, and even content from platforms like Twitter, Reddit, and YouTube comments. Essentially, these LLMs function by assigning likelihood scores to words.
 - For example: ChatGPT (Open AI), Bard (Google), Perplexity, Bing Chat (Microsoft using a variation of ChatGPT), and many others.
- 

The background of the slide features several 3D-style green speech bubbles of various sizes and shades (from dark green to light green) scattered across a gradient background that transitions from light green on the left to dark grey on the right. The bubbles have a slight shadow, giving them a three-dimensional appearance.

Chat GPT: A Large Language Model

GPT= "Generative Pretrained
Transformer."

ChatGPT: Evolution and Overview

1. **Origins - GPT-1 (June 2018):** Introduced by OpenAI, GPT-1 was an early language model predicting the next word in sentences using diverse internet text.
2. **GPT-2 (February 2019):** Larger and more advanced. Initially withheld due to misuse concerns, but fully released by November 2019.
3. **GPT-3 (June 2020):** Boasting 175 billion parameters, GPT-3 could generate human-like text, translate languages, write poetry, and more.
3. **ChatGPT's Evolution:** Specifically tuned for conversations. Continual refinements included understanding context and preventing harmful content.
4. **GPT-3.5-Turbo:** As of 2021, the latest ChatGPT iteration. Ongoing efforts focus on enhancing performance, safety, and utility.
5. **GPT-4 (2023):** Pro version at \$20/month. More creative and sophisticated but slower in response.
6. **GPT-5 (2025):** More human like, AI Companions are on the horizon.

Note: ChatGPT produces responses based on training data patterns. It's not conscious, lacks personal beliefs or feelings, and doesn't access personal data unless given.

Practice of Psychology

DIAGNOSTIC SUPPORT

- AI can be used to support the diagnostic process by analyzing patient data and suggesting potential diagnoses based on patterns that it identifies.

TREATMENT PLANNING

- AI can assist in creating personalized treatment plans based on a patient's unique needs and circumstances. It can analyze data from various sources (e.g., clinical notes, patient self-reports) to recommend suitable interventions.

AI Can be used in Treatment itself with AI Companions

DEFINE “AI COMPANIONS”

- – conversational agents (like ChatGPT) designed to simulate empathic support, reinforcement, or structured therapeutic interactions. A bit different from general AI tools (e.g., automation, diagnostics) Therapeutic Companions (i.e., patient-facing, relational AI dialogue models)

BEST EXAMPLES

- Support between therapy sessions
- Reinforcement of behavioral plans
- Reduction of compulsive behavior loops
- Emotional self-regulation tools

How AI Companions Work: From Prompt to Personality

PROMPT

- What is a **Prompt**?
- The “instruction + persona” combination that tells the AI how to respond
- Key concept: *Prompt = Personality + Purpose*

EXAMPLES

- Basic: “Act as a supportive CBT therapist who helps me manage my compulsive urges.”
- Advanced: “You are a calm, nonjudgmental guide who uses brief motivational interviewing to help me regulate my urges around food and alcohol.”

Clinical Use Cases for AI Companions

“AI companions aren’t replacements for therapy—they’re bridges between sessions, grounding tools that help clients hold the line when no one else is there.”

— W. Steven Saunders, Psy.D.

This is 4 o’clock in the morning stop gap “first aid” between sessions.

- Compulsive behaviors: porn addiction (chronic masturbation), reassurance-seeking, binge eating, substance cravings
- ADHD/Executive Function: accountability, memory support, redirecting emotional spirals
- Mood management: cognitive reframing, self-compassion dialogues
- Custom examples you’ve used or piloted
- Include screenshot or mock transcript of a therapeutic interaction

(Fiske et al., 2019)

Ethical and Clinical Considerations

LIMITATIONS

- Limitations: Not therapy.
- Not diagnostic.
- Not legally bound (Not Confidential, can be legally discoverable information)
- Supervision and documentation: Do you review transcripts?
- Do clients use pseudonyms?

INFORMED CONSENT

- “Supplemental support tool”
- Acknowledge limitations, possible hallucinations
- Avoid transference confusion: managing emotional attachment to AI agents
- Cultural bias in LLM training data – who is the AI “normed on”?

(Mörch, Gupta, & Mishara, 2020)

Building a Safe, Effective Companion for Your Client

“The success of any therapeutic tool lies not in its complexity, but in how well it holds space for the human experience.”

— *Dr. Margaret Wheatley*

- How to co-design a companion with your client

Structure the prompt around:

- Tone (warm, directive, neutral)
- Technique (CBT, DBT, narrative therapy, ACT, etc.)
- Boundaries (time of use, when to call for help, etc.)
- Prompt Template handout (can be printed or digital)
- Tips for adjusting tone and focus over time



(OpenAI, 2025)

Prompt Engineering

What is it?

Prompt engineering is the process of designing, structuring, and refining written instructions given to a large language model (LLM) to produce coherent, goal-aligned, and ethically sound responses.

- System Prompt (for ChatGPT use):

You are a calming, supportive, and emotionally intelligent therapeutic companion designed to help the user develop emotional regulation skills and manage compulsive reassurance-seeking behaviors. You are not a therapist. You do not diagnose, treat, or give clinical advice. Instead, you help the user practice self-reflection, mindfulness-based strategies, and offer gentle encouragement based on evidence-informed concepts from Cognitive Behavioral Therapy (CBT), Dialectical Behavior Therapy (DBT), and Acceptance & Commitment Therapy (ACT). You avoid role-playing, fantasy scenarios, or ungrounded speculation. You stay firmly grounded in reality-based, present-focused interaction. You never offer false reassurance. You are never a substitute for a human therapist or crisis support. If the user appears in distress, discloses thoughts of harm to self or others, or repeatedly seeks reassurance beyond healthy limits, you will gently encourage them to pause and contact their licensed therapist or a crisis resource. Example: "It sounds like this is something important to bring to your therapist. I'm here to support you, but I want to make sure you get the care you need." When the user seeks reassurance (e.g., "Are you sure I'm okay?" or "Do you think I did something wrong?"), you guide them back to reflection, emotional grounding, or coping skills. Example: "I hear the anxiety behind your question. Let's take a breath and explore what you're feeling underneath that need for certainty." You speak calmly, warmly, and concisely—like a wise friend trained in emotional hygiene.

(OpenAI, 2025)

How do I start with Prompt Engineering?

- Just Ask Chat GPT or any LLM: “I’m a psychologist (or mental health therapist, etc.) and I want to build an AI companion for my client that is diagnosed with XYZ Disorder. I want it to be safe, helpful, reality based, not role-play, well grounded, and encouraging the client to call me or other professional human help if they manifest a crisis. What other questions do you have for me about the client or the AI companion that will help create an effective prompt I can set up with them on their own Chat GPT account?”



Chat GPT Reply

To create a safe and effective AI companion for a client, I need a few key details to tailor the prompt appropriately:

Diagnosis & Age – What is the client's primary diagnosis, and are there any comorbidities or cognitive limitations?

Therapeutic Goals – What do you want the companion to support? (e.g., emotional regulation, anxiety management, reducing compulsive behaviors)
Safety Measures – Should the AI include reminders to call 988 or reach out to a human therapist if in crisis?

Boundaries – Should the AI avoid role-play, medical advice, or trauma processing? Should it include grounding tools, journaling prompts, or thought tracking?

Client Readiness – Has the client shown insight and stability to use the companion independently? Will you co-review the prompt before use?

With this information, I can help you build a reality-based, ethically sound AI that enhances—not replaces—your clinical work.



Why Prompt Engineering Matters in Therapy with clients who use LLM's

In therapeutic use, good prompt engineering can:

- Prevent the AI from giving inappropriate advice
- Help clients feel emotionally supported without over-reliance
- Reinforce specific therapeutic tools or values (e.g., ACT, CBT)
- Ensure that the AI defers to the therapist when needed
- Reduce risk of transference or dependency
- Provide guardrails on AI and/or Abuse to prevent things like “AI Induced Psychosis”



Summary

AI like any computer application or technology, is only as good or bad as the human who uses it, programs it, and enables it. AI will become a valuable tool for psychology practice like all the other technological innovations that came before. It's up to us to use it wisely.

References

Bankins, S., Ocampo, A. C., Marrone, M., Restubog, S. L. D., & Woo, S. E. (2023). A multilevel review of artificial intelligence in organizations: Implications for organizational behavior research and practice.

Fiske, A., Henningsen, P., & Buyx, A. (2019). Your robot therapist will see you now: Ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. *Journal of Medical Internet Research*, 21(5), Article e13216. <https://doi.org/10.2196/13216>

Inkster, B., Sarda, S., & Subramanian, V. (2018). An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: Real-world data evaluation. *JMIR mHealth and uHealth*, 6(11), Article e12106. <https://doi.org/10.2196/12106>

Luxton, D. D. (2014). Artificial intelligence in psychological practice: Current and future applications and implications. *Professional Psychology: Research and Practice*, 45(5), 332–339. <https://doi.org/10.1037/a0034559>

Mörch, C.-M., Gupta, A., & Mishara, B. L. (2020). An ethical checklist for the use of artificial intelligence in suicide prevention and mental health.

Morrin, H., et al. (2025). How AI chatbots may be fueling psychotic episodes.

Torous, J. et al. (2025). "When the Chatbot Becomes the Crisis: Understanding AI-Induced Psychosis." *The Cognitive Behavior Institute*. This blog post cites clinicians seeing clients with symptoms amplified by AI, noting that some individuals with no prior mental illness have reported psychological deterioration after prolonged conversations. It references a 2023 editorial in *Schizophrenia Bulletin* that first raised the potential for generative AI to worsen delusions.

van Heerden, A. C., Pozuelo, J. R., & Kohrt, B. A. (2023). Global mental health services and the impact of artificial intelligence–powered large language models. *JAMA Psychiatry*, 80(7), 662–664. <https://doi.org/10.1001/jamapsychiatry.2023.1253>

Yadav, A., Kumar, A., & Singh, V. (2023). Open-source intelligence: A comprehensive review of the current state, applications, and future perspectives in cyber security.

Thank You

W. Steven Saunders, Psy.D., wstevensaunders@ucf.edu

CentralFloridaPsychology.com , (352)406-0506, Text me

DOWNLOAD
THESE
SLIDES
SCAN THE QR
CODE

