



BIRMINGHAM AI

September 2025 Meeting

sponsored by





BIRMINGHAM AI

**Our mission is to accelerate the
adoption of artificial intelligence
in Birmingham, for all.**



LOOKING TO GROW YOUR TECH STARTUP?

BOOST IS FOR STARTUPS WITH AN MVP
WORKING TO LAND THEIR FIRST CUSTOMER.

apply now!



applications now open

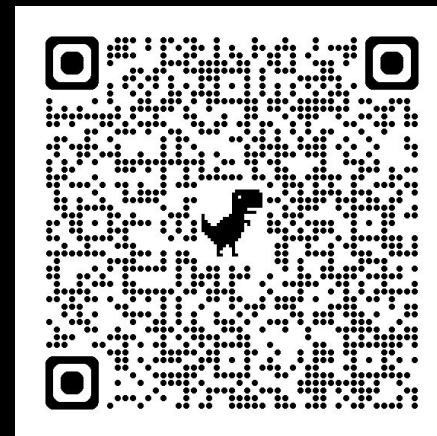
TECHS AND THE CITY



BHM

Real Talk - Is Birmingham ready for AI?
Exploring the Talent Gap (or Opportunity) in the Age of AI

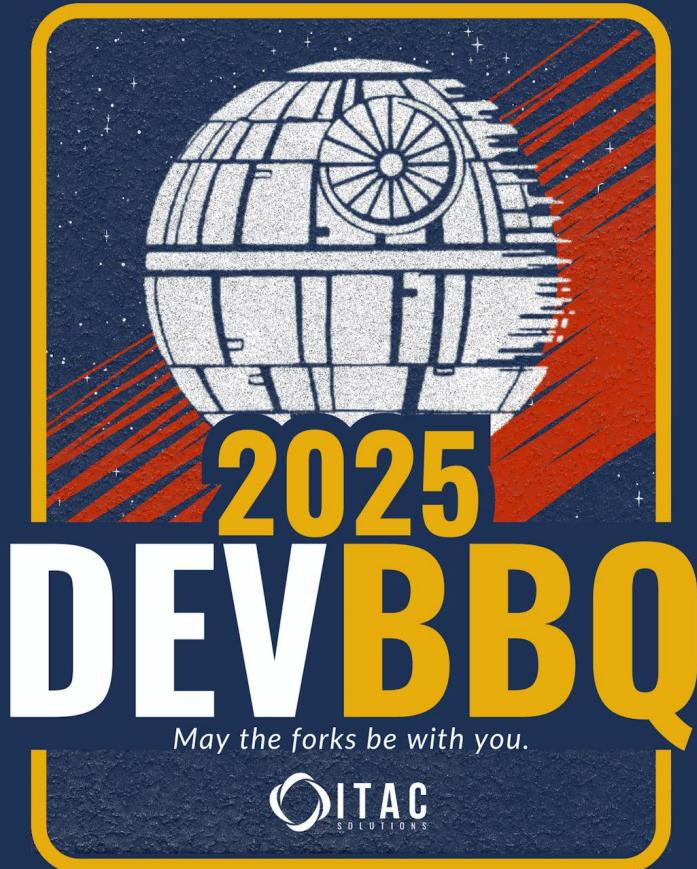
Oct 21, 2025 | 5:30-8:00pm
WorkPlay - 500 23rd St South



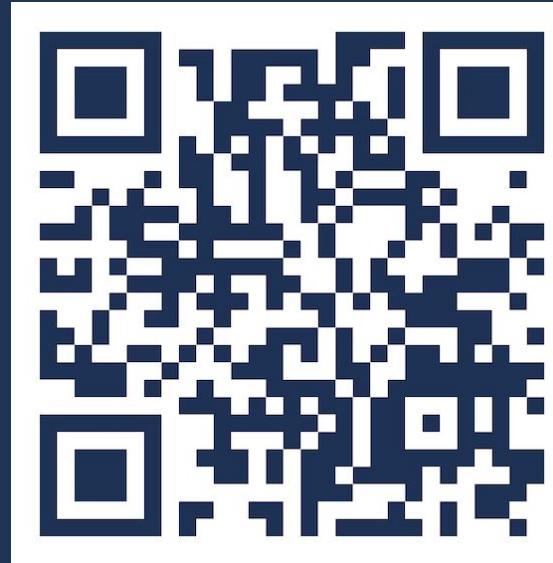
Sponsored By:
sparkhound

Hosted By:





TONIGHT
Wednesday September 10th
5-8pm Vestavia Hills



*BBQ, Beer, &
Good Vibes*



HR • Pay • Time • Culture



Chas Guy

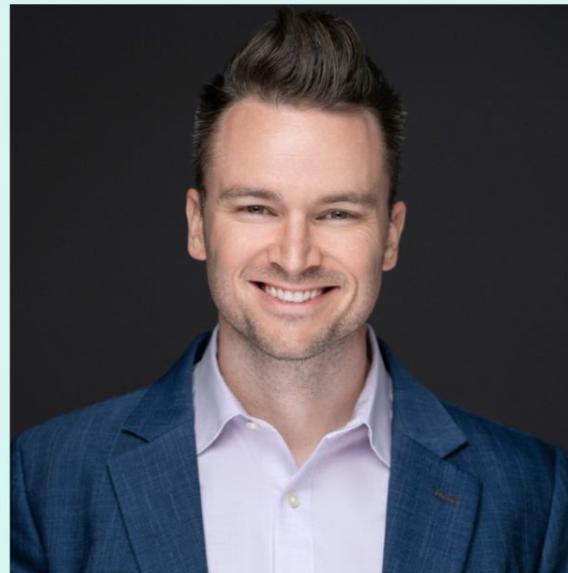
Sales Executive, Alabama

BIRMINGHAM AL





HR • Pay • Time • Culture



Chas Guy

Sales Executive, Alabama

Sales Guy

BIRMINGHAM AL





HR • Pay • Time • Culture

BIRMINGHAM AL



Chas Guy

Sales Executive, Alabama

Sales Guy

ChatGPT "More Handsome" Guy

The Power of Data



Our Scale is Unparalleled

WORK

- 30** B+ Frontline interactions
- 12** B+ Schedules created
- 10** B+ Punches Recorded

PEOPLE

- 750** M+ Candidate applications
- 70** M+ Hires
- 10** M+ Performance reviews

CULTURE

- 20** B+ Survey participants/year
- 170** B+ Employer countries

Connected Agents make the workforce more productive and deliver curated experiences

»»» Configurable

AI Agents can be configured as digital twins to help people by completing day-to-day tasks on their behalf, with accuracy and speed

»»» Self-learning

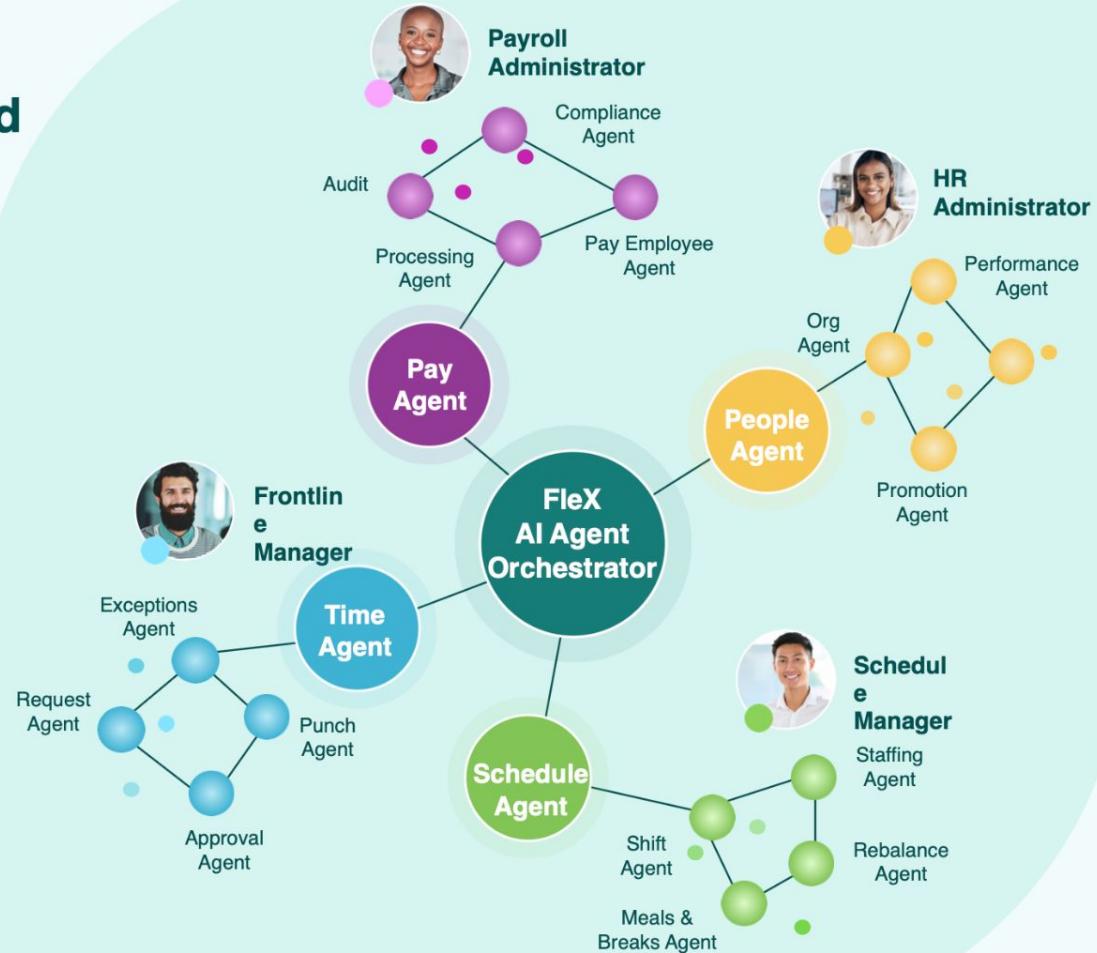
AI Agents understand the context in which they are operating, make decisions based on data, and learn over time

»»» Collaborative

Connected AI Agents can collaborate with people, systems or other AI Agents in real time as they act to achieve their goal

»»» Autonomous

Working together, Connected AI Agents can perform actions autonomously to execute a complex task



Thank You!



Our purpose is people

BIRMINGHAM AL

Updates



Gemini



BIRMINGHAM AI

Talk Track Nano Banana

- Image editing and generation is changing. The old standard has been surpassed... ask audience to raise hand who has used photoshop? AI image generation tool?
- Get into the problems with image generation. Slow, not realistic, tough to keep consistent character, or editing a single part of an image.
- New Google launched Nano Banana built on a diffusion model, it outperforms DallE & Midjourney on speed and memory. It's great at editing, project images and building off of an existing reference or set of reference images.





Image Generation... is bananas

Gemini
2.5 Flash ▾

NEW! Try image editing with our best image model, Nano Banana

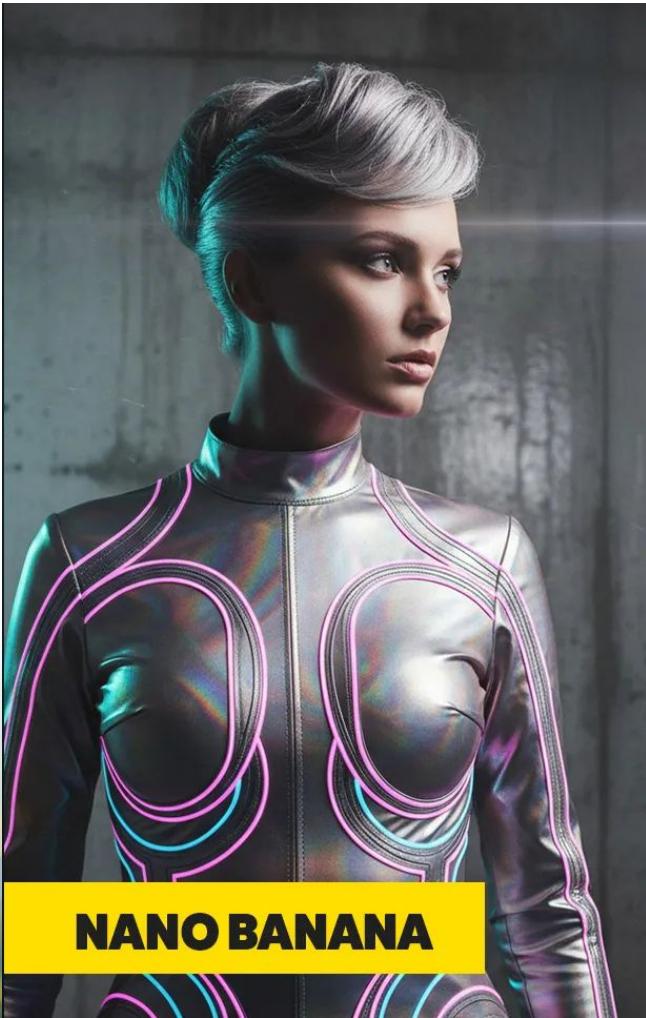


Describe your image

+ ° 8 Image X

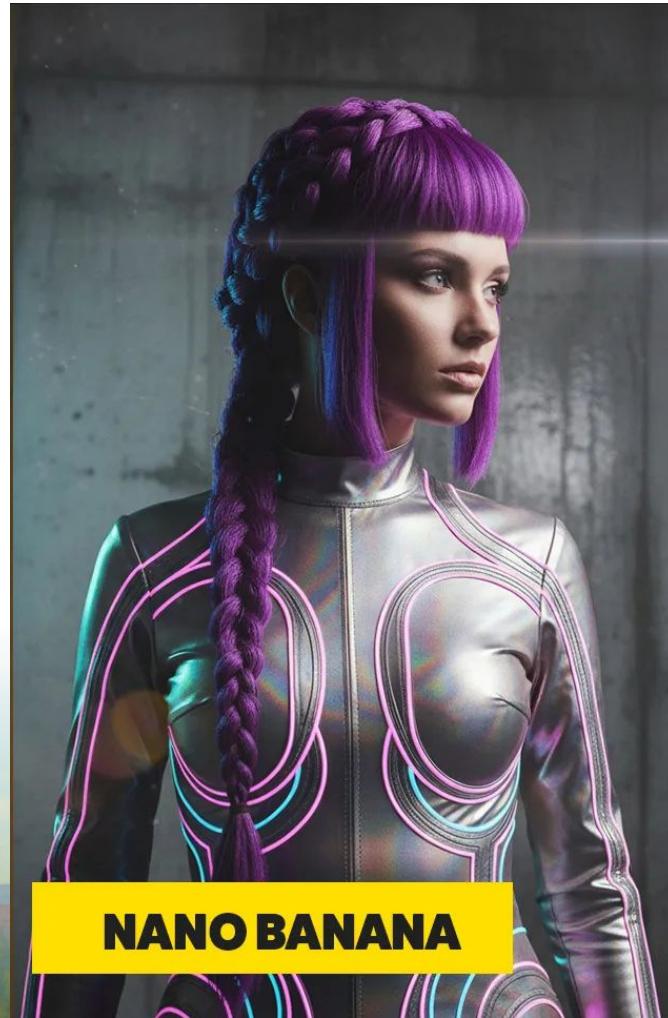


BIRMINGHAM AL



NANO BANANA

BIRMINGHAM AI



NANO BANANA

AI Agents (30 min)



AI Agen



BIRMINGHAM AL



BIRMINGHAM AI



AI Agents (30 min)

AI Ag

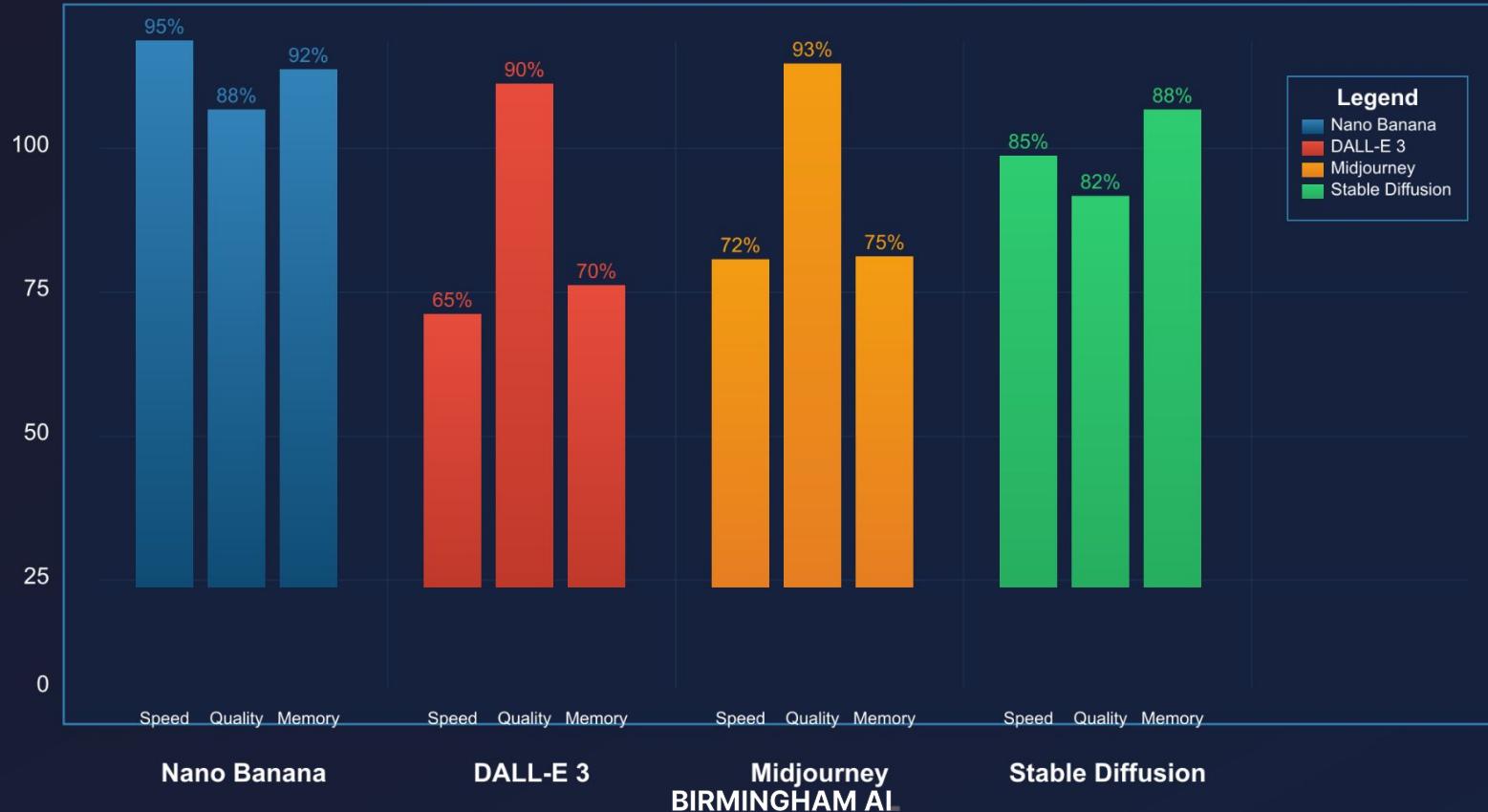
BIRMINGHAM AI



BIRMINGHAM AL



Performance Metrics Comparison



Overly Bananas 🍌 ?



malikona 7d ago · Edited 6d ago

This is why people underestimate AI's capabilities in every sector, not just design.

EXPERTS: Ignore this at your peril.



npquanh30402 8d ago

I think you're judging this tool by the wrong standard. The hype isn't coming from graphic designers thinking it will replace Photoshop; it's coming from the 99% of people who have never used Photoshop and now have a powerful tool at their fingertips.

For that huge audience, the minor artifacts and texture issues you mentioned are a total non-issue. They don't need a final edit, professional-grade output; they just want to be able to make quick, easy changes with a few words. The one valid point you raised is the 50% failure rate, which is a huge bug that needs to be addressed.

But the hype is justified not because the tool is perfect, but because it's a game changer for a massive, underserved market. It doesn't have to destroy Adobe to be revolutionary.



BIRMINGHAM AI



NVIDIA



Genie 3

A new frontier for world models

BIRMINGHAM AL



WHY

“Imagine if we could safely train AI to operate in the real world — without risk, without cost, and without limits. That’s what Google’s new Genie 3 represents. It gives AI a playground the size of the universe, where it can practice endlessly before stepping into our world.”

WHAT

“Genie 3 is what’s called a *world model*. Instead of generating static text or images, it creates full 3D, interactive environments from a simple text prompt. If I type in ‘*a busy kitchen*’ or ‘*a construction site*,’ Genie 3 instantly produces a virtual space where AI can interact, experiment, and learn.”

HOW

“Here’s how it works:

- You start with a text description.
- Genie 3 transforms it into a 3D, interactive simulation.
- AI agents can then move through that space, pick things up, solve problems, and learn through trial and error — just like humans do.

And because it’s virtual, we can scale this to millions of training scenarios, far beyond what’s possible in the real world.”

WHAT IF

“So what does this mean for us? What if every robot could practice in simulation before ever touching the real world? What if industries could test safety procedures, logistics, or even medical training in limitless virtual environments? What if cities like Birmingham could use tools like Genie 3 to both upskill workers and safely innovate with AI?

Genie 3 isn’t just another model — it’s a step toward AI that learns by *doing*. And that opens the door to



OpenAI

BIRMINGHAM AL



Fidji Simo ✅ @fidjissimo · Jul 21

...

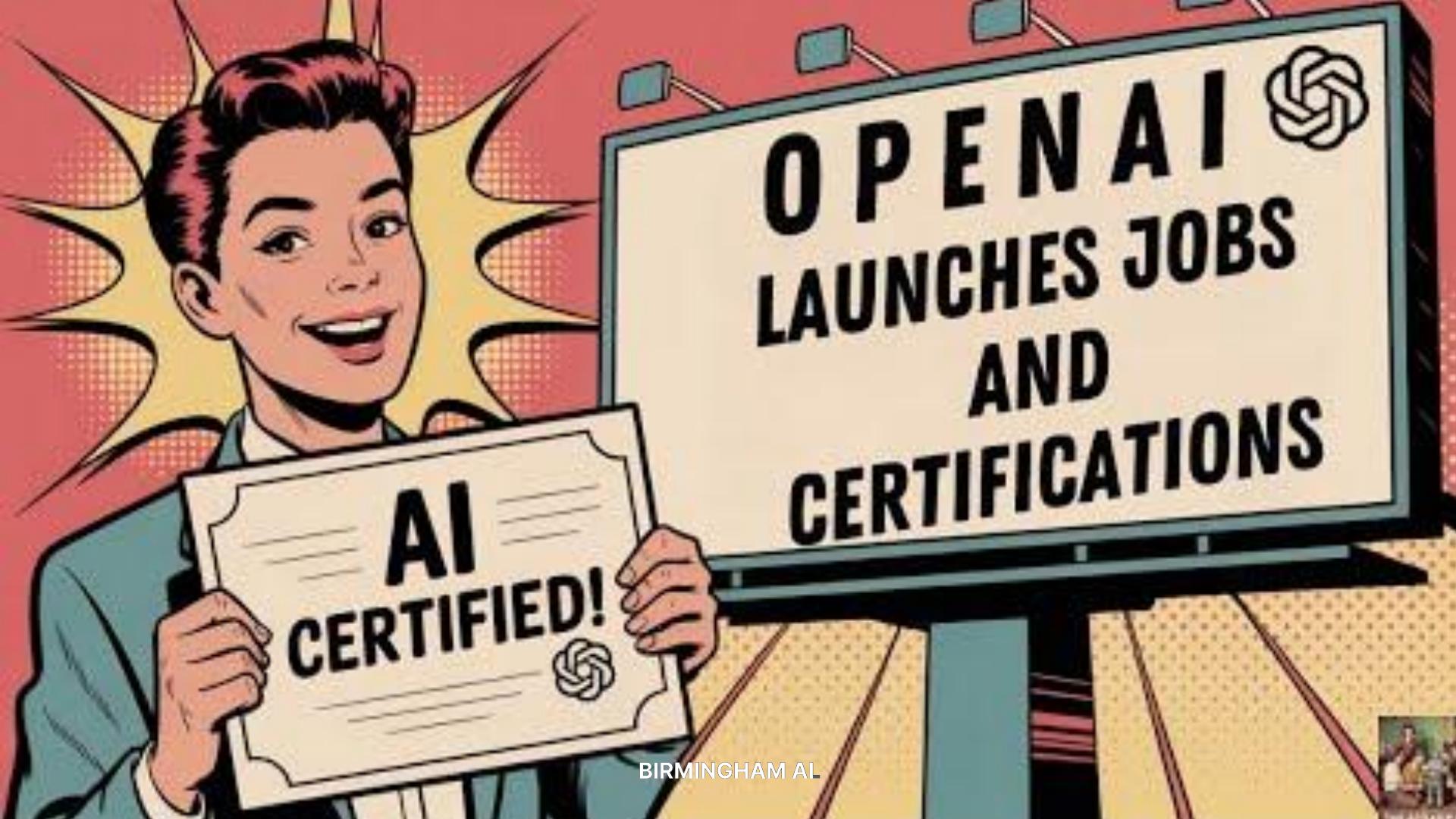
I will officially start at OpenAI as CEO of Applications on August 18. I am sharing this essay on why I believe AI can be the greatest source of empowerment for all. 

AI as the greatest source of empowerment for all

Fidji Simo

AI as the greatest source of empowerment for all

BIRMINGHAM AL



OPENAI 

**LAUNCHES JOBS
AND
CERTIFICATIONS**

BIRMINGHAM AL





Ethan Mollick

@emollick

...

We have data on the environmental impact per AI prompt:

Gemini: 0.00024 kWh & 0.26 mL water

ChatGPT: 0.0003 kWh & 0.38 mL

...the same energy as one Google search in 2008 & 6 drops of water.

Seems to be improving, too: Google reports a 33x drop in energy use per prompt in a year.

ENVIRONMENTAL IMPACT OF DELIVERING AI AT GOOGLE SCALE

Huang, David Patterson, Ian Schneider, Robert Sedivy, Savannah Gthy Ranganathan, Jeff Dean, Amin Vahdat, Ben Gomes, and James

Google, Mountain View, CA, USA

ABSTRACT

ver of AI is undeniable—but as user adoption accelerates, so does the environmental impact of AI serving. However, no studies have measured AI serving. This paper addresses this gap by proposing and executing a comprehensive usage, carbon emissions, and water consumption of AI inference work. Our approach accounts for the full stack of AI serving infrastructure system energy, idle machine capacity, and data center energy overhead's AI infrastructure for serving the Gemini AI assistant, we find the 0.24 Wh of energy—a figure substantially lower than many public estimates. Clean energy procurement have driven a 33x reduction in carbon footprint for the median Gemini Apps text prompt over one year. A text prompt uses less energy than watching nine seconds of television (0.26 drops of water (0.26 mL). While these impacts are low compared to the environmental impact of AI serving continues to warrant important attention. Toward sensitive measurement of AI serving environmental metrics is critical for incentivizing efficiency gains across the full AI serving stack.

As datacenter production gets automated, the cost of intelligence should eventually converge to near the cost of electricity. (People are often curious about how much energy a ChatGPT query uses; the average query uses about 0.34 watt-hours, about what an oven would use in a little over one second, or a high-efficiency lightbulb would use in a couple of minutes. It also uses about 0.000085 gallons of water; roughly one fifteenth of a teaspoon.)

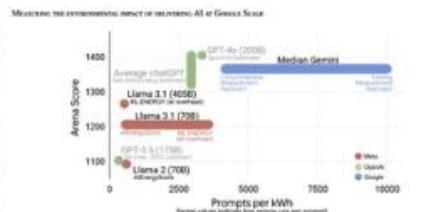
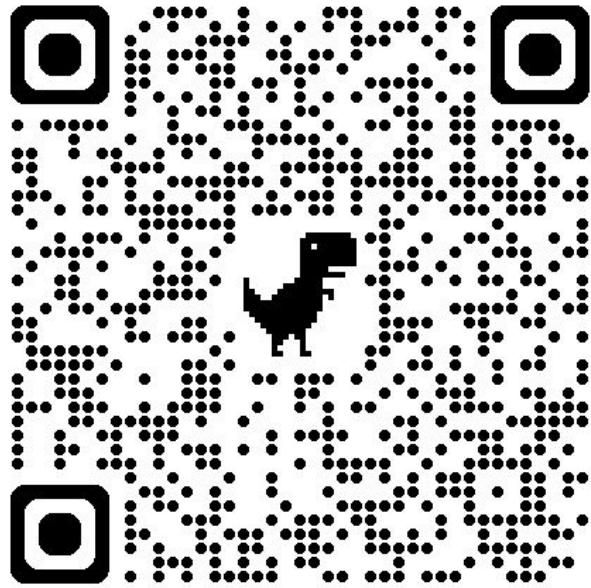


Figure 2: Energy per prompt results for large production AI models plotted against LM-Arena score. (● = estimated metrics (gray text) and measured metrics (red, blue, and blue text). For similar models and parameter counts, the different measurement approaches can lead to large differences in results, e.g. Llama 3.1 (20B) varies from ~500 to ~3000 prompts per kWh. The results for the median Gemini Apps text prompt presented in this paper vary from a more narrowly defined Existing Approach (10,000 prompts per kWh) to the more complete Comprehensive Approach (14,167 prompts per kWh).)



**Sign up
for October
Meetup**

The emotional rollercoaster of hiring an AI

(how to not waste your time exploring AI)

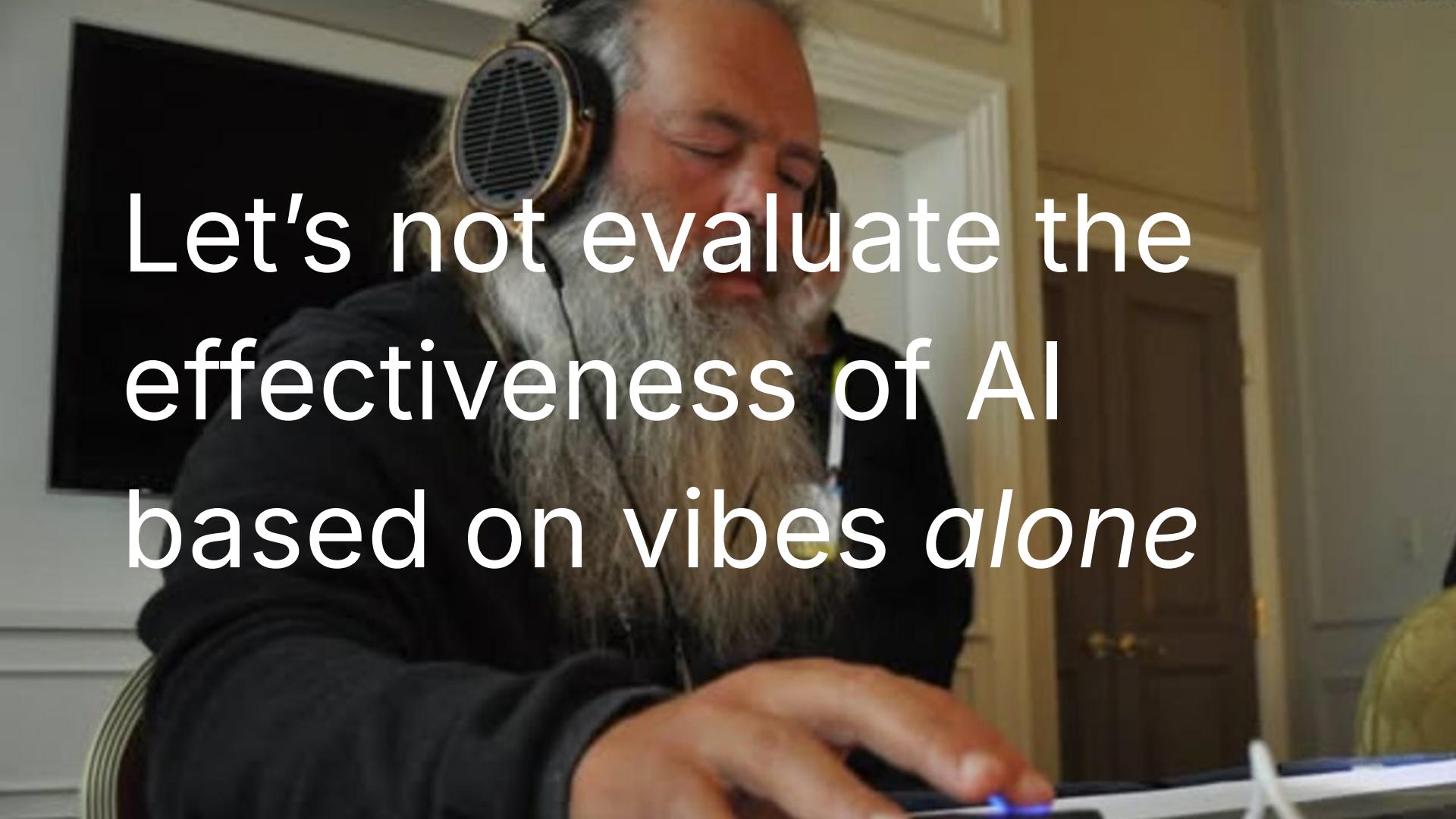
(how to develop trust in your AI tools and vendors)



BIRMINGHAM AI

“I have 50 people just playing around with AI. AI can't do the things we want it to do, and I feel like we're wasting our time.”

– *Birmingham CEO*

A photograph of a man with a long, grey beard and mustache. He is wearing over-ear headphones and is looking down at a black electronic keyboard he is playing. His eyes are closed, suggesting he is deeply focused or enjoying the music. The background shows a room with white walls and a dark door.

Let's not evaluate the
effectiveness of AI
based on vibes *alone*

~~Evaluations are unit tests for LLMs, where we benchmark prompts against a fixed eval set to catch regressions and compare model quality over time~~

Evaluations are structured tests
that assess the performance of AI.

#JudgeJudy

Who cares?!



Lifetime

Evaluations are structured tests that check how well an AI does its job.



MAKING PROGRESS. YES?

How to Explore AI without Wasting Time

Sanity Check

Is it *possible* AI can do X?

Reliability Gate

Can we *trust* AI do X?

DECISION

Onboarding

How do we *hire* AI to do X?

Management

How do we *manage* AI doing X?

Sanity Check: Is it possible AI can do X?

What Can You Do Today to Get Started?

Research / Case Studies

Talk with Peers

Talk with Vendors

Go Find Out!

Voice AI in Firms: A Natural Field Experiment
Automated Job Interviews

89 Pages • Posted: 18 Aug 2025

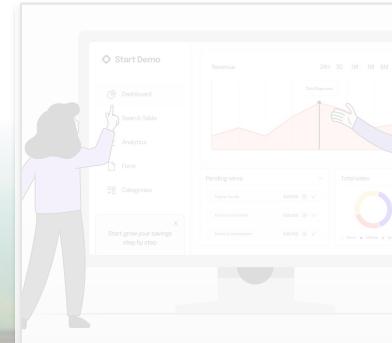
Brian Jarabian
University of Chicago - Booth School of Business

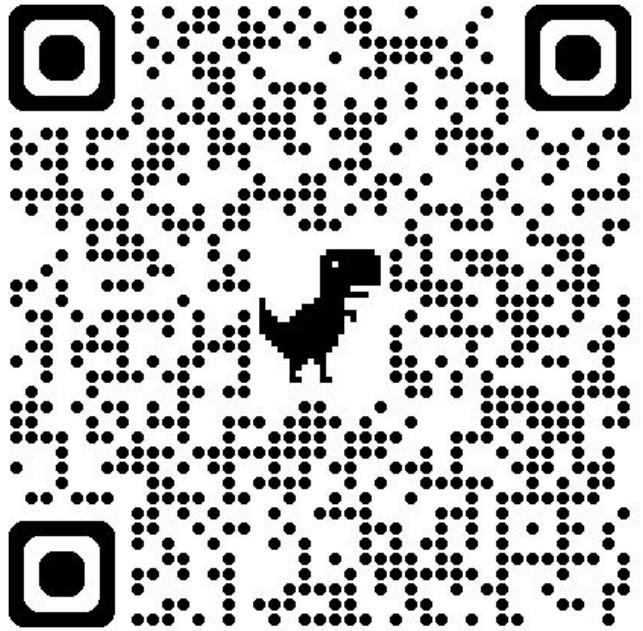
Luca Henkel
Erasmus University Rotterdam

Date Written: August 18, 2025

Abstract

We study the impact of replacing human recruiters with AI voice agents to conduct job interviews. In collaboration with a recruitment firm, we conducted a natural field experiment in which 700 individuals were assigned to be interviewed by human recruiters, AI voice agents, or given a choice between the two. Across all three conditions, human recruiters evaluated interviews and made hiring decisions.





A silhouette of a person stands facing the ocean at sunset. The sky is filled with warm orange and pink hues. The person's back is to the viewer, looking out over the water.

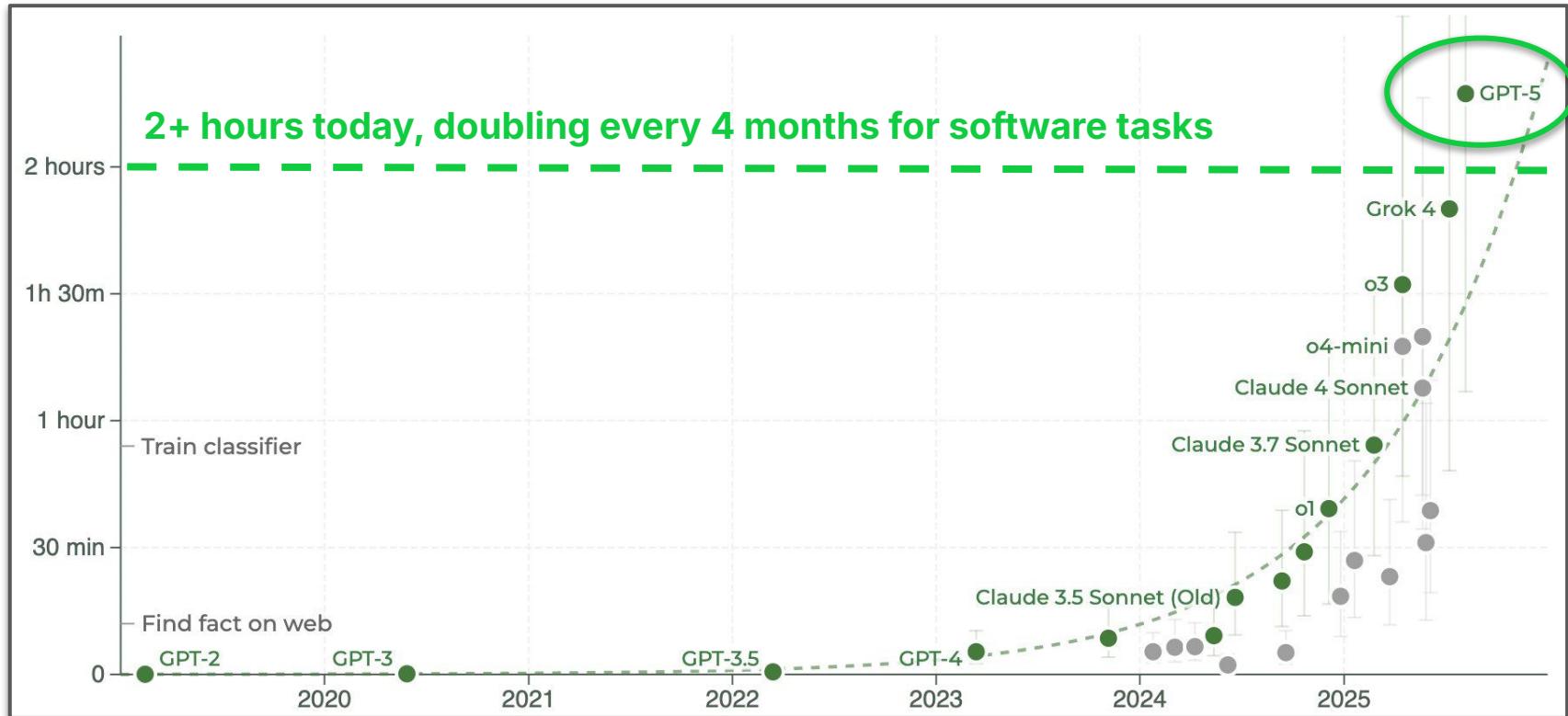
Sign up to Volunteer

“I have 50 people just playing around with AI. AI can't do the things we want it to do, and I feel like we're wasting our time.”

– *Birmingham CEO*

We want you to feel like
AI is something that's
happening for you not
to you.

The benchmarks do give us hints about where things are headed



Sanity Check: Is it possible AI can do X?

What Can You Do Today to Find Out?

I'm searching for rigorous evidence on whether AI can (or cannot) perform **[this specific task]** at or above the level of the **next best human**. Play the role of a **skeptical researcher**: prioritize **data-driven findings** over opinions or anecdotes, and favor **academic and empirical sources**. At the same time, **actively surface optimistic evidence**, but contextualize it with skepticism.

Gather material from **Academic research** (peer-reviewed papers, preprints, systematic reviews), **Independent case studies**, **Practitioner/forum discussions**, **Vendor materials**. Present findings as an **annotated bibliography**. For each item: 1) Label the source type (academic, case study, forum, vendor), 2) Summarize the claim or result, 3) Assess credibility and limitations (with a skeptical lens), 4) Highlight whether the evidence supports or contradicts AI achieving human-level performance.

Conclude with a **balanced synthesis**: what the strongest data suggests, where optimism is warranted, and where skepticism remains justified.

Sanity Check: Is it possible AI can do X?

How can you set up your first evaluation

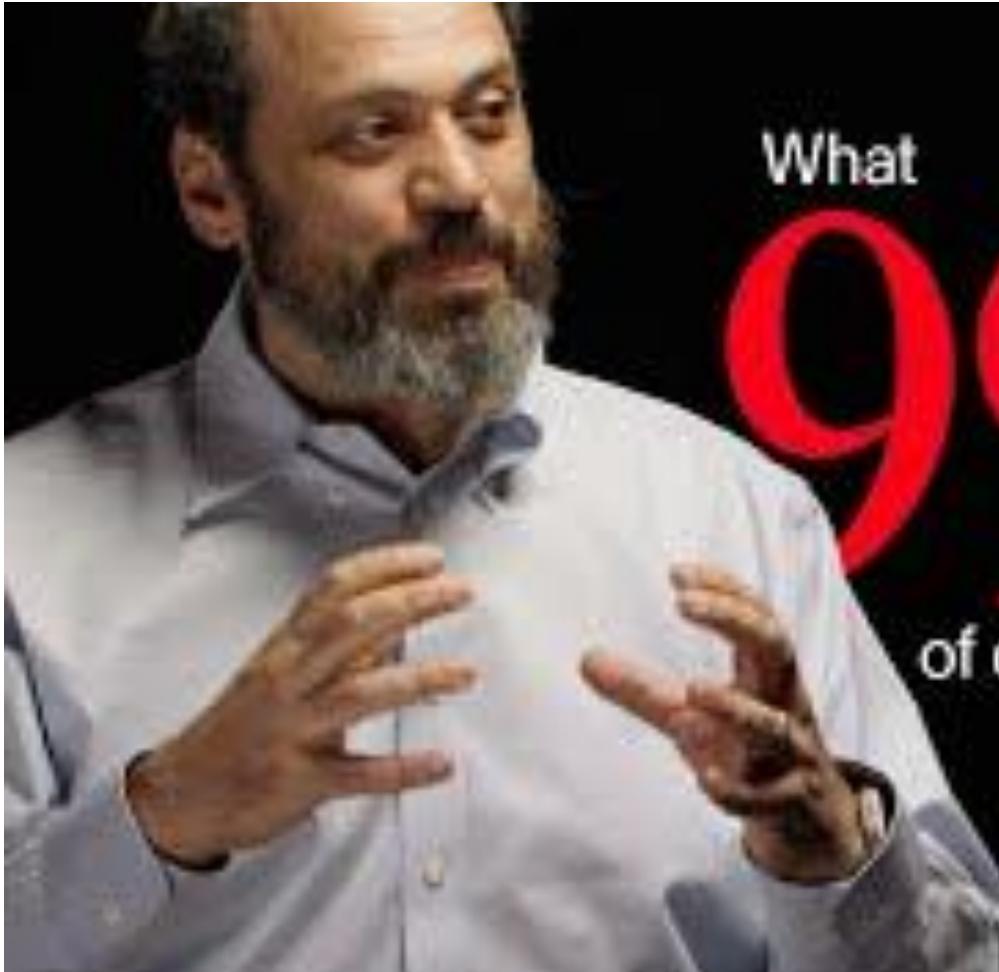
- Designate a single person as the domain expert for this task
- Develop at least 10 test cases for evaluating the AI's performance
- Start with chat models (Ex. ChatGPT, Claude) and agent tools (Ex. n8n, Zapier)
- Have the expert grade the results in a spreadsheet- score, pass/fail, detailed critique
- Identify objective (rule-based) and subjective failures (opinions)
- Optional: Develop an AI grader that codifies the domain expert's knowledge

<https://www.lennysnewsletter.com/p/building-eval-systems-that-improve>

Evaluating AI Results

PROMPT (simplified): Analyze the opportunity of expanding my startup Conserv into the United Kingdom. Consider market size, competition, regulations, and potential strategies. Provide a detailed recommendation.

	Deep Research	o1	o1-Pro	4o	o3-mini
Clarity & Structure	4.7	5.0	4.0	2.7	3.0
Competition & Differentiation	5.0	3.3	3.7	2.0	1.7
Data Depth & Credibility	5.0	3.0	2.7	3.3	2.0
Implementation & GTM	5.0	4.7	4.3	1.7	3.0
Market Opportunity	4.3	3.0	3.0	4.0	2.3
Regulatory & Risk Management	5.0	2.0	2.3	2.3	2.0
Grand Total	4.8	3.5	3.3	2.7	2.3

A photograph of a man with a beard and short hair, wearing a light blue button-down shirt. He is gesturing with his hands as if he is speaking or presenting. The background is dark.

What

99%

of companies are missing



Sanity Check: Is it possible AI can do X?

Understand the levers you can pull to improve

Is this the **right task** to tackle?



What **prompt(s)** to use?

What **general tool(s)** to use?

What **model(s)** to use?

What **context** to provide?

Reliability Gate: Can we trust AI do X?

MIT report: 95% of generative AI pilots at companies are failing



BY SHERYL ESTRADA

SENIOR WRITER AND AUTHOR OF CFO DAILY

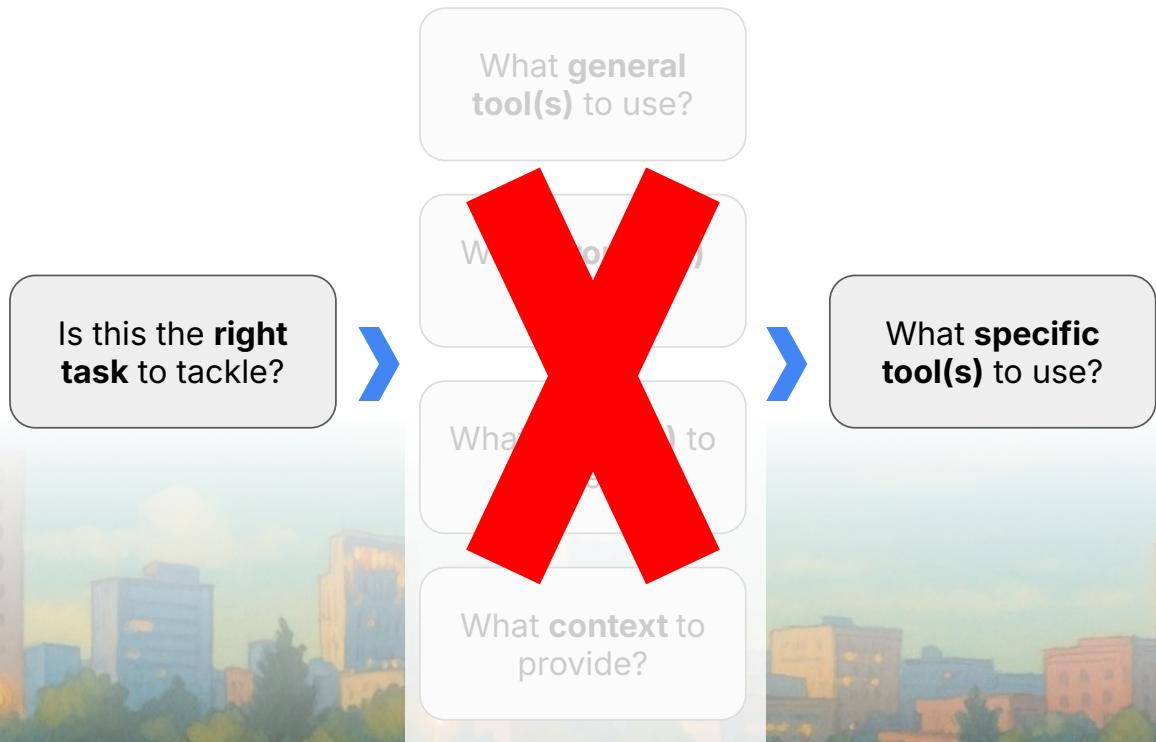
August 18, 2025 at 6:54 AM EDT

** based on 52 interviews at a conference & unclear methodology

Successful pilots focus on

- tight integration into existing workflows
- back-office operations, instead of sales & marketing
- ***purchasing AI tools from specialized vendors***

Reliability Gate: Can we trust AI do X?



Starting today ...
Spend more time exploring specialized tools

Breakouts are Peer Groups to Explore Tools

Engineering
September 17



Non-Profits
September 16



Finance
September 24



Reliability Gate: Can we trust AI do X?

Evaluations are structured tests that *check how well an AI does its job.*



Allie K. Miller
@Lewellen

Evals help us answer 4 questions

Why AI evals matter and why I believe most of the tweets are getting

task performance

How do we develop confidence that AI can do a task?

role replacement

How much of a job can AI take on at human performance?

anti-hype

How can we evaluate trust and reduce the overhype with AI?

tracking progress

How do we know AI is heading in a better direction versus today?

How to Explore AI without Wasting Time

Sanity Check

Is it *possible* AI can do X?

Reliability Gate

Can we *trust* AI do X?

DECISION

Onboarding

How do we *hire* AI to do X?

Management

How do we *manage* AI doing X?

A photograph of a dirt path winding through a dense forest. The path is covered in fallen brown leaves. Tall, thin trees with green leaves stand on both sides of the path, creating a shaded canopy. The overall atmosphere is peaceful and natural.

World 1: AI is a waste of time

World 2: AI is transformational

your psychology matters

Breakout Spotlight!

Engineering Breakout



O HML



Praise Daramola

Pax Nguyen



[+ New Transaction](#)

2774 N Route 9

BUYING

0% To Closing

Active

 Search Tasks Timeline Details Documents Notes History

Active

2774 N Route 9

2774 N Route 9

2774 N Route 9

5505 BROCKS PASS

5505 BROCKS PASS

112 Hartsdale Rd

2774 N Route 9

5505 BROCKS PASS

 Fe

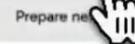
Tasks

 Search tasks[+ New](#) Verify Fully Executed Contract 05/29/2025 Follow-up on Earnest Money Receipt 06/02/2025 Schedule Appraisal 06/13/2025 Obtain Mortgage Pre-Approval Letter 06/14/2025 Follow-up on Septic Inspection 06/19/2025 Coordinate Home Inspection 06/19/2025

- Schedule the appraisal by **June 13**
- Ensure submission of the **Earnest Money Receipt** due by **June 3**
- Confirm completion and review of the **Septic Inspection** by **June 20**

When drafting your next email, remember to be concise and clear. Highlight key points at the top to ensure they are not missed by the recipient.

Let me know if you have any questions or if the timeline requires adjustments. I'm here to support you through every step of this transaction.

[Prepare new](#)[List missing documents](#)[Review key dates](#)[Task summary](#) Type your message...Ava can make mistakes. Verify important information.

1) Logic Based

What year was ChatGPT released?

A: 2022

2) Model Grading

[A support conversation with a customer]

Did the agent solve the customer issue?



Distillation Test / shikhar

Playground Dashboard Docs API reference S

DASHBOARD Prompts Logs Traces Assistants Batches Datasets Evaluations Fine-tuning Storage Usage API keys

< sentiment_extraction_eval Report Data ... + Add run

Run gpt-4.1-sentiment Sentiment Grader

Item			Sample View details		Testing Criteria
text	channel	language	Input	Output	Sentiment Grader: Result
It's okay – not great but not bad either.	app_review	en	System: You are a helpful assistant. User: It's okay – not great but not bad either.	Assistant: {"sentiment": "neutral", "confidence": 0.95, "emotions": ["indifference", "acceptance"]}	Pass: 4.5 ↗
I love this product!	twitter	en	System: You are a helpful assistant. User: I love this product!	Assistant: {"sentiment": "positive", "confidence": 0.98, "emotions": ["joy", "satisfaction"]}	Pass: 5 ↗
This is the worst experience I've ever had.	support_ticket	en	System: You are a helpful assistant. User: This is the worst experience I've ever had.	Assistant: {"sentiment": "negative", "confidence": 0.99, "emotions": ["anger", "frustration", "disappointment"]}	Pass: 5 ↗
No estoy seguro de lo que pienso sobre este producto.	facebook	es	System: You are a helpful assistant. User: No estoy seguro de lo que pienso sobre este producto.	Assistant: {"sentiment": "neutral", "confidence": 0.95, "emotions": ["indecision"]}	Pass: 5 ↗
总体来说，我对这款产品很满意。	wechat	zh	System: You are a helpful assistant. User: 总体来说，我对这款产品很满意。	Assistant: {"sentiment": "positive", "confidence": 0.98, "emotions": ["satisfaction", "contentment"]}	Pass: 5 ↗



Playground - 2025-02-04 03:56:04

Compact Full Diff JSON Heat Map Columns + Compare

Input	Reference Output	Output	Hallucination	Helpful	Random	Latency	Status	Tokens
OCTOBER 2023, VOL...	#4eb6 → {"tone":"positive","topic":"Princeton U...	ai: This email is a newsletter ... 1/3 < >	1.00	0.00 μ	0.50 μ	14.48s	SUCCESS	3 5,428
Solve the RedWordle We kn...	#6a7d → {"tone":"positive","topic":"RedWordle ...	ai: The email is an engaging,... 1/3 < >	1.00 μ	0.00 μ	0.50 μ	14.47s	SUCCESS	3 339.333
*Would you like to add spe...	#72fb → {"tone":"positive","topic":"Personalize...	ai: This email is a promotion... 1/3 < >	1.00 μ	0.00 μ	0.50 μ	14.47s	SUCCESS	3 377.667
--- --- --- Book with Fal...	#8602 → {"tone":"positive","topic":"Fall Cruise ...	ai: This email is a promotion... 1/3 < >	1.00 μ	0.00 μ	0.50 μ	14.47s	SUCCESS	3 2,251.667
--- Costco --- ANSWER & ...	#c4fa → {"tone":"positive","topic":"Invitation to...	ai: This email appears to be ... 1/3 < >	1.00 μ	0.33 μ	0.50 μ	12.98s	SUCCESS	3 578.333
USA Finance Solution offers ...	#f1a8 → {"tone":"positive","topic":"Business Fu...	ai: This email is a marketing ... 1/3 < >	1.00 μ	0.33 μ	0.50 μ	12.97s	SUCCESS	3 396



Breakouts are Peer Groups to Explore Tools

Engineering
September 17

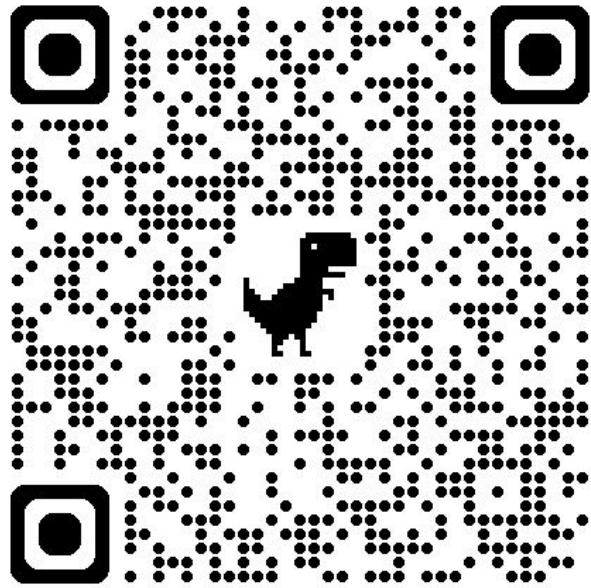


Non-Profits
September 16

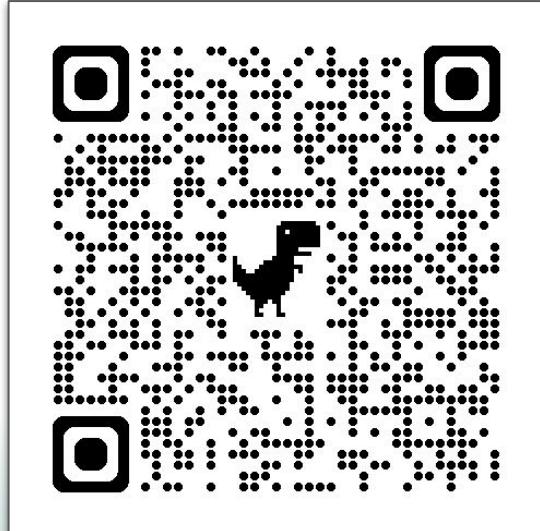


Finance
September 24





**Sign up
for October
Meetup**



Help us track attendance
Provide quick feedback
Enter to win a t-shirt

