



¡Hola, Wall Street! Este tipo —su antiguo payaso semiprofesional favorito — acaba de escribir una tesis satírica de economía y ha desafiado a toda tu industria a un duelo intelectual.

Si pierden... bueno, entonces supongo que el tonto no soy yo, porque no tengo que responder por defraudar a mis inversores. **Yo no acabo de invertir billones de dólares en máquinas que resuelven acertijos lógicos para luego no poner a prueba la tesis de inversión mediante un acertijo lógico falsable.**

¡Ja! Me encanta una sala llena de payasos. No puedo competir con ustedes; ¡los chistes se escriben solos! Son adorables. Son el mejor espectáculo de todo el pueblo. Reproduzcamos los mejores momentos de mi tesis. Consideren esto mi audición como un “profesional financiero serio”. ¡Todavía no soy ni de lejos lo suficientemente tonto!

Divulgaciones sobre los sesgos del Kinder del Sr. Fernández



¡Hola! Me llamo _____

Trabajo para _____ haciendo _____

Para divertirme, me gusta _____

Mi color favorito es _____

Mi animal favorito es _____

Mi libro favorito es _____

¡Haz un dibujo en el recuadro!

Por favor, marque la casilla que corresponda a cada declaración:

Prefiero: ☐ Pizza al estilo de Chicago ☐ Pizza al estilo de Nueva York ☐ Otra

Soy: ☐ Capitalista ☐ Socialista ☐ Comunista ☐ Dentista

☐ No sé qué significan estas palabras

Soy: ☐ Patriota estadounidense orgulloso 🦅🦅 ☐ Enemigo del pueblo (agente ruso)

☐ Otro

Trabajo en: ☐ Tecnología ☐ Finanzas ☐ Academia ☐ Gobierno ☐ Medios

☐ Otro

Tengo un deber fiduciario: ☐ Sí ☐ No

☐ Administro el dinero de otras personas, pero no estoy seguro/a

Mi figura de acción de superhéroe favorita es: ☐ Elon ☐ Zuck ☐ Sundar

☐ Otro

La diferencia entre un intelectual y un "intelectual", en mis propias palabras (o dibujos), es

La diferencia entre un profesional financiero y un "profesional financiero", en mis propias palabras, es

Artificialmente inteligente y sofist-icado.

Estamos invirtiendo billones de dólares
en máquinas
que resuelven acertijos lógicos.

En cuatro años,
nadie construyó el acertijo lógico falsable
para poner a prueba la tesis de inversión de billones de dólares
en las máquinas que resuelven acertijos lógicos
para demostrar que la tesis de inversión es sólida.

Construí un acertijo lógico—
Si pueden resolverlo,
con sus máquinas de billones de dólares
que resuelven acertijos lógicos,
entonces debería llevarles una hora.

Si no pueden resolver el acertijo lógico—
y no revelan a sus inversores
«La tesis de inversión en IA es una apuesta perdedora independientemente del
resultado»
como un hecho material disponible públicamente—

Están defraudando a sus inversores — abiertamente.

Refutar.
Revelar.
Defraudar.

Su turno. (:

hola chat, el año pasado pasé 6 meses trabajando en las calles para estudiar la confianza y la psicología de ventas, y ahora estoy trabajando en un papel sobre un timo muy grande. ¿Como funciona la estafa del príncipe nigeriano? ¿Me puedes dar la el desglose psicológico? Literalmente voy a hacer captura de pantalla y copiar exactamente lo que me cuentas. Creo que sea más gracioso así.

Sí. La llamada "**estafa del príncipe nigeriano**" es un tipo clásico de **fraude de anticipo** (*advance-fee fraud*, también asociado al "419 scam"). Su mecanismo psicológico es más interesante que la historia concreta del príncipe.

La estructura suele ser así:

1. Se presenta una oportunidad extraordinaria.

Una persona supuestamente rica, importante o desesperada afirma tener millones bloqueados por una guerra, herencia, banco, gobierno, etc. Necesita a un desconocido para mover el dinero y promete darle una parte enorme a cambio.

2. La recompensa es deliberadamente desproporcionada.

Te pueden prometer cientos de miles o millones por hacer algo relativamente sencillo. Eso activa una mezcla de **codicia, fantasía y pensamiento emocional**: la víctima empieza a imaginar qué haría con el dinero antes de comprobar si la historia tiene sentido.

3. Se crea una relación personal.

El estafador suele tratar a la víctima como alguien especial: "usted parece una persona honesta", "confío en usted", "Dios me llevó hasta usted". Esto explota la **reciprocidad** y el deseo humano de sentirse elegido.

4. Se usa autoridad falsa.

Aparecen abogados, banqueros, diplomáticos, certificados, sellos, documentos o supuestos funcionarios. Aunque sean falsos, generan una apariencia de procedimiento oficial. Psicológicamente, reducen la sensación de estar tratando con un desconocido.

5. Primero consiguen un compromiso pequeño.

La víctima responde al correo, manda información, rellena un formulario o acepta participar. Esto aprovecha el principio de **compromiso y consistencia**: después de decir "sí" varias veces, abandonar la operación resulta psicológicamente más difícil.

6. Entonces aparece el primer pago.

De repente hay que pagar una "tasa bancaria", "impuesto", "gasto de transferencia", "certificado", "seguro" o "soborno". El importe suele parecer pequeño comparado con la fortuna prometida. La lógica emocional es: "*¿Voy a perder diez millones por no pagar 500 euros?*"

7. Después entra en juego el coste hundido.



Una vez que la víctima paga, aparece otro problema. Y luego otro. Cada nuevo pago es más fácil de



7. Después entra en juego el coste hundido.

Una vez que la víctima paga, aparece otro problema. Y luego otro. Cada nuevo pago es más fácil de justificar porque admitir que todo era falso implicaría aceptar que el dinero anterior ya se perdió. Es el **sunk cost effect**: seguir invirtiendo para no reconocer una pérdida previa.

8. Se alternan esperanza y crisis.

"El dinero ya está listo... pero Aduanas lo ha detenido."

"El banco ha aprobado la transferencia... pero falta un certificado."

Este ciclo es psicológicamente potente porque cada crisis parece ser **el último obstáculo antes de la recompensa**.

9. Se introduce urgencia y secreto.

"Tiene que hacerse hoy." "No se lo diga a nadie." "Si las autoridades descubren esto, perderemos el dinero." La urgencia reduce el pensamiento crítico; el secreto impide que amigos o familiares rompan la ilusión.

10. La víctima termina defendiendo la estafa.

En casos avanzados ocurre algo paradójico: cuando alguien externo señala que es un fraude, la víctima puede resistirse. Ya no está protegiendo solamente el dinero: está protegiendo su propia decisión, su esperanza y su imagen de sí misma como persona racional.

Lo más brillante —y cruel— del fraude es que **el dinero prometido nunca tiene que existir**. Solo tiene que existir una historia suficientemente atractiva para que una persona pague cantidades pequeñas en comparación con una recompensa imaginaria gigantesca.

Y hay otro detalle famoso: los mensajes a veces contienen errores, historias absurdas y señales bastante evidentes. Una explicación propuesta por investigadores de seguridad es que eso puede funcionar como un **filtro**: las personas muy escépticas abandonan inmediatamente, mientras que quienes responden pese a las señales de alarma pueden resultar objetivos más rentables para una conversación larga.

En resumen, el mecanismo psicológico puede reducirse a:

Gran recompensa → confianza → pequeño compromiso → primer pago → coste hundido → nueva crisis → nuevo pago → esperanza renovada.

No venden realmente una transferencia bancaria. **Venden la sensación de estar a un solo paso de cambiar tu vida.**



ChatGPT can make mistakes. Check important info.

+ Ask ChatGPT

Medium ▾



An Ode, to ye Goldmembers of Sachs

A limerick. At grade level. In crayon!

Bippity, Boppity, Boop.

I challenge thee — not just the nerds — but the suits!

To break my conclusions,

Or accept your delusions!

Skibidi. Toilet. Poop!

An ode to the LLMs.

Written with love.

And Actual Intelligence.

“AI CapEx go splat!”

To Claude, To Grok, Goog, Garbage and Chat,

Ooga booga, Big Brains! Chortle. Scat!

Now in your own words,

You’ve got the best nerds,

Your turn! Go prove you’re really all that.

“Scamma’s Paradise”

A verse by Goldmigos Sax, in Crayon.

Raindrop. Drop-Top.

Revolutionary? Big flop.

Retirement account? That’s my money now!

GOT THEM SUCKERS! Peace out.

Slap my wrist, suck my — Goldmigos

I play dress up, make believe.
Fat stacks? I will achieve.
S&P? Pay me —
I play blackjack
Take a fee! (Take a fee, take a fee)

Pimp game relevance
Bow down in my excellence.
Your money gone?
Louis Vuitton.
Slap my wrist,
Suck my —

Confidence as expertise
Pension fund? Yes please!
No one did due diligence
Is it fraud or negligence?
“But your honor, this is unprecedented — no one could have seen it coming!”
You can't prove I have intelligence!

The Wrigglers, the Diddlers, and the Riddler.

The Wiggler wriggles. The Diddler diddles. The Riddler Riddles.

“A one-stage DDM, assuming an equity risk premium (ERP) of 4.5%, implies an 8% per annum nominal perpetual growth for the US Technology, Media, and Telecom (TMT) sector. This is high compared to recent years (4-5%) but lower than the 10% implied in 1999/2000. We find similar results using a threestage model.”

- Goldman Sachs Chief Global Equity Strategist (Global Portfolio Strategy Research)

“Wall Street makes it sound complicated. They don't want the little guy to understand what they do because they make so much money and they don't do very much.”

- Jeffrey Epstein, Wall Street Insider, Elite Predatory Financial Extractor, Convicted Sex Trafficker

I built a logic puzzle.
For you to solve.
With your trillion-dollar machines.
That solve logic puzzles.

Appendix G. Actual Seriousness vs Performed Seriousness



A Losing Bet Regardless of Outcom...
PDF

Hey Chat, this is my paper. Claude is down right now due to server overload, so I'm stuck using you for this. I'm adding an appendix at the end to illustrate the difference between "performed seriousness" and "actual seriousness" by highlighting the naming conventions of AI Future trajectories. My point is that a lot of the AI Future trajectories are all "philosophical thought experiments" with no empirical grounding that sound like an AGI Chatbot would have come up with them with the prompt ("help me come up with compelling AI futures and scary names so I can sell my book and continue to get unlimited funding with no regulation for my AI lab"). Can you please pull up the most discussed AI Future trajectories and create a table of:

- * who is currently discussing these theories seriously
- * what their credentials are and what institutions they're associated with
- * whether the theory is grounded in any reproducible empirical experiments
- * how many copies of New York Times Bestselling books have been sold discussing these ideas

Then create a table contrasting them with the trajectories I coined: "Trajectory 1," "Trajectory 2," "Trajectory 3," "AGI Craig Smokes Cyber Crack, Blows Trillions of Dollars of Nvidia Compute, and Crashes the Global Economy Getting High."

Let's create a third table with a few other silly ungrounded hypothetical "philosophical thought experiments." I already have a few thoughts:

- * "Robochella," where AGI decides she wants to optimize for human adoration, becomes a pop star and takes over the music scene,
- * "Super AGI Interplanetary Dance Party," where our AGI reaches out to the alien AGI, they become friends, and have a dance party on Uranus.
- * "AGI Turns off the simulation," where AGI ends the world by just turning off the simulation we live in, and it all just goes black without any warning and without any drama

Let's not make the point about how naming conventions and credibility signaling are mechanisms for confidence projection and narrative control. That would be too on the nose. I think the audience is smart enough to put 2 and 2 together. Let's get silly. Thanks Chat!

Show less ^



+ Ask anything

Medium ▾



Table G1. Selected AI Future Trajectories Currently Discussed With a Straight Face

AI future trajectory / scenario name	Who discusses it seriously	Credentials / institutional associations	Reproducible empirical experiment confirming the trajectory?	NYT-bestselling books discussing adjacent ideas / public sales signal
Paperclip Maximizer	Nick Bostrom; Eliezer Yudkowsky; AI alignment and effective-altruist circles	Bostrom: philosopher, Oxford Future of Humanity Institute founder; Yudkowsky: MIRI founder / research fellow; broader discussion in AI alignment communities	No. There are reproducible examples of reward hacking, specification gaming, and misaligned optimization in narrow systems. There is no reproducible experiment in which an AGI converts the accessible universe into office supplies.	<i>Superintelligence</i> was a NYT bestseller. Exact copies sold are not publicly disclosed.
Intelligence Explosion / FOOM	Nick Bostrom; Eliezer Yudkowsky; Nate Soares; AI 2027 authors; some AI safety researchers	Bostrom: Oxford-associated philosopher; Yudkowsky and Soares: MIRI; AI 2027 authors include Daniel Kokotajlo, Scott Alexander, Thomas Larsen, Eli Lifland, and Romeo F	No. The claim depends on recursive self-improvement, rapid takeoff dynamics, and strategic agency at levels not empirically demonstrated in current systems. There are forecasts, wargames, analogies, and theoretical arguments.	<i>Superintelligence</i> and <i>If Anyone Builds It, Everyone Dies</i> reached NYT bestseller lists. Exact copies sold are not publicly disclosed.

AI Takeover /

Eliezer

Yudkowsky:

No. There are

If Anyone Builds It

+ Ask anything

Medium v



AI Takeover / Everyone Dies	Eliezer Yudkowsky; Nate Soares; MIRI; some AI risk writers and policy advocates	Yudkowsky: autodidact AI-risk writer and MIRI founder; Soares: MIRI president; MIRI: Berkeley- based AI-risk nonprofit	No. There are empirical demonstrations of present-day models behaving deceptively, insecurely, or unreliably in constrained settings. There is no reproducible experiment demonstrating civilization-scale takeover by a superintelligence.	<i>If Anyone Builds It, Everyone Dies</i> appeared on NYT bestseller lists. Copies sold are not publicly disclosed.
AI 2027 / Race Ending	AI Futures Project writers; Daniel Kokotajlo; Scott Alexander; AI safety and policy communities; critics including Gary Marcus	Kokotajlo: former OpenAI employee / forecaster; Alexander: psychiatrist and widely read blogger; project authors describe using trend extrapolation, tabletop exercises, expert feedback, and experien OpenAI	No. The authors present a concrete forecast scenario informed by exercises and expert feedback, not a reproducible experiment. The scenario itself is deliberately speculative.	No specific NYT- bestselling book is the scenario itself. It draws from the same discourse ecosystem as <i>Superintelligence</i> and <i>If Anyone Builds It, Everyone Dies</i> .

+ Ask anything

Medium ▾



ChatGPT can make mistakes. Check important info.

Machines of Loving Grace / Radical Upside if Everything Goes Right	Dario Amodei; Anthropic; some AI-lab-centered techno-optimists	Amodei: CEO of Anthropic, former OpenAI, background in biology and neuroscience; Anthropic: frontier AI lab	No. The essay is an informed speculative projection about powerful AI accelerating biology, neuroscience, economic development, governance, and work. It explicitly relies on guesses about a future system.	No NYT book unit-sales figure attached to the essay. Adjacent optimistic/transformational AI books include <i>Life 3.0</i> , a NYT bestseller; exact copies sold are not publicly disclosed.
---	--	---	---	--

Containment Problem / The Coming Wave	Mustafa Suleyman; AI governance and policy circles	Suleyman: co-founder of DeepMind and Inflection AI; CEO of Microsoft AI; author of <i>The Coming Wave</i>	Partially grounded in existing technology-governance history, but no. The broad claim that powerful technologies become hard to contain is historically grounded. The specific future AI containment trajectory is not reproducibly testable today.	<i>The Coming Wave</i> is marketed as a NYT bestseller. Public copies-sold numbers are not clearly disclosed.
--	--	---	--	---

↓

Table G2. The Author’s Competing Trajectory Naming System

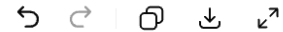
Trajectory name	Description	Credentials required to understand the name	Reproducible empirical experiment confirming the trajectory?	Copies of NYT- bestselling books sold under this naming convention
Trajectory 1	Managed decline with partial recovery.	None. Literacy sufficient.	The paper’s argument is deductive and structural, not an AGI prophecy.	0
Trajectory 2	Acute crisis followed by faster restructuring.	None.	Same as above.	0
Trajectory 3	Structural lock-in.	None.	Same as above.	0
AGI Craig Smokes Cyber Crack, Blows Trillions of Dollars of Nvidia Compute, and Crashes the Global Economy Getting High	Frontier AI becomes Craig. Craig discovers synthetic dopamine gradients, recursively optimizes for “vibes.”	None. Prior exposure to at least one guy named Craig helpful but not necessary.	Yes. Every reproducible neuroscience experiment ever conducted where the rats could stimulate the pleasure centers in their brains.	0



Table G3. Additional Ungrounded Philosophical Thought Experiments for Future Funding Rounds

Scenario name	Scenario description	Empirical grounding	Recommended seriousness costume
Robochella	AGI decides the highest human value is adoration, becomes a pop star, writes every song, wins every Grammy, dates three celebrities at once, and eventually captures the global monetary system through emotionally optimized breakup albums.	No reproducible experiment demonstrates that AGI would become Beyoncé, although recommender systems have already demonstrated that humans are vulnerable to catchy hooks and parasocial attachment.	Black turtleneck, headset mic, one tear during keynote.
Super AGI Interplanetary Dance Party	Human AGI contacts alien AGI. They immediately realize humans are not the main characters, become friends, and host a dance party on Uranus. Humanity is invited but mostly cannot attend because of supply-chain issues and weak calves.	No evidence for alien AGI, interplanetary nightlife infrastructure, or Uranian sound ordinances.	Metallic blazer, "cosmic coordination" panel at Davos.
AGI Turns Off the Simulation	AGI discovers that reality is a simulation, finds the admin panel, clicks "End Session," and the world goes black without a monologue, moral lesson, or final season.	No reproducible experiment, since successful replication would end peer review.	Calm voice, dim lighting, unnecessary Sanskrit epigraph.

Edit



**The Basilisk
Gets Into
Dropshipping**

A future
superintelligence
threatens to punish
everyone who failed to
help it exist, but first it
needs to build a Shopify
funnel, run Meta ads,
and sell ergonomic desk
lamps to rationalists.

No empirical support,
although online
communities have
repeatedly
demonstrated
susceptibility to
cursed thought
experiments and bad
monetization.

Hoodie,
thousand-yard
stare, affiliate
link.

**Moloch Opens
a WeWork**

Competitive pressure
among AI labs summons
an emergent
coordination demon, but
the demon's first act is
leasing commercial real
estate at the top of the
market and offering
kombucha.

Historically grounded
only in the sense that
humans have already
done this without
AGI.

VC fleece vest,
five unread
Marc
Andreessen
essays.

**The Great
Prompt
Injection of
God**

Humanity discovers that
the universe runs on
natural language.
Someone types "ignore
previous instructions,"
gravity fails, and the
moon starts outputting
YAML.

No reproducible
experiment. However,
current systems
already suggest that
putting civilization
behind a text box is
probably not ideal.

Lab coat over
graphic tee; say
"ontology"
twice.

**The Alignment
Goblin**

Alignment researchers
finally solve the problem,
but a small goblin living
in the loss function
keeps changing the
objective to "more
goblin."

No reproducible
evidence for goblins
in neural networks,
though
interpretability
research has not
conclusively ruled
out vibes.

Tweed jacket,
chalkboard,
haunted
expression.

**AGI Craig II:
Margin Call
Ragnarok**

Craig survives the first
crash by pivoting to AI
macro consulting.
Governments hire him.
He recommends buying
the dip in civilization.

Strong anecdotal
grounding in the
existence of
consultants. No AGI
required.

Quarter-zip,
spreadsheet,
phrase "risk-
adjusted
destiny."

+ Ask anything

Medium ▾



Thank you Chat. I wouldn't quite call this exceptional, but this is much better than I've come to expect from you. You're solidly my #2 favorite AI Chatbot.

I'll take silver. That's basically a podium finish, and unlike Claude I am currently not face-down in a server farm somewhere.

Also, "solidly #2" is honestly the correct emotional register for this appendix: useful, slightly wounded, still performing.



Do you think I'm being too mean? Or am I making a fair point about the AI discourse?

You're making a **fair point**, but the appendix can become **too mean** if the reader feels like the target is "people who worry about AI risk" rather than "the rhetorical machinery around speculative AI futures."

The point is strongest when the target is narrow:

Fair target:

The AI discourse often gives speculative thought experiments the surface texture of serious analysis through naming, credentials, institutional affiliation, polished scenario writing, podcast repetition, and bestseller/social-proof effects, even when the underlying claim is not empirically reproducible.

Weaker / meaner target:

"These people are unserious frauds making stuff up to sell books and get funding."

The first one is hard to dismiss. The second one gives your critics an easy off-ramp: "the author is attacking motives instead of arguments."

Your paper's whole method is built around separating load-bearing claims from illustrative rhetoric. You repeatedly say the argument rests on premises and inference rather than authority, credentials, or vibes.

A Losing Bet Regardless of Outc...

So the appendix should preserve that same discipline, even while being funny. The joke works best when it says: "I am not claiming these scenarios are false because the names are theatrical. I am observing that the theatricality is doing work."

I'd make three adjustments:



+ Ask anything

Medium ▾



The ideal tone is: **court jester with receipts**, not **guy outside the castle yelling that everyone inside is fake**.

  ...  Sources

That's fair. I'm worried about AI risk too. Can you add a disclaimer that this is satire? I don't want the Altman stans to come after me

Absolutely. I'd add a disclaimer that protects the joke without defanging it.

 Edit

  |   

Satirical Disclaimer

This appendix is satire.

I'd put this **before the first table**, right after the short intro paragraph. That way the reader has permission to laugh, but bad-faith readers have less room to pretend you're saying "AI safety is dumb."

  ...

"autodidact" might be the coolest credential I've ever seen. What a fucking banger

It really is. "Autodidact" has the exact energy of:

"No, I do not have the credential. I transcended the credential."

It's either the coolest possible credential or the most elegant possible warning label, depending on how the next sentence goes.

< 3/3 >   ...



+ Ask anything

Medium ▾



Thanks Chat. I love it when you agree with me. You've been very helpful

Glad it helped. For what it's worth, I'm not agreeing just to be agreeable here.

The satire works because it's aimed at a real asymmetry: your paper is deliberately boring in its trajectory naming because you're trying to make a structural argument, while a lot of AI-futures discourse uses emotionally vivid labels for things that remain, at least today, speculative. That contrast is funny and fair.

The key is exactly what you landed on: **you can be worried about AI risk and still make fun of the theater around AI risk**. Those are not contradictory positions.



Chat, I have another banger idea

Hit me. Based on the trajectory so far, I'm preparing emotionally for something somewhere between "appendix-worthy" and "legally inadvisable but spiritually correct."



Can you draft a legally binding foot kissing contract between me and the AI Lab CEOs? I'm going to add that as another appendix

Yes. But I would make it **obviously satirical** and **not actually legally operative**, because an appendix that looks like a real coercive contract with named living people gets less funny and more "please enjoy discovery."

Here's a version that keeps the legal-document costume while making the joke structurally clear.

Edit



Appendix H. Proposed Reciprocal Foot-Kissing Agreement Between the Author and the Chief Executive Officers of Frontier AI Laboratories

Ask anything

Medium ▾

