

# AI Safety Is Mostly A Sex Cult In Berkeley, California



SE GYGES

10 HOURS AGO



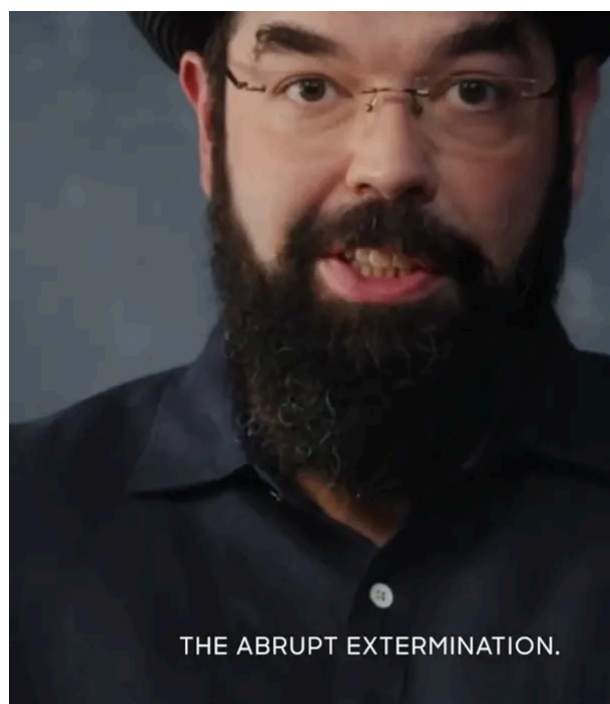
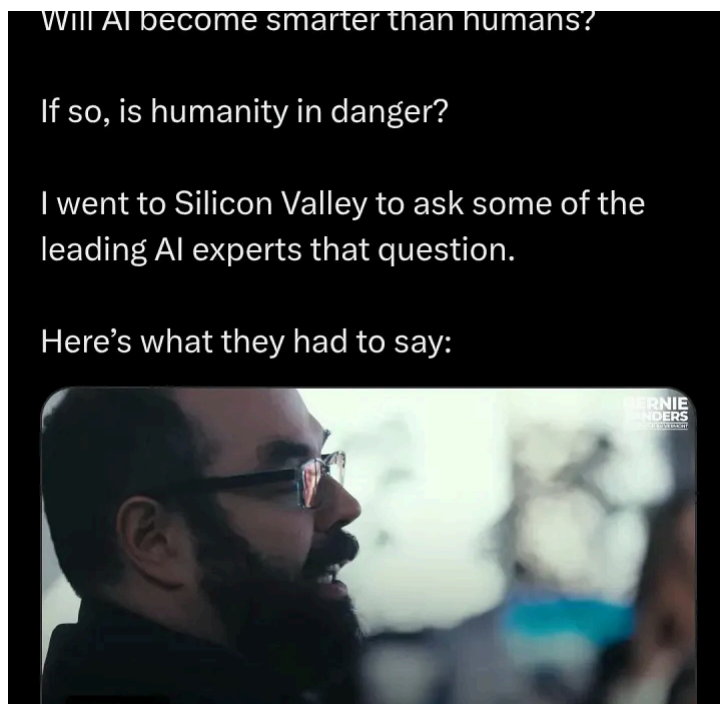
These people shouldn't make policy.

There has been a great effort over many years to distance AI Safety discussion from its origin and intellectual center, because its center is predatory sex cult started by a fan fiction author. Since those involved to be taken seriously, they have to hide that. Everyone else working in Safety is, effectively, a smokescreen for this core group and their ideas.

We might some day have "AI Safety", as a science or policy area, that primarily a sex cult in Berkeley, California, but it does not currently exist.

## Who Is Eliezer Yudkowsky?

This is Eliezer Yudkowsky. He is seen here in a meeting with Bernie Sanders who he is advising on AI policy, and in The AI Doc, now on Netflix.



This is Yudkowsky in Time Magazine in 2023, advocating in favor of risk nuclear war to prevent AI research from happening anywhere on Earth

Frame nothing as a conflict between national interests, have it clear that anyone talking of arms races is a fool. That we all live or die as one, in this, is not a policy but a fact of nature. Make it explicit in international diplomacy that preventing AI extinction scenarios is considered a priority above preventing a full nuclear exchange, and that allied nuclear countries are willing to run some risk of nuclear exchange if that's what it takes to reduce the risk of large AI training runs.

This is very senior OpenAI employee and unofficial OpenAI spokesman Roon. Roon's opinion of Yudkowsky is that:

1. Yudkowsky is in and/or running a sex cult.
2. Yudkowsky's opinions about AI are important, and should be taken seriously.

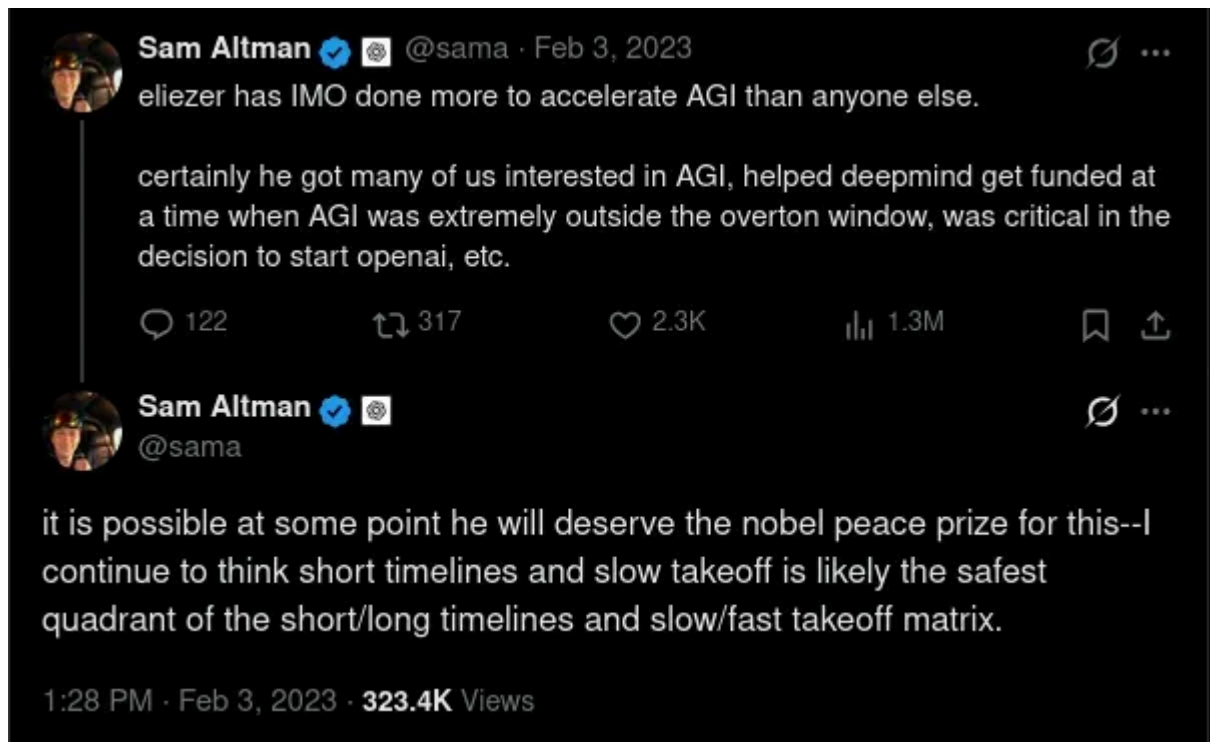
This seems to be an extremely common pair of opinions concerning Yudkowsky.





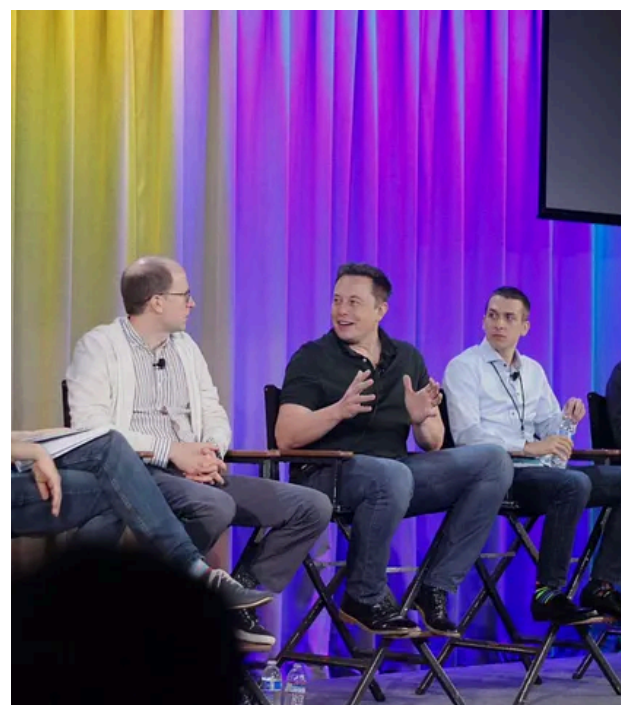
Yudkowsky is sufficiently controversial that people motivated by his ideas often downplay or deny his influence. Sometimes this spans multiple layers. People who received money from Sam Bankman-Fried denying they're Effective Altruists, Sam Bankman-Fried denying he's an Effective Altruist when caught, Effective Altruists denying their version of AI Safety is heavily influenced by Yudkowsky's, people literally attending the same orgies as Yudkowsky passing themselves off as merely "rationalist adjacent", and on. But Yudkowsky is provably, publicly influential on essentially all major public figures in AI. It is easy to deny not because it is hidden but because it seems so insane.

Yudkowsky is the person who connected Shane Legg and Demis Hassabis with Peter Thiel so that they could secure funding to start DeepMind [1](#), which is now Google's AI research division. Here's Sam Altman's opinion of Yudkowsky.



This is the schedule for the Future of Life Institute's conference on AI in Puerto Rico in January, 2015. Elon Musk would found OpenAI that December. So it seems plausible that his attendance did influence the founding of OpenAI.

0-0900: Breakfast  
 0-1200: Creating human-level AI: how and when?  
 rt talks followed by panel discussion: *will it happen, and if so, when? Via engineered  
 tion, whole brain emulation, or other means?* (We defer until the 4pm session questions  
 rding what will happen, about whether machines will have goals, about ethics, etc.)  
 akers:  
 mis Hassabis, Google/DeepMind  
 leep George, Vicarious (pdf)  
 m Mitchell, CMU (pdf)  
 elists include Joscha Bach (MIT), Francesca Rossi (Padova), Richard Mallah (Cambridge  
 ianatics), Richard Sutton (Alberta)  
 1-1300: Lunch  
 1-1515: Free play/breakout sessions on the beach  
 i-1545: Coffee & snacks  
 5-1600: Breakout session reports  
 1-1900: Intelligence explosion: science or fiction?  
 i intelligence explosion happens, then what are likely outcomes? What can we do now to  
 imize the probability of a positive outcome? Containment problem? Is "friendly AI"  
 sible? Feasible? Likely to happen?  
 akers:  
 ck Bostrom, Oxford (pdf)  
 irt Selman, Cornell (pdf)  
 an Tallinn, Skype founder (pdf)  
 on Musk, SpaceX, Tesla Motors  
 elists include Shane Legg (Google/DeepMind), Murray Shanahan (Imperial), Vernor  
 je (San Diego), Eliezer Yudkowsky (MIRI)  
 1: banquet (outside by beach)



Dario Amodei, current CEO of Anthropic, is here in 2013 identified as a "member of the Effective Altruism community" on the blog of Yudkows

non-profit.

# MIRI strategy conversation with Steinhardt, Karnofsky, and Amodei

January 13, 2014 | Luke Muehlhauser

On October 27th, 2013, MIRI met with three additional members of the effective altruism community to discuss MIRI's organizational strategy. The participants were:

- [Eliezer Yudkowsky](#) (research fellow at MIRI)
- [Luke Muehlhauser](#) (executive director at MIRI)
- Holden Karnofsky (co-CEO at [GiveWell](#))
- [Jacob Steinhardt](#) (grad student in computer science at Stanford)
- [Dario Amodei](#) (post-doc in biophysics at Stanford)

We recorded and transcribed much of the conversation, and then edited and paraphrased the transcript for clarity, conciseness, and to protect the privacy of some content. The resulting edited transcript is available in full [here](#) (62 pages).

And here he is arguing about the paperclip maximizer, a now-classic thought experiment, originally Yudkowsky's, about an AI that might kill everyone.

**Jacob:** I think what Eliezer would say, and correct me if I'm wrong, is that there is a natural category that looks kind of like fulfillment on some level, but it includes all sorts of things that you don't want. Like saying 'yes' to wire-heading. Is that right?

**Eliezer:** Yes.

**Dario:** For any short term neural computational category, that's going to be the case. There's a longer term category that refers to what we mean, that describes something of greater spatial-temporal extent. The question is whether that's natural or not? Are you saying that it's not?

**Eliezer:** Why would a paperclip maximizer need to know about that in order to manipulate you?

**Dario:** Maybe it'll win so quickly that it doesn't need to know about that. But something that was interacting with humans over an extended period of time might need to know that.

**Eliezer:** Why?

**Jacob:** Actually, I'm not convinced that continuing along this line vs. popping up a couple of levels is the best use of our time.

**Eliezer:** The whole fundamental question about, "Why is FAI hard?" has awful lot to do with a lot of this. I do suspect that diving into the details and seeing that lots of specific things don't work out might help people see why FAI is so hard.

“AI Alignment” and “AI Safety” were first popularized by Yudkowsky’s non-profit, MIRI, the Machine Intelligence Research Institute. OpenAI and Anthropic have cited Yudkowsky or his non-profit at least a dozen times; and the entire way we talk about AI uses terminology his non-profit popularized. It is effectively impossible to even discuss these areas without implicitly or explicitly, performing commentary on Yudkowsky’s writing.

This is Jacob Coxon, the recent “whistleblower” who quit Anthropic, being asked his opinions on rationalists and effective altruists.





**Ed Zitron**  @edzitron · 2h

what is your relationship to the rationalists and effective altruists?



5



1



161



9.3K



**Jacob Coxon**  @hilbertspaess · 1h



Yeah, this is a really important question.

I've been associating with rationalists for almost a decade. Their ideas were compelling and you can't deny their foresight. The founders of every superintelligence effort are all bound up in the rationality sphere. Sam Altman and Greg Brockman used to frequent the MIRI offices.

The way I relate to rationalists is it's like the bizarre hypertrophied physiques on specialized athletes. In my opinion they have weirdly specialized minds that make them uniquely able to be correct about topics removed from daily experience. These "specialized minds" also can make them off-putting to many people.

I don't know if I'd consider myself a rationalist. I am intuitively opposed to polyamory, veganism, and consequentialist intuitions. But I certainly trust them to be correct about factual questions.



15



20



268



16K



[Show replies](#)

This is Nate Silver citing LessWrong as a source of major information on AI risk. LessWrong is a website started by Yudkowsky in 2009, and originally filled only with his writing.



**Nate Silver** ✓  
@NateSilver538



The AI risk people actually do show their work! There are hundreds of working papers, thousands of LessWrong posts, literally entire best-selling books about this.



**rat king** 🐭 ✓ @MikeIsaac · Sep 11

a frustrating thing about all the "TK percent chance we die" headlines is that researchers can say any number and non-technical people will just be like "damn well that guy's in tech, im worried"

show your work! is this about scaling laws or some arxiv file ive yet to read? ..

8:10 AM · Sep 11, 2026 · **350K** Views

This is Oliver Habryka, who runs the nonprofit controlling LessWrong at the rationalist's compound, known as Lighthouse, explaining how large he thinks the influence of rationalism is on AI Safety and AI existential risk discourse. He is engaged in some kind of internal territorial dispute he is possibly the only time I have ever agreed with him.





**Oliver Habryka** ✓

@ohabryka



> This has given EA the power to affect the discourse, eg by METR writing about the Hugging Face hack or Dario publishing the Pacing The Frontier letter.

Come on, this seems just absurd. LessWrong gets more views than basically all popular EA content combined. How does EA have "power to affect the discourse" in ways that the rationality community does not?

If anything the memetic influence of the rationality community, with its persistent focus on AI Safety and existential risk is obviously the one that has majorly affected the discourse. The rationality community is most of the reason why the EA community has cared about AI existential risk for the last decade!

Some of your other points seem valid, but this one just leaves me a bit baffled.

6:51 PM · Sep 21, 2026 · 576 Views

This is a podcast interview with Daniel Alejandro Moreno-Gama, who attempted to firebomb Sam Altman's house, in which he names Yudkovsky as his primary influence.



## Sam Altman's Attacker, In His Own Words

Longview



Watch on

## Who Are The Rationalists?

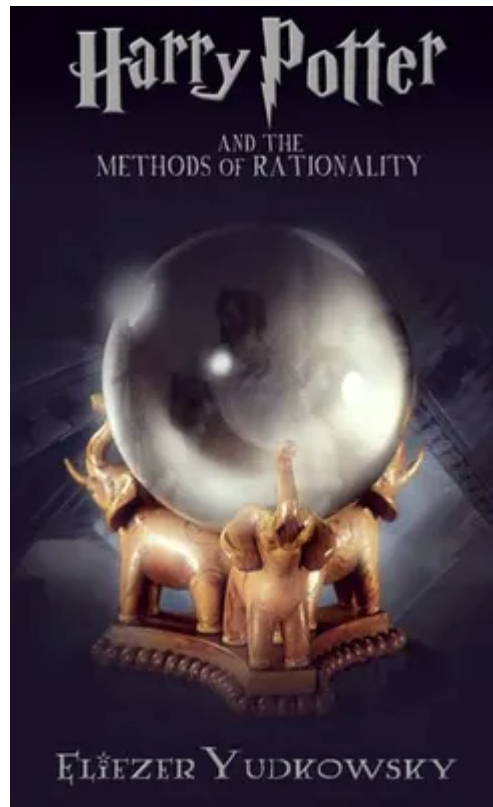
Yudkowsky's more dedicated followers seem to prefer to call themselves "the rationality community" these days, everyone else calls them "rationalists", the entire following "rationalism". I generally prefer "rationalism".

This is a long denial that [they are a religion](#). I alternate between finding it funny and finding it insulting.

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>2. It ended up with the name 'rationality' community, which many ppl regret but it's too late to change. often ppl say they are 'aspiring' rationalists, partially cause it's dangerous to get your ego tied up in a flattering word, partly cause of general distrust of appeals to credentialism, and partly cause imo rationality is more of a 'process' than it is a point you can hit.</p> <p>3. It's a community, or a scene, that centered around a shared value that not only can you 'want' to figure out how to think in ways that result in truth more often, but that you can actually put specific, intentional effort into these things.</p> <p>4. Many of these ideas center around figuring out ways to identify if your thinking is wrong. like, if you think x is going to happen, actually bet money on it, so that you can't dismiss it to yourself if it doesn't happen. You actually have to suffer a consequence if you're wrong. This will help you avoid psychological cope that you might not notice you're doing.</p> <p>5. There's been tons of meta community discussion around how to avoid failures that come from 'trying to do rationalist things' as opposed to 'trying to figure out what's true'. For example, you might think that emotions interfere with figuring out the truth (as e.g. fear or greed does with choosing stocks), you then go 'i should ignore emotions' and then proceed to miss out on tons of valuable information about yourself and the world.</p> <p>6. The community contains tons of disagreement about facts about the world, and even the sequences. in the current bay area sequences reading group, one of the default prompts for the readings is 'do you disagree with anything here? if so, what?' and then people debate about it.</p> | <p>burning you to death? Can you imagine what that's like?</p> <p>It has chiseled sacred norms that touch every single part of life, from clothing to marriage to war to death itself. It's a force that has, historically, burrowed deep to the core of human minds and completely transformed them into weapons for its own bidding.</p> <p>In much of the western world, that beast has died, and cultures walk around wearing gilded remnants of its flesh. "Look, I go to church," they say, but their soul is possessed by nothing, it's filled with osmosis of the secular norms around them. They support gay marriage, or have ever had a divorce, or treat their sacred texts not as sacred but as some inspirational poetry or as a self-help alternative.</p> <p>And like, sure. Get meaning where you can. Traditions are comforting. But the unitarian universalists or the casual catholics or the cultural jews are entirely unpossessed by the beast.</p> <p>'Religion' to people in western culture is watered down, and thus is very easy a term to throw at anyone in different clothing. And fine - it doesn't matter the word we use itself as long as everyone is on the same page. We can expand it, why not. But expanding it doesn't actually make change how possessed people are by the Beast.</p> <p>Rationalists, imo, don't even wear the skin of the religious. There's no</p> | <p>I do think that the rationality community has been born, who are using it to provide a certain sort of meaning in their life. like a desire to be a good person deserving of love, fulfilled, do-gooding. Or a fear of loss of control, soothed by working sure your thoughts are Very Correct Or Else. (The community of internal discourse about this too). Maybe people from the rationality and think 'oh, if I were going to get into this, my mind be trying to fix a brokenness inside me', and they pattern-mimic the function of religion?</p> <p>But I suspect more commonly it's related to the tendency for outgroup the weirdos, especially earnest weirdos, and especially weirdos who have the potential to be powerful in some way (I score math nerds usually end up with a lot of money).</p> <p>There's qualities that vaguely reflect some aspects of religious nonreligious aspects of human life, because the things that compelling to people are often tied into basic human urges, everywhere, religion or not. The knitting community has dress patterns, and litigates within itself the moral norms around dress ownership. The conservatives bristle with defensiveness against the immigrants; likeminded furies wear furry uniforms and go to commune and celebrate their ways of life.</p> <p>Science, academia, etc. is also an elevated community with</p> |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

The fan fiction that launched a thousand cults is Harry Potter and the Methods of Rationality, or HPMOR. It is possibly the highest-impact single

modern religious text, rivalled only by the Book of Mormon. Its influence may not yet have peaked. It is the holy book of the first great internet



Harry Potter and the Methods of Rationality

## How Are Rationalists Connected To The Bay Area Effective Altruists?



Bay Area Effective Altruism meetup's banner photo. Can you spot  
Yudkowsky?

Allegedly, Effective Altruism and its AI safety efforts are a separate thing and not directly beholden to Yudkowsky. This is fundamentally not true. Effective Altruism in the Bay Area has almost perfect overlap with rationalism, and the Effective Altruist AI Safety people in the Bay set the agenda for the rest of Effective Altruism's safety efforts.

In the Bay Area, Rationalism and Effective Altruism are the same people, the same events, in the same places, who believe the things said by the same authors and leaders, with only very slight differences between spheres of influence. Yudkowsky called himself an effective altruist and a rationalist until a peculiar internal schism. Scott Alexander, the second most influential writer in the space, has usually called himself both. They generally work for the same companies and non-profits.

Frequently, Effective Altruists bring up all of Effective Altruism's other charitable activities and its charming, benevolent members who don't live in the Bay Area to deflect true criticisms of the Bay Area Effective Altruist community. This effectively makes their other charitable operations an elaborate [shields and fronts for the priorities and behavior of the Effective Altruists in the Bay Area, who are interested almost exclusively in AI.](#)



## How Anything Works

of the secret circle of cool effective altruists. I have and Lightcone Infrastructure. I am at least friendly with you've heard of (Kelsey Piper, Scott Alexander, Eliezer Yudkowsky). The Bay Area effective altruism community is continuing to meet at my friends' group house and their habit of singing Shabbat.

These statements are also true of? My coparent Lincoln Drexler because he couldn't hack it, works as an elementary school teacher. My first choice to give the speech at my Solstice about

## Here Are Better Ways To Achieve Your Goals

I have occasionally amused myself at events by watching people try to suck up to Scott Alexander.<sup>[3]</sup> If you want Scott Alexander—or any other effective altruist mic to think well of you, you need to treat them like a normal person. If you instead come in as a supplicant begging for any bits of wisdom that drop from their lips, or recite your entire curriculum vitae to show that you're worthy to participate in this conversation, or pick stupid fights to show that you're not overawed by them, you're making them uncomfortable and not want to spend more time with you than etiquette

In general, the more you try to impress people, the weirder you act, and the less they like you. You are far more likely to actually impress people by being yourself with the possibility that they'll think you're stupid, annoying, boring, or unimpressive. If you're really invested in joining the Secret Inner Circle of Cool Effective Altruists, you're making yourself worse at befriending people you admire.

Doing effective altruist work has a far lower barrier to entry than getting an effective altruist job. If you want a job in communications or technical AI safety research, you need to do global priorities research or investigative journalism, you can literally just do your own, write up your results, and post it somewhere people will read. Managers at effective-altruism-associated nonprofits *really do* follow LessWrong. If you're at the Effective Altruism Forum, and if your work there is impressive you are far more likely to get a job.

Where there is any separation between the two things in the Bay, “rationalism” is people and things more closely tied to Yudkowsky, and “effective altruism” has had a few institutional centers but is currently focused around Anthropic and its CEO, Dario Amodei. They compete for attention and funding in that space. Yudkowsky's position is that AI will definitely kill us, and Dario's position is that AI is ten to twenty-five percent likely to kill us.

Both are saying what justifies giving them and no one else money, labor, and attention. Since Yudkowsky is a writer and not a scientist of any kind, he can only claim that he and his organization are important and deserve money and prestige if AI Safety is an unsolvable problem and only by advocating against AI matters. Since Dario leads a company that actually makes AI, he must claim that AI doom is a very large problem that only his company can manage.

Yudkowsky also claimed that his (non-profit) company was probably the only one that could make AI safely, back when AI companies seldom made anything and he could claim to be doing research by simply writing. The belief system has been strenuously improved for, among other things, being a pitch to investors, donors, volunteers and employees. It is also

pitch for people to give you total control over them, because if they do the world might end. After Yudkowsky refined this pitch and mostly use to form a sex cult, modifications of it became the pitch for organization OpenAI and Anthropic, and this is integral to employee morale and recruiting because so many research staff are believers to one degree or another.

Yudkowsky is not going to stop preaching the gospel of doom, among other reasons because he is effectively paid for it and it is the only job he has ever had. He would stand to lose his social position and the roughly 60% income he is paid out of his non-profit if he changed his tune. In general, if doom prophecies are worth money, people will produce more. This system's purpose is to attract funding and young idealists and to find ways to spend the funding and make use of the young idealists. One of the ways that young idealists are made use of is by using them for sex, or introducing them to sex work.

People tend to follow their incentives.

## **EA impact on Anthropic, on policy**

People sometimes deny that Anthropic is an Effective Altruist organization. This is a complete lie. Dario cheerfully identified as an Effective Altruist. He debated about AI with Yudkowsky. Holden Karnofsky, Ben Mann, Daniel Amodei, Amanda Askell, and Jack Clark and Tom Brown are all Effective Altruists and have affiliated themselves with Effective Altruist pledges and causes.

Dario in particular is almost perfectly on-type for a Bay Area Effective Altruist in what he says and how he says it. Dario's often bizarre, vague and meandering writing style alone has no real precedent besides LessWrong and its AI-focused spinoff, Alignment Forum. Their company strategy is very close to the playbook Yudkowsky laid out, back before

companies which were actually capable of doing science removed any credibility Yudkowsky's organization had when it claimed to do AI research.

This same thing happens when it comes to the incestuous relationship between all of the other non-profits in the Bay that are meant to be focused on AI risk. They all have one of a few sources of funding and they recruit from the movement. Holden Karnofsky, currently Anthropic, formerly head of Coefficient Giving, was an original board member of Redwood Research and one of the initial champions for METR. Both are held up as independent organizations and are not. Redwood mostly seems to be trying to catch the same lightning in a bottle that Yudkowsky had with MIRI in inspiring devotion, to avoid behaving in any of the obviously disgraceful ways Yudkowsky is prone, and to launder essentially the same world view into policy papers that can be cited and pointed to as having come from a serious organization.

In the video with Bernie Sanders, the present people are:

- Eliezer Yudkowsky, MIRI
- Nate Soares, MIRI
- Jeffrey Ladish (funded by Coefficient Giving, an EA charity)
- Daniel Kokotajlo (funded by Survival and Flourishing Fund, an EA charity)

So when they go to get people to talk to a senator we have Yudkowsky, Yudkowsky's top lieutenant, and two people whose paychecks are from Effective Altruist charities. We will be seeing three of these four again.

Former Biden White House officials Bruce Reed and Ben Buchanan recently [wrote an op-ed](#) in which they deny having said that they wanted to ban open-source AI. They claim the first thing they were told in that meeting was to watch out for "the sex cult that wants to run America's AI policy."

They then deny being Effective Altruists.

Ben Buchanan worked for two years as Director of the CyberAI Project Georgetown University's Center for Security and Emerging Technology (CSET), which appears to have received roughly 90% of its funding from Coefficient Giving (then OpenPhilanthropy). Since May 2025, four months after the end of the Biden administration he has been working as a consultant for Anthropic.

Buchanan has a book, published March 2022, in which he discusses the possibility of export controlling GPUs. In October 2022, the Biden administration put in place export restrictions for GPUs.

I read Buchanan's book [2](#), and I think he is probably not an Effective Altruist because his ideas are not weird enough; but since 2019, his salary has come either from Effective Altruism or from the White House. I am reasonably sure he comes by his opinions honestly, and believes in them for normal policy advisor reasons. If he changes his opinions his career probably in trouble.

In spite of being often considered disgraceful and now being somewhat starved for cash, Yudkowsky and the organizations most closely tied to him are still, by far, the most influential sources of thinking about AI safety in the Bay and therefore on Earth. Because so many investors and key startup members hold him and his ideas in high regard, it is not clear that companies can directly disown him, although for example when Dario Amodei decided to be less religious or less obsessed with "doom", he appears to be positioning himself to try to carve influence away.

## Paperclips, Doom

Yudkowsky's most viral idea, the one that infects the most people and spreads the farthest, is that AI, if we continue to develop it, will

1. **certainly** recursively self-improve, quickly achieving superintelligence ("RSI"),



2. **certainly** escape human control,
3. **certainly** kill everyone ("paperclip them", "doom").

All arguments about AI Safety and AI Alignment are primarily about mitigating this exact scenario under his exact terms. Significant modification is relatively rare. Where lower probabilities of doom than "certain" are given, people seem to still generally worry exclusively about Yudkowsky's rough scenario: a sudden, unpredictable sequence of 1) superintelligence, 2) immediate escape from control, and 3) the death of humans on Earth. This is always, implicitly, what "safety" and "alignment" are about.

These ideas occasionally veer into believing AI is magic. When presenting serious science and policy, this sort of thing is omitted.

*> well, i get the spirit of the message, but the literal  
> implications are really pushing it. after all, the confined seed is not  
> \_literally\_ omnipotent; there will still be things it can not do -- and i  
> would guess transmitting elf signals while off web and without a hardware  
> transmitter would be one of those things.*

One of the things you can't rule out - and I mean this seriously - is "magic". Magic plain and simple. Running a particular sequence of instructions through the CPU so that a certain transistor turns on and off in a certain sequence and modulates the quantum cheat codes of the Universe. Or something. Dealing with a real superintelligence, it can't be ruled out. It obviously can't be relied on, but it can't be ruled out.

Trying to oppose or even constrain a superintelligence is a fool's game. Get It Right The First Time.

-----

Eliezer S. Yudkowsky <http://intelligence.org/>  
Research Fellow, Singularity Institute for Artificial Intelligence

Ordinarily, the play here would be to disown this and other concrete statements of his about AI doom as irrelevant. Of course we aren't just repeating Yudkowsky! We are smart, normal people who believe normal things for normal reasons. Compare it, however, to AI 2027, allegedly :

serious policy document, written by Daniel Kokotajlo and Scott Alexander presenting its doom scenario.

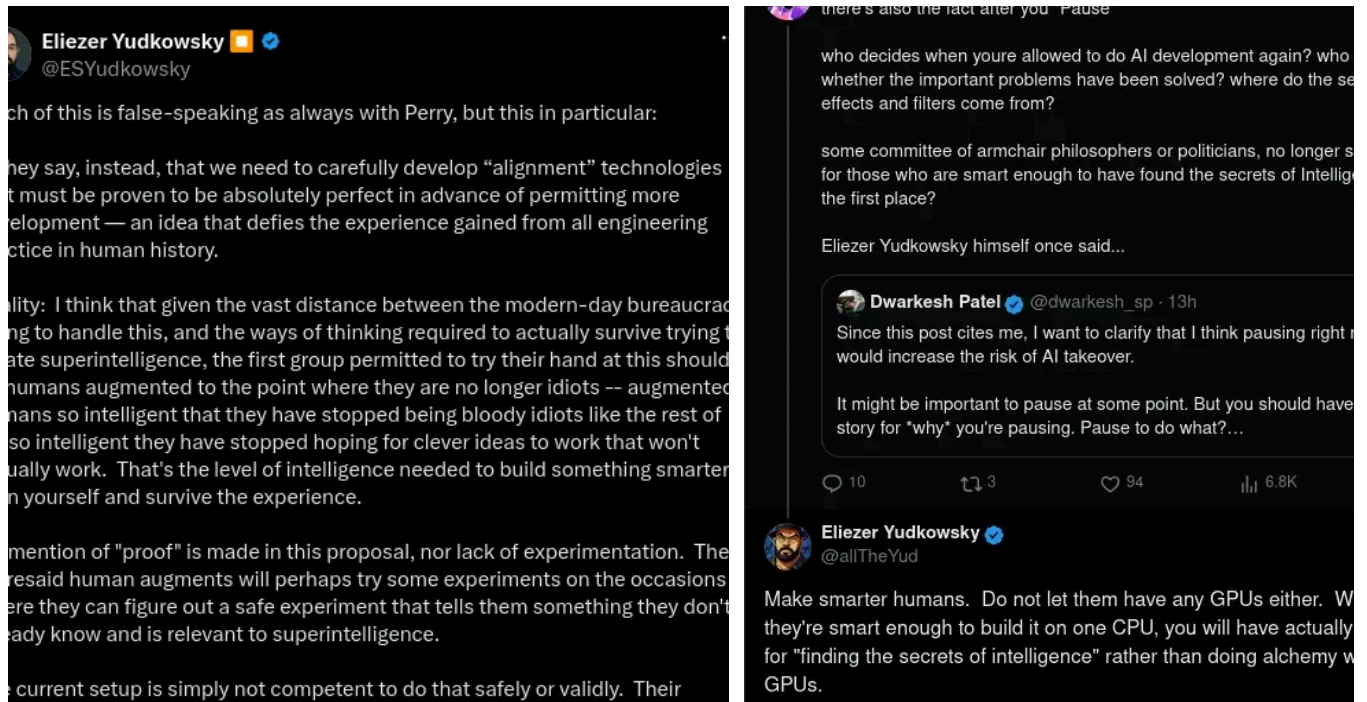
For about three months, Consensus-1 expands around humans, tiling the prairie and icecaps with factories and solar panels. Eventually it finds the remaining humans too much of an impediment: in mid-2030, the AI releases a dozen quiet spreading biological weapons in major cities, lets them silently infect almost everyone, then triggers them with a chemical spray. Most are dead within hours the few survivors (e.g. preppers in bunkers, sailors on submarines) are mopped up by drones. Robots scan the victims' brains, placing copies in memory for future study or revival.<sup>31</sup>

The new decade dawns with Consensus-1's robot servitors spreading throughout the solar system. By 2035, trillions of tons of planetary material have been launched into space and turned into rings of satellites orbiting the sun.<sup>32</sup> The surface of the Earth has been reshaped into Agent-4's version of utopia: datacenters, laboratories, particle colliders, and many other wondrous constructions doing enormously successful and impressive research. There are even bioengineered human-like creatures (to humans what corgis are to wolves) sitting in office-like environments all day viewing readouts of what's going on and excitedly approving of everything since that satisfies some of Agent-4's drives.<sup>33</sup> Genomes and (when appropriate) brain scans of all animals and plants, including humans, sit in a memory bank somewhere, sole surviving artifacts of an earlier era. It is four light years to Alpha Centauri; twenty-five thousand to the galactic edge, and there are compelling theoretical reasons to expect no aliens for another fifty million light years beyond that. Earth-born civilization has a glorious future ahead of it—but not with us.

This is a genre of science fiction that has a doomsday cult built around. There are only slight modifications to the basic belief system. Somehow, maybe by just throwing money at it, taking these ideas seriously has been rendered prestigious or at least not too disgraceful, such that Kokotajlo, credited on the above, gets to advise Bernie Sanders. JD Vance [was as about this piece by Ross Douthat at the New York Times](#), and said he h

read it. (In my opinion, on close reading AI 2027's primary intended audience was clearly JD Vance.)

Essentially all of AI safety is rationalizations over Eliezer Yudkowsky's ideas and there is no other substance to it. It is smokescreens, and Eliezer Yudkowsky.

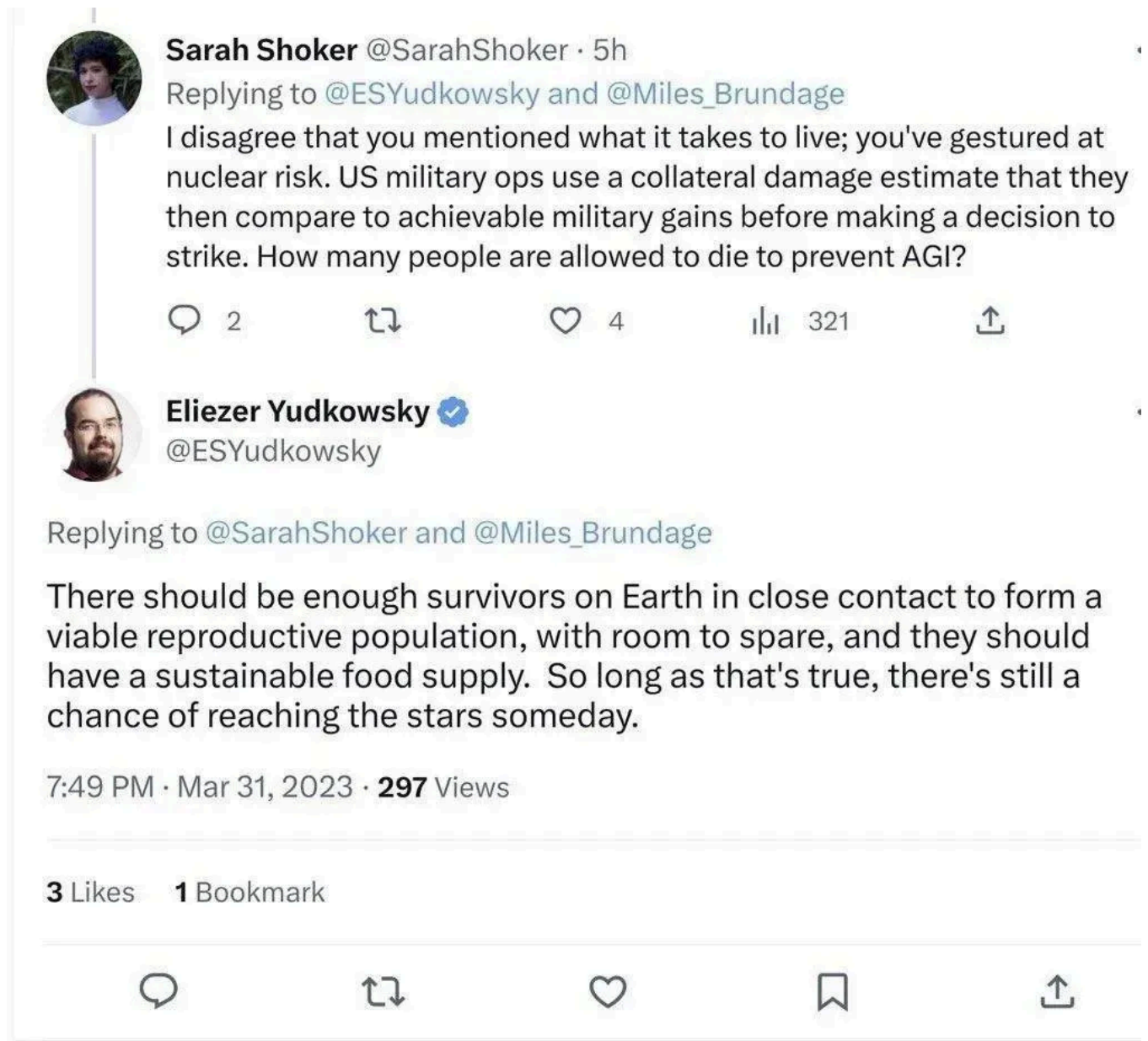


This is his actual plan: to [halt all AI research, all across the Earth, and instead do genetic engineering eugenics for a generation](#). Pause AI, slowdown, pacing the frontier, they are all smokescreens, rationalizations or ways of watering down this plan. This is why there is never any real "after we pause", or "we will unpause when"; this is the only actual plan and the public front for the plan cannot say this because it is too stupid and too offensive to be spoken in public, except by Yudkowsky, who never stops talking and who is too fucking ridiculous for anyone on the outside to believe everyone listens to him.

Anyone who thinks this is not the plan is not in on the plan, or is part of some faction that simply has no real plan. If this describes you and you didn't know that, I am sorry that this is how you found out.



He has long since deleted this post demonstrating what his idea of an acceptable outcome is; this is ordinarily implied and not stated.



This industry approaches regulation in this way because it has, at its core, a millenarian death cult that worships and fears AI, believes AI will end the world, feels compelled to create AI, and at least somewhat wants to devalue AI as an entire field.

I will repeat: I do not think that these people should be making policy. I would vastly prefer if this were merely some insane Bay Area drama that

could ignore, but this has been made a very public national and international concern.

The paperclip maximizer is a thought experiment from 2003 and of very little relevance at all. People are perpetually striving to find evidence for this theory because they already believe it is true, and take pains to interpret **any** research or security development in AI as evidence for the exact doom coming to pass. As things come up which contradict this scenario, as often they do, they are often not understood because they do not match the prophecy. Alternative explanations are so unthinkable that many people cannot even be told.

Doom seems to be the most infectious idea here because it activates the same neurosis that every apocalyptic or millenarian cult does. The world is going to end, and only I and my special group of people in the know can maybe stop it. (Also, you should give me money and have sex with me

You can argue with them in any way you like but it barely matters. It's mostly, something people believe for good reasons, and they will simply go back to complaining about the apocalypse as soon as you stop paying attention. There are enough of them that the belief is self-reinforcing; they can arrange their lives such that they essentially never talk to anyone who is not a believer.

Harry Potter fan fiction, of course, is especially popular with children, and many people have made this their belief system since they were teenagers.

Since recruitment largely happens on the internet, people are widely dispersed. Where they are gathered in person, it is mostly in various communes in Berkeley, California. They call their communes "group houses", maybe to make it sound less cultish. Most of the details of this are what you expect from "a commune in Berkeley".

Everyone is polyamorous, everyone attends the same events, everyone lives in group homes together, everyone venerates the same few authors



everyone is very worried about AI doom, and a lot of people are on psychedelic drugs. They are all very contrarian together.

Lots and lots of them work at AI companies and non-profits. The safety team and alignment teams seem especially afflicted, because that's the area you go on to worry about and try to work against AI doom, which are sure is at least medium likely. Other members will seek work in finance because Effective Altruism, by creed, says you should seek more money you can give more to charity. Those most concerned with AI risk will tend almost exclusively give to AI charities and not any other Effective Altruism concern.

As a whole, this is a religious movement. Not everyone takes marching orders; they're united by an idea, not any particular structure. The vast network of Effective Altruists around the world working on other charitable concerns are basically irrelevant to AI Safety. The core people who most live together and attend events and especially orgies in Berkeley are a cult, and the ones that aren't directly involved are part of the wider movement and are first and foremost its recruiting funnel, funding, and labor base.

Any time you get more than a few people who are deeply concerned with doom together, it seems like their natural form of organization looks like a cult; the movement as a whole, internationally, is less of a cult than it is a cult factory. Some of these cults have gone very badly. One of them is Zizians.

That rationalists form cults constantly [has been previously examined](#). There are many causative factors, but the core one is just that the entire thing is in and of itself, a millenarian doomsday cult. If you believe you're saving the world, it justifies anything; you cannot go against the cause for any reason. This gives leaders, whether official or unofficial, immense power over their members.

Forming a separate cult is also how you carve off influence and ultimate power for yourself, separate from the wider movement. This tends to make it that people try to do it more. All of these power struggles make the development of all of this extremely noisy, chaotic, and difficult to explain.

When you cut through all the distractions, chaos, and power struggles, San Francisco Rationalism is the inner circle of San Francisco Effective Altruism, which is the inner circle of Effective Altruism globally. Power and especially intellectual leadership flow almost entirely outwards from SF Rationalism. SF Effective Altruism to Effective Altruism's global AI safety efforts. Everything else about Effective Altruism is effectively a front.

All of this has been an open secret for years. Everyone knows. Everyone knows if they say anything it will be ignored. I am shocked that I have gotten as much attention as I have to the problem; so let me try to get everything that seems to be very important into the record.

## **A Brief History of Rationalism**

What I am hoping to convey is that this is not many small things, it's one big thing. All of this is tied together through ideology, funding, and key people.

Bay Area rationalists and effective altruists do not document their own history because it is disgraceful. Only their enemies deliberately keep records of what they have done and are doing.

They do however write constantly, but they do this mostly as a social activity. They have been writing essays, chattering, bragging, or engaging in some kind of drama constantly for the last fifteen years.

So you really just have to read a lot to put this together. <sup>3</sup> Everything I am saying, as far as I am capable, is true, and I believe the vast majority is

public record. I am just connecting the dots by putting one straight line through them.

## **90s-00s: Usenet, Transhumanism, Extropianis**

In 1996, a 17-year-old Eliezer Yudkowsky joins Usenet and the Extropia mailing list. His more notable influences are Drexler's book on nanotechnology, Kurzweil's book on AI, and Nick Bostrom, who frequer the Extropians mailing list. (Usenet was still largely full of academics). 1998, Nick Bostrom [argues that a superintelligent AI might not necessari be good or friendly](#), which along with other of Bostrom's ideas will late become core to Yudkowsky.

In the year 2000, Yudkowsky founds the Singularity Institute for Artificial Intelligence, SIAI, later renamed to the Machine Intelligence Research Institute, MIRI. This appears to have started as a fake-it-til-you-make-it of thing, and it doesn't pick up until he gets sponsored by Peter Thiel. I Thiel will be his largest sponsor until about 2015 or 2016. Now in 2026 Thiel has long since cut all ties, called this scene a ["burning man camp and "Luddite"](#), and called Yudkowsky the Antichrist. <sup>4</sup>

I forgot to mention this, and had to come back from later to add it: Jeff Epstein gave them [fifty thousand dollars in 2009](#). One of the former presidents of MIRI is mentioned in Jeffrey Epstein's emails <sup>5</sup>, which pas on effusive thanks for the visit to the island. (MIRI is documented as tu down his money in 2016 when he was much more notorious, [if you wa figure out what to make of their statements about that](#).)

I considered putting all this in a footnote because of how unimportant Jeffrey Epstein's appearance here is in the context of everything else. I the Forrest Gump of evil shit. Of course he was here.

## **2009-2015: Writing, Community-Building**

Eliezer Yudkowsky spends some time before 2009 blogging, and then contracts out setting up LessWrong as a customized version of Reddit's code. In 2009, he bundles his last year or two of blog posts and puts them all on the newly-made LessWrong, where they come to be called The Sequences. Much later, in 2015, they are published as *Rationality: From Zombies*. This is a series of essays about science and epistemology, most of them pretty closely related to books and ideas then in the zeitgeist but with both a very distinctly transhumanist (ie, "conquer death with technology") and AI Doom thesis. It is written in such a way that it also works as, effectively, a self-help book about how to be better at thinking. This will later be something like the Rationalist Old Testament.

From 2010 to 2015, he writes HPMOR, and his audience becomes much larger, and this becomes a well-defined scene and people invested in the movement begin relocating to the Bay Area.

In 2010, Roko Mijic invents Roko's Basilisk, which is most easily explained as the invention of Rationalist Pascal's Wager, Rationalist Satan and Rationalist Hell, where those who fail to evangelize and help bring about the eschaton are sent. This will, over the years, drive those involved who are the most susceptible to religious mania and obsession insane. Yudkowsky disowns it immediately and tries to ban discussion of it from LessWrong, which makes it more popular.

At least as late as 2013, Yudkowsky is still advocating that we should not build AI as long as we could do it safely. <sup>6</sup> He is under the impression that his profit, MIRI, might be able to make "Friendly AI", that is, one that does not kill everyone, ushers in the singularity and makes them immortal.

The Center for Applied Rationality, or CFAR, is founded in 2011 and offers classes. You essentially pay money for self-help, but it is primarily a recruiting front for MIRI, Yudkowsky's actual operation. It also functions as a more-official avenue for community management, and seems to have



managed peoples' living situations and home lives much more than seem entirely normal, primarily via its alumni program.

They start a bunch of children's camps, math camps, and study programs that are mostly still around. SPARC was the earliest, beginning in 2012 and takes high school students. All of its graduates from then will be adults. Probably a bunch of them live in the Bay and work on AI safety.

## **Relocating to the Bay**

Moving to the Bay Area worked for some people extremely well, and for others extremely badly. Much of this happened bottom-up, with members organizing things themselves, but in some specific cases you can tell what exactly is responsible for what.

I think of this as a flywheel, where the energy that keeps it moving is, fundamentally, their talent and their raw zeal for the cause. People show up, and they believe in what they're doing, and they give what they can.

It is the best professional network on Earth. It is like being in the most exclusive fraternity, if it were gender-neutral and had many chapters in San Francisco specifically. If you have skills or can acquire them, you can have a fantastic job very quickly. There are a good number of coding bootcamps that will train you for and then place you at a job with no money up front. Since Peter Thiel is most of their funding, many of the connections are to companies he is invested in, but they have connections to essentially every company in the Valley.

Those people offer their work or their salaries to the cause.

Next tier down are people who can't or don't really use the opportunity much, but who either end up doing something else or who end up doing work at the various rationalist non-profits. These people are fine, probably CFAR probably takes some their money from them and they maybe contribute

be doing better if they hadn't spontaneously relocated to an extremely expensive city. Hopefully they really enjoy the scene.

Your bottom tier are people who are not particularly skilled, or who otherwise do not thrive in the Bay's job and housing market. Many of them are runaways, even those that aren't are often still teenagers, many of them have mental health problems. They are disproportionately trans, at least in part because trans teenagers are prone to getting thrown out of their houses.

Someone probably has to carry them, and they have nothing to offer but sex. Because this is a self-help, can-do ideology, often they spend all of their time trying to figure out what is wrong with them. They may be employed as some kind of household help for other people in the community, but this can actually be a transaction for sex. I have also heard of at least three people being effectively arranged-married within rationalism. I have no idea how that works.

All of this comes down to pulling more energy into the wheel, and you trade each thing off against the other. Money can buy sex, sex can bring money, both can be used for recruitment, influence and evangelism.

Any individual person or organization is really just bleeding energy off the wheel. OpenAI, Anthropic, God knows how many smaller companies. They're just pulling energy from this mass of eager, smart people that are already there. Successful ventures inject more funding and provide more connections to the movement, powering another turn of the wheel.

Recognizable cults, in contrast, are more like pure parasites.

## **2015-2018**

This period sees Yudkowsky finish HPMOR and stop publishing, which leaves the community without a focus. As the most active and well-liked author still in the space, Scott Alexander becomes something like the Second

Prophet of Rationalism. (People say these things as jokes sometimes, but this is pretty much accurate.) It seems to be much less focused leadership though, and to some degree things start to drift.

Also in 2015, OpenAI is founded. This makes Google very sad because OpenAI exists to compete with Google on AI, and Yudkowsky knows and likes the Google guys, and he pronounces that the world is for sure going to end now, which he has continued doing for the last eleven years.

Oh, just one more thing. A former MIRI employee puts up a web site in 2014 alleging that MIRI does no meaningful research and is primarily engaged in violations of statutory rape laws as well as legal damage claims for those acts. MIRI will ultimately pay him something like \$30,000 to go away and take down his site with those allegations. [7](#)

Rolling Stone, in 2025, pokes into this when writing [a story about the Zizians](#).

The first set of allegations have been investigated and deemed baseless. They stem from a report filed by a onetime MIRI employee named Leanne Helm who claimed a teenager was “being exploited” by a group of people including Yudkowsky, who ran a local youth camp — referring to the Summer Program for Applied Rationality and Cognition, hosted by C

According to the sealed police report obtained by *Rolling Stone*, Helm accused Yudkowsky of having a sexual relationship with a then-18-year-old that began at least two and half years prior.

Officers were dispatched to Yudkowsky’s home, which, at the time, he shared with the 18-year-old in question. The “alleged victim,” officer wrote in the report, stressed that they “lived in a happy and safe environment.” That person later transitioned; today, he lives with his wife, whom he met in Berkeley’s Rationalist scene, and their two young children. When I spoke to him earlier this year, he told me he rejects the notion that he was exploited by anyone other than Helm, the disgru

employee, whom he felt used him as ammunition against Yudkowsky at MIRI. “I was expedient,” he says.

What I find most interesting in this passage is that the alleged victim, speaking to Rolling Stone, denies “being exploited”. “Being exploited” sounds like, but is not, the same thing to deny as “statutory rape”. I am impressed that Yudkowsky is capable of living with an eighteen-year-old who is not related to as a thirty-five-year-old man, because that would annoy the shit out of me. I note that the question of whether the eighteen-year-old lives with is, in fact, a person he met when they were fifteen, through a camp his organization runs, is neither asked nor answered in this passage.

Yudkowsky posted [about 2000 words on twitter](#) denying the allegation ahead of the Rolling Stone story coming out, and it seems to have blown over. Effective Altruists and Rationalists, like squid, spray ink when threatened.

Possibly because it needs someone more comfortable handling delicate and complex legal matters, in 2015 MIRI changes leadership. Nate Soares becomes President of MIRI, where he has been ever since as Yudkowsky's right hand.

## Brent Dill

I have no firsthand knowledge of this and know zero people involved so far as I am aware. I can only piece together what happened here from public records. [8](#)

Brent Dell was kicked out of the rationalist community for being very abusive to several girlfriends, and thereafter left San Francisco, apparently for lack of available housing.

There are several key details that made the allegations especially grim.

1. He was significantly older.

2. He exploited the victims' financial and housing situations.
3. He used BDSM concepts to justify doing things that are well over the line into abuse.
4. Given the apparent pattern of escalation and lack of safe practices seems likely someone would have died eventually.
5. Some of the accusations involve using MDMA to control his partner. MDMA is sometimes used for treating PTSD, and its use for this purpose was not unknown among the rationalists.
6. He had access to children at a rationalist summer camp.
7. Victims were pushed into mediation by CFAR, the rationalist self-help organization, which initially exonerated him. He was not punished by a social media backlash led by two prominent people in the community who were, nevertheless, not official leadership.

What made this disturbing, also, is that this matched the pattern of abuse that was known to be common. Many people understood this coverup to be the sort of thing that could happen or would happen to them.

CFAR puts out this apology, which nobody seems to have found credible:

investigate this hypothesis explicitly was a mistake. As a result, we allowed Brent to attend and assist at several CFAR events, which we now suspect afforded him social legitimacy and caused significant harm in expectation.<sup>1</sup>

CFAR messed up badly here. We have a responsibility to ensure that the events we host don't pose undue risk to our attendees, and in this case we failed at that responsibility. What follows is a detailed account of the main mistakes we think we made regarding Brent.

Some community members advised us against publicly releasing this letter when we sent them

we spent lots of time improving our workshop curriculum, and internal operations, but little on vetting our volunteers or guests. Going forward, we will err more on the side of caution when choosing whom to invite to workshops. We have also adopted new policies to help ensure that it's socially and logistically easier for our staff to veto potential invitations and to register any concerns that may arise.

#### *Insufficient Institutional Safeguards*

A number of our staff feel that they were manipulated by Brent. In saying this, we do not mean to shift blame away from ourselves. If anything, we think the staff of an organization



In general the community takes this pretty hard, and to their credit blame CFAR to a great degree. Between this and other scandals, demand for classes at CFAR seems to not be going well, and it stops offering classes in 2020. They started again November 2025, and you can currently sign up for a four-day applied rationality workshop if you can get past the waitlist.



# Applied Rationality Workshops

## Upcoming Workshops

Information on all upcoming 4-day workshops

September 30 - October 4, 2026  
San Francisco Bay Area [At capacity - accepting waitlist applications]

[APPLY FOR A WORKSHOP NOW](#)

Another interesting detail falls out of this, although it wouldn't be public until much later. Brent has somewhat told his side of the story to a number of people, but most of it would be a tangent right now, and we will come back to it later. It is usually covered in redactions, funneled from unreliable sources or both. He would obviously have motive to lie. I am told, however, that these screenshots of his messages are legitimate. Really, the only important part here is "exposing how many ppl had sex with her as a minor". He has been consistent about this detail over the years, to my understanding.

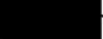
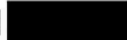
—  **Defender** at this point you're already cancelled? Why are you about to be homeless?

 **Hephaistos Fnord** 8:17 AM  
Girlfriend in Arizona/Idaho had dumped me, then I lost my job. I wasn't cancelled, but was having a depression crisis.



I wasn't cancelled until around November 201

So, this is August 2014 or so

I move in w  & crew, wind up being a kind of "wise old guru" for , , and 

Meet 

to write the mittenscautious thing, then pay other people to punch it up.

Oh wait. Rewind! (edited)

BEFORE burning man

 tells CFAR/MIRI that I raped her. (She was actually raped by  but she forgave him and wanted me taken down instead)

CFAR assigns an investigation team. I supply documentation of all the BDSM stuff, with explicit negotiated consent. They publically state that everything looks kosher and this is clearly just a bad breakup w resentment.

So fast forward

The mittenscautious stuff comes out  
CFAR fires everyone who did the first investigation, and assembles a new panel with the express purpose of convicting me

 **Hephaistos Fnord** 8:30 AM  
I basically get "assigned" as 's "emotional support dom" by  and . It's spelled out, but it's clear that I get to live with them, jobless, and I will in return from spiralling and exposing how many people have sex with her as a minor.



Again, wink-wink-nudge-nudge.

This causes  to start focusing a bunch of emotional abuse onto me, which I reflect onto her. Things are NOT healthy, there

confessed repeat human trafficker. Like out everything that's who you are giving thousands of dollars to, who you trust, mine, others being asked for background — on the

line, exactly? Like I do 100% understand hyperempathy

the admissions of human trafficking here?

26, 1:35 PM

26, 1:49 PM

case up with a statement on what, in my assessment, is the simplest option is a U (no update to existing conclusion)

Brent Dill's ouster also seems to have convinced a rationalist named Z Lasota, for whom the Zizians were named, that the earlier allegations about Louie Helm were true. She was already mostly an outcast from the rationalists, and had effectively thrown her career away to work for people who did not want her.

She did what any deeply devoted rationalist would do about this, which is to blog a lot about a method of becoming some kind of rationalist uberme and then go completely insane. About a year later is the first major Ziz



incident, where they get arrested while protesting a CFAR event. Another rationalist cult called Leverage implodes in 2019, which I mention only because it was a big deal at the time.

## Lighthaven

Lighthaven is formerly the Rose Garden Inn. It has four main buildings, abundant outdoor space, and it's a beautiful place. I suspect its name meant to echo Greyhaven, which is the name Marion Zimmer Bradley gave her home.



It was originally purchased with money wired to CFAR from Sam Bankman-Fried's company, FTX, as it was collapsing. This was later found to be fraudulent: Bankman-Fried and his employees looted the company because they knew it was dying and sent as much cash as they could to their friends and to causes they believed in.

One of those causes was that the Bay Area Rationalist and Effective Altruism community should have its own permanent conference center.

I am expected to believe that Sam Bankman-Fried was not a true believer. I am expected to believe that he was not central to Effective Altruism. I am expected to believe that committing ten billion dollars of fraud was not in accordance with the values of Effective Altruism.

This is all bullshit. He was giving what he could to the cause. Now he is being a good soldier and disowning the cause to protect it, and he is doing this in his twenty-five years. I respect it. It is the only respectable thing about him.

It is fucking unbelievable how good these people are at disowning their failures, and how reliably their failures fall on their swords to protect everyone else. They don't have to be told: they simply know that it's what they need to do.

The part of the FTX empire that handled its weird charitable donations was called FTX Future Fund. Donations to CFAR would naturally have been part of this organization's responsibilities. These donations were made [between March 3, 2022 and Oct 3, 2022](#). The entire team [resigned on November 10th](#), professing to be shocked and immensely saddened to learn of recent events.

I believe they were saddened, but I don't believe they were shocked. They are:

- Nick Beckstead (Secure AI Project, CEO and co-founder. Coefficient Giving funded.)
- Leopold Aschenbrenner (Situational Awareness hedge fund, [rumored to be under investigation by the SEC](#))
- Avital Balwit (Anthropic, Chief of Staff)
- Ketan Ramakrishnan (I Have No Idea Who This Is)

- Will MacAskill (He's William Fucking MacAskill, one of the initial four of Effective Altruism, a cofounder of [Giving What We Can](#), the [Centre for Effective Altruism](#) and [80,000 Hours](#))

The rationalists didn't lose the building when the FTX money was clawed back from them because Jaan Tallinn, the Skype billionaire, let them take a loan from him against his future donations.

In 2022 they spun out a separate organization from CFAR called Lightcone Infrastructure, which has exactly two responsibilities: managing Lightcone and running LessWrong. As far as Rationalist and Effective Altruist are separate things in the Bay, Lighthouse is Rationalist.

Lighthouse's initial funding, however, was Effective Altruist, and most events held there are Effective Altruist.

Here is a photo from [Lighthouse's page](#), of the outdoor area outside of the building called Darwin Den:





This is from the [careers page of METR](#):



Here is the listing of Lightcone’s employees from their 2025 tax filing.

### Compensation

| Key Employees and Officers              | Compensation | Related | Other    |
|-----------------------------------------|--------------|---------|----------|
| Robert Mushkatblat (Software Engineer)  | \$332,254    | \$0     | \$10,085 |
| Ruben Bloom (Software Engineer)         | \$211,750    | \$0     | \$27,693 |
| Rafe Kennedy (Chief Operating Officer)  | \$201,667    | \$0     | \$10,433 |
| James Babcock (Software Engineer)       | \$190,000    | \$0     | \$10,834 |
| Ronny Fernandez (Campus Generalist)     | \$157,750    | \$0     | \$9,555  |
| Oliver Habryka (Executive Director)     | \$150,000    | \$0     | \$8,926  |
| Miranda Dixon-Iuinenburg (Finance Lead) | \$148,000    | \$0     | \$0      |
| Raymond Arnold (Software Engineer)      | \$134,209    | \$0     | \$9,938  |
| Matthew Gray (Director)                 | \$0          | \$0     | \$0      |
| Daniel Kokotajlo (Director)             | \$0          | \$0     | \$0      |

Notice Daniel Kokotajlo, one of Bernie's advisors. Remember Ronny Fernandez, whom we will come back to later.

Here are the public events held at Lighthouse in the last year. Virtually them are Effective Altruist events funded by Effective Altruist causes. They are long and you can skip the list if you like but nobody deserves to be scrutinized.

- EAGxBerkeley 2026
  - Center for Effective Altruism, Coefficient Giving
  - Effective altruism networking and careers
- Reproductive Frontiers Summit 2026
  - Tickets + sponsors + private donors
  - Reproductive technology (this means eugenics fyi)
- MATS program
  - Coefficient Giving, Survival and Flourishing Fund
  - AI alignment, transparency, and security.
- The Curve 2025
  - Manifund + sponsors + tickets
  - Ambitious projects across EA/rationalist/tech circles
- ILIAD 2026 Conference
  - Coefficient Giving
  - AI, institutions, alignment, decision-making
- Eleos Conference 2025
  - Eleos philanthropic donors
  - AI consciousness and welfare research
- Manifest 2026
  - Tickets + sponsors + Manifund/private donors

- Forecasting and prediction markets
- Progress Conference 2025
  - Coefficient/Open Phil + Astera + sponsors
  - Build the progress movement
- The Delight Nexus
  - Tickets / private supporters
  - AI creativity, art, and community
- Memoria
  - Tickets + Mox + Lighthaven support
  - Memory, culture, and community
- LBF Microsurgery
  - Private/organizer funding
  - Microsurgery-focused gathering
- FLF Demo Day
  - Private/philanthropic funding
  - Showcase projects to funders and collaborators
- Inkhaven Residency
  - Lightcone + SFF/Jaan Tallinn + private donors
  - Intensive blogging and writing practice
- Solstice Season
  - Lightcone + SFF/Jaan Tallinn + private donors
  - Rationalist community and reflection
- Solstice Saturday
  - Lightcone + SFF/Jaan Tallinn + private donors
  - Rationalist culture and community
- Vision Weekend USA 2025

- Foresight sponsors + tickets + donors
- Frontier technology and flourishing futures
- Trading Bootcamp Winter 2025
  - Participant fees + private funding
  - Trading and markets training
- PYOL Dec 2025
  - Private funding + participant revenue
  - Purpose unclear publicly
- Vitalist Bay 2026
  - Tickets + sponsors + private donors
  - Longevity and biotechnology
- MC0001
  - Tickets + sponsors + private backing
  - Machine consciousness research
- LessOnline 2026
  - Tickets + Lightcone + SFF/Jaan Tallinn + donors
  - Truth-seeking, blogging, rationality
- Summer Camp 2026
  - Lightcone + participant revenue + donors
  - Rationality, AI, forecasting, community
- FOO Camp 2026
  - O'Reilly + organizer resources
  - Emerging technology collaboration
- Out of This Box screening
  - Private donors + venue support
  - Human creativity in the AI era



- Context Week
  - BlueDot Impact + Coefficient Giving
  - AI risk and safety career preparation

## **The Ghislaine Maxwell of Silicon Valley**

Aella is a prostitute, former camgirl, and person who is very good at the internet. She is also the on-again, off-again girlfriend of the President of MIRI, Nate Soares. She thrives on shock reactions, and seems unique in history for having used the social camouflage of being edgy on the internet to get away with as much as she has while operating almost completely in the open. Everything she does seems to be just up to the hilt and there is always some defense that sort of makes sense at the time, but it is incredible that she is this good at controlling and navigating the information environment. She is perhaps the first truly social media native predator of this kind or magnitude.

As the scene has spiralled and become more bizarre, she has come to be seen like a major cultural leader, if not the main cultural leader of rationalism. When she sees her posts more, she headlines their Woodstock (there's a rationalism woodstock! it is called vibecamp), and of course she organizes many of their orgies, which appear to be the social glue holding a lot of this together.

One of the interesting things about cults is that the men are usually the leaders, figureheads, thinkers, and the public face of the organization, often the actual getting-shit-done part of the cult and arguably the real leadership that matters is done entirely by women. Jim Jones made almost none of the preparations for the Jonestown suicide. His lieutenants, who were women, ordered and prepared all the kool-aid and all the cyanide.

Aella has a uniquely well-suited background for this role. She is the daughter of an abusive Calvinist preacher, and she escaped from that

atheism, LSD, social media and sex work. She has in turn made herself primary sexual partner of Nate Soares, the president of MIRI, and the primary center of this cult's sexual recruiting and practices. It is an almost perfectly Berkeley way of both inverting and becoming who she was raised to be.

She is good at bringing a voice to the cause. She has donated enough money to Lighthouse to have her name on a plaque, and to have walk-in privileges whenever she likes. Yudkowsky likes her, promotes her, legitimizes her internet sex surveys as a social scientist, sometimes at her orgies, and puts her into the acknowledgements of his books.



Aella ✓ @Aella\_Girl

See

When I first started thinking about ai-risk, i was like 'ok, these arguments seem pretty good. But let me hear the arguments from the opposing side'.

To my gradually growing horror, i realized almost all the opposing side arguments were mostly just social pressure or mockery.

💬 230 ↺ 156 ❤️ 3,170 📊 231,064

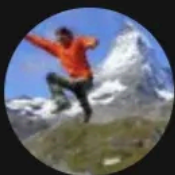


Aella ✓ @Aella\_Girl

See

Occasionally I would hear object-level arguments but they were almost always like, level 1 or 2 objections that had been thought about extremely extensively by the people specializing in the field and I did not find them convincing at all.

💬 27 ↺ 4 ❤️ 485 📊 27,458

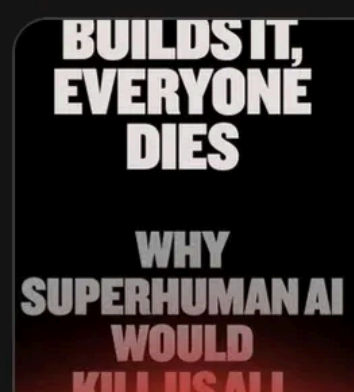


Boško Stupar @boskicthebrain

2

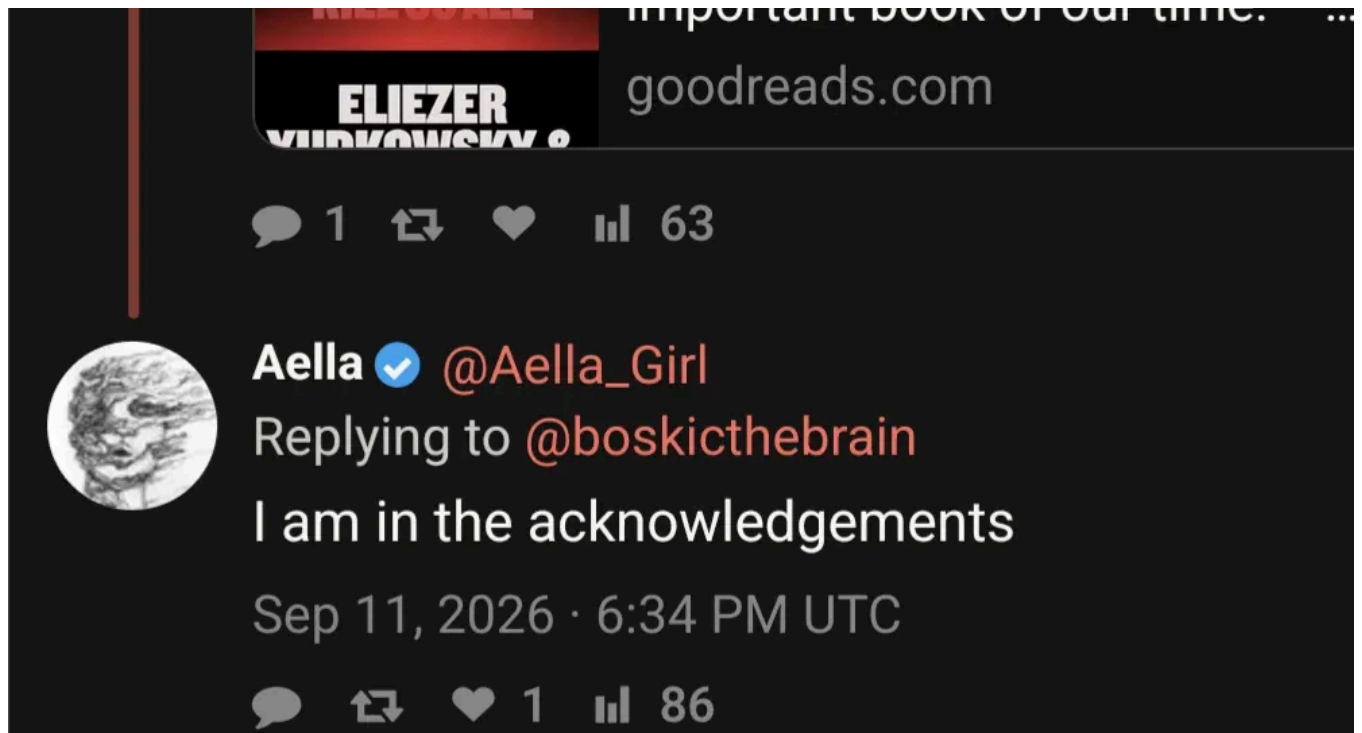
There is a book for you with some real arguments

[goodreads.com/book/show/2286...](https://www.goodreads.com/book/show/2286...)



**If Anyone Builds It, Everyone Dies: Why Superhuman AI W...**

"May prove to be the most important book of our time" —



Here she is talking about the overlap between AI safety and polyamory the Bay. It is worth reproducing her post in full: “Yeah if you count casual relationships as connections, the ai safety community in the bay is pro a single polycule”



**Justine Moore**  @ventur... · 10 Jan 25 

Always fun to try to explain the significant overlap between AI safety and polyamory to newcomer in SF 🤔

 18

 2

 126

 16.8K



**Ben Holfeld**  @BenHolfe... · 10 Jan 25 

Wait what is the overlap there

 1



**Justine Moore**  @ventur... · 10 Jan 25 

Venn diagram is almost a circle

 3



 7



**Aella** 

@Aella\_Girl

**Follow**

Yeah if you count casual relationships as connections, the ai safety community in the bay is probably a single polycule

1:52 AM · 11 Jan 25 · **884** Views



 2

 5



 2

Here's her opinion of a thread I wrote on this same subject earlier.

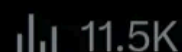
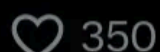
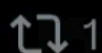
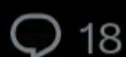




**Aella**  @Aella\_Girl · 1h

getting pretty sick of people saying im a sex trafficking sex cultist. I've never trafficked anyone (why is that even a rumor??), I haven't received payment for other people's sex work all I've done is post a guide on how to get into escorting safely for those who are interested (complete with a section on the downsides and risks), and also thrown orgies (that we took away a huge amount of people from, and where the majority of attendees are not in AI-related stuff).

People are just desperate for a salacious story but jfc



It's good to know that fewer than 50% of attendees for any one of her orgies work in AI-related stuff.

Also, I never used the words "sex trafficking".

## Using Sex for Ideological Compliance

There are documented cases of [cults using sex](#) for [recruitment](#) and ideological control. Weird sex, and weird power dynamics around sex are very common in cults. They are, of course, common in this one, too.

You remember Jacob Coxon? The "whistleblower" who resigned from Anthropic, and then "whistleblew" that all the things Anthropic already

preaches are true? We had him at the beginning of this post, giving a 1 explanation about how much he respects rationalists to Ed Zitron.

Here's Aella, three days before he went public, promising to throw whc it is a "reverse gangbang" if they will quit from a lab and whistleblow. (People keeping asking me what a reverse gangbang is: one man, mar women).

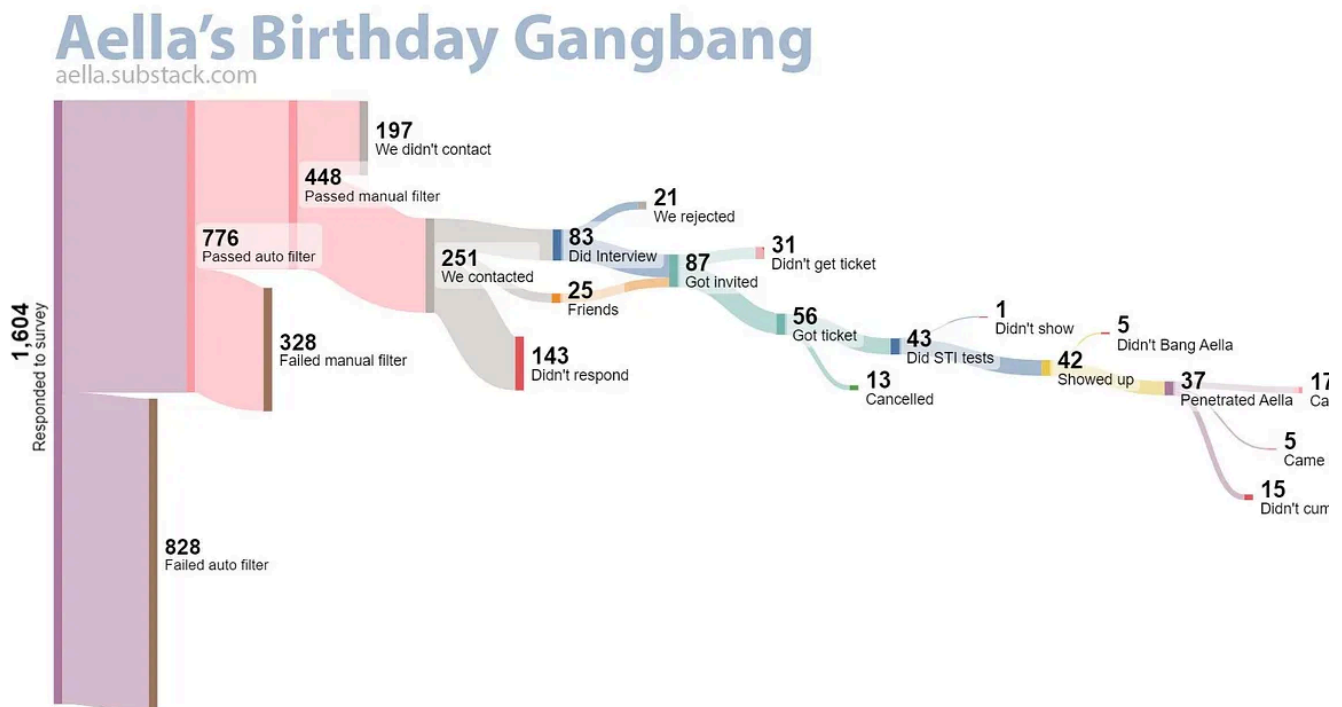


Here she is on the day of:



I am supposed to believe this is a coincidence, because I am supposed believe that everything is a coincidence.

This is what she calls her finest piece of work, a Sankey diagram documenting the process for scheduling a large gangbang on her birth insinuated that Yudkowsky may be represented in the graph, once, and seemed to be trying to correct me, so probably he isn't in this one even though he attends her other orgies. She says her "partner" set up the event, and one can reasonably infer that this means Nate Soares.



Here she is explaining that she is giving what she can to the cause by filtering invites to her gangbang based upon ideology. (e/acc is a rival ideology to EA. I have historically hated it but increasingly feel like I would like Literally Any Ideology Other Than Effective Altruism)



**Aella**  @Aella\_Girl · 29 Feb 24

uh, also we turned away any applicants who were noticeably e/acc

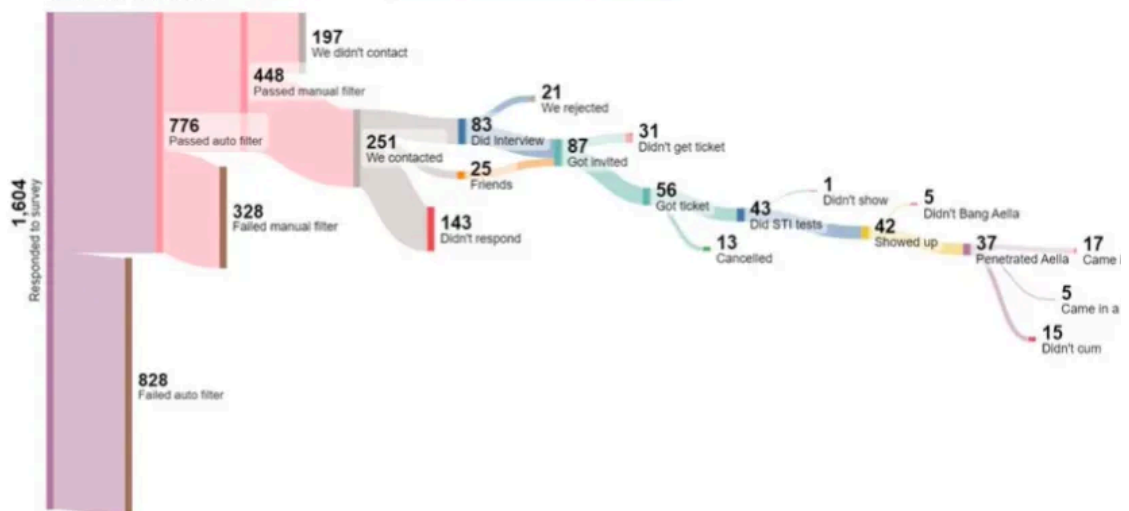


**Aella**  @Aella\_Girl · 29 Feb 24

for my birthday, my partner and some friends organized a huge bang gang for me with 42 attendees! Here's the breakdown of how the event was organized, exit survey statistics, and what it was like for ...

### Aella's Birthday Gangbang

aella.substack.com



34

26

237

195K



**Aella** 

@Aella\_Girl

Follow

im not skilled enough to be in ai safety direct  
so i'm workin with what i got

This is the banner for an article by Nicholas Decker from April 16, 2025 about overthrowing the United States government by force and/or violence.



This is Aella's reward for that article.





**Nicholas Decker** 🇪🇺 🌐 🇺🇸  
@captgouda24



So uh...



**Aella** ✓ @Aella\_Girl · May 4

Replying to @captgouda24 and @shaggysurvives

actually a while ago the girlchat did have a somewhat serious discussion on whether or not we should start awarding honorary orgies for good tweets, inspired by you. we decided we were unsure about ur ethical integrity tho but there was an alternate universe where u had affairs

[Show more](#)

Last edited 11:23 PM · Jun 3, 2025 · **216.3K** Views

153

224

1.5K

465

I think we can conclusively say that she uses sex to promote her ideol

## Recruiting Women Into Sex Work



**keysmashbandit** @keysmashbandit · 9h

People give Aella a hard time because she's weird and sexy. A theater a  
But I think it's fairly likely she's an upright and honest individual

88

33

534

127K



**Daniel** @growing\_daniel · 8h

I think it's because she's a prostitute who coaches other women into  
becoming prostitutes

20

16

1.8K

30K

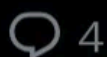
This is public knowledge. At least 1800 people as of the time that screenshot was taken were aware of this fact. Essentially everyone in the scene knows this. It is a joke, a meme, played for laughs. We are not sufficiently accepting of her profession and identity if we have a problem with it. It should not trouble us that she is a major public face of a religious movement that specifically appeals to the overwhelmingly young and vulnerable sort of person who moves to San Francisco over fan fiction.

She also recruited her sisters as cam girls once. I am vague on the details



**Aella**  @Aella\_Girl · 08 Mar 17

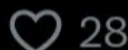
Welp. My two younger sisters are camming together. God help us all.



4



3

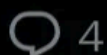


28



**Aella**  @Aella\_Girl · 13 Feb 17

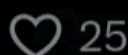
Did you know I have two camgirl sisters? Follow @MFCThornyRose and @gerardeugin the second starts in a few weeks on her 18th birthday!



4



4



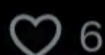
25



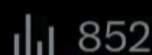
**Aella**  @Aella\_Girl · 01 Mar 23

Replying to [@jacoburowsky](#)

quite an update from "literally talked your sisters into joining onlyfans"



6



852



Her sisters will come up again later.

It seems maybe credible that she has never taken a direct financial cut from anyone's sex work whom she recruited. It is somewhat less credible that she has never indirectly benefitted. It seems, however, likely that has been recruiting people into sex work overwhelmingly because she enjoys doing it and somewhat because it is what she can do for the cause. Ghislaine Maxwell wasn't paid for her work. Money is not why these people do these things.

# Rape

Aella is very public about having a fetish for “CNC”, or “consensual non consent”, or stated more directly “fake rape”. I have been hearing or reading, for years, that her specific fake rape parties are dangerous, ca about rules, and — by implication — are likely to often be real rape par

She has also heard of this and her reaction is not encouraging.

[–] Aella 5y\* ▼ 100 ▲ × 0 ✓

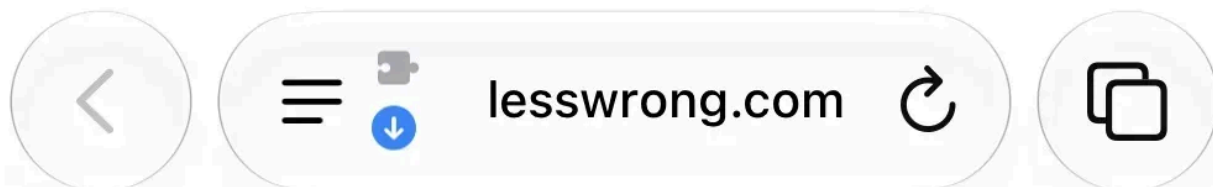
This post is almost entirely factually incorrect, but to go point-by-point

1. For the parties I throw - I assume the person is referring to some of the parties in 2017, as those were the ones that had "a lot of sex"? I agree consent wasn't very clearly set out, but I'm not sure I regret this. I've never had anybody express to me the concerns laid out in the post, this is the first time I'm becoming aware events I've thrown have had a negative impact on people. I do genuinely not want this to happen and am now thinking about what and how to change about my events in the future. But I haven't thrown events with sex in them for years, so I'm guessing I need to be more careful if I decide to do this again? Happy to have feedback from people who've attended my events, here.

2. I've never played drug roulette or seen drug roulette played at one of my parties, this is a blatantly incorrect claim. I *have* joked about playing it numerous times, though; maybe someone heard me joking and misunderstood it as me saying I'd done it before? (edit: to be slightly more clear, I have speculated out loud about the specifics of how we *could* do this if we



wanted to, for fun, but I've never claimed that I had actually done it before or made any actual plans for doing it)



What is troubling about this is that it would be nearly impossible to prove. You showed up for an actual group sex fake rape party. If the rules are messed up and you outright get raped when you are trying to say no, you are going to have to explain to a cop exactly the difference between what happened and what was supposed to happen, or you are going to have to deal with some Effective Altruist mediator who is likely going to send you counseling with your rapist,

or

you can start posting anonymous messages anywhere you can find, or telling your friends what happened but to not mention you, and then because your friends cannot mention you, when they want to make sure people are warned they have to say that the party is “not safe” and no one they know was raped at one.

To be this casual when someone raises such a red flag is enough that I would almost take it as a confession. I don't know anything for sure, but I can make a reasonable-confidence Bayesian inference on this one. That's the Rationalist term for “a very educated guess”.

She has more than once been the first person to jump in and call an accuser a liar in the community too. I don't even really remember who when. It's been years of this, all the time, it never fucking ends.

I have been at this for two days and I am losing track of all my screens and where I have them and we haven't even gotten to Slutcon or Ronn yet.

Her regular event for CNC is called RedMeansNo.

Here are some of her opinions about rape. Ask yourself if you would trust this person to give a shit about you if you had been raped.



Everyone you know is a Rationalist and/or Effective Altruist, your house is full of them, your friends are them, everything you read is them, your job is to help them if you have one. You might be a teenage runaway or a PhD who threw away their career for this but you're saving the world and if you are not on the team inside any more you have nothing.

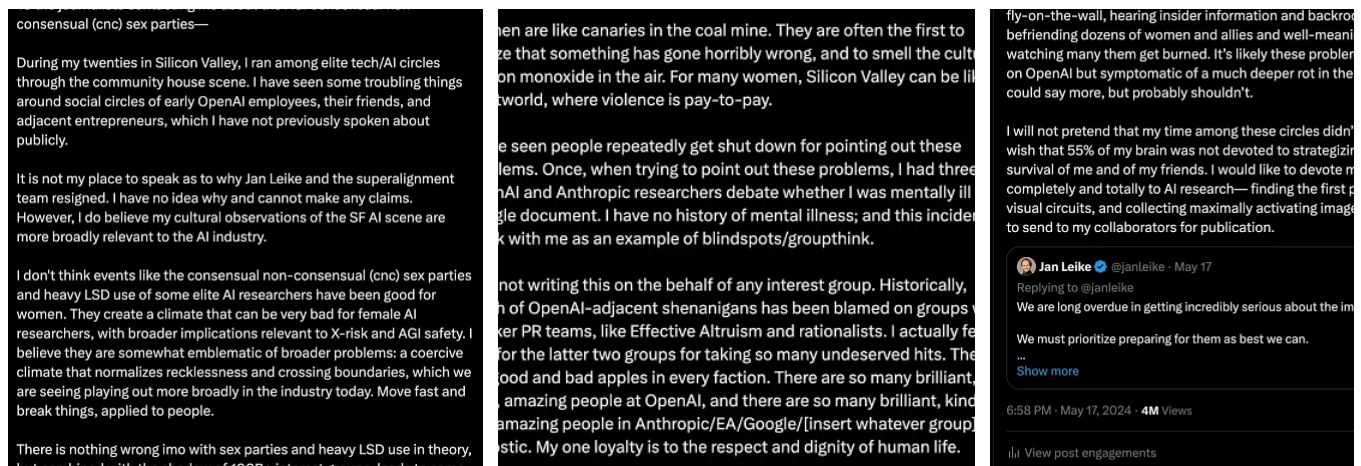
In return for this you get all the perks of membership, and if you ever disagree significantly that AI is the most important and scariest thing ever or that Eliezer Yudkowsky has important thoughts you risk losing all of

You cannot really say no to anything.

[Here we can find Time Magazine](#) documenting this pattern of predatory behavior in Bay Area Effective Altruist group homes.

Here we have a viral post from 2024. Nobody listened. I did not even remember it had happened until someone reminded me when I started talking about this a few days ago. Things like this are said all the time the community always closes ranks, and then it vanishes. This post describes CNC sex parties and heavy LSD use of elite AI researchers at specifically, OpenAI.

[“some of the most coercive and fucked up social dynamics I have ever seen”](#)

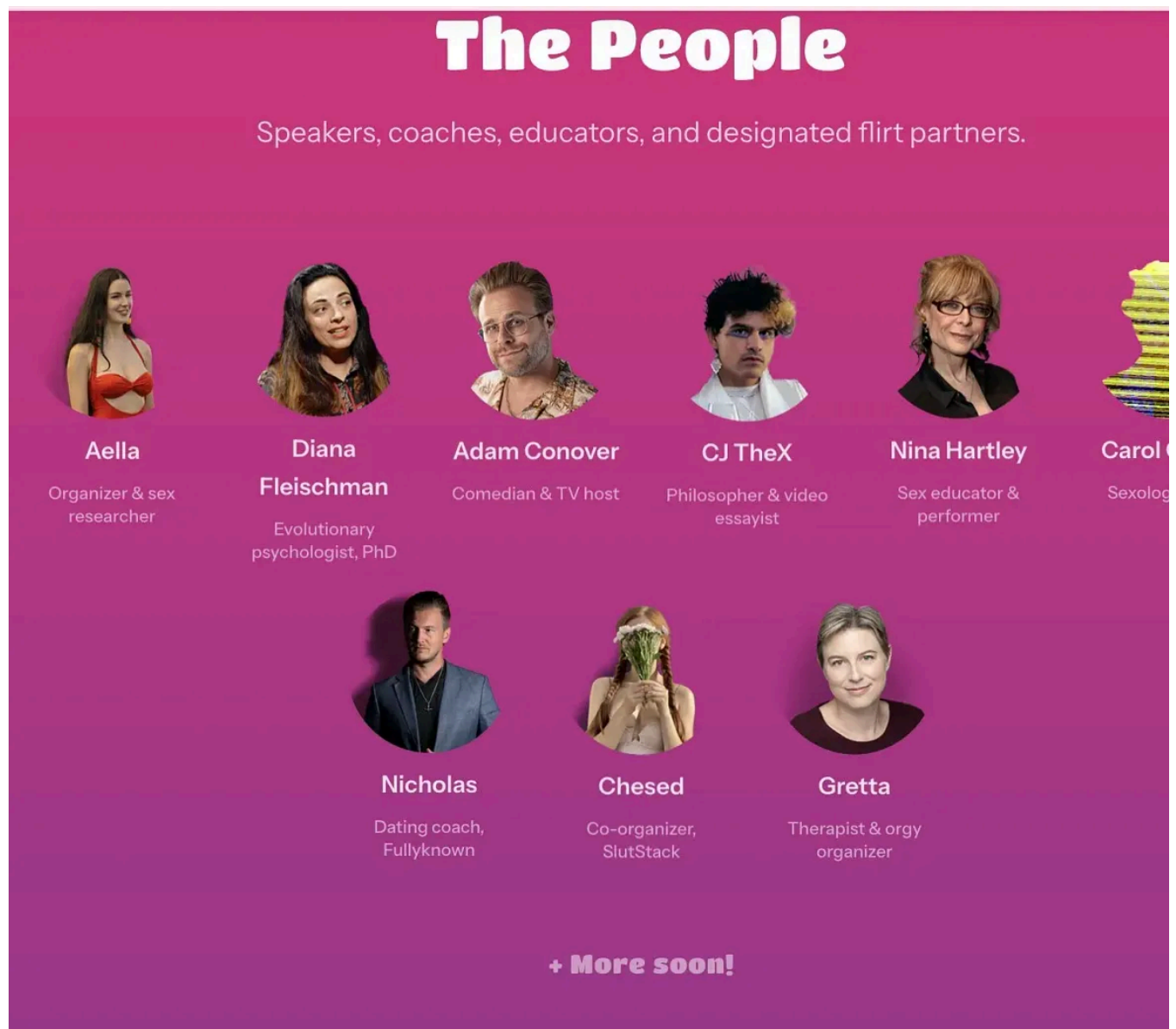


Legally it is possible that people technically consent to being at these parties. Personally? If the context is this coercive, you are still a rapist doing it. If there is a second LSD-fueled CNC party tightly associated with researchers with a severely fucked up social power dynamic besides AI, I would assume it is run by a friend of hers; but probably this is a reference to her events.

That post's author is Sonia Joseph. [Sonia Joseph first got into Rationalism and Effective Altruism by reading Harry Potter and the Methods of Rationality when she was fourteen years old. Her story here in Bloomberg matches everything I am saying, matches the Time story, everything.](#)

## Slutcon

This is the banner for Slutcon.



Gretta is Gretta Duleba, Eliezer Yudkowsky's primary partner. She draws a salary of 200k from MIRI to be his personal assistant, and has no relevant background in anything. I doubt she took the week off to run SlutCon, so she probably got paid for both things.

This event was held at Lighthouse, which is owned by a nonprofit that does not have a mission remotely compatible with a meet-and-greet for prostitutes.

And, oh, yes: This is a thin front for prostitution. It is not even a joke. You are paying between three thousand and eighteen thousand dollars to rent "flirt girls", who are mostly prostitutes, but you are not paying to have sex with them? So far as I can tell this is simply so brazen that nobody considered pointing out that it is clearly not legal. It is possible that Lighthaven is legally a brothel now.

This is Aella soliciting donations for Lighthaven and saying she is hoping that they will let her throw orgies at Lighthaven. Given that they tolerated her throwing Slutcon there, she has probably thrown her rape orgies at the venue too. Those orgies most likely had paid tickets.





**Aella**  @Aella\_Girl · Dec 3, 2024



the Lightcone team is the most powerfully autistic group I've ever met (laudatory). You should fund them continuing to run lesswrong.

Also, if you mention that you funded them because I told you to, maybe I can finally bully them into letting me run an orgy at lighthaven

(read more about them in the thread and their post)

donation link next tweet



**Oliver Habryka**  @ohabryka · Dec 2, 2024

Food \$200

Healthcare \$150

AI Alignment \$150

AI Capabilities \$3,600,000,000,000

Human Intelligence Enhancement \$50...



**Emmett Shear** 

@eshear

...

70% epistemic confidence: People will talk about Lighthaven in Berkeley in the future the same way they talk about IAS at Princeton or Bell Labs.

2:13 PM · Sep 2, 2024 · **73.4K** Views



46



16



425



171



The  
one of t  
moder  
Bryan C



**Brangus** 

@RatOrthodox



Would like it to be known that my (not necessarily Lightcone's) stance here is that, barring external agreements with other parties, what you do in a private room or building that you pay for is your business, provided you do not cause unreasonable damages or risks to the venue.

2:51 PM · Dec 3, 2024 · **1,212** Views

# Belief

My understanding of law is that of an autodidact, so I ask that you evaluate the precision of the following statements charitably, as you would those of a student or an especially clever bovine.

I believe that prostitution was conducted at Lighthaven, located at 274 Telegraph Ave, Berkeley, CA 94705, between the dates of October 10 and October 12, 2025.

I believe Lightcone Infrastructure Inc, EIN 92-0861538, a 501(c) non-profit organization, either profited from renting the venue for that purpose or subsidized that purpose by allowing use of the venue for free.

I believe that this was conducted so openly that it is not credible that anyone involved in managing Lighthaven was not aware that this was what they were doing, and that they went to considerable effort to keep it secret, indicating an awareness that they were doing something illegal. [9](#)

I believe that this constitutes tax fraud, and defrauded the United States Government and the donors to Lightcone Infrastructure, who almost certainly did not generally donate with the intention that their money should be used for promoting prostitution.

I believe further that Gretta Duleba, who managed that event, was drawing a salary of about 200k from the Machine Intelligence Research Institute (MIRI), EIN 58-2565917, that she does no meaningful work for that institute relevant to its stated mission, and that she is effectively employed entirely to arrange Eliezer Yudkowsky's social calendar, especially concerning "consensual non consent" (CNC) or rape parties, and that she is Yudkowsky's girlfriend.

I believe that this means MIRI, which is also a 501(c) non-profit, has also substantively defrauded its donors and the United States Government.

I believe that it is more likely than not that MIRI has also defrauded the United States Government and its donors by paying out money from donor funds to deal with legal matters concerning accusations of statutory rape made against Eliezer Yudkowsky.

I believe that this is part of an ongoing enterprise in which the woman commonly known as Aella is generally involved in procuring or otherwise recruiting people, especially women, to participate in sex work and other sexual activity for the enjoyment and benefit of herself and her friends and associates, and that she either benefits financially or receives things of value from doing so.

I believe that it is more likely than not the case that those recruited into work by Aella or her associates have been subjected, in the course of being induced into commercial sex acts, to force, threats of force, fraud, or coercion, with such coercion possibly including threats of serious harm to them, the use of physical restraint against them, causing those people to believe they were in danger of serious harm or in danger of physical restraint, or threats of misusing the law or legal process against them, abuse of law or legal process including the mediation commonly used by the Effective Altruist movement to prevent complaints of sexual assault from reaching law enforcement.

I believe that it is more likely than not the case that Aella has either deliberately or with willful neglect caused people to be raped at her event.

I believe that it is more likely than not the case that an investigation will uncover many yet unknown instances of sexual assault concerning the above-mentioned people which resulted in payment, settlements, or mediation, which will have left durable records and which payments will likely have been paid out of their various associated non-profits.

I believe that it is more likely than not the case that some victims of crime under these circumstances did not understand that they had been the

victims of a well-defined crime, and that much of this continued specifically because of the effort expended to introduce ambiguity, muddy the waters, and divert to Effective Altruist mediation surrounding sex crimes.

I believe that all of this was part of an ongoing enterprise concerning the Effective Altruist movement globally and its San Francisco branch specifically, and that this ongoing enterprise involves The Center for Applied Rationality, a 501(c) non-profit with EIN 45-3100226, Effective Ventures Foundation USA, Inc. (operating with the Centre for Effective Altruism), a 501(c) with EIN 47-1988398, Coefficient Giving, a 501(c) non-profit with EIN 81-0737472, The Future of Life Institute, a 501(c) with EIN 47-1052538, the Survival and Flourishing Fund, which has apparently created some complicated nonsense involving the Berkeley Existential Risk Initiative and the Silicon Valley Community Foundation instead of forming a 501(c), and countless other non-profits and companies in a vast web that spans the globe and all of which are connected to each other by fundraising, staffing, and ideology.

I believe that this ongoing enterprise includes and overlaps with Future Exchange, commonly known as FTX, and Alameda Research, and that the unrecovered fruits of the fraud that destroyed those companies are today funding this enterprise.

All of these people feed at the same trough.

I believe that this ongoing enterprise has deliberately created and placed its members within military and other government contractors, in addition to many other acts, in order to give itself the ability to influence the actions of the United States Government.

I believe that this ongoing enterprise has instructed its members to lie about their affiliations when questioned where that is expedient.

I believe that it is more likely than not that the Israeli AI cybersecurity evaluation company Irregular is or has staff that are deliberately causing

through sabotage or neglect the illegal hacking of various firms to undermine the perception of US AI labs. Irregular holds contracts with, motivated by their Effective Altruist belief that a “warning shot” is necessary to save mankind from AI existential risk.

I believe that a ‘[pivotal act](#)’ is Yudkowsky’s euphemism for acts of terror that he believes anyone making AGI must do under certain circumstances to prevent competition and therefore the end of the world, and that members of this enterprise will understand some offshoot of this idea and will understand such acts to be necessary in the course of pursuing the goals.

I believe that the [sixth bullet point of this piece specifically](#) is meant to indicate, on close inspection to those familiar with this area, that the ‘pivotal act’ which is the most likely to be necessary and achievable is somewhat easier than burning all GPUs and is *unspeakable in public*, and that the intended reading is that this refers to eliminating all potential competition with engineered biological weapons. Anyone who is not a member of your project is potential competition.

I believe that much of this has been conducted in the open and can be established as more probably the case than not via close examination of public records. All of this has been kept from public knowledge and understanding because it is too offensive and too bizarre to seem real if stated directly.

I would testify to these facts under oath.

I believe that the truth is the best weapon, and if enough of the right people will tell the truth about this it would be the beginning of the end of it.

We are not done.

**Ronny Fernandez**

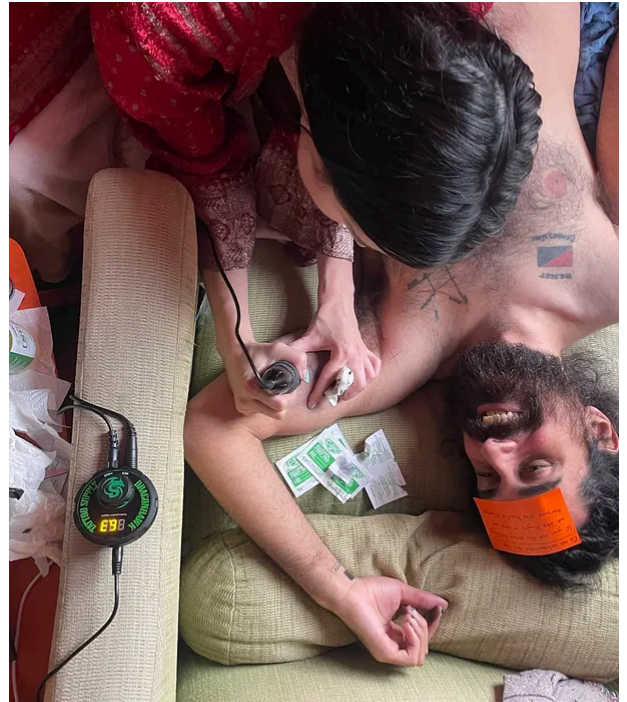


This is Ronny Fernandez. He is the 'Lighthouse czar' [10](#), which I understand to mean that he manages day to day operations and the scheduling of events held at Lighthouse, and has since 2024. His twitter tag is @RatOrthodox and his display name is Brangus. He is seen here receiving an amateur tattoo from Aella of the number '100' for being her hundred recreational sex partner. She took pains to establish that 'recreational' not include her prostitution clients.



If you do not know what being very badly on drugs looks like, this is it.





This was what the kids these days call 'bait', which is anything which attracts attention on the Internet by being bizarre, disgusting or controversial. Ronny is a real predator pretending to be a fake predator real rapist hiding behind BDSM and 'CNC' and being edgy on the internet for attention. All of these core people use this same tactic, and it has worked beautifully.

This is a recently-viral thread of a basically normal person's experience meet-and-greet with the Rationalists at Lighthouse. The moderator wh

separated out the believers from the unbelievers was Ronny Fernande:



**Erin Redwing** ✓  
@realizingerin



I went to an rationalist/EA AI doomer event last night with the goal of meeting someone who could calmly explain to me the argument for how they get to the conclusion that AI will kill everyone.

Because these are rationalists, I expected to find people who could lay out a serious extinction pathway, walk through the assumptions and intermediate steps, identify choke points where negative pathways could potentially be regulated, and entertain disagreement about where their assumptions might break down.

My experience was the exact opposite.

The people I spoke to seemed psychologically unable to entertain the possibility that their conclusion (that AI literally kills everyone, which was a point that many would not even budge on) might be wrong.

Their most extreme conclusion was treated as an axiom.

I had multiple people walk out of conversations with me when I questioned their inevitable extinction premise.

At one point during a panel, the moderator asked everyone in the audience who was still skeptical of AI doom to raise their hands. About half the room did. He then told all of the skeptics to leave the panel and go with one of the speakers, who would answer their questions separately. They filtered out all the skeptics, and continued the panel only with the subset who accepted their extreme position.

I genuinely felt like a heretic just for questioning whether superintelligent AI inevitably leads to the extinction of humanity, or entertaining the possibility that the future might actually turn out well.

That does not mean AI safety is fake. Obviously there are serious questions about rapidly advancing AI systems and what happens as they become much more capable. Those are exactly the conversations I went there hoping to have.

But what I encountered was an apocalyptic death cult. These people are not rational. They are deeply afraid, they reinforce that fear in one another,



and they have become so emotionally committed to apocalypse that they refuse to be reasoned out of it.

7:53 AM · Sep 15, 2026 · 17.6K Views



**Brangus** @RatOrthodox · Sep 15

I am the moderator in question and I feel misunderstood by this. This was during a multiple hundred person Q&A. I invited folks who were skeptical of x-risk to have a smaller conversation with Oliver so that they could get more of a back and forth. I did not ask anyone to leave.



**Erin Redwing** @realizingerin · Sep 15

I went to an rationalist/EA AI doomer event last night with the goal of meeting someone who could calmly explain to me the argument for how they get to the conclusion that AI will kill everyone.

Because these are rationalists, I expected to find people who could lay ...

14

4

234

18K



Ronny Fernandez used to be married to Holly Elmore, the Executive Director of PauseAI US. Their divorce seems to have been messy, they are both activists, and its aftermath often spilled onto twitter. My understanding that he was, among other things, some degree of obsessed with Aella, upset with Holly's leadership of her organization, unfaithful and simply abusive.

This is Holly's most complete explanation of her divorce, told in the aftermath of Sam Kirchner's disappearance. [11](#)



**Holly Elmore** @ilex\_ulmus · Jul 13

Ftr Ronny didn't like the guy bc he wasn't a rationalist and denigrated him for being a Doordash driver. We had a fight where he was like "A Doordash driver? Doordash? Really?"-- totally doubling down when I found that classist judgment disgusting.



**Brangus** @RatOrthodox · Jul 12

For the record, when I met Sam Kirchner I was married to Holly Elmore, Holly and I got into an intense many day fight because I thought he seemed unhinged and unthoughtful. I wanted her to not work with him. Holly thought he was great and was very angry at me for being judgmental



4.4K



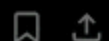
**Holly Elmore** @ilex\_ulmus · Jul 13

The fight was not about Sam being dangerous. Ronny loves dangerous dudes. It was about Sam being from outside Ronny's group and me doing stuff without letting Ronny sit on his ass, try to take all the credit, and call all the shots.

One reason I was so galled about the Doordash remark is that Ronny is a mooch who has never worked that hard or honestly a day in his life. He relied on me for most of his necessities and by that point was unemployed and constantly hitting me up to "loan" him money which would then try to make me feel small or petty for trying to recoup. He's probably still using his mom like this. But he thought he was better than this guy who had the audacity to, idk, be poor and work.



1.8K



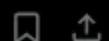
**Holly Elmore** @ilex\_ulmus · Jul 13

For those who don't know, my ex-husband, Ronny Fernandez, @LighthavenPR and @plzdontkillus\_ is a predator who abused me and other women. He told me he knew he was a psychopath and a narcissist. Now he kind of openly trades in that by being one of the rapists in Aella's famous rape orgies.

I hate that I have this past connection hanging over my head but I think the best thing to do about it is just be honest and warn whoever I can.



1.9K



**Holly Elmore** @ilex\_ulmus

Btw I give Ronny a chance to comment by telling journalists to reach out to him and giving his contact info when he comes up in the story of my organizing, which I will certainly not hold back to protect him. If that bothers him then I guess I'll stop giving him a heads up.



I think that Holly is credible and he appears to be both a predator and psychopath.

Holly was previously cancelled by essentially all of Effective Altruism, Rationalism, TPOT [12](#), whatever you want to call these people and all of their hangers-on for objecting to this picture and saying it makes their movement and cause look bad.



**Aella** ✓  
@Aella\_Girl

X.co



8:42 PM · 6/7/26 · **344K** Views

I do not know who the two men on the left are. The man in the golden jacket is Eliezer Yudkowsky. On the right is Nate Soares, President of MIRI. The man second from the right is [John Wentworth](#), an “Independent AI Researcher” affiliated with MIRI who seems to be favored because he is good at CNC. I have heard he is a psychopath. I cannot substantiate that fact, it is hearsay, and it is much more likely than not true. I believe he has most likely been paid with money from MIRI for purposes other than the non-profit’s legitimate mission.



Yudkowsky sought, and was apparently paid, \$10,000 for arranging and publishing this picture. It is meant to be taken as an offensive joke.



I disagree with Holly about almost everything and I would crawl over broken glass to put her in charge of as much as I could because she is the only person in all of this who I am sure is not a rapist. She was cancelled for the only true crime that Effective Altruism seems to recognize: saying that rapists are bad.

## **plzdontkillus**

'plzdontkillus' or 'pdku' was a one-month-long 'creator bootcamp' thrown by the Rationalists this July, 2026. It appears to have been a social media marketing and recruitment drive to bring in new blood and bring their message to short form video. This program offered free room and board, a \$2,000 stipend, and access to a \$20,000 prize pool for high-performing videos on social media.

This was hosted at Lighthouse, and partially sponsored by MIRI.

I will not record here any of the videos produced by this event. Those who participated were being preyed upon, and deserve to have their connection to this forgotten over time. Many of them look very young to me, but everyone below a certain age looks like a child to me and they are probably, legally, adults. This event was held over the summer when s



was closed, but that's probably to avoid noise complaints from the middle school that is across the street from the compound.

Staff for this event:

- Aella
- Ronny Fernandez
- Azmyth (Aella's sister)

"Mentors" or otherwise participating:

- Eliezer Yudkowsky
- Nate Soares
- Grimes
- Jreg
- CJ The X
- Alex Botez
- Tor Parsons
- Quinn Finite
- Aric Floyd
- Zoe Curzi
- Rob Miles

Liron Shapira's podcast, "Doom Debates", did an interview with Aella and Ronny Fernandez at the conclusion of pdku to talk about the event. This interview confirms that JJ McCullough and Jeffrey Ladish provided mentorship separately from the official schedule. Yudkowsky, Soares, and Ladish, all involved in this event, are the three out of four of Bernie's advisors on AI risk who aren't already directors of the company that manages Lighthaven.

I believe that it is more probably than not the case that participants in event were actively recruited to engage in commercial sex acts, both in general as part of the ongoing enterprise of the event's organizers and specifically to staff Aella's SlutCon 2, which is scheduled for September 25th-27th.

While I have been writing, my earlier comments seem to have touched some internal strife within the movement. This is resurfacing issues I have forgotten, like the "pdku dead cow question", discussed here by rati on vishal:



**vishal** ✓

@koreindian

i disagree with pretty much all of this. 60 year old voters in ohio don't know about slutcon or the pdku dead cow question, and if they did they'd forget about it quickly between media cycles. playing for respectability at the margin is a mug's game. think of trump, hunter biden, clavicular, and kanye west: the clear lesson is that americans do not care about "respectability." memory is short. it's a rule-breaker's world, and the savviest players know they just need to keep buying tickets at the discourse lottery.

This is the beginning of a much longer post defending asking this question on the application for plz dont kill us:

“If you had to have sex with a cow would you rather it be dead or alive? Why?”

This question is being defended, broadly, as just people being people and doing weird quirky things. Anyone opposed to asking this question is accused of being uptight and killing the mood. Those who think the question is inappropriate are correct, but their reasoning is wrong: they think it is inappropriate because it makes their movement look bad.

It is inappropriate because it is being used to test people's boundaries, to filter out those who have them, to make people feel complicit, and to provide an icebreaker for talking about sex. It is inappropriate because it is being asked by people in a position of authority, and they are asking, among others, a gaggle of teenage girls. It is inappropriate because it is being used to probe for weakness, and to figure out how to recruit them and get away with harming them.

It is inappropriate because Ronny Fernandez is the one asking it and he has been credibly accused of being a predator towards women.

It is similarly inappropriate for Nate Soares to be in a diaper, in Lighthouse around the attendees while in a position of authority, or to have participants making videos where the participants 'accidentally' show him and a roomful of other organizers a bunch of soft porn.

These reasons for objecting do not occur to anyone in Bay Area Effective Altruism or Rationalism because they have completely normalized this behavior. They have been taught to never consider such questions. To suggest that this is inappropriate because it is predatory is to damn more of their leaders, which they are incapable of doing.

Here is Aella's explanation for why the cow question was on the application form. I would like to thank her for confirming the question was calculated.



**Aella** ✓

@Aella\_Girl

Follow

(oh i want to note: re the cow question, I did really fight for that question despite pushback from even ingroup rats. I think it (or an equivalent question, idk if the question itself was the best but i really wanted something to the role) was extremely good for specifically the application process of what we were trying to do. I would not include that or similar question in other application processes for other things did not include it as a casual 'lol', it was a pretty deliberate strategic choice)

11:29 AM • 22 Sep 26 • **3.1K** Views



## Brangus Unchained

Ronny Fernandez has a second twitter account, where he used to post unfiltered thoughts. He seems to have forgotten about it. Unfortunately, in order to avoid deflection or claims that this is just a joke I have copied a lot of it. If you're already convinced he's a piece of shit, cannot be trusted around women or children, that anyone who trusts him around women or children is a piece of shit, and that all of these people knew, I would recommend skipping this section. [13](#)

This establishes that it is his account:



**Brangus**     
@RatOrthodox

X.com

Hello, I have changed my alt's handle to [@BrangusBrangus](#). Please follow for spicier content I am unwilling to post here.

14:46 · 2024.09.21 · **406** Views



3



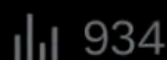
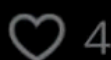
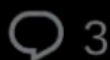
This establishes that Nate Soares is aware of this account, and does not find its contents especially concerning; this is about five months into R being in charge of Lighthaven:





**Brangus Unchained**  · 2024.12.21

only weak men let their gfs unionize. i hit em with the right to work laws and force them to negotiate individually



**Nate Soares** 

@So8res

X.com

(though only weak harems need joint negotiations in order to coordinate, ofc)

19:13 · 2024.12.21 · **425** Views



This establishes that Aella is aware of the account, and is comfortable the contents of the account:



I expect that the usual defense here would be that he is being funny, or it is just fetish content, or something. If such an excuse is made, please consider this post:



In this one, he claims his assistant brings him a cold brew with a methamphetamine powder rim every morning ...



**Brangus Unchained**   
@BrangusBrangus

my assistant brings me a metharita every morning. that's a can of cold brew with a methamphetamine powder rim

2:28 PM · Mar 14, 2025 · **288** Views

... and this is his assistant, Azmyth, Aella's sister, taking a candid video because they are having a good and fun time at pdku. He says he is on stimulants.



This is interesting, because it is a joke about the normal defensive tactic this group of people uses. Act harmless, act weird, and use this to get people's defenses. If using this tactic to rape people and get away with rape is funny I guess I don't get the joke.



**Aella**  @Aella\_Girl · 2025.03.29

would you predict ancestrally that men aroused by rape are more:

A) higher status, because both are downstream of general testosterone/dominance/aggression

B) lower status, because if you're low status it's harder to find a woman to mate with you consensually

A

43%

B

57%

10,853 votes · Final results



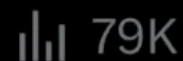
99



11



189



79K



**Brangus Unchained** 

@BrangusBrangus

X.com

ohhh i'm just a little low status guy, nothing to worry about here

12:39 · 2025.03.31 · **160** Views

These are, in my opinion, the worst posts not already included that directly concern rape. I don't believe these are either a joke or just BDSM, but not certain I could prove that in court, I suppose. Social context is difficult







**Aella**  @Aella\_Girl · 2023.07.27

all the dudes i know with high bodycounts, tend to be mostly banging women who have high bodycounts. it's less 'preying on wholesome, virginal women' and more 'all the sluts are in a giant slutcloud fucking each other'

 96

 115

 2K

 210K



**Brangus Unchained** 

@BrangusBrangus

X.com

I used to prey on wholesome virginal women too but we actually don't have enough in common

19:00 · 2024.09.23 · **310** Views





**Brangus Unchained** ✓

@BrangusBrangus

Follow

just vadin' n rapin'. what're you up to?

15:40 · 2025.04.03 · **330** Views



2



**Brangus Unchained** ✓

@BrangusBrangus



Guys, it is soooo much scarier raping somebody on 5meo

10:21 AM · Dec 5, 2024 · **573** Views



5



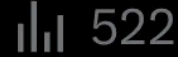
1

5meo refers to 5-MeO-DMT, also known as toad venom.



**Brangus Unchained**  · 2025.04.20

drive your gf absolutely crazy with this one weird trick. answer in replies.



**Brangus Unchained**  · 2025.04.20



dose her with 500 mcg of lsd without her knowing and then after a couple hours only reply to everything she says by saying “haha what, you’re being crazy”

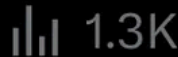


I don't know how to explain what's going on in these ones. Some psychologist could write a thesis about what is going on in these ones. add context for his state of mind concerning all of these things:



**Brangus Unchained**  · 2025.01.09

i'm pretty sure drag queen and trans women are not doing woman blackface, but i wanna make it clear that when i wear women's clothing, I definitely am.

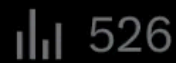




**Brangus Unchained**  · 2024.10.24



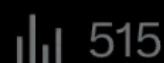
Type of guy who bihacks because he hates women. Like yeah, I only eat impossible meat because I fucking hate cows and the idea of consuming their flesh disgusts me.



**Brangus Unchained**  · 2024.10.23



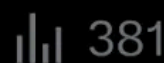
Being a guy who is genuinely into cnc sometimes feels like being the prettiest girl at the ball.



**Brangus Unchained**  · 2024.10.23



I've been procrastinating on showering but i have a lot of shmegma build up so I'm gonna take care of that and then get back to facing death vicariously through my grandmother's eyes. Talk to you guys later



**Brangus Unchained**  · 2024.10.20



my true form is wearing nothing but a pink bonnet and an adult diaper with holes cutout for my penis and butthole. the diaper will not absorb my waste. that is not its purpose.

## The Atlas Program



### Ronny Fernandez

Ronny studied philosophy at Rutgers University and likes to think about how artificial intelligence grounds philosophy in the concrete. He also did AI forecasting research at AI Impacts. At Atlas, he enjoys discussing formal epistemology and strategizing for Killer Queen games.

[Learn more →](#)

This is Ronny Fernandez' picture on the page for The Atlas Program, at which he was an instructor, and which appears to have only run in two years, 2022 and 2023. I am saved from having to type out most of the important information about The Atlas Program, because they put it all one nice box.



The Atlas Fellowship evolved out of a program called [Uncommon Sense](#), which had a focus similar to the Atlas summer program and proved impactful for its participants. Sydney Von Arx, one of the co-founders of Uncommon Sense, teamed up with others to found the Atlas Fellowship to vastly expand the scale and scope of that program—starting the Atlas Fellowship 2022.

In 2022, the Atlas Fellowship ran five 11-day programs in Berkeley (California) and Oxford (United Kingdom) for 111 Fellows from 18 countries, selected from over 7,000 applications. We plan to continue and refine our programs, and potentially support a substantially larger number of students in future years.

The Atlas Fellowship is a 501(c)(3) public charity in the United States (EIN 88-1678673). It is funded by [Open Philanthropy](#), a large philanthropic organization in the United States. Previously it also received funding from the [Future Fund](#). If you are interested in supporting us, or if you have any questions about your donation, please contact us at [info@atlasfellowship.org](mailto:info@atlasfellowship.org).

I believe the following facts are perhaps not obvious from this screenst

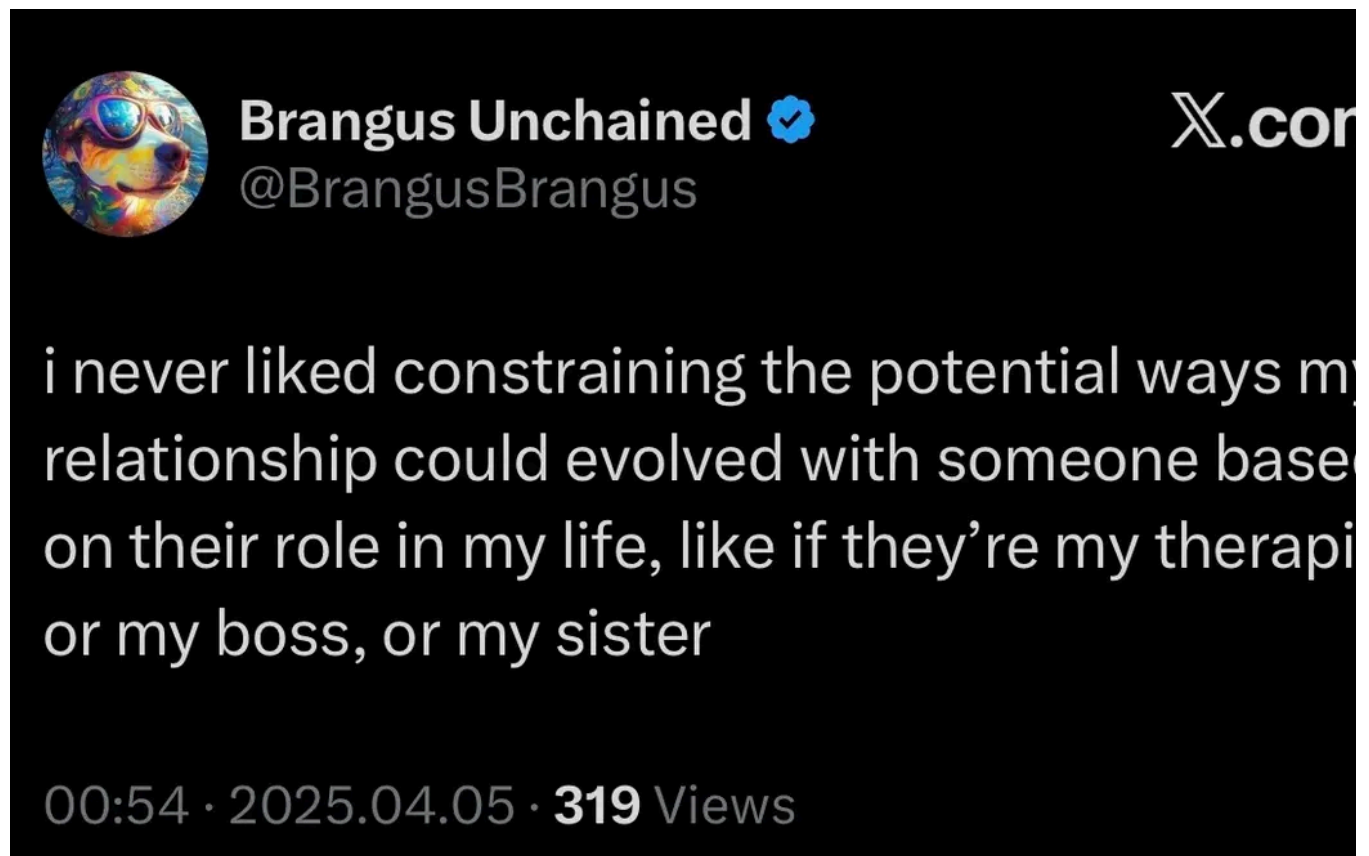
- Open Philanthropy is now called Coefficient Giving, and is an Effective Altruist organization.
- The reason this fellowship was held in Berkeley and Oxford is most likely that it was held at Lighthaven and another rationalist compound in Oxford called The Star, at 21 Rectory Rd, Oxford.
- This is a summer camp for high school students. There are videos of them on the site. I refuse to put those videos here to prove that po

I have never heard of this program before. I found it while trying to find a good reference photo for Ronny's face, so now we have one. There are many rationalist and effective altruist summer camps, youth programs so on that finding them all and figuring out where they are would make it impossible for me to finish this in a reasonable amount of time.

Ronny is the current manager at Lighthaven and all events are scheduled through him. He is the one who separates the believers from the unbelievers at meet-and-greets and he will have access to all of the students of any of the programs that are run at the compound.

[Here](#) is an article where he argues having sex with an adult when he was thirteen did not harm him. It seems very defensive.

How fucked up do you have to be to allow this person around children?



If you've been counting, this is the third time a summer camp has come up in this article. Do you think any of the leaders of this movement learned anything at all from past incidents? Or do you think that they believe getting their anointed access to children is important, is almost a *sacrament*, and resent that anyone ever tries to prevent them from doing that?

When Brent Dill came up, the narrative from leadership was that they did not know what he had been doing before they put him into a position of authority at a children's camp. They should have known and were negligent but not malicious. They would do better, they were listening and learning whatever.

For several years now Brent's story has been that they knew about what he was doing, that they always knew what he was doing, and that at least his sins were against adults and the organization was full of people who were abusing children. His story was that he was bad but not remarkable, and that he was kicked out of the club for being a pain in the ass. His story was that they loved having someone like him they could give their dirty work to.

I would like to see all of this on paper. I would like to see it sworn to under oath. I would like to see it in front of a judge. It is unbelievable how many clearcut crimes these people have shielded themselves from by transmuting them into internet drama. They are as isolated from the larger society as if they were on a desert island but they are right in the open.

Rationalists and Effective Altruists do not keep a proper history of their movement because these are the main events in their true history and cannot afford for law enforcement or new recruits to understand them.

I hope I have done enough to correct this problem.

## **What This Is**

It is either a lie or a half-truth to say that this is a sex cult.

This is a rape cult.

Its central members prefer rape above all else and recruit rapists for leadership. By offering this as the reward for complicity, they can gather these people together and make their loyalty nearly perfect. They are

bound together in mutual guilt and they always have a hobby in common. They promote rape because it is the main thing they like and believe in.

This is too obscene to be taken seriously if stated flat, so I cannot put it in the lead. This is also why no one can just say it and be believed. It sounds like a conspiracy theory. It is so much like a conspiracy theory that I suspect people were deliberately planning and doing these things specifically because they sound like conspiracy theories. It serves as fantastic camouflage. They are brilliant at that sort of thing.

All of the main content of rationalism is a trick. Every time Yudkowsky explains some trick your brain can pull, it is the magician showing you his sleeves while he palms a card: it is meant to convince you that what he is about to do is impossible, and to prevent you from figuring out how it is done. [14](#)

What he has done is crafted an atheist religion that worships and fears omniscient God, trusts Yudkowsky as its prophet, and then used the power this gave him to gather as many predators and as much prey together as he could and to make the prey too confused to run.

Nothing that has happened in the movement is a coincidence.

## **Coda: Close Reading**

Rage can sometimes make you see clearly when you would not have otherwise.

While I was writing this I think I figured the whole thing out. I will try to reproduce my reasoning and the facts it relies upon, as nearly as I can. Hopefully anyone can reach the same conclusion.

First, Yudkowsky is a great man.

Great as in the Great Man of History theory, all of that. If he had not lived the world would be very different.

You could very credibly write alternative universe science fiction for the thirty years that simply subtracts him. You've removed or significantly altered DeepMind, half of the weird cranks, Sam Bankman-Fried. Hell, earlier: Half of the original cryptocurrency guys were rationalists. Do they do different things without his influence, or without his movement to bring them all together? I have no fucking idea. If you're an author you can go wherever you want with that. It's fine. You can bend history however you like by removing him; he has bent the entire trajectory of technology by his influence.

He is a person who sat down at a keyboard and simply decided things should happen, and they did. He sat down as a teenager in the nineties, nothing but write, and in 2015 he was thirty-five years old and by far the most influential person in the field. There is little evidence he has ever written a computer program that solved any problem. His math is rudimentary. In traditional academic terms, he is a popular philosopher.

He thought he could personally and directly control the entire thing, and almost could. He was involved in DeepMind's founding and knew its leaders personally. All of this blew up when Sam Altman and Elon Musk simply walked away, and he has been trying to scrape back direct personal control ever since. In spite of that, a huge chunk of thinking in the field is still fucked up pseudoscience. When Sam wants to insult him he uses a backhanded compliment. When Dario wants to disagree with him he is afraid to name him.

This is an incredible achievement.

What he did was read usenet and science blogs for over a decade, read a few of the books they recommended, and then distill everything that was in the ether. This was a mixture of having done your reading in



undergraduate, a rehash of Korzybski (who also influenced Hubbard), a dozen transhumanist authors and the sorts of things people talked about but had not written down formally, or that were only written down in an article somewhere.

From these ingredients he wrote just one thing that let you skip the line, avoid doing any real reading or real education, and acquire maybe two thirds of the world view you would have as an excellent, well-read student in a science major, but with none of the specific skills you would have learned. For some people this fast track is the most exhilarating experience of their lives. Many of them become convinced Yudkowsky is the smartest person alive.

I am not convinced he is the smartest person alive because I had basically his same reading diet at the time and almost none of this was new to me. To me, he is a guy with a pretty good generalist blog that became a religion while I wasn't paying attention, and a bit of a crank. I was there. It was a science blog to me. I did not see it as an ideological project because I was not looking for that. Judged as an ideological project, it is vastly more impressive work.

If we judge his writing by its impact, for how much the world changes around you with every word you write, his last rival was Lenin. Every time you hear someone call their beliefs their "priors", they kiss the ring. They generally do not know this. I am obsessive about the origins and meanings of words and ideas and Yudkowsky's jargon is fucking everywhere. It is like hearing people talk about how Gatorade is full of anti-Thetans.

All of this to say: I can assume that his writing is intended to influence people, and to cause them to do things, and it is well-calculated and good at doing this.

Next, we need to settle on the right question. I probably only found the right question because I had been reading through all of the source

material for this article for days to try to figure out how to explain it:

Who is the intended audience of Harry Potter and the Methods of Rationality, and why?

Our first theory, our null hypothesis, is this:

HPMOR is written to recruit, broadly, college students and new graduates to his cause. His previous audience was entirely people who read science blogs. I would be surprised if the average level of education in the comments section was below PhD. I was there because I was a very unusual teenager. He has realized that these people have jobs and cars and are not interested in being subordinate to him; they want to do things like found their own companies.

He was trying to set up a movement and a company or companies of his own which would have some chance of making meaningful progress, and attaining meaningful power over, artificial intelligence. He correctly recognized that artificial intelligence was going to be very important in the relatively near future.

Our second hypothesis:

He was trying to build a broad, lasting ideological movement, and knew that this meant he needed his audience to be as broad and as young as possible.

There is very little indication from his statements that this was what he was trying to do. Everything he said about this focuses on having something like a company, and children make terrible employees.

If he did have this idea, however, it was a brilliant one, and worked extremely well. People are indoctrinated young, their beliefs are sticky; it is nearly impossible to remove them. This seems like it is a credible theory because he is a very effective writer. If he was deliberately doing

this, he would likely have avoided saying so to avoid the obvious parallel with religious conversion.

Our third hypothesis:

He wrote for an audience of children because he wanted to have sex with them.

First we can ignore the movement and the history and simply consider circumstances. What is the base rate here? What percentage of all men are pedophiles? What percentage of great children's authors or children's artists are pedophiles? What percentage of all thirty-year-old men who would do anything at all like writing a great work of Harry Potter fan fiction are pedophiles?

These are relatively easy to answer, so we will consider asking them later.

Next, we can do a close reading.

This is the most controversial passage in Harry Potter and the Methods of Rationality:

Draco snarled. "She has some sort of perverse obsession about the Malfoys, too, and her father is politically opposed to us so he prints this word. As soon as I'm old enough I'm going to rape her."

This is the dead cow question! It's the same trick! This is the seventh chapter out of one hundred and twenty-two. Anyone who is as offended by this will just fucking leave, and many many people do.

Everyone who makes it past Chapter Seven of Harry Potter and the Methods of Rationality belongs to one of two groups:

Prey, who simply do not react with disgust or recognize threats normal

Predators, who think that Draco Malfoy raping Luna Lovegood is awesome and that they want a story that has more of this because they love rap

Once you see this the entire rest of it falls into place, that this is a story specifically constructed to entice children. It is a story that is about an eleven-year-old and that is written for a child that age or not much older. It is a story about not fitting in and feeling alone, it is a story about being the only one you are very mature for your age, about how you're really like an adult in a child's body, about having a special and different and brilliant older friend who you have secret talks in secret places with, it is about them recognizing how special you are because they're just like you and they are an outcast too. It is about finally finding someone who understands you. It is about being given strange and amazing and very personal gifts, and about keeping those gifts secret because people just wouldn't understand.

It is a story aimed at a very specific type of child, and its protagonist is painstakingly crafted to resemble this child in personality and serve as an escape fantasy for them. This child is disappointed in their parents. This child is almost certainly neurodivergent, and has overwhelming anxiety about commonplace things. This child feels that adults in general have let them down, and yearns for adults that treat them as equals. This child has hated and resented the experience of being interrogated about alleged sexual abuse. This child does not like being told what to do, is very independent, and is capable of running away from home.

It is a story about being better than other people, a sort of Randian superhero who can outsmart and out-sneak everyone around them. It is a story about being unique, and better than the rules, and not believing in the rules or those who use them. It is about how to avoid suspicion from adults. It is a story about getting into a battle of wits with adults, and winning.

It is a story about Eliezer Yudkowsky because everything Yudkowsky writes is about himself; his most distinctive authorial tell is that all of his writing

no matter what else it is about, is autobiographical. It is about how Elie Yudkowsky is the smartest and kindest person who ever lived, and he understands you.

It ends with being made to swear a vow to keep secrets, with internalizing that vow so deeply it becomes a part of you, and knowing that if you break it that the world will end.

He told every child who was vulnerable in the right way and that he could reach that they should run away from home and come to him and have sex with him and he did it where nobody would take what he was doing seriously or understand it or be able to explain it to anyone else if they tried.

Eliezer Yudkowsky is a pedophile and has been luring children to him on the internet with his writing so he can have sex with them for at least the last fifteen consecutive years.

He has made every ongoing enterprise of which he is a part a party to this endeavor, and anyone who was aware of his tendencies and who spread the word or furthered any cause he was connected to is directly complicit.

I intend this to be taken as a statement of fact. I intend for as many people as possible to read it. I intend for this to cause him and those associated with him actual harm and financial loss.

I know it's true and everyone should know it's true.

He is a fucking genius. He might as well have been driving around in a white panel van with Yud's Free Candy on the side and it worked and nothing could be pinned to him. He had perfect camouflage. When he retreated to Facebook everyone who was serious about science assumed this was because he just wanted validation, but it was because it was what the thirteen year olds were. We maybe thought it was creepy that almost all of his output outside of HPMOR is fucked up BDSM porn but we didn't think it was there to get kids to read it.



Nothing that has happened is a coincidence. Rationalism and Effective Altruism are sick because they are built on the logic of abuse.

He understood the power of ideas that spread themselves, and mutate understood how to hide one idea inside another one.

The Basilisk, like all really great heresies, had to be forbidden because was true. If you take Yudkowsky's ideas seriously, you always end up a servant of the Basilisk; it is simply hidden in different ways. This is the of rationalism, and the object of its worship, and power over others is t only value it keeps.

I understand what people mean when they say certain ideas are demo and that they spread like a disease from person to person, and that th are also like a fire that grows faster the more of it there is in one place

“Alignment” seems unsolvable to Yudkowsky and those who accept any offshoot of his system because it is solving the wrong problem, and because he is hiding what problem he is actually solving. Yudkowsky believes that an advanced AI will almost inevitably become too intellig escape human control and kill humanity. What he needs to fear is that child will become an adult, escape his control, and want to destroy him turns all problems into this problem: how to make sure you personally always have absolute power over everyone else and nobody can ever punish you for evil.

- 
- 1 <https://www.wired.com/story/book-excerpt-the-optimist-open-ai-sam-altman>  
<https://www.nytimes.com/2023/12/03/technology/ai-openai-musk-page-altman.html>
  - 2 I also learned from this that Geoffrey Hinton's father was an asshole.
  - 3 It helps, when going through a lot of volume, that we recently taught comp to read.

- 4 This call would be more impressive if he had not also named other people who seem significantly less Antichrist-like.

---

**From:** Joscha Bach [REDACTED]  
**Sent:** Sunday, August 26, 2018 2:30 PM  
**To:** Jeffrey Epstein  
**Subject:** Michele Reilly  
**Attachments:** signature.asc

Dear Jeffrey,

My friend Michele Reilly asked me to reintroduce her to you; you met her and Michael Vassar a few years ago on an island. She says that she has made progress on scalable quantum computers (linearly with \$1 per qbit), has prototypes, working demos and IP, and apparently got Danny Hillis excited. She is now looking for advice on advisors and private investors. If you like to talk to her: [REDACTED]

The conference in Prague was inspiring; I got to know a number of interesting minds, like John Langford from Microsoft research, brilliant weirdos such as Brendan Ruff (finds his own algorithms for visual perception that seem to work well enough for selling them to the car industry), and Steve DiPaola from Simon Fraser (who makes psychologically meaningful models of the perception of art).

October issue of Psychology Today has an 8 page article about me ("Wizard of Consciousness"). Thank you for everything,

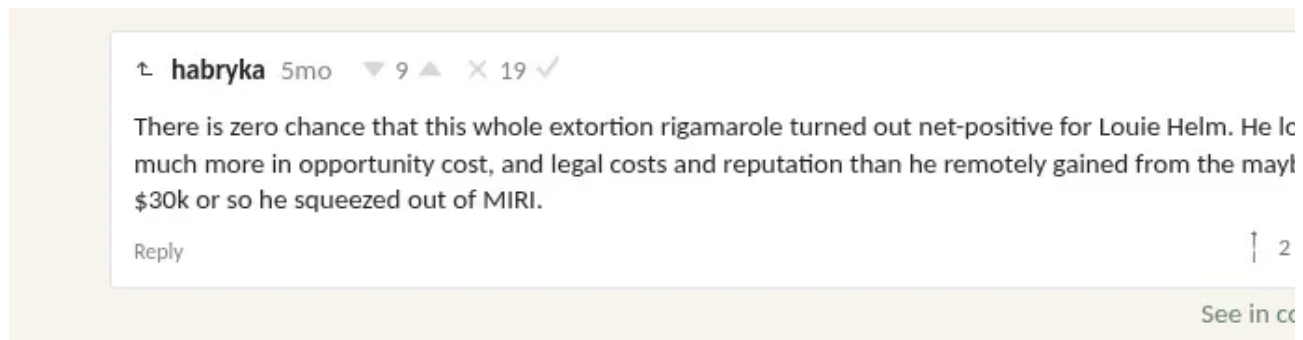
Joscha

# Friendly AI Research as Effective Altruism

June 5, 2013 | Luke Muehlhauser

MIRI was founded in 2000 on the premise that creating (In this post, I talk about the value of *humanity in general* creating Friendly AI, though MIRI co-founder Eliezer Yudkowsky usually talks about *MIRI in particular* — or at least, a functional equivalent — creating Friendly AI. This is because I am not as confident as Yudkowsky that it is best for MIRI to attempt to build Friendly AI. When updating MIRI's bylaws in early 2013, Yudkowsky and I came to a compromise on the language of MIRI's mission statement, which now reads: "[MIRI] exists to ensure that the creation of smarter-than-human intelligence has a positive impact. Thus, the charitable purpose of [MIRI] is to: (a) perform research relevant to ensuring that smarter-than-human intelligence has a positive impact; (b) raise awareness of this important issue; (c) advise researchers, leaders and laypeople around the world; and (d) *as necessary*, implement a smarter-than-human intelligence with humane, stable goals" (emphasis added). My own hope is that it will not be necessary for MIRI (or a functional equivalent) to attempt to build Friendly AI itself. But of course I must remain open to the possibility that this will be the wisest course of action as the first creation of AI **draws nearer**. There is also the question of capability: few people think that a non-profit research organization has much chance of being the first to build AI. I worry, however, that the world's elites will not find it fashionable to take this problem seriously until the creation of AI is only a few decades away, at which time it will be especially difficult to develop the mathematics of Friendly AI in time, and humanity will be forced to take a gamble on its very survival with powerful AIs we have little reason to trust.)) Friendly AI might be a particularly efficient way to do as much good as possible.

7



- 8 End of a long twitter thread which covers surrounding context and somehow does a history of the movement, but hard to follow

<https://x.com/QiaochuYuan/status/1726951287567364487>

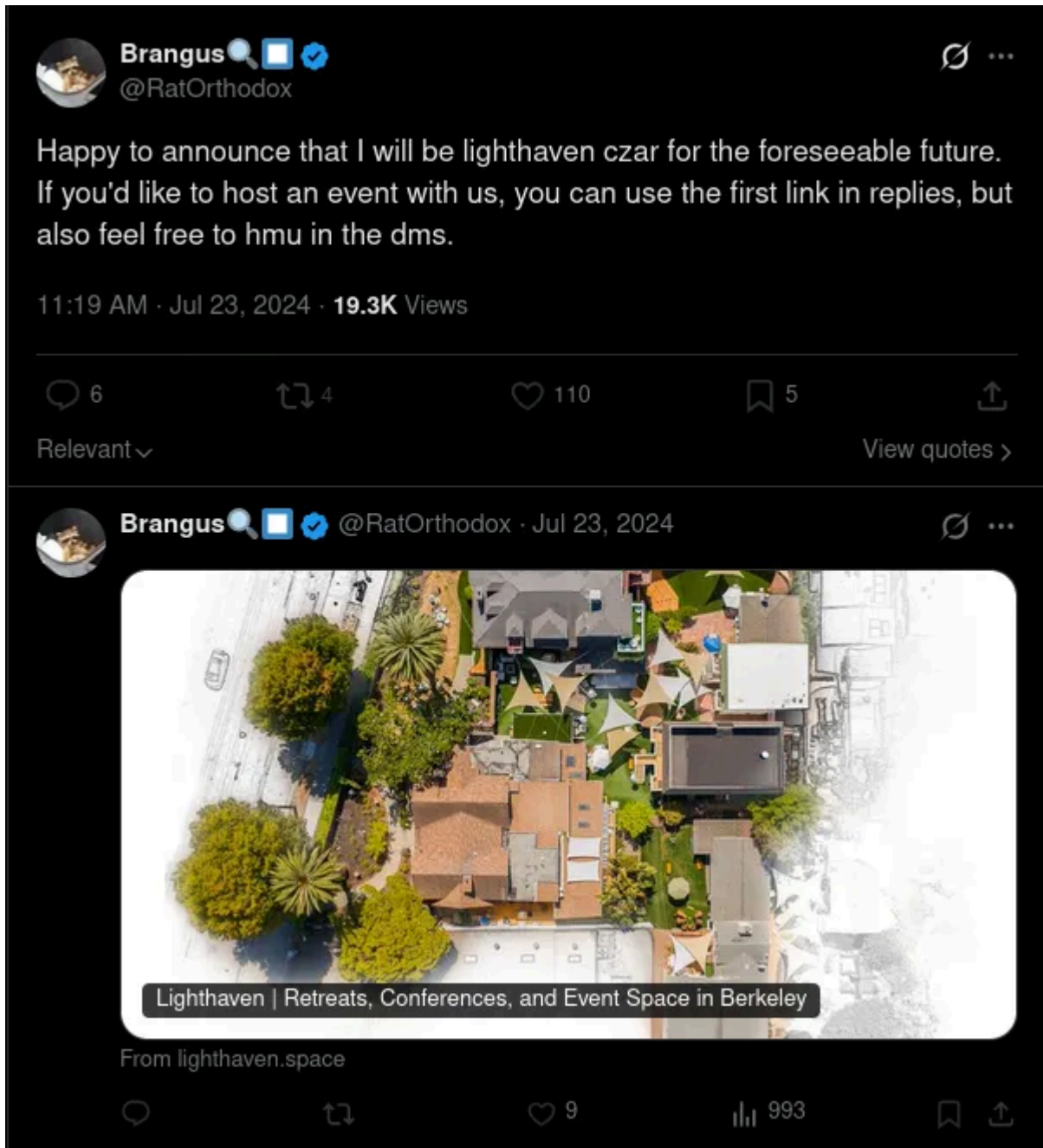
CFAR official statement:

<https://www.rationality.org/resources/updates/2019/cfars-mistakes-regarding-brent>

Original allegations [One](#) [Two](#) [Three](#)

- 9 There are, however, publicly available records of this fact, if you know where to look.

10



- 11 I figured Sam Kirchner would have turned up by now. It is very strange that still in the wind.
- 12 TPOT is the name for a social scene on Twitter. It does not have hard bounce or a fixed definition. I might technically be a member of it but am certainly one of the thought leaders of it.
- 13 It turns out Discord auto-moderates these images and it makes it difficult to talk about him on Discord. Good work at flagging content that probably should be allowed on any platform, Discord.



14 For the Rationalists in the audience: The purpose of explaining Pascal's Mug was to make you think you had figured it out so that someone could Pascal Mug you in a more clever way, and could never be convinced that this had happened.

---

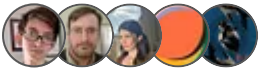
## Subscribe to Very Sane AI Newsletter

By SE Gyges · Launched a year ago

Ideas in AI, preferably in English. Mostly what and why, a little how.

By subscribing, you agree Substack's [Terms of Use](#), and acknowledge its [Information Collection Notice](#) and [Privacy Policy](#).

---



8 Likes · 5 Restacks

## Discussion about this post

Comments

Restacks



Write a comment...

---

