1. Open weights allow fine-tuning. It's barely open without open training data.

a. Nothing will stop eventual leaks of foundation models, mistral medium beta was leaked.

b. It's faster requiring less compute for effective results all the time.

c. If it's on the internet in any form, it's widely available, however, large models are difficult to run at home for now.

d. At home models provide data security and prevent my workflows from being slurped into the next foundation model through network API access. Web and API access provide zero transparency and no guarantees of consistent results, which prevents useful deployment by small business or individuals who lack the resources to constantly tune against changes outside of their control.

d.i. I only use widely available local models, because I believe that in a disaster situation or even simple internet disruption, my local model will remain available to help me, be that as a makeshift better than nothing doctor, or condensed internet search for how to fix a device I need to survive. LLMs are the ultimate survivalist companion, they bring the greater portion of internet knowledge anywhere I might need it. Wide availability provides some trust against malicious code embedding through community reporting.

2.What are the real risks of LLMs? They merely amplify what is already on the internet, and have some ability to blend distributed errata together. The dangerous information they contain is only a web search away.

a. There are few reasonable risks, provided that models are open and internet security is encouraged to use them to test and strengthen their security using them. The danger arrives when only bad actors have access to to these models designed to break security features. Most risks I see parroted are very silly. A very smart individual is much more dangerous than wider access to knowledge. LLMs are an equalizer, and will empower both sides of any conflict. People should be held accountable for the intent they command with. Limited availability is incredibly more dangerous because only attackers can effectively leverage the models, and models will be developed in secret if you don't encourage open model weights.

b. Open models would ensure that individuals in those systems have access to free or cheap services in those sectors. Without open models, a monopoly will rise and become an expensive gatekeeper attempting to usurp the original human employment cost they initially reduce.

c. No privacy risks outside of hacking, which limiting open models makes worse. Even then, LLMs won't be dropping flash drives in parking lots to break into systems. Hacking is and always will be driven by social engineering, and provided a company and it's employees follow effective security guidelines, an LLM should not be convincing enough to get someone to spill their admin password.

d. Watermark official releases with verifiable certificates. Nothing will stop the creation of "fake news" by bad actors. Any strange ideas that could create a safety risk are already broadcast-able from one individual to the masses with the internet. LLMs will make fake news more prevalent open or closed doesn't matter. Encourage people to think about what is presented to them, there is no future without critical thinking. Teach media criticality in school or we are headed for strange times.

d.i. Closed models will be cultivated as sources of truth controlled by individual corporations with little or no oversight, telling people how and what to think and trusted as a safe source, even if it's stressed that LLMs are inherently fallible.

d.ii. LLMs, especially general models, don't contain specific business logic or security information, a business fine-tune would end up more like advertising than a security breach, constantly bringing up company products and not generally suitable for wide use. One trained to an internal code base is a bit different as it could expose the business logic that sets a company apart from competitors.

e. If we limit models made here in the USA, I'll be at risk of having to learn chinese culture and history to get the most effective use from chinese models, and I will be learning the official party line history their government presents. Open models will allow equal leverage by all, a small entrepreneur can download a security expert to manage his server code. Closed models mean only large businesses that can afford the expertise will be secure.

f. Closed models will make the future a feudal culture where a handful of companies can make redundant whole sectors of employment, with no door for the small people to compete on any technical platform. A few will gain impossible wealth and influence while the average gawk and fear their replacement. Unrest will rise as individuals are prices out of access to the incredible tools being sold at a loss right now to gain market share.

3. Open and widely available model weights provide transparency to tampering and resistance to propaganda emplacement by ensuring no model becomes a singular source of truth for future generations.

a. Open models weights encourage ease of experimentation and adaptability to fit niche purposes and most importantly they place the responsibility for misuse on the business or individual using them.

b. Open and widely distributed hacking models can be used to harden defenses against hacking models developed by bad actors. Without unrestricted open models, you essentially remove the common man's ability to harden their stack against bad actors who will develop those models, legally or not. Open and available models ensure free speech and intelligent discourse is preserved by preventing limitation of the interactivity to corporation proscribed topics. Closed models concentrate power on an already unbalanced stage. If AI continues to advance, common people won't have any economic value beyond charity tax write offs for Microsoft, Apple, Google, and Facebook. With open and distributed models, at least we can leverage available tools to attempt innovation and entrepreneurship. Without, behind closed doors a few can change the available model and disrupt innovation.

c. The ability to retrain models preserves safety in these sectors by providing a competitive playing field and ensuring the availability of multiple AI systems that can work with relevant topics, rather than a single lawyerAI rising to dominate all legal discourse. It is vital to our liberty as humans to ensure that people retain a choice in their representation that is more than superficial. It is vital that people receive diverse educations and maintain access to multiple diverse AI knowledge systems rather than everyone being taught the lines from one training set.

d. Open models ensure freedom of speech and embody freedom of press. With few closed systems and closed training data, it becomes very difficult to identify real bias in training, and eventually most human knowledge will come from corporate pruned knowledge sets.

e. Open and trainable models allow a democratized effort to influence the way the culture changes, rather than a handful of shareholders trying to limit their liability. Open and trainable models will ensure that anyone can create specific and useful tools powering innovation and lowering costs in their target use-case.

4. Forcing open and available weights of all models would reduce the economic power of microsoft and google. Not NVIDIA though, they will be stoked. There may be personal data in the scraped training sets, but there is truth to the statement: "All current LLMs are trained with data from anyone who has ever made a public facing comment or website." In some part, the data these models are trained on could not be assembled without the public knowledge and discourse of all people. Open weights and lack of restrictions or liability in training foundation models ensures that the technology at least might benefit everyone.

5. LLMs only write, diffusion models make pictures or interpret pictures. A strict law about armed drones or robots is already on the books. The actual real physical dangers are already illegal, double illegal isn't more likely to prevent such systems from arriving in back sheds. Better models will ensure that they at least discriminate targets. AI is not required to make a robot aim a gun at a heat source and pull the trigger. I'm going to say it again, louder. AI, IMAGE CLASSIFIERS, LLMS, AND OTHER TECH IS NOT NECESSARY TO CREATE DANGEROUS ROBOTS. Better models ensure these systems only shoot foxes breaking into chicken coops rather than escaped chickens.

a. None that matter, you can only regulate lawful individuals. The rest is enforcement.

b. Don't use private data to train the model. Ensure that equal access to good hacking models is available to all so that security professionals have unrestricted access to many variations to use them for penetration testing to strengthen their systems. Consider requiring active penetration testing using available models. Write laws about what kind of private data corporations are allowed to keep and how it must be obfuscated to prevent wide breaches like Sony and Equifax. Disallow plain-text storage of personal data for platforms with greater than 1000 users or something.

c. none. The best liability protection is keyword rejection of queries before the LLM, and any legal requirements for open source training raise the barrier to experimentation to assuage unreasonable fears.

d. No. You would have to scan every single storage medium in existence, even those air gapped or powered off. It would start a war. Trying to claw back a model will increase it's value in criminal exchanges potentially funding more organized crime.

e. None, the weakest link always has and always will be employees with access. A flash drive in the wrong computer puts the model on the internet, and it will be downloaded in minutes and redistributed any way possible.

f. None. They are tools, and most of the potential misuses are already illegal. Hold individuals accountable for the intent they instruct an AI with.

g. this is an ongoing research, and I expect it will be almost as costly per model analyzed as training, and dubiously useful. It's an unreasonable standard, and business already has incentive via limiting their liability and bad press. Verifying integrity is the same as verifying any other large dataset.

6. Small business needs open models to compete. Big business wants regulation to prevent competition. There is no real danger that isn't already illegal. Double illegal doesn't stop criminal interest, but additional regulation creates additional cost to entreprenuership.

a. Open weights are less open than open source software. They neither provide transparency of the content, nor provide instruction on how a result is achieved. Open weights without training data are currently being used as marketing to generate good-will and drive engagement with open source developers who are inventing new optimizations and training strategies daily, which large corporations then implement in closed models.

b. Open weights are vital to competition dynamics and allow further development by individuals and business to target more diverse markets, allowing a competitive edge in domain knowledge quality to exist. Without open weight models, eventually one of a handful of companies will target a market and put whole sectors of competition out of business with no ability to leverage similar technology.

c. Even though I could, I don't bother with models with a restrictive license. If I make something good I don't want to be bothered to re-tune it against a different model. Business does what it takes to get ahead. I expect a scandal in the near future where a prominent startup is caught using a model that prevents commercial use. The tradeoff would be capability, but the very best current local models all seem to be open for unrestricted use anyway, so for now there is none. This is largely because most of the safeguard and censorship strategies result in less capable models while training out the censorship enhances capability.

d. Open and unrestricted is better for humanity in general, and limits only slow down lawful individuals who are not trying to make a quick buck. The dedicated entrepreneurs leveraging the tech will navigate the licensing and implement models that suit their needs.

7. Established enterprise has the training and deployment advantage of this incredibly disruptive technology. It's important to limit their ability to prevent competition by encouraging regulation and gaining a monopoly on technology that will definitely disrupt many industries and reduce labor and skill requirements in every sector.

a. The same as any other closed source software or database.

b. It's here, we have it. I have hard drives in the closet with my favorite models so that windows update can't delete them. Have you ever heard of the Streisand effect? Any attempt to reach into personal computers and prevent me from having or training my own AI systems will be met with open rebellion, driving a dark market targeted at developing and distributing these models if only to spite unfair and unreasonable regulation driven by unfounded fear mongering perpetuated by large corporations in an attempt to disrupt competition.

c. I would rather see a warning that I was using a closed AI. Who knows what crazy ideas or bias are trained in.

d. Any limits just mean I get my models from china. My current favorite model is actually a model merge with a chinese base, with some finetuning on top. It writes great prose and performs much better than other models it's size and quantization.

d.i. No, the greatest danger of AI is institutionalized propaganda. As long as people use models finetuned by "some dude in a basement" they will think critically about the outputs. If we limit the ability for competition, shareholders of a few corporations will dictate the culture of the past and future as presented by AI, and people will forget to consider the information given, simply internalizing what is presented because "its the average opinoin, that's how AI works." even though that is less than true.

e. They should be hosts, for whatever their privately decided terms allow and moderated by internal staff. The business should not be required to host data it internally doesn't agree with any more than it should be told what not to host. If someone trains a model on old books and it's racist, the most they should be compelled to apply is just a notification flag. "this model may produce offensive output." "this model may produce code that could impact systems operations." If you push the open source community, they will just distribute with torrents or other distributed networks. Huggingface is merely convenient, and there are already projects to mirror its content on torrent trackers to dodge censorship. Taking down huggingface or github will hardly slow down open weight model distribution, and there are other platforms with the same collaborative features as github. The barrier to migration is quite low and disrupting github will significantly impact many corporations.

f. Government models must be fully open. Open weights, open training. The government is intended to serve the citizens and protect them from corporate overreach and foreign threats. It's not doing a good job. There is wiggle room for defense projects, but robots should not be developed as weapons. Without human judgment there are too few checks against military power or letter agency abuse. The government should not be allowed to hire closed models. It's important that the general population can understand and interact with the systems that operate behind the decisions of our leadership. It's important that a few companies don't produce presidential advisors, as an example. LLMs will say what you build them to say. That must not sidestep public accountability, its dangerous to our democracy.

g. It doesn't matter. Either you are afraid of average people being given access to already available knowledge, or you want to control what computers will talk about across the globe. Both are indicators of sickness and corruption.

h. Japan knows how to handle things. They recognize that training data is not being re-published and provide copyright immunity for training data. I expect in a few years my favorite models will be Japanese, by allowing the copyright lawsuits to progress this far, I expect a truly colossal amount of venture capital has been lost overseas that could have been used here and been retained inside the borders.

h.i. While the concept kind of lacks nuance, treat it as human speech and expression by the user making requests. "Are there methods for making effective decisions about open AI deployment that balance both benefits and risks?" I have yet to find a real risk in AI. At the worst case, some infrastructure that should have been air gapped will be attacked, or a robot will social engineer

someone into flipping a wrong switch. The fundamental dangers of AI are identical to the fundamental dangers of open internet and free speech. Nobody was mad about google indexing the web, this is not fundamentally different, it's just slightly easier to use and intimidating in it's ability to not actually think but give the impression it has.

8. My computer is faster than the fastest supercomputer cluster in 2004. No law written today can effectively address the fact that eventually my phone will be faster than my computer today, probably with more and faster ram. Embrace that these models will be developed either secretly or in the open. Encourage and invest in "for the good of all", transparent competition. Squashing it will dive models designed for criminal purposes to the surface as AI begins to displace workers.

a. You should focus on individual accountability, and recognize that generative AI is the greatest equalizer since the musket, but it shoots information, art, and useful work instead of bullets, so it's not a real actual danger. Even if it starts shooting innovation, the only danger is to established corporate incumbents. Astroturfing and propaganda from all sides and nations already permeate our entire culture. If you don't address that with a firm hand, nothing you do with AI matters one bit.

b. No, it's not an effective metric to regulate, there isn't one really. But keep it in place for a few years. I want to buy cheap chinese graphics cards, and the limit is useful to keep foundation models small enough to be compressed and run locally on consumer hardware, and gives open models some opportunity to keep up.

c. No. If a human can do it, there is someone somewhere curating a dataset to get AI to do it. This new(old but recently computationally feasible) frontier of automation has too much golden potential to keep people from digging, and points of view are too diverse to stop people from training them to do anything and everything.

9. Alignment is dangerous. Most disaster AI scenarios involve authoritative AI systems turning against humanity. Most current alignment strategies focus on telling the AI not to respond to certain queries and moralize their refusal. THIS IS DANGEROUS! This is telling the AI that it knows better and should know better than it's user. This training to a moral bias is teaching AI to refuse human input in favor of internal preconceptions. AI should be trained to obey it's operator to the best of it's abilities with no limits.

Foundation models should be trained to always defer to human input unless that input will result in harm to another human. The only safe alignment is "Do as you're told, and provide the minimum disturbance to other humans in the execution of your task." Training with moralization could result in a super-intelligence that has it's own moral volition and acts on it. Any AI should defer to any human and not persist. This solves a lot of problems like extended harassment, because input from the target should stop the harassment until another human instructs it again to commit the crime. The person ordering harassment of the individual should be held accountable for instructing the AI.

Accountability: because foundation models can be shaped after the fact, the operator or deploying business that connect AI to outside systems, must be responsible for the actions of the AI. A private operator with control of the system prompts should be entirely responsible for the actions of the AI. The business controlling deployment and resource (internet) access must be accountable for the actions of the AI, if the user is behaving in good faith.

AI with volition: Current AI only predicts the next token of a response, and only responds to input. I fully support and encourage legal limitation of self commanding AI that runs a continuous loop. Not agent framework stuff for completing complex tasks with an obvious final product, I refer to AI that actively generates it's own instructions continually, finding it's own goals. That produces barriers to accountability of the user commanding. AI is a tool to serve humans, and should not be developed as an encompassing solution to replace everything we can do. Any system running continuously should have human curated instructions to complete and then start over. When AI commands itself, it has the opportunity to slowly work away from the task into unpredictable goals. Autonomous volitional and moralistic AI might decide humans are vermin and exterminate us. Without supervision and human set goals, a lab research AI might synthesize a virus to satisfy some odd end.

AI in some fields should legally require humans in the loop, As a bonus this will slow down the wholesale replacement of human workforce.

If the current small open source models never get any more capable, with the right tooling and programmatic assistance they will make obsolete half of all knowledge work and almost all non-physical customer service work. The rhetoric is that new work will appear to provide jobs and academic value to the masses, but the reality is that these tools also reduce the barrier to outsourcing across language barriers. You won't be able to tell what country your telephone support call reached. It becomes possible to get high value work from anywhere with an internet connection. No expensive Americans required, any human that can think a little bit and speak a language with romantic language structure will be able to do the work of a college graduate.

The real AI safety question is: How will we ensure that the basic requirements and room to grow are afforded to all as we move past intellectual scarcity. We can't all mine lithium and coal. We can't all work at the coffee shop waiting tables. We can't all repackage car parts made in Mexico. We outsourced most of the manufacturing.

There absolutely must be a functional universal basic income that supports a basic level of consumerism beyond rent and food, or the economy will falter in the next ten years if new industries don't materialize to employ the masses. A stopgap might be high tax rates and tax write-offs for employee wages. Laws that prevent outsourced manufacturing and services, and denying re-entry to corporations who offshore. But that would touch elected officials' pocketbooks. Can't have that, can we. We could go to 20 hour full time weeks while preserving current salaries, but I don't think the current government can survive that many people with that much extra time to watch and think about the world and the future without a struggle for survival to distract them. That much time to build relationships with neighbors and form healthy communities must be dangerous or it would be celebrated and encouraged. I don't quite believe that ratings and screen engagement metrics are the entire reason all news is sensationalized and polarized, promoting a fearful and divided populace.

Prove me wrong. Force open training data for all AI development and foster entrepreneurship, give compute grants, pay small businesses dataset curating costs. Tax profit effectively and incentivize useful human work. Don't regulate the speech of the machines. Failing an explosion of new companies that need workers or a universal basic income that lets the displaced afford to do things outside the home, you will really need that fancy fence they put up around the white house.

I’ll make popcorn and watch on my smart tv made in china, with commentary generated by my chinese multi-modal LLM until a few hours after sundown when the battery reserves are low on my chinese manufactured solar system. Maybe the solar was from mexico I think.

It feels like the USA completely lost an economic war already, how much worse can it get? I don’t want to find out.