website: www.stopai.info
email: contact@stopai.info
socials: @stopai_info

Stop AI Proposal

Overview

What Does Stop AI Want?

Stop AI wants governments to permanently ban the development of smarter-than-human AI, also known as Artificial General Intelligence (AGI)[1].

Introduction

Stop AI is a non-violent civil resistance organization working to permanently ban the development of AGI to prevent human extinction, mass job loss, and many other problems. We believe human thought should always remain at the forefront of intelligence. Unfortunately though, the AI tech companies in San Francisco are trying to build Artificial General Intelligence (AGI), an AI that is generally more intelligent[2] and economically productive[3] than any one human across all domains, which can quickly morph into Artificial Superintelligence (ASI)[4], an AI that vastly outperforms all of human intelligence, in other words, digital AI god.

Our intent with protests and civil-resistance is to raise awareness, among the general public, about the fact that the AI tech companies in San Francisco say they are trying to build AGI,

even though they themselves admit AGI has a real chance of causing human extinction, and some AI experts say AGI could be here this year. AI is dangerous and AI safety is an illusion; there is no way to prove experimentally that the AGI will never want something that will lead to our extinction, similar to how we have caused the extinction of many less intelligent species. AGI also has the power to permanently destroy democracy, which the AI companies are not quick to admit. They do however admit that AGI will take most jobs, which will cause a lot of people who enjoyed their work to lose purpose in life and cause all the other problems that mass unemployment causes.

Some people working in AI might say we should build AGI/ASI so we can ask it to cure cancer and fix climate change for us, but it does not make sense to build AGI/ASI for those benefits if we can not have proof that it will stay safe forever and will never want something that will lead to our extinction. It is impossible to have experimental evidence showing AGI/ASI will stay safe forever, so we should never build it.

What Are Our Demands To Governments?

1. We want governments to force AI companies to shut down everything related to the creation of general-purpose AI models, destroy any existing general-purpose AI model, and permanently ban their development. We currently define "general-purpose" as any model with capabilities equivalent to or exceeding Open AI's Chat GPT-4, due to this model's ability to perform multi-domain tasks[6]. In other words, it can learn to perform tasks that it was not specifically programmed to perform, which could in principle lead to Recursive Self-Improvement (RSI), i.e., the ability to automate AI research & development, which could quickly lead to ASI.
   a. As a start, any artificial neural network that was trained, or is being trained, using a sum of computing power that amounts to more than 10^{25} Floating Point Operations (FLOPS) should be immediately destroyed and banned from being created. Chat GPT-4 was trained on about this much compute power[10]. Using FLOPS as a metric for compute and then regulating based on that metric is called a "FLOP limit" or "compute cap" and will eventually become obsolete. Improvements in algorithmic efficiency will eventually force the compute cap of 10^{25} FLOPs to have to come down, or we will have to find some other metric, such as weight number limits, to base regulation on AI around. As a preemptive protection against algorithmic efficiency improvements, the compute cap for systems using artificial neural nets should be set to 10^{12} FLOPs, or 1 teraFLOP, which is similar in power to the IBM Watson computer that won Jeopardy in 2011 with 80 teraFLOPs.[11] This will allow some innovation of domain-specific AI tools while keeping a safe distance from general-purpose AI tools (at least for a little while). Any systems using neural nets that were trained using more than this amount of compute should be slowly phased out to prevent infrastructure collapse, but infrastructure collapse is better than human extinction, so a swift implementation of the 10^{12} FLOP limit for anything using neural nets, is better than a slow implementation.

2. Disband any company solely devoted to the creation of general-purpose AI, and any company with a subdivision devoted to the creation of general-purpose AI should have that subdivision disbanded.

3. Track computer hardware the same way we track supplies to build nuclear weapons. Prohibit a company's access to hardware if they try to build general-purpose AI.

4. Create a US federal agency solely devoted to protecting the American people from the harms of AI.
   a. The members of this agency must not benefit from AI's development.
   b. The members of this agency must be technically able to understand AI development at an expert level without an AI company needing to interpret a new development for them, which will end the government's need to rely on AI companies to self-report their developments. Part of the current problem is companies are spitting out AI developments way faster than people in government, who are often AI novices, can realistically understand those developments and then authorize their deployment.
   c. Most importantly, the members of this agency must believe the AI Control Problem is impossible to solve, i.e., the problem that tries to figure out how to control the behavior of an AGI/ASI. If the members of this new agency do not have this characteristic, then they will fail to take the appropriate actions to permanently ban AGI's development.

5. Ban the following current harms from AI:
   a. Ban the use of domain-specific and general-purpose AI tools in a given profession if the majority of professionals in a given field vote to not use an AI tool for a given purpose to preserve their job. Ideally, every profession should unionize to be able to take this vote.
   b. Ban the use and creation of AI-generated images, AI-generated video, AI-generated audio, and AI-generated text entirely. Destroy all existing AI-generated content everywhere. This will eliminate many problems with deepfake technology (another word for generative AI technology). A few of these problems include:
      i. Deep fake technology creates a world where we have no way of knowing what is real anymore online, on TV, in articles, on our phones, or on any other information source. You don't even know if this proposal is AI-generated or not. Generative AI can create fake videos[12], images[13], audio[14], or text[15] of anyone saying anything.
      ii. Generative AI can create pornography where the face of an underage person, or a nonconsenting person, is seamlessly pasted onto a naked body.[16]
      iii. Generative AI can create seamless scams that are nearly impossible to tell if they are fake or not. It will eventually become impossible to tell.[17]
      iv. Generative AI can create fake videos and messages of politicians,[18] causing the democratic majority to not know who they are voting for, which will undermine democracy.
      v. AI-generated content also blatantly ignores all intellectual property and copyright law by using protected material to train the AI. Generative AI blatantly steals the work of writers,[20] actors,[21] and artist[22] by training the model on their work, and then produces an output that is identical, or nearly identical, to the original.
   c. Ban the use of AI in mass surveillance and censorship.[23]
   d. Ban the open-source proliferation of large AI models.[24]

e. Ban the use of AI in specific domains where there is a larger net negative than net benefit of using AI in that domain due to bias in a model's training data, causing the model to misclassify people. An example of this would be using AI for insurance claims where the AI misclassifying someone's claim based on race, age, sex, or overall health, could falsely deny them service with no recourse.[26]

6. Preserve democracy by ensuring AI never infringes on the right to vote, right to assemble, or right to free speech.

7. Finally, we want governments to just listen to the people's expressed will as shown in preliminary polls by the AI Policy Institute (AIPI) that reveal the public majority already wants AI development to stop.
   a. AIPI conducted bipartisan polls in 2023 that showed 82% of Americans don't trust AI companies to self-regulate[27], 72% of Americans want the development of AI to slow down[28], and most importantly **63% of Americans said, "Yes, regulators should aim to actively prevent AI superintelligence."**[29]

How Does Stop AI Plan To Achieve A Permanent Ban On AGI?

To give a little context, A study[30] showed that engaging 3.5% of a country's population in active peaceful protest has a very high success rate at overthrowing a tyrannical government. This study also showed that peaceful protest is nearly twice as effective as armed revolt when attempting to overthrow a tyrannical government. And make no mistake, our government in the US is tyrannical because they refuse to protect us against the harms of AI. This 3.5% statistic is only meant to show the power of peaceful protest. We prefer to have amicable relationships with governments and government officials.

So our plan is very simple. We plan to hit that 3.5% population threshold by spreading the word with peaceful civil disobedience and traditional protests. The previously mentioned polls by AIPI make this goal actually seem possible. If we get that 3.5% population threshold we can stop AI development.

It's also important to note that a group called Just Stop Oil in the UK, who use roadblocks, company entrance barricades, and other civil disobedience tactics just won their first demand in the UK parliament. They successfully pressured the UK parliament into not issuing any new oil and gas licenses to UK oil companies. They accomplished this victory with way less than 1% of the UK population, so the tactic of using roadblocks and company entrance barricades, which we also use, does work in 2024 for achieving high level political demands.

We could very easily stop the development of Artificial General Intelligence (AGI) if a small group of people repeatedly barricaded entrances at AI company offices and data centers. our criminal records doesn't matter if we're all dead.

We will never turn violent. If someone does do something violent we will condemn them. Violence will not work and peaceful civil disobedience may not work either, but it is worth trying if we are all going to die. The study by Erica Chenoweth also showed that non-violence is twice as effective at achieving high level political demands compared to using violence to achieve the same goal.

You may be wondering about China and Russia. Surprisingly there was a poll done by AIPI that said 67% of Americans said, "Yes, we should restrict how powerful models can become... [even if] doing so risks making American companies fall behind China."[31] Stop AI plans to hit the previously mentioned 3.5% threshold with peaceful protest in the US first, then focus on China and Russia. All the companies leading the AI race are here in the US, so we have to stop it here first. That's not to say we don't have a plan for China and Russia. We know that the people of China and Russia do not want to be killed by AGI, live under an AI dictatorship, or live in a world where they can't feel useful. We have faith that they will want what we want. Once we stop AI here in the US, which will be difficult, we plan to then do everything in our power to help the Chinese and Russian people peacefully resist their governments if said governments fail to listen to their people. We should be prepared to die in that peaceful resistance. We are not guaranteed to succeed, but we have to try if the development of AGI continues in China because it is going to cause our extinction. We are currently working to establish contacts in these countries, and we could use help on this front.

Does Stop AI Want to Ban All AI?

Not necessarily, just whatever is necessary to keep humanity alive.

Time Remaining Until AGI's Arrival

Here is a quote from an interview with ex Open AI CTO Mira Murati on June 20th, 2024, which sums up the terrifying rate of AI progress, "If you look at the trajectory of improvements, systems like GPT-3 [released June 2020] were maybe toddler-level intelligence and then systems like GPT-4 [released March 2023] are more like smart high-schooler intelligence, and then in the next couple of years, we're looking at Ph.D intelligence for specific tasks. Things are improving pretty rapidly." Here are some quotes from AI experts on their expected timelines for the arrival of AGI:

- Elon Musk, in a New York Times interview at minute 52:00, said, **"Smarter than any human, less than 3 years"** (*said in Nov 2023)[61].
- Elon Musk, in an interview on X with Norway wealth fund CEO Nicolai Tangen at minute 1:30, said, **"Smarter than any one human next year, and then smarter than all humans in 5 years"** (*said April 9th, 2024)[62].
- Sam Altman, CEO of OpenAI, in a Wall Street Journal interview at minute 4:00, said, **"Less than a decade"** (*said October 21st, 2023)[63].
- Sam Altman, CEO of OpenAI, in a Y Combinator video when asked what he is excited for _in 2025_ he said, "AGI, I'm excited for that […] we basically know what to go do." (*said Nov 8th 2024)[63].
- Shane Legg, lead scientist at Google DeepMind, said, **"50% chance of AGI arrival by 2028"** (*said in 2011)[64].

We, and the experts above, may be wrong in predicting that AGI will arrive in the next year or two, if we do nothing to stop it, but the important thing to remember here is we can be wrong on long timelines, 'timelines' meaning time left until AGI's arrival, but we can not be wrong on short timelines. If five years goes by and we don't have AGI, then that is fine, but if AGI is actually going to be here in a year or two then we have to do everything possible, short of violence, to

stop it now because there is no turning off a superintelligence. The point of no return, or the metaphorical cliff edge, is when AI can invent new AI on its own, better than the best human AI R&D expert. This can lead to an intelligence explosion where the AI recursively self improves (RSI) it's own design. This is what makes AI go from AGI to ASI. According to AI experts AI is already at the 50th percentile on AI R&D tasks compared to human AI R&D experts[65] and OpenAI's o3 model is the 175th best programmer in the world[66]. It seems like AI could start doing RSI any moment now. We are on the edge of the cliff and are about to step off. We have to stop.

Acronyms

AI: Artificial Intelligence
AGI: Artificial General Intelligence
ASI: Artificial Superintelligence
AIPI: AI Policy Institute
FLOPS: Floating Point Operations
GPT: Generative Pre-trained Transformer
RSI: Recursive Self Improvement

References

1. http://www.scholarpedia.org/article/Artificial_General_Intelligence
2. https://openai.com/about/
3. https://openai.com/charter/
4. Nick Bostrom's 2014 book "Superintelligence" page 23 https://books.google.com/books/about/Superintelligence.html?id=7_H8AwAAQBAJ
   a. https://www.ibm.com/topics/artificial-superintelligence
5. https://www.ibm.com/topics/neural-networks
6. https://community.openai.com/t/chatgpt-goes-multimodal-sound-and-vision-is-rolling-out-on-chatgpt/394501
   a. https://hgs.cx/blog/multimodal-chatgpt-what-can-it-do-and-what-are-its-real-world-applications/
   b. https://en.wikipedia.org/wiki/GPT-4#:~:text=It%20was%20launched%20on%20March,the%20free%20chatbot%20Microsoft%20Copilot.
7. https://www1.udel.edu/educ/gottfredson/reprints/1998generalintelligencefactor.pdf
8. https://en.wikipedia.org/wiki/G_factor_(psychometrics)
9. https://www.anthropic.com/news/claude-3-family
10. https://epochai.org/blog/tracking-large-scale-ai-models
11. -
12. https://x.com/MoistCr1TiKaL/status/1758329918683779411
13. https://www.cbsnews.com/news/pope-francis-puffer-jacket-fake-photos-deepfake-power-peril-of-ai/
14. https://www.youtube.com/watch?v=pJZYd_65xs4
15. https://chatgpt.com/
16. https://x.com/ai_ctrl/status/1737171648275423322?t=ZJWzyayh6ilJM4gqTqRHcA&s=19
17. https://www.youtube.com/watch?v=pJZYd_65xs4
18. https://video.ucdavis.edu/media/Deep+Fake+of+Barack+Obama/1_6zmvebuf
    a. https://www.youtube.com/watch?v=4GDvuMkMOXM

19. -
20. https://www.theverge.com/2023/9/20/23882140/george-r-r-martin-lawsuit-openai-copyright-infringement
21. https://www.hollywoodreporter.com/business/business-news/scarlett-johansson-ai-legal-threat-1235905899/#
22. https://x.com/StopAI_Info/status/1784457621635817955?t=LZPsQIaAhAybEDaOn6F_lw&s=19
23. https://edri.org/our-work/how-to-fight-biometric-mass-surveillance-after-the-ai-act-a-legal-and-practical-guide/#:~:text=One%20of%20the%20most%20high,track%20people%20across%20public%20spaces.
24. https://metaprotest.org/
25. -
26. https://www.prolific.com/resources/shocking-ai-bias
27. https://www.vox.com/future-perfect/2023/8/18/23836362/ai-slow-down-poll-regulation
28. https://www.vox.com/future-perfect/2023/8/18/23836362/ai-slow-down-poll-regulation
29. AIPI Sep 02, 2023, poll crosstabs, page 39:
    a. https://drive.google.com/file/d/1n0pXDBuIcb01tW4TQdP1Mb5aAiFDvWk0/view?usp=drive_link
    b. https://www.vox.com/future-perfect/2023/9/19/23879648/americans-artificial-general-intelligence-ai-policy-poll
    c. https://theaipi.org/poll-shows-voters-oppose-open-sourcing-ai-models-support-regulatory-representation-on-boards-and-say-ai-risks-outweigh-benefits-2/
30. https://www.nonviolent-conflict.org/resource/success-nonviolent-civil-resistance/
31. AIPI Sep 02, 2023, poll crosstabs, page 31
    https://drive.google.com/file/d/1n0pXDBuIcb01tW4TQdP1Mb5aAiFDvWk0/view
32. -
33. https://www.larksuite.com/en_us/topics/ai-glossary/interpretability-in-ai-and-why-does-it-matter#
34. https://aiimpacts.org/2022-expert-survey-on-progress-in-ai/
35. -
36. https://en.wikipedia.org/wiki/Instrumental_convergence
37. https://www.huffpost.com/entry/artificial-intelligence-oxford_n_5689858
38. https://abcnews.go.com/Politics/emerging-ai-technologies-make-easier-bad-actors-conceptualize/story?id=109749169
39. https://encyclopedia.pub/entry/35791
40. https://www.youtube.com/watch?v=3TYT1QfdfsM
41. https://en.wikipedia.org/wiki/Three-body_problem
42. https://en.wikipedia.org/wiki/Chaos_theory
43. Nick Bostrom's 2014 book "Superintelligence" page 62
    https://books.google.com/books/about/Superintelligence.html?id=7_H8AwAAQBAJ
44. https://openai.com/index/introducing-superalignment/
45. https://www.youtube.com/watch?v=ZeecOKBus3Q
46. https://www.youtube.com/watch?v=bJLcIBixGj8
47. https://www.youtube.com/watch?v=myR882u2izg
    https://tonyzhaozh.github.io/aloha/
48. https://koopingshung.com/blog/turing-test-is-obsolete-bring-in-coffee-test/
49. Nick Bostrom's 2014 book "Superintelligence" page 62
    https://books.google.com/books/about/Superintelligence.html?id=7_H8AwAAQBAJ
50. https://www.dhs.gov/news/2024/04/26/over-20-technology-and-critical-infrastructure-executives-civil-rights-leaders
51. https://en.m.wikipedia.org/wiki/Existential_risk_from_artificial_general_intelligence

52. https://www.youtube.com/watch?v=PbUCu1HBMjE
53. https://openai.com/about/
54. https://openai.com/charter/
55. https://community.openai.com/t/sam-altman-s-universal-basic-income-experiment-openresearch/877705
56. https://youtu.be/M24Nnd6ihFA?si=CUD31LTXDamwU-7p
57. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6392452/#:~:text=In%20a%20systematic%20review%2C%20Roelfs,net%20of%20covariates%20%20%5B23%5D.
58. Iain M. Banks' 1987 book Consider Phlebas page 452
59. https://www.lesswrong.com/tag/coherent-extrapolated-volition
    a. https://www.lesswrong.com/tag/coherent-extrapolated-volition/history
60. https://x.com/tsarnick/status/1803901130130497952
    a. Full video interview: https://www.youtube.com/watch?v=yUoj9B8OpR8
61. https://youtu.be/2BfMuHDfGJI?si=jO-KAH13hmR_MIKK
62. https://youtu.be/_rQBZ3vKRA0?si=SDZ4YXQWpJ9jJsiF
63. https://youtu.be/byYlC2cagLw?si=Uk1_tiBN7kwFIhu2
    a. https://youtu.be/xXCBz_8hM9w?si=hb1kwLHvKZV_vH8G
64. https://futurism.com/google-deepmind-agi-5-years
65. https://x.com/eli_lifland/status/1860087262849171797?t=PZet1VlAQGoXgPYBIK39OQ&s=19
66. https://www.reddit.com/r/singularity/s/rxju6L8ko7

Written: 08/29/2024
Edited: 02/02/25

website: www.stopai.info
email: contact@stopai.info
socials: @stopai_info