

AI INCONVENIENT DOOM

A FILM BY NIKI DROZDOWSKI

WHY AI COULD KILL US ALL
(AND WHY NOBODY WANTS TO BELIEVE THAT)





2023 and 2024 will be by far the most important and consequential years in the entire history of the human species. Because how we approach and handle AI today will determine, if we are about to enter a new era of unimaginable possibilities and prosperity – or if we will simply cease to exist.

A cinematic, dystopian cityscape under a hazy, grey sky. Several large, dark, multi-rotor drones or flying vehicles are suspended in the air, some with red lights visible. The ground is a vast, flat, industrial or military-style area with various structures, vehicles, and patches of green grass. In the background, a city skyline with tall, futuristic buildings is visible through the haze. The word "PREMISE" is written in large, bold, red capital letters on the left side of the image.

PREMISE

There have been a lot of movies about AI in the past – but quite ironically the topic of existential AI risk has so far been explored to much greater length in Hollywood sci-fi movies than in the few serious documentaries.

And even if existential AI risk was mentioned in non-fictional works, it was often treated as “one of the many risks” or even as “a distraction from the real problems”. Unfortunately, this sentiment is more and more echoed by the press recently.

Even though AI scientist signed a dire statement about the existential dangers, the warning voices of the AI safety community can so far only be heard in disconnected YouTube interviews, presentations or podcasts. Therefore a documentary that takes x-risk seriously and gives it its deserved center stage is sorely needed.

“An Inconvenient Doom” is this documentary.



FOCUS

One of the challenges of x-risk is that it is extremely hard to explain. The documentary could easily spend all of its 90 min. going through the important problems and subproblems of AI safety and would still only scratch the surface. So it will take a different approach.

There still needs to be some explanations, but these will be more of an overview and invite the audience to explore the topics in greater detail on the project's website, where the interviews with the experts are shown in full and where links lead to more detailed explanation videos.

Besides the explanations the film will focus on three areas: The emotional toll this topic has on the scientists and experts [who in some cases have been warning about x-risk in vain for more than 10 years], the challenges for the necessary international discussion about x-risk [that already manifest in a strange fight between AI safety and AI ethics experts] and a look at the future and the possible solutions for humanity. Or as a snappy alliteration: **E**xplanation, **E**motion, **E**ntrenchment, **E**xtinction [?]

Explanation (45%)

The AI experts touch on the basics of x-risk in laymen's term: The alignment problem, interpretability and the black box, convergent instrumental goals, the stop button problem, deception, timelines specification gaming, emergent capabilities, mesa-optimizers (maybe), problematic humanization, and simple solutions that don't work. The explanations are accompanied by the CGI graphics and animations to convey the complex concepts.

Emotion (20%)

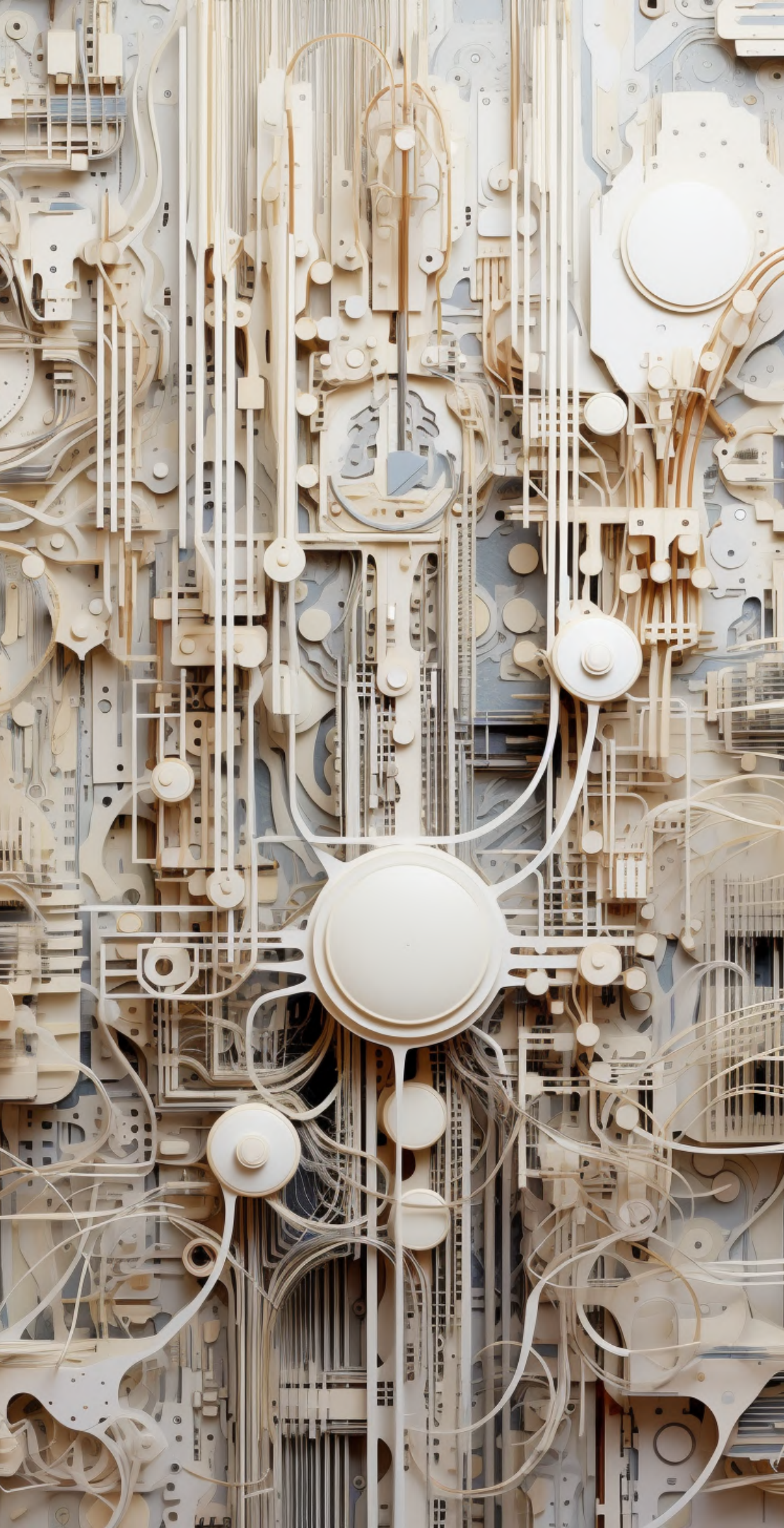
With such an extremely theoretical and technical topic the audience needs an emotional anchor to make it "real". Therefore the AI experts talk about their hopes and fears for the future (e.g. the first AI safety summit in the UK in November) and the toll of their Cassandra-like experience of not having their warnings taken seriously. They also share which moment in time made x-risk tangible for them and show us how they like to take a break from all the impending doom.

Entrenchment (25%)

The beginning polarization of the public AI discussion is preemptively put into context for the audience (AI safety vs. AI ethics) and the different sides of the argument are explained. Also sociologist and social psychologists talk about the reluctance of the human mind to think and talk about problems on this enormous scale, crisis fatigue, and elaborate on the mechanism of avoidance and denial – and what can be done to counter them.

Extinction? (10%)

At the end of the film the AI experts and the sociologists take a look into the near and far future and ponder on the critical importance of the next 3 years and the uphill battle to make AI safe or stop it altogether before it reaches dangerous levels of intelligence (-> Overton window). Addressing the camera (and the audience) directly they make a hopeful call-to-action that time is running out and humanity needs to work together now.



STRUCTURE

To navigate all these different themes the documentary will employ two structuring methods: A narrator/presenter and a guiding statement at the beginning of the movie.

The narrator will only be used sparsely but is needed to introduce the audience to the topic in general, provide segues between segments and provide an overview of the public discussions. If budget allows it the narrator could be a famous actor/actress and show up in person as a presenter. Ideally someone whose former roles have a connection to sci-fi or apocalypse: Benedict Cumberbatch (Marvel), Keanu Reeves (Matrix) or Leonardo di Caprio and Jennifer Lawrence (Don't Look Up). But also a cast member from "The Big Bang Theory" could work wonders for credibility and publicity.

The guiding statement at the beginning is *"This movie will be difficult!"* Because the topic of x-risk is difficult on various levels: It is difficult

1. to bring up
2. to consider possible
3. to explain & to understand (-> Explain)
4. to emotionally process (-> Emotion)
5. to agree on (-> Entrenchment)
6. to act on (-> Extinction)

These difficulties work as a chapter structure for the movie which the narrator uses to guide the audience through the process of understanding x-risk. #1 and #2 are an introduction and are mostly handled by the narrator alone, #3 is mainly done by the AI safety experts while the other points bring in the sociologists as well.

PRODUCTION & DISTRIBUTION

The production company has secured an initial funding of 20.000 \$ which guarantees that the film can already be made in a basic form. Additional funding is planned to come from three sources: German film funding (federal production and postproduction funding), a minimum guarantee for a distribution company and crowdfunding platforms.

Additional funds can be used for making the film better through (in no specific order):

- a more prominent narrator
- more archive material & news footage licences
- better CGI
- custom-made music
- 5.1 sound mixing
- better color grading
- drone footage
- paid ads & marketing

A free release to YouTube was considered, but the film would suffer a heavy credibility loss from it, since there are also free documentaries about flat earth and chemtrails on YouTube. Therefore the goal is a worldwide distribution through a classical sales agent ideally aiming at a theatrical release.



INTERVIEW QUESTION EXAMPLES

(FOR AI EXPERTS, NON-EXHAUSTIVE)

Explanation

- What is the alignment problem?
- How does a neural network learn?
- Why is AI a black box?
- What are convergent instrumental goals?
- Why is the stop button problem important?
- How dangerous are emergent capabilities?
- Why shouldn't we compare AI to humans?
- Why don't simple solutions work?
-

Emotion

- When was the first moment when it hit you that x-risk is real?
- Have you understood x-risk emotionally?
- What is your P[doom]?
- How is your daily life looking like?
- What has changed about your outlook of the future?
- Do you allow yourself to be sad? Or hopeful?
- How often do you think about it on any given day?
- What do you do if you need a timeout from it all?

Entrenchment

- What kind of doubts you face?
- What is the biggest hurdle for people to understand x-risk?
- Are you happy with the public discussion / media coverage?
- Why are AI safety and AI ethics not getting along?
- Could you be wrong? Do you wish to be wrong?
- What is TESCREAL? Do they have a point?
- Are you a Luddite?
- How could both sides come together?

Extinction?

- What are your hopes and fears regarding multi-modal models?
- Is alignment possible?
- What is the best current alignment strategy in your opinion? Why?
- What is the most dangerous alignment strategy right now?
- What would convince you, that alignment has been solved?
- What needs to happen so humanity has a chance?
- What can the audience do right now?



DIRECTOR'S STATEMENT

Please take this pitch document with a grain of salt. This is the first version, and I didn't want to loose any valuable time over-thinking or over-designing it. So a lot of aspects – especially on the art direction side – are still under construction. As it is normal for documentaries the project will undergo quite a lot of changes in the editing process as new ideas and patterns form, new information from the interviews influence the narrative and even new developments in AI science emerge.

Therefore the exact interview partners, the balance between the four main narration avenues and especially the art style and presentation will change in the final movie.

But rest assured I won't change the general direction of the project: I want to create a documentary that

- alerts the public to the problem.
- is not aiming to be "balanced", but clearly sides with "x-risk is real and urgent"
- tries to convey the problem not only by explaining the science but also by highlighting the emotional impact on the scientists and by framing the expected hurdles in public discussion.
- in the end is a call to action for the public to engage in this discussion and take democratic action.

CONTACT US

EMAIL

niki@aninconvenientdoom.com

MOBILE

+49 163 1645437



LOOK UP FILMS

