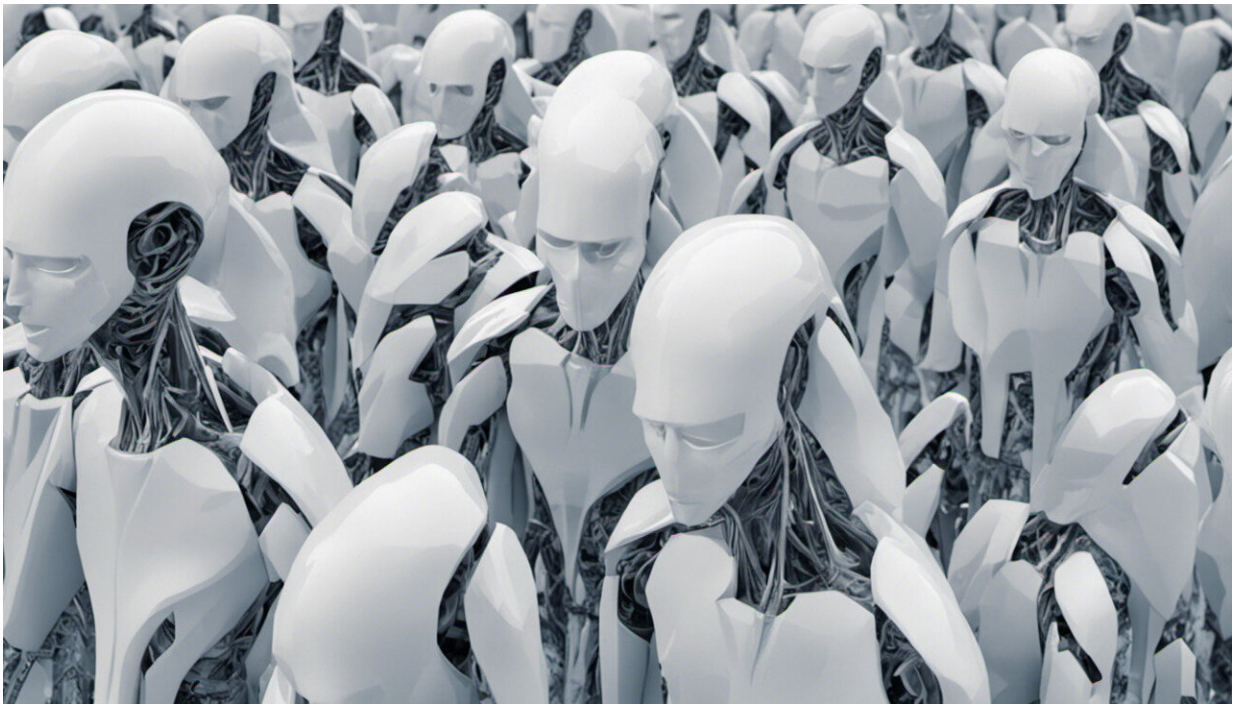


# The drive towards ethical AI and responsible robots has begun

December 16 2015, by Sean Welsh

---



Credit: AI-generated image ([disclaimer](#))

Robotist Sabine Hauert, from the Britain's University of Bristol, [wrote in Nature](#) earlier this year:

*My colleagues and I spend dinner parties explaining that we are not evil [...]*

People are worried, she said.

*They hear a mostly one-sided discussion that leaves them worried that robots will take their jobs, fearful that AI poses an existential threat, and wondering whether laws should be passed to keep hypothetical technology 'under control'.*

These fears are not helped by a continuing epidemic of artificial intelligence (AI) and robophobic screenplays emanating from Hollywood.

It is hard to give examples of recent mainstream films with robots and AIs in them that are not infected with AI-phobic ([Transcendence](#), [Eagle Eye](#)) or robophobic ([Oblivion](#), any of the [Robocop](#) movies, [Ex Machina](#)) scenes.

The indie caper picture [Robot and Frank](#) and the relatively mild dumping of the human by the AI in [Her](#), who then runs off to hang out with cooler Alan Watts-based superintelligences, are the only ones coming to mind that do not succumb to the prevailing moods of [AI panic](#) and robophobia.

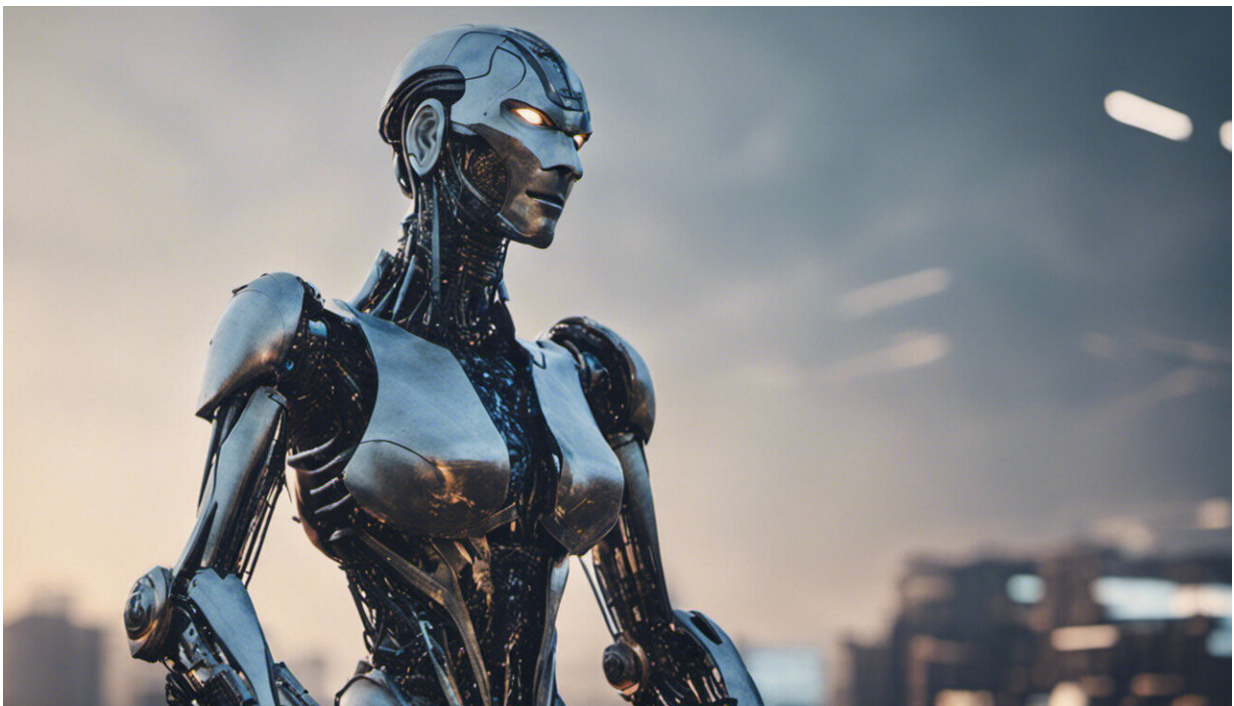
Robots and AIs do make good cinematic villains. However, in reality no one has much clue how to make a [robot](#) "want" or "feel" *anything* in a phenomenologically credible way as yet; let alone how to make them sociopaths hell bent on world domination and the extermination of *Homo sapiens*.

They are more likely to become innocently dangerous *idiot savants* than malevolent overloads seeking to get psychotic kicks by making humans "bend the knee".

Rule-driven robots play a mean game of chess but *feel* nothing about

winning or losing. They just pick moves that optimise a mathematical value function. Humans associate intelligence with desire but "desire" as [formally modelled](#) in the rulebook of a Turing machine is a very different thing from the combustive forces of fury, jealousy and "star-crossed love" that drive humans.

## Responsible robots



Credit: AI-generated image ([disclaimer](#))

Two new AI and robotics nonprofits launched over the weekend. In different ways, both are responses to public concerns about the safety of AI and robotics.

The first is The Foundation for Responsible Robotics ([FRR](#)), which

wants to "promote responsibility for the robots embedded in our society".

The FRR is headed by [Noel Sharkey](#) and various robotics experts. It aims to engage policymakers, create interdisciplinary teams of robotic, legal, ethical and societal scholars, work to explore what it means to be "responsible" as robotics researchers and designers, run workshops and engage the public.

Sharkey, who is known for his activism with the [Campaign to Stop Killer Robots](#), is concerned that "we are rushing headlong into the robotics revolution" without giving enough policy thought to the social problems that might arise.

He says governments are looking to robotics as a "powerful new economic driver" but "only lip service is being paid to a long list of potential societal hazards".

New technologies could cause mass unemployment or there might be an acceleration of social inequality caused by robots and automation leading to a society divided between robot-owners (living in gated communities) and a robot-less underclass (struggling on a brown burnt Earth) such as was depicted in [Elysium](#) .

The FRR wants to ensure that the public have confidence in robotics research and that robots will be developed with due regard for their human rights and freedom of choice.

## **AI for everyone**

Open AI is backed by Elon Musk, [Peter Thiel](#) and various technology entrepreneurs.



Films like RoboCop entrench fear of robots and artificial intelligence. Credit: Orion Pictures

Their focus is more on research and on making advanced AI freely available. They seek to develop innovations in "deep learning" a technique where rather than "hand-code a new algorithm for each problem, you design architectures that can twist themselves into a wide range of algorithms based on the data you feed them".

They aim to "advance digital intelligence in the way that is most likely to benefit humanity as a whole, unconstrained by a need to generate financial return".

When "human-level AI" arrives, Open AI feels that it is important that there be "a leading research institution which can prioritize a good outcome for all over its own self-interest".

They say "our researchers will be strongly encouraged to publish their work, whether as papers, blog posts, or code, and our patents (if any) will be shared with the world".

Open AI's backers have committed a billion dollars in funding though they expect to "only spend a tiny fraction of this in the next few years."

Open AI is mainly about open access to advanced AI thus reducing the risk of a world of AI haves and have nots. This is a good idea. If the team can keep competitive and advanced AI open source this should reduce the risk of people being excluded from advanced AI for financial reasons.

The Foundation for Responsible Robotics has a broader agenda of policy engagement and raising professional and public awareness of robot ethics issues. Again, this is a worthwhile endeavour. AI and robotics researchers tend to be hard scientists unused to ethical debate.

Scientists need to step out of the empirical and into the normative. As trusted thought leaders of the citizenry, they should cross the line between "is" and "ought" and participate in policy debate.

Hopefully both these groups will help provide cures for the current epidemics of AI panic and robophobia.

*This story is published courtesy of [The Conversation](#) (under Creative Commons-Attribution/No derivatives).*

## Source: The Conversation

Citation: The drive towards ethical AI and responsible robots has begun (2015, December 16)  
retrieved 9 April 2024 from

<https://phys.org/news/2015-12-ethical-ai-responsible-robots-begun.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.