**Bold, text background color, font and justified left aligned**

**Another common misconception is that the only people harboring concerns about AI and advocating AI safety research are luddites who don’t know much about AI. When Stuart Russell, author of the**[**standard AI textbook**](https://www.amazon.com/Artificial-Intelligence-Modern-Approach-3rd/dp/0136042597)**, mentioned this during**[**his Puerto Rico talk**](https://futureoflife.org/data/PDF/stuart_russell.pdf?x72900)**, the audience laughed loudly. A related misconception is that supporting AI safety research is hugely controversial. In fact, to support a modest investment in AI safety research, people don’t need to be convinced that risks are high, merely non-negligible — just as a modest investment in home insurance is justified by a non-negligible probability of the home burning down.[[1]](#footnote-1)**

*Italic, text background color, font and justified right aligned*

*Many AI researchers roll their eyes when seeing*[*this headline*](http://tinyurl.com/hawkingbots)*: “*Stephen Hawking warns that rise of robots may be disastrous for mankind.” *And as many have lost count of how many similar articles they’ve seen. Typically, these articles are accompanied by an evil-looking robot carrying a weapon, and they suggest we should worry about robots rising up and killing us because they’ve become conscious and/or evil. On a lighter note, such articles are actually rather impressive, because they succinctly summarize the scenario that AI researchers*don’t*worry about. That scenario combines as many as three separate misconceptions: concern about*consciousness*,*evil,*and*robots*.[[2]](#footnote-2)*

Underline, text background color, font and justified center aligned

If you drive down the road, you have a subjective experience of colors, sounds, etc. But does a self-driving car have a subjective experience? Does it feel like anything at all to be a self-driving car? Although this mystery of consciousness is interesting in its own right, it’s irrelevant to AI risk. If you get struck by a driverless car, it makes no difference to you whether it subjectively feels conscious. In the same way, what will affect us humans is what super intelligent AI does, not how it subjectively feels.[[3]](#footnote-3)

The fear of machines turning evil is another red herring. The real worry isn’t malevolence, but competence. A super intelligent AI is by definition very good at attaining its goals, whatever they may be, so we need to ensure that its goals are aligned with ours. Humans don’t generally hate ants, but we’re more intelligent than they are – so if we want to build a hydroelectric dam and there’s an anthill there, too bad for the ants. The beneficial-AI movement wants to avoid placing humanity in the position of those ants.[[4]](#footnote-4)

The robot misconception is related to the myth that machines can’t control humans. Intelligence enables control: humans control tigers not because we are stronger, but because we are smarter. This means that if we cede our position as smartest on our planet, it’s possible that we might also cede control.[[5]](#footnote-5)

Bookmarks [Q1](Question1.docx) [Q2](Question2.docx) [Q3](Question3.docx) [Q4](Question4.docx)

Pakistan (Auto Formatted)

1. Misconception about AI [↑](#footnote-ref-1)
2. Researchers Headline [↑](#footnote-ref-2)
3. Researchers Explanation [↑](#footnote-ref-3)
4. Machines are evil.? [↑](#footnote-ref-4)
5. Human vs Robot [↑](#footnote-ref-5)