Are we destined to be the house pets of the robot revolutionaries?

Steve Wozniak, co-founder of Apple, has recently given voice to his belief that humans will be treated like cherished house pets when the robot revolution finally takes hold. This view puts him firmly in the optimists' camp, alongside Nick Bostrom (philosopher and futurist), who has argued that strong artificial intelligence is the last invention humans will ever need to make.

To be fair to Bostrom, he is forthright in outlining the dangers inherent to developing such technology. During his TED talk in March, he warned that the conditions for creating strong AI would need to be set up "just right", to ensure a "controlled detonation". Without this, the artificial super intelligence might go rogue, with potentially damning consequences for our species. He cites the fable of King Midas as a "strong optimisation process" that achieves its end result via a ruthless application of command inputs, without concern for the wider repercussions, and without recourse to shutting the process down, once it is underway.

How to avoid such a nightmare situation? Assuming that someone, somewhere, will one day invent strong, fully self aware artificial intelligence (which remains a highly controversial idea), Bostrom argues that we must put in the groundwork now, to create robust safety mechanisms to ensure it works *for* us, rather than *against* us. The best way to do that, Bostrom contends, is to instill our most cherished values into the nascent mega brain.

Which begs the question: what *are* our values, exactly?

Back to Wozniak's idea of the pampered house pet. We all love our pet cats and dogs (and lizards, parrots and tarantulas, no doubt). It is indeed a beguiling idea that we might create, as our last invention, a being that takes care of our every whim, and provides for us unquestioningly, with kindness, and bottomless reserves of patience. A being that is not only *attentive* to our needs, but which is able to *anticipate* our needs; to know us better than we know ourselves. That is the future to which Wozniak refers, when he observes, with the lazy contentment of an overfed Siamese cat: "We're just going to have the easy life".

Sounds great. I, for one, am all in favour of *that* - if it pans out that way. There are of course alternatives to this beatific vision.

For clues as to where and how Wozniak's dream could turn into something more sinister, let us unpick the pampered pet analogy. It is true that we are good to our pets, by and large, and that we offer them a kind of unconditional love usually reserved for our human children. But we are also pragmatic beasts: My girlfriend, Claire, told me a story from her early childhood, about one of the sheep her dad used to keep in his field - of which she was fond - called Peg Leg. One day, little known

to Claire, her dad killed the animal, brought it home, and served it for dinner. He didn't tell her until years later that they had all eaten her woolly friend, and by then, of course, it didn't really matter. Now, it is an amusing anecdote.

See where this is going? How can we have the confidence that a super intelligent machine, inculcated by our world view and "values", yet unfathomably more complex than we are, will interpret our own contradictory and illogical moral programming any more successfully than we are able to? Nick Bostrom suggests that we should let the super intelligence learn from us, so that it might successfully integrate our values into its source code. Is that *really* such a good idea? Do we want it picking up this wryly comic tale about poor old Peg Leg and then deciding to use it as material for its own improvised routine, involving the sudden, inexplicable eradication of huge swathes of the human race?

It seems dangerous that we should position ourselves as moral arbiters in the instruction of machine super intelligence, and yet, what is the alternative? I would suggest - if this technology ever becomes available to us - we deploy it in highly specialised ways; "nodes" of intelligence that are responsible for certain things, such as overseeing manufacturing processes, or governing traffic flow at an airport. These individual machines would lack the holistic cognisance of a fully self aware AI, and be allowed to develop only in ways that are helpful in executing the prescribed function. Humans would remain in control.

Of course, in an interconnected world, such systems are likely to make contact with each other at some point, regardless of the limiting safety protocols imposed on them. Maybe we are destined to be replaced. But let's face the new challenges imposed on us by our own technology with clear vision, and without sentimentality. Because, whatever the intelligence potential lurking within microprocessors, and matter more generally, it is probably of the cool, pragmatic variety.

July 2015