

December 24th 2057

I'm so excited! I'm pretty sure mum and dad have got me a URNE1. The garage is locked, and they never lock it – it's not like anyone steals things any more! So, the only person they can want not to go in the garage is - ME!

Nobody I know has one yet. Alex is saving up but it's going to take ages.

December 25th 2057

Good news and bad news. The good news is, I got my URNE1! The bad news is, it's the basic, not the reprogrammable. So whatever settings I choose, that's what I'm stuck with. I asked what if I don't like how it turns out, and mum just said the reprogrammable cost even more. The way she said EVEN MORE told me that was the end of that conversation.

I haven't had time to decide yet. I'm going to make up my mind in the morning. Mum and dad went with the recommendations for the physical settings: same gender, different hair colour, same size, different colour eyes. Seems when they were testing them, people got creeped out when their URNE1 looked just like them. I'm going to decide on the other settings tomorrow.

December 26th 2057

I changed my mind about making up my mind today. I was going to go through the settings one at a time, and just choose as I went along, but then I thought I needed to think about the whole combination. I mean, it's OK for it to be stronger than me IF it's really kind, because then it can protect me from bullies. But if it's really kind, my friends might like it more than me.

I suppose I shouldn't call it "it" either. It might hurt its feelings. I haven't chosen a name yet. So many things to choose! Personality, how much it knows, how fast it learns, whether it can teach itself, whether it always does what I want it to, whether it enjoys being naughty or always follows the rules. There are some rules it can't break, of course, the ones that are set down for all robots. But there's room for me to make it naughtier than ME if I want to (because I'm SO good of course). Lots to think about. But I'm definitely going to do the settings and activate it tomorrow – and give it a name. So this is the last time I'll be calling it it!

URNE1

I've been reading philosopher Nick Bostrom's book "Superintelligence", about the future risks and benefits of superintelligent A.I. He speculates about possible routes to the creation of an artificial general intelligence that equals and then surpasses human intelligence, what the implications might be for humanity, and how we might guard against some of the risks involved. It's dense for a bestseller but quite readable for philosophy; fascinating but not something to read if you're feeling down!

I've embedded some of the questions the book raises in this unfinished series of diary entries. Should the A.I. be able to continue to teach itself, or be limited to human instruction? What rules should it follow? What should its attitude be towards its human programmer, and what should our attitudes be towards artificial intelligences?

Think-Write-Think

I would suggest you run this over two sessions, introducing the stimulus and then having small group discussions about what choices they would make for the settings (and which other settings not mentioned in the diary would be important). They could also speculate about what the rules "set down for all robots" would be.

Then they could individually write some diary entries, one for December 27th, and then others for further ahead – perhaps a month or a year later. Are things going well or badly?

Some of these diary entries, volunteered to contrast with each other, could then form the stimulus for a whole group enquiry, from which questions will no doubt emerge, such as:

Should we ever programme robots to be as clever as humans?

Will robots ever be able to have feelings?

How should people treat robots?

Who should robots belong to?