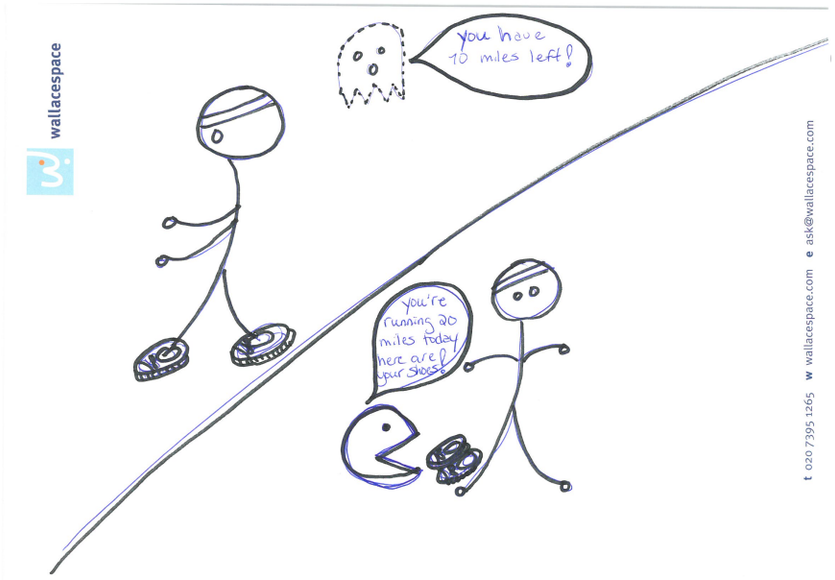
The Robot Code of Conduct

A 5-point guide for designers of
connected, autonomous intelligence

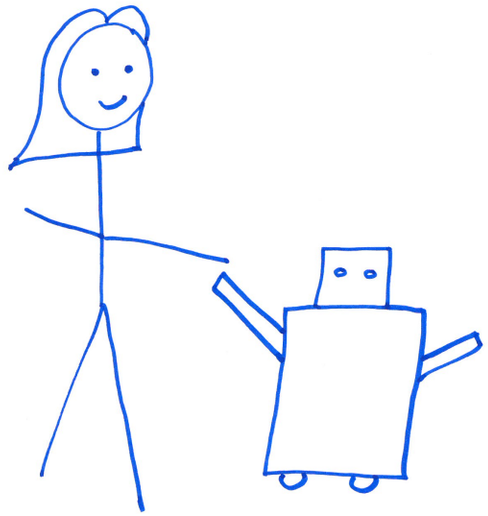
The process we underwent was meant to highlight some of the problems and challenges with connected autonomous intelligences (robots) in a very organic way, rather than running on assumption or prejudice.
The first session, being fairly blue-sky, uncovered not only the security challenges which need to be addressed in the design of these things, but also how little is actually known about how it all works together (or, in some cases, doesn't work at all).
The second session, being more mundane, let people get stuck in at a much deeper level and uncovered rather a lot of 'lines'. These are places where the balance tipped on questions such as value and convenience against privacy or the balance of effort required to manage the technology against the value it brought. These were often more detailed expressions of feelings people had at the end of the first session.
We began the third session by doing a brain dump of all the problems we'd uncovered in the first 2. They included everything from 'forgetting our humanity' to 'dealing with chaos and variance' to 'semantic frameworks'.
We then worked together as a group to come up with ways to counter these challenges. We clustered like with like and that brought us to the 5 rules discussed in the team presentation.
Interestingly, they cover only 2 or 3 of the Laws of Robotics. The idea of self-preservation for the robot was never explicitly addressed in our sessions, but I do think our 5 rules below give a bit more of a bridge to application.



Know yourself & know your place.
Keep the technology invisible or in the background as much as possible.
When physical, make sure it's clearly not human, yet still relatable.

Stay in the background
Doesn't HAVE to be physical - invisible unless necessary
Better to be unrealistic than too realistic - clear that it's not human or meant to be
Don't try to emulate or replace humans
Don't over-anthropomorphise
BUT should be comforting/comfortable/relatable

wallacespace



t 020 7395 1265 w wallacespace.com e ask@wallacespace.com

Play well with others.

Connect and collaborate with other technologies to assist the human in accomplishing their goals. Learn from the human's behaviours & responses.

Collaborate with the human (not competition)
Collaborate with other technologies
Learn from human, teach human

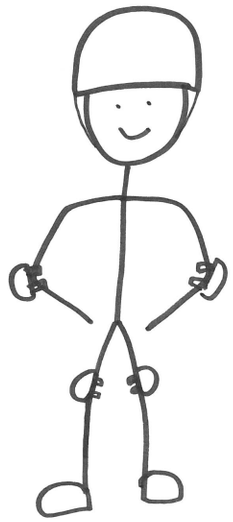
wallacespace

t 020 7395 1265 w wallacespace.com e ask@wallacespace.com

Communicate with, and understand, the human.

Provide appropriate feedback on what's happening. Be transparent about why. Fail gracefully.
Let the human decide how much info is enough. Speak in language the human understands.

A variety of levels of communication/contact, appropriate to the matter at hand
Clear feedback and graceful failure so I know what's happening
Communicate what's happening enough for me to understand it, but not so much that I'm annoyed. Tell me as much as I want to know about whatever, but then still let me make my own decisions/override
Stay in the background



Protect the human.

Preserve privacy, prevent unauthorised access and report faults to maintain trust. Let the human override at any level they like, including the ‘kill switch.

Keep personal data safe
Respect & abide by governmental, societal and individual ethics
Let the human correct/override at every level
PANIC BUTTON

Enable the Human.
Make life easier without taking over. Reduce friction
without being mysterious. Make more suggestions, less decisions.

Make life easier without taking over
Reduce friction without becoming mysterious
Make suggestions not decisions
The human is the superhero, the tech gives them superpowers!