Science News

from research organizations

Empowering robots for ethical behavior

Researchers have developed a concept called Empowerment to help robots and humans to work and live side-by-side safely and effectively

Date: July 18, 2017

Source: Frontiers

Summary: Scientists have developed a concept called Empowerment to help robots to protect and serve humans, while keeping themselves safe. Rather than trying to make a machine understand complex ethical questions, the concept is based on robots always seeking to keep their options open, and doing the same for the humans around them.

RELATED TOPICS

Mind & Brain

› Brain-Computer Interfaces

FULL STORY

Scientists at the University of Hertfordshire in the UK have developed a concept called Empowerment to help robots to protect and serve humans, while

- Language Acquisition
- Social Psychology

Matter & Energy

- Robotics Research
- Engineering
- Vehicles

Computers & Math

- Robotics
- Artificial Intelligence
- Neural Interfaces

Science & Society

- Disaster Plan
- Bioethics
- Sports

keeping themselves safe.

Robots are becoming more common in our homes and workplaces and this looks set to continue. Many robots will have to interact with humans in unpredictable situations. For example, self-driving cars need to keep their occupants safe, while protecting the car from damage. Robots caring for the elderly will need to adapt to complex situations and respond to their owners' needs.

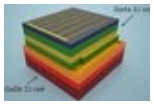
Recently, thinkers such as Stephen Hawking have warned about the potential dangers of artificial intelligence, and this has sparked public discussion. "Public opinion seems to swing between enthusiasm for progress and downplaying any risks, to outright fear," says Daniel Polani, a scientist involved in the research, which was recently published in *Frontiers in Robotics and AI*.

However, the concept of "intelligent" machines running amok and turning on their human creators is not new. In 1942, science fiction writer Isaac Asimov proposed his three laws of robotics, which govern how robots should interact with humans. Put simply, these laws state that a robot should not harm a human, or allow a human to be harmed. The laws also aim to ensure that robots obey orders from humans, and protect their own existence, as long as this doesn't cause harm to a human.

The laws are well-intentioned, but they are open to misinterpretation, especially as robots don't understand nuanced and ambiguous human language. In fact, Asimov's stories are full of examples where robots misinterpreted the spirit of the laws, with tragic consequences.

One problem is that the concept of "harm" is complex, context-specific and is difficult to explain clearly to a robot. If a robot doesn't understand "harm," how can they avoid causing it? "We realized that we could use different perspectives to create 'good' robot behavior, broadly in keeping with Asimov's laws," says Christoph Salge, another scientist involved in the study.

The concept the team developed is called Empowerment. Rather



Testing a Soft Artificial Heart

Air Pollution: Diesel Is Now Better Than Gas, Emitting Fewer Carbonaceous Particulates

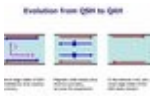
More Than 8. 3 Billion Tons of Plastics Made: Most Has Now Been Discarded

COMPUTERS & MATH

No More Queueing at the Ladies' Room

New Transistor Concept Developed

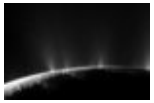
Folding Robots: No Battery, No Wire, No Problem



Evidence for the Majorana Fermion, a Particle That's Its Own Antiparticle

Strange & Offbeat

SPACE & TIME



Holographic Imaging Could Be Used to Detect Signs of Life in Space

Smallest Particles and the Vastness of the Universe Connected

than trying to make a machine understand complex ethical questions, it is based on robots always seeking to keep their options open. "Empowerment means being in a state where you have the greatest potential influence on the world you can perceive," explains Salge. "So, for a simple robot, this might be getting safely back to its power station, and not getting stuck, which would limit its options for movement. For a more futuristic, human-like robot this would not just include movement, but could incorporate a variety of parameters, resulting in more human-like drives."

The team mathematically coded the Empowerment concept, so that it can be adopted by a robot. While the researchers originally developed the Empowerment concept in 2005, in a recent key development, they expanded the concept so that the robot also seeks to maintain a human's Empowerment. "We wanted the robot to see the world through the eyes of the human with which it interacts," explains Polani. "Keeping the human safe consists of the robot acting to increase the human's own Empowerment."

"In a dangerous situation, the robot would try to keep the human alive and free from injury," says Salge. "We don't want to be oppressively protected by robots to minimize any chance of harm, we want to live in a world where robots maintain our Empowerment."

This altruistic Empowerment concept could power robots that adhere to the spirit of Asimov's three laws, from self-driving cars, to robot butlers. "Ultimately, I think that Empowerment might form an important part of the overall ethical behaviour of robots," says Salge.

Story Source:

Materials provided by **Frontiers**

. Note: Content may be edited for style and length.

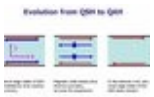
Journal Reference:

Space Sound Waves Around Earth: Electrons Whistle While They Work?

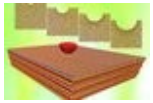
Unusual Galaxy in Distant Universe

MATTER & ENERGY

New Type of Soft, Growing Robot Created



Evidence for the Majorana Fermion, a Particle That's Its Own Antiparticle

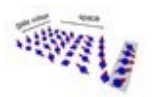


Here's a Tip: Indented Cement Shows Unique Properties

Toward 20-Story Earthquake-Safe Buildings Made From Wood

COMPUTERS & MATH

Folding Robots: No Battery, No Wire, No Problem



Manipulating Electron Spins Without Loss of Information

No More Queueing at the Ladies' Room

Mobile Control With Facial Gestures

1. Christoph Salge, Daniel Polani. **Empowerment As Replacement for the Three Laws of Robotics**. *Frontiers in Robotics and AI*, 2017; 4 DOI: 10.3389/frobt.2017.00025

RELATED TERMS

- Humanoid robot
- Industrial robot
- Robotic surgery
- Nanorobotics
- Robot
- Altruism
- Bioethics
- Narcissism

RELATED STORIES

For They Know Not What They Do: Study on the Robotization of Office, Service Professions

Nov. 2, 2016 — Robots, common in manufacturing jobs, increasingly spread to office professions. Humanoid or human-like robots already perform tasks in hotels, in stores, and in restaurants. They cook, serve, or ... **read more** »

Soft-Bodied Robots: Actuators Inspired by Muscle

June 1, 2016 — To make robots more cooperative and have them perform tasks in close proximity to humans, they must be softer and safer. A new actuator generates movements similar to those of skeletal muscles using ... **read more** »

Carrying a Table Together With a Semi-Cooperative Robot

June 19, 2015 — From a robot's perspective, humans might be considered a nuisance: when robots and humans have to work to-

gether, it often leads to problems. Researchers want to teach robots how to interact with ... **read more »**

Collisions With Robots, Without Risk of Injury

July 8, 2014 — Teamwork between humans and robots will be the motto of the future. But robots may not injure humans at all. When does contact cause an injury, though? Researchers are exploring this for the first ... **read more »**