

## Ethical Robots: Do We Have 30 Years to Get Them Right?

James Gips  
Egan Professor of Computer Science  
Boston College  
Chestnut Hill, MA 02467 USA  
gips@bc.edu  
<http://www.cs.bc.edu/~gips>

In the movie *I, Robot* it's 2035 when the new NS-5 robots, more, uh, spirited, more sentient, more prone to violence, are rolled out.

From *The New York Times* ("A New Model Army Soldier Rolls Closer to the Battlefield" by Tim Weiner, February 15, 2005):

Even the strongest advocates of automatons say war will always be a human endeavor, with death and disaster. And supporters like Robert Finkelstein, president of Robotic Technology in Potomac, Md., are telling the Pentagon it could take until 2035 to develop a robot that looks, thinks and fights like a soldier. The Pentagon's "goal is there," he said, "but the path is not totally clear."

...

"The lawyers tell me there are no prohibitions against robots making life-or-death decisions," said Mr. Johnson, who leads robotics efforts at the Joint Forces Command research center in Suffolk, Va. "I have been asked what happens if the robot destroys a school bus rather than a tank parked nearby. We will not entrust a robot with that decision until we are confident they can make it."

Creating ethical robots – robots that behave ethically in complex situations and perhaps reason ethically – is not just an exercise in academic speculation or science fiction.

Computers continue to become more powerful: Moore's Law has held for the last 60 years and may well hold for the next 40. A century of exponential improvement may yield astonishing results.

We see the beginnings of the field of robotics in Mars landers, Sony robot pets, military Unmanned Aerial/Autonomous Vehicles (UAVs), robot vacuum cleaners, Honda's walking Asimo and other humanoid robots being developed in Japan and Korea. Our automobiles have multiple microprocessors and our airplanes spend most of their trips under computer control.

We have thousands of years of thinking on what it means for a human to behave ethically. Do we have 30 years to create top-level control programs or overall control architectures that allow (force?) complex robots to behave ethically in the complex situations they are bound to encounter in a complex world? It's an extraordinarily difficult problem and we have hardly begun.