Like that emotional kid David, played by Haley Joel Osment, in the movie A.I. Or WALL•E, who obviously had feelings for EVE-uh. The robot in Lost in Space sounded pretty emotional whenever warning Will Robinson of danger. Not to mention all those emotional train-wreck, wackadoodle robots on Westworld.  
  
But in real life robots have no more feelings than a rock submerged in novocaine.

There might be a way, though, to give robots feelings, say neuroscientists Kingson Man and Antonio Damasio. Simply build the robot with the ability to sense peril to its own existence. It would then have to develop feelings to guide the behaviors needed to ensure its own survival.

“Today’s robots lack feelings,” Man and Damasio write [in a new paper](https://www.nature.com/articles/s42256-019-0103-7" \t "_blank) (subscription required) in Nature Machine Intelligence. “They are not designed to represent the internal state of their operations in a way that would permit them to experience that state in a mental space.”

So Man and Damasio propose a strategy for imbuing machines (such as robots or humanlike androids) with the “artificial equivalent of feeling.” At its core, this proposal calls for machines designed to observe the biological principle of homeostasis. That’s the idea that life must regulate itself to remain within a narrow range of suitable conditions — like keeping temperature and chemical balances within the limits of viability. An intelligent machine’s awareness of analogous features of its internal state would amount to the robotic version of feelings.

Such feelings would not only motivate self-preserving behavior, Man and Damasio believe, but also inspire artificial intelligence to more closely emulate the real thing.

Typical “intelligent” machines are designed to perform a specific task, like diagnosing diseases, driving a car, playing Go or winning at Jeopardy! But intelligence in one arena isn’t the same as the more general humanlike intelligence that can be deployed to cope with all sorts of situations, even those never before encountered. Researchers have long sought the secret recipe for making robots smart in a more general way.

In Man and Damasio’s view, feelings are the missing ingredient.

Feelings arise from the need to survive. When humans maintain a robot in a viable state (wires all connected, right amount of electric current, comfy temperature), the robot has no need to worry about its own self-preservation. So it has no need for feelings — signals that something is in need of repair.

Feelings motivate living things to seek optimum states for survival, helping to ensure that behaviors maintain the necessary homeostatic balance. An intelligent machine with a sense of its own vulnerability should similarly act in a way that would minimize threats to its existence.

To perceive such threats, though, a robot must be designed to understand its own internal state.

Man and Damasio, of the University of Southern California, say the prospects for building machines with feelings have been enhanced by recent developments in two key research fields: soft robotics and deep learning. Progress in soft robotics could provide the raw materials for machines with feelings. Deep learning methods could enable the sophisticated computation needed to translate those feelings into existence-sustaining behaviors.

Deep learning is a modern descendant of the old idea of artificial neural networks — sets of connected computing elements that mimic the nerve cells at work in a living brain. Inputs into the neural network modify the strengths of the links between the artificial neurons, enabling the network to detect patterns in the inputs.

Deep learning requires multiple neural network layers. Patterns in one layer exposed to external input are passed on to the next layer and then on to the next, enabling the machine to discern patterns in the patterns. Deep learning can classify those patterns into categories, identifying objects (like cats) or determining whether a CT scan reveals signs of cancer or some other malady.

An intelligent robot, of course, would need to identify lots of features in its environment, while also keeping track of its own internal condition. By representing environmental states computationally, a deep learning machine could merge different inputs into a coherent assessment of its situation. Such a smart machine, Man and Damasio note, could “bridge across sensory modalities” — learning, for instance, how lip movements (visual modality) correspond to vocal sounds (auditory modality).

Similarly, that robot could relate external situations to its internal conditions — its feelings, if it had any. Linking external and internal conditions “provides a crucial piece of the puzzle of how to intertwine a system’s internal homeostatic states with its external perceptions and behavior,” Man and Damasio note.

Ability to sense internal states wouldn’t matter much, though, unless the viability of those states is vulnerable to assaults from the environment. Robots made of metal do not worry about mosquito bites, paper cuts or indigestion. But if made from proper soft materials embedded with electronic sensors, a robot could detect such dangers — say, a cut through its “skin” threatening its innards — and engage a program to repair the injury.

A robot capable of perceiving existential risks might learn to devise novel methods for its protection, instead of relying on preprogrammed solutions.

“Rather than having to hard-code a robot for every eventuality or equip it with a limited set of behavioral policies, a robot concerned with its own survival might creatively solve the challenges that it encounters,” Man and Damasio suspect. “Basic goals and values would be organically discovered, rather than being extrinsically designed.”

Devising novel self-protection capabilities might also lead to enhanced thinking skills. Man and Damasio believe advanced human thought may have developed in that way: Maintaining viable internal states (homeostasis) required the evolution of better brain power. “We regard high-level cognition as an outgrowth of resources that originated to solve the ancient biological problem of homeostasis,” Man and Damasio write.

Protecting its own existence might therefore be just the motivation a robot needs to eventually emulate human general intelligence. That motivation is reminiscent of Isaac Asimov’s famous [laws of robotics](https://theleonardo.org/fascinating-history-behind-asimovs-three-laws-robotics/): Robots must protect humans, robots must obey humans, robots must protect themselves. In Asimov’s fiction, self-protection was subordinate to the first two laws. In real-life future robots, then, some precautions might be needed to protect people from self-protecting robots.

“Stories about robots often end poorly for their human creators,” Man and Damasio acknowledge. But would a supersmart robot (with feelings) really pose Terminator-type dangers? “We suggest not,” they say, “provided, for example, that in addition to having access to its own feelings, it would be able to know about the feelings of others — that is, if it would be endowed with empathy.”

And so Man and Damasio suggest their own rules for robots: 1. Feel good. 2. Feel empathy.

“Assuming a robot already capable of genuine feeling, an obligatory link between its feelings and those of others would result in its ethical and sociable behavior,” the neuroscientists contend.

That might just seem a bit optimistic. But if it’s possible, maybe there’s hope for a better future. If scientists do succeed in instilling empathy in robots, maybe that would suggest a way for doing it in humans, too.