

# To stop the machines taking over we need to think about fuzzy logic

March 10 2015, by Simon James

---



A model of the Terminator from the popular movie series where machines take over the world. Credit: Flickr/Edwin Montufar, CC BY-NC-SA

Amid all the dire warnings that machines run by artificial intelligence (AI) will one day take over from humans we need to think more about how we program them in the first place.

The technology may be too far off to seriously entertain these worries – for now – but much of the distrust surrounding AI arises from misunderstandings in what it means to say a machine is "thinking".

One of the current aims of AI research is to design machines,

algorithms, input/output processes or [mathematical functions](#) that can mimic human thinking as much as possible.

We want to better understand what goes on in human thinking, especially when it comes to decisions that cannot be justified other than by drawing on our "intuition" and "gut-feelings" – the decisions we can only make after learning from experience.

Consider the human that hires you after first comparing you to other job-applicants in terms of your work history, skills and presentation. This human-manager is able to make a decision identifying the successful candidate.

If we can design a computer program that takes exactly the same inputs as the human-manager and can reproduce its outputs, then we can make inferences about what the human-manager really values, even if he or she cannot articulate their decision on who to appoint other than to say "it comes down to experience".

This kind of research is being [carried out today](#) and applied to understand risk-aversion and risk-seeking behaviour of financial consultants. It's also being looked at in the field of [medical diagnosis](#).

These human-emulating systems are not yet being asked to make decisions, but they are certainly being used to help guide human decisions and reduce the level of human error and inconsistency.

## **Fuzzy sets and AI**

One promising area of research is to utilise the framework of [fuzzy sets](#). Fuzzy sets and fuzzy logic were formalised by [Lotfi Zadeh](#) in 1965 and can be used to mathematically represent our knowledge pertaining to a given subject.

In everyday language what we mean when accusing someone of "[fuzzy logic](#)" or "fuzzy thinking" is that their ideas are contradictory, biased or perhaps just not very well thought out.

But in mathematics and logic, "fuzzy" is a name for a research area that has quite a sound and straightforward basis.

The starting point for fuzzy sets is this: many decision processes that can be managed by computers traditionally involve truth values that are binary: something is true or false, and any action is based on the answer (in computing this is typically encoded by 0 or 1).

For example, our human-manager from the earlier example may say to human resources:

- IF the job applicant is aged 25 to 30
- AND has a qualification in philosophy OR literature
- THEN arrange an interview.

This information can all be written into a hiring algorithm, based on true or false answers, because an applicant either is between 25 and 30 or is not, they either do have the qualification or they do not.

But what if the human-manager is somewhat more vague in expressing their requirements? Instead, the human-manager says:

- IF the applicant is tall
- AND attractive
- THEN the salary offered should be higher.

The problem HR faces in encoding these requests into the hiring algorithm is that it involves a number of subjective concepts. Even though height is something we can objectively measure, how tall should

someone be before we call them tall?

Attractiveness is also subjective, even if we only account for the taste of the single human-manager.

## **Grey areas and fuzzy sets**

In fuzzy sets research we say that such characteristics are fuzzy. By this we mean that whether something belongs to a set or not, whether a statement is true or false, can gradually increase from 0 to 1 over a given range of values.

One of the hardest things in any fuzzy-based software application is how best to convert observed inputs (someone's height) into a fuzzy degree of membership, and then further establish the rules governing the use of connectives such as AND and OR for that fuzzy set.

To this day, and likely in years or decades into the future, the rules for this transition are human-defined. For example, to specify how tall someone is, I could design a function that says a 190cm person is tall (with a truth value of 1) and a 140cm person is not tall (or tall with a truth value of 0).

Then from 140cm, for every increase of 5cm in height the truth value increases by 0.1. So a key feature of any AI system is that we, normal old humans, still govern all the rules concerning how values or words are defined. More importantly, we define all the actions that the AI system can take – the "THEN" statements.

## **Human–robot symbiosis**

An area called [computing with words](#), takes the idea further by aiming

for seamless communication between a human user and an AI computer algorithm.

For the moment, we still need to come up with mathematical representations of subjective terms such as "tall", "attractive", "good" and "fast". Then we need to design a function for combining such comments or commands, followed by another mathematical definition for turning the result we get back into an output like "yes he is tall".

In conceiving the idea of computing with words, researchers envisage a time where we might have more access to base-level expressions of these terms, such as the brain activity and readings when we use the term "tall".

This would be an amazing leap, although mainly in terms of the technology required to observe such phenomena (the number of neurons in the brain, let alone synapses between them, is somewhere near the number of galaxies in the universe).

Even so, designing machines and algorithms that can emulate human behaviour to the point of mimicking communication with us is still a long way off.

In the end, any system we design will behave as it is expected to, according to the rules we have designed and program that governs it.

## **An irrational fear?**

This brings us back to the big fear of AI machines turning on us in the future.

The real danger is not in the birth of genuine artificial intelligence — that we will somehow manage to create a program that can become self-

aware such as HAL 9000 in the movie 2001: A Space Odyssey or Skynet in the Terminator series.

The real danger is that we make errors in encoding our algorithms or that we put machines in situations without properly considering how they will interact with their environment.

These risks, however, are the same that come with any human-made system or object.

So if we were to entrust, say, the decision to fire a weapon to AI algorithms (rather than just the guidance system), then we might have something to fear.

Not a fear that these intelligent weapons will one day turn on us, but rather that we programmed them – given a series of subjective options – to decide the wrong thing and turn on us.

Even if there is some uncertainty about the future of "thinking" machines and what role they will have in our society, a sure thing is that we will be making the final decisions about what they are capable of.

When programming [artificial intelligence](#), the onus is on us (as it is when we design skyscrapers, build machinery, develop pharmaceutical drugs or draft civil laws), to make sure it will do what we really want it to.

*This story is published courtesy of [The Conversation](#) (under Creative Commons-Attribution/No derivatives).*

Source: The Conversation

Citation: To stop the machines taking over we need to think about fuzzy logic (2015, March 10)

retrieved 1 May 2024 from <https://phys.org/news/2015-03-machines-fuzzy-logic.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.