# Chapter 4
# Why Not Fuzzy Logic?

**Ivan Restović**

**Abstract** Fuzzy logic is an approach to AI which focuses on the mechanization of natural language. It has long been proposed by Zadeh, its originator, as another paradigm for AI and the correct way to achieve "human level machine intelligence". To present day, this approach hasn't prevailed, but in the light of some recent tendencies in AI development it can gain traction. The "black-box property" of the currently predominant method—deep learning—has recently sparked a movement called "explainable artificial intelligence", a quest for AI that can explain its decisions in a way understandable and acceptable to humans. As it has been recognized, a natural way to provide explanations to users is to use natural language, embedded in the fuzzy logic paradigm. However, to model natural language fuzzy logic uses the notion of "partial truth", which has brought some philosophical concerns. The very core tenets of fuzzy logic have often been described as counterintuitive. In this text, we provide philosophical support for fuzzy logic by providing possible answers to the two most common critiques raised about it, as well as by offering independent philosophical motivation for endorsing it.

## 4.1 A Problem and a Movement

Machine learning with deep neural networks is the predominant method in AI. However, there are some concerns about one of its widely recognized properties—many of the results of deep learning algorithms remain intransparent to humans. They reach a certain decision, but cannot provide reasons for it. This is usually called "the black-box property". Now, in certain areas it becomes "the black-box *problem*", e.g., in the field of medicine or in the financial sector.

I. Restović (✉)
Institute of Philosophy, Zagreb, Croatia
e-mail: ivan@ifzg.hr

This has sparked a movement in AI in 2016[1] coined "eXplainable Artificial Intelligence" or XAI, proposed by Gunning on behalf of the USA Defense Advanced Research Projects Agency (DARPA). The present situation is described as such:

> The current generation of AI systems offer[s] tremendous benefits, but their effectiveness will be limited by the machine's inability to explain its decisions and actions to users. Explainable AI will be essential if users are to understand, appropriately trust, and effectively manage this incoming generation of artificially intelligent partners. [6, p. 2]

The proposal for XAI provides a summary of some existing AI techniques based on two features: performance versus explainability. Deep learning scores the highest on performance, but has very low explainability. Bayesian belief networks offer better explainability, but lag behind on performance. The best explainability is provided by decision trees, but there we also find the lowest performance. The desideratum is, of course, more explainability without loss in performance.

Although not mentioned by DARPA's proposal, many researchers have recognized the potential of fuzzy logic paradigm to assist XAI [1, 2, 7, 11]. As Alonso puts it, "interpretability is deeply rooted in the fundamentals of fuzzy logic" [1, p. 245]. This logic with its supporting theories and implementations has long been proposed as another paradigm for AI, most vigorously by its originator, Lotfi A. Zadeh. But is it itself understandable and acceptable to humans?

## 4.2  Zadeh's Proposal

Throughout his career, Zadeh argued for a paradigm shift in AI development. His position can be illustrated by this often paraphrased place:

> Humans have many remarkable capabilities; there are two that stand out in importance. First, the capability to reason, converse and make rational decisions in an environment of imprecision, uncertainty, incompleteness of information, partiality of truth and possibility. And second, the capability to perform a wide variety of physical and mental tasks without any measurements and any computations. A prerequisite to achievement of human level machine intelligence is mechanization of these capabilities and, in particular, *mechanization of natural language understanding*. In my view, mechanization of these capabilities is beyond the reach of the armamentarioum of AI – an armamentarioum which in large measure is based on classical, Aristotelian, bivalent logic and bivalent-logic-based probability theory. [22, p. 11, added emphasis]

Zadeh talks about the "achievement of human level machine intelligence". In the present context, we will slightly specify his claim. What we are looking for is "human *understandable* machine intelligence". Zadeh's original term may be misleading because it can be argued that some machines already surpassed the human level of intelligence. AI outperforms humans on a variety of tasks. There are forms of artificial intelligence *alien* to us—this makes the problem of XAI all the more urgent.

---

[1]According to [1, p. 244]. We use an updated version from 2017 [6].

In the context of transparent AI, one of the more substantial claims made in the above quote is one about natural language. Humans reason (mostly) in natural language. In Zadeh's opinion, this amounts to saying that we take as inputs sentences in natural language and after some "computation" output a conclusion, also in natural language. Now, wouldn't it be nice if computers reasoned in natural language in the way humans do? Zadeh's working assumptions on natural language can be seen here:

> Much of human knowledge is expressed in natural language. [...] The problem is that natural languages are intrinsically imprecise. Imprecision of natural languages is rooted in imprecision of perceptions. A natural language is basically a system for describing perceptions. Perceptions are intrinsically imprecise, reflecting the bounded ability of human sensory organs, and ultimately the brain, to resolve detail and store information. Imprecision of perceptions is passed on to natural languages. [21, p. 2769]

This passage seems to imply that natural language is enough to store human knowledge based on perception. In other words, that there is nothing in perceptions which cannot be expressed in the natural language. This is clearly an even more substantial and also a controversial claim. However, in the context of XAI, we don't need to fully endorse it. Maybe there is something "lost in translation" but it is lost on both sides since the question posed by a human is itself in natural language. This text is only about "linguistic explanations" given by a logical system designed to resemble human reasoning.

Zadeh proposed several closely connected theories for implementing his above-described motivation, some of which we mention here. For modeling the underlying perceptions, he proposes the Computational Theory of Perception (CTP) wherein perceptions and queries are expressed as propositions in natural language. Having perceptions thus modeled, we can use CTP's underlying methodology of Computing With Words (CWW) to yield answers to queries [20]. Computing With Words in turn is a branch of fuzzy logic in the broad sense [19], but it is also based on fuzzy logic in the narrow sense, a logic of approximate reasoning [17]. More on this ambiguity shortly.

Fuzzy logic employs a nonclassical set of truth values: they are considered as belonging to the unit interval [0, 1], in accordance with the notion of fuzzy sets Zadeh introduced in [15]. The basic notion of this set theory is *partial elementhood*. In fuzzy logic, there is partial elementhood in the "set" of truth values.

Partiality of truth was introduced to capture the intuition that for some concepts there are no clear boundaries. In [15] Zadeh wonders if bacteria are animals. The answer might be—partly. In fuzzy logic atomic propositions are assigned truth values in the interval [0, 1]. Truth conditions for connectives are taken from Łukasiewicz[2]:

$v(\neg p) =_{def} 1 - v(p)$
$v(p \wedge q) =_{def} min(v(p), v(q))$
$v(p \vee q) =_{def} max(v(p), v(q))$
$v(p \rightarrow q) =_{def} min(1, 1 - v(p) + v(q)).$

---

[2]Appearing in his paper from 1930. For an English translation see [9].

What brings fuzzy logic closer to natural language is its use of *linguistic variables* [16] for truth values. Even though there is an underlying computation, we wouldn't get an answer like "*Bacteria are animals* is 0.892 true" since it would be far from natural language. In [17, p. 410] Zadeh uses the countable set {*true, false, not true, very true, not very true, more or less true, rather true, not very true and not very false, ...* }. For instance, we can label as "true" those propositions the value of which exceeds 0.5. The threshold for "very true" can be 0.7, and so on.

"Fuzzy logic" can mean different things. In the broad sense, it includes all the theories Zadeh proposes for mechanization of natural language, some of which are more specialized than the others. So, when Zadeh opts for a paradigm shift toward fuzzy *logic*, he doesn't mean that the whole work in a field as broad as AI has to be done solely within logic as a subfield of pure mathematics or of philosophy. Then, in the narrow sense, "fuzzy logic" signifies such a subfield, i.e., the logical system with truth values in the unit interval and with linguistic variables for such values.

Because of its focus on (computing with) natural language, fuzzy logic has been recognized as a viable approach to XAI [1, 7, 11]. Even if Zadeh's insistence on the use of natural language was exaggerated, this feature is now extremely useful for giving *logical* explanations acceptable to humans.

For instance, Hagras states:

> [...] FRBS [fuzzy rule-based system] generates if-then rules using linguistic labels (which can better handle the uncertainty in information). So, for example, when a bank reviews a lending application, a rule might be: if income is *high* and *home owner* and *time* in *address* is *high*, then the application is deemed to be from a good customer. Such rules can be read by any user or analyst. More importantly, such rules get the data to speak the same language as humans. [7, p. 35]

We will not debate the understandability of Hagras' example rule. Notice just that italicized words represent variables, some of which are fuzzy terms. Consider "high income". It would not be useful for a bank to classify incomes only according to two categories: high versus low. Two incomes $a$ and $b$ can both be low, but one can still be higher than the other. In fuzzy logic this amounts to saying that the sentence "Income $a$ is high" is more true than the sentence "Income $b$ is high". Similarly, Alonso [1] sees Zadeh's CWW especially relevant to XAI since humans are used to explanations in natural language.

However, a host of philosophical critiques are raised against fuzzy logic. A great deal of them attacks even its fundamental tenets, like the very notion of partial truth. In the following section, we provide philosophical support for fuzzy logic. First, we analyze the philosophical setting in which fuzzy logic is often proposed—the sorites paradox. Then we describe the two common concerns raised about the viability of fuzzy logic and outline possible answers. The last part offers an intermediate position, tenable even if some critiques against fuzzy-set-theoretic treatment of truth values are left unanswered.

## 4.3  Philosophical Concerns

### 4.3.1  *Fuzzy Logic and the Sorites Paradox*

In philosophy, fuzzy logic is often considered as a special solution for a more general problem—vagueness, or the possibility of a concept to have borderline cases. Vagueness is problematic because it invites the famous sorites (heap) paradox.

Let's illustrate this by using the most popular predicate in the literature about fuzzy logic, "tall". Consider Sandy Allen, the American actress who was 231 cm tall. Now, everyone would agree that the proposition "The person standing at 231 cm is tall". Also, it seems plausible to affirm the conditional: "If a person standing at $x$ cm is tall, so is the person standing at $x$ cm − 1 mm". In other words, if there was a person only 1 mm shorter than Allen, they would still be considered tall; a tenth of a centimeter doesn't make a difference.

But if we were to line up actors and actresses by their height starting with Sandy Allen and apply the conditional a number of times, we would get counterintuitive results. For instance, it would follow that Danny DeVito, standing at 147 cm, is tall. This is clearly not the right result, despite the premises and the rules of inference being acceptable. How to make DeVito short again? Similarly, we can start with DeVito and the conditional "If a person standing at $x$ cm is short, so is the person standing at $x$ cm + 1 mm", which would in turn make Allen short.

Introducing partiality of truth can help us solve the paradox. Allen is clearly tall. DeVito is clearly not. Fuzzy logic can get us to the right conclusion. Just as the height of people in our lineup decreases, so do the truth values of height ascriptions to people down the sorites. Also, the conditionals are all almost fully true. There is nothing paradoxical about sorites in the fuzzy logic paradigm. The paradox appears only when we use bivalent definitions for fuzzy concepts [22]. Tall is a fuzzy concept, and it should be modeled as such.

Consider again our example. To make comparisons one-dimensional, we will only speak about heights of actresses. Now, we all agree that:
$$v(\text{Sandy Allen (231 cm) is tall}) = 1 \tag{i}$$
But obviously, she is not the only one who deserves to be classified as "fully tall", i.e., for whom the truth value of height ascription proposition is 1.[3] Let's decide that the last actress to be fully tall is Geena Davis:
$$v(\text{Geena Davis (183 cm) is tall}) = 1 \tag{ii}$$
So, anybody shorter than her would have $v < 1$ as a truth value of the proposition ascribing height.

On the other end of the spectrum we have some actresses that are clearly not tall. Call this "short". We now have to decide on the tallest "fully not tall" or the tallest "absolutely short" actress on the list. We decide on:

---

[3]Firstly, there could be an even taller actress. Since we cannot give a value above 1, she would get the same truth value as Allen. Note that, in the present case, it would not make sense for the arrival of a newcomer to lower the truth value of the initial "truest" element.

$v$(Judy Garland (151 cm) is tall) $= 0$                                              (iii)

So, anybody taller than her would have $v > 0$ as a truth value of the proposition ascribing height.

Let's now calculate the intermediate case. It is someone of a height that is in between the two cutoff points. We will characterize it thus:

$v$("Judy Davis" (167 cm) is tall) $= 0.5$                                     (iv)

This would put Meryl Streep (168 cm) slightly on the taller side. Let's approximate:

$v$(Meryl Streep (168 cm) is tall) $= 0.55$                                    (v)

Assume that there is no one between "Davis" and Streep. In that case, the latter actress is the last person in our sorites series to whom tallness may be ascribed more than shortness. So, not everybody is tall—we don't get the counterintuitive result as in the case of bivalent logic.

The presented scale looks useful, but as the reader might have noticed, we have made some questionable assumptions. We had to draw two borders, both of which seem arbitrary. And that in turn resulted in an also seemingly arbitrary intermediate case between the two. Had Geena Davis not been absolutely tall, Meryl Streep might have ended on the other side of the boundary.

### 4.3.2   The Problem of Higher-Order Vagueness

What we have encountered is the problem of higher-order vagueness or arbitrary precision (cf. [14, Chap. 4], [8, Chaps. 4–5]). As Keefe puts if, "what could determine which is the correct function, settling that my coat is red to degree 0.322 rather than 0.321?" [8, p. 114]. She argues that a function from measurements to truth values for every fuzzy concept should be unique. Otherwise, we lose the ordering relation between sentences which was supposed to be an asset of fuzzy logic. If the same coat is also blue to the degree 0.321, is it now more red than blue? However, this uniqueness is unwarranted since there is no clear-cut answer on how to acquire the initial truth values.

Similarly, Williamson argues that in fact, sentence like (v) above are vague rather than exact. So, although truth by numbers looks like a more precise and nuanced account than the classical picture, it doesn't resolve the original problem. Is the sentence (v) *absolutely* true? "Even if statistical surveys of native speaker judgements were relevant to deciding [...], the results would be vague. It would often be unclear whom to include in the survey, and how to classify the responses" [14, p. 128].

Both Keefe and Williamson propose different accounts of truth value ascription to vague statements. Williamson proposes a position now called "epistemicism". Vagueness is just ignorance, there is nothing vague or fuzzy about the world. Every sentence is either true or false. Even the seemingly borderline sentence (iv). On that view, making the notion of truth more nuanced doesn't help our lack of knowledge, as it is shown by the problem of higher-order vagueness [14, Chaps. 7–8]. Keefe, on the other hand, argues for "supervaluationism". On that view, there are nonclassical

values, but these values are not truth-functional. Some sentences fall in a "truth-value gap" [8, Chaps. 7–8].

We will now outline two possible ways of alleviating the problem of higher-order vagueness for a proponent of fuzzy logic. First one more philosophical, the other more mathematical.

Take Keefe's question about finding out the exact value of redness for her coat. Smith [12] argues that this is not the job for fuzzy logic. Fuzzy logic is a calculus of fuzzy truth values. Given such values for atomic propositions, we use logical laws to infer other truths. What these truths are is a matter for another discipline:

> Classical logic countenances only two truth values [...]. This does *not* make it correct, however, to say that it is a commitment of classical logic (model theory) that every statement is either true or false. Such a commitment comes into play only when one seeks to use classical logic to shed light on the semantics of some language (e.g., natural language, or the language of mathematics). It is thus a commitment not of pure classical logic (model theory) – considered as a branch of mathematics – but of model-theoretic semantics (MTS). [...] Pure model theory tells us only that a wff is true on this model and false on that one (etc.). [12, p. 2]

Of course, the problem is here not solved, just relocated. Smith [12] recognizes that and offers possible answers. But from the standpoint of pure logic, one can lessen the concern by endorsing a working assumption of logical pluralism: different logics can be used for different purposes. Remember intuitionism in mathematics. Brouwer [4] argued that real mathematics doesn't conform to some laws of classical logic, most famously the principle of excluded middle. But in other domains, such as reasoning about our everyday finite domains, there is no fault in using classical logic. However, it took a different kind of research to come to the true nature of mathematics and its corresponding logic. The research into the correct fuzzy truth values may turn out to be in the scope of fuzzy logic in the broad sense, but this shouldn't hinder the progress of fuzzy logic in the narrow sense.

Especially since if truth *simpliciter* is a logical or mathematical problem, it is not so just for fuzzy logic. Even classical predicate logic cannot decide on a truth value of the proposition "Bacteria are animals". It can only say what follows from that proposition. Every logic is about valid reasoning, arriving at true conclusions *given* true premises, which often come from other areas of knowledge. Fuzzy logic, we argue, can be the right way to describe some phenomena.

The mathematical way to combat higher-order vagueness is to admit that in some cases the truth value of a proposition is not unique, but that this can be accounted for set-theoretically. Along with "regular" fuzzy sets, Zadeh [16] proposed fuzzy sets with fuzzy membership functions. In that way, we can model second order, as well as higher-order vagueness.

> A fuzzy set is of type $n$, $n = 2, 3, ...$, if its membership function ranges over fuzzy sets of type $n − 1$. The membership function of a fuzzy set of type 1 ranges over the interval [0, 1]. [16, p. 242]

Let's illustrate this within our example. The average height for an actress was 167 cm. But there seem to be other appropriate ascriptions. Let's say we have several

authorities on height, who don't all agree. The lowest proposed average height is 162 cm, and the largest 170. Now we can fuzzify the concept of average height. It is not exactly 167 cm, but somewhere in between—it becomes an interval, rather than a point. There is a "footprint of uncertainty" [7, p. 34] on the scale. Note that the actresses over 170 cm tall are still clearly above average.

All this being said, one might still claim that some terms like "beautiful" seem to resist mathematical treatment. Height is easy to capture since there is only one variable to measure. But how to find out which variables to consider in the correct "beauty-function"? Here it would be useful to introduce a notion of a "prototype". Something may be said to be beautiful to such-and-such degree of truth depending on its closeness to some prototype(s). In the area of psychology of concepts, this is a well-known approach and some of the groundbreaking work was influenced by Zadeh himself. See [3] for a discussion about fuzzy logic in this area.

### 4.3.3 The Problem with Contradictions

Putting aside the problem with arriving at the initial truth values, there is yet another concern often raised against fuzzy logic, one that actually *is* in its providence—it allows for true contradictions. As stated above, in fuzzy logic, the truth value of $\neg p$ ($v(\neg p)$) is defined as $1 - v(p)$. Also, conjunction assumes the same value as the lowest conjunct ($min$ function).

Previously, we defined "short" as the negation of "tall". Considering the clear cases of tallness, we can assert (see proposition (ii)):

$v(\text{Geena Davis is tall and short}) = 0$                              (vi)

Davis is fully tall and not at all short. The conjunction takes the lesser value and turns out totally false, just like in classical logic. However, the problem appears among intermediate cases. For we have:

$v(\text{"Judy Davis" is tall and short}) = 0.5$                     (vii)

$v(\text{Meryl Streep is tall and short}) = 0.45$                 (viii)

Numerous authors have criticized fuzzy logic for this feature. So much so that Smith calls it "the undead argument" [13]. He outlines several lines of response to this argument, coming from several disciplines.

Philosophers usually label the sentences (vii–viii) as *counterintuitive*: the principle of non-contradiction is the undisputed logical axiom and should (fully) hold in all theories. However, this may be circular. Fuzzy logic is accused of not following the classical principles. But it is exactly the inability of classical logic to model the "real world" that was the motivation for a nonclassical approach, such as fuzzy logic. Zadeh simply has different intuitions about bivalence, as we saw from his proposal for a paradigm shift.

Also, note that there are no blatant contradictions in fuzzy logic. Contradictions can be at most half-true. Nothing is both a triangle and a circle, both in classical and in fuzzy logic. This is because such concepts are not vague. Not-absolutely-false

contradictions appear only with vague concepts, which classical logic cannot model in the first place.

Returning to different intuitions, consider again the putatively controversial proposition (viii). From it we can infer:

Meryl Streep is more tall than short.                                             (ix)

We don't consider this proposition neither blatantly false nor meaningless, even if it rests on a contradiction. Streep *is* both tall and short but she is also more tall than short, which can be seen as just another way of saying that her height is slightly above average.

The situation seems to be even more clear in the case of our "most true contradiction", proposition (vii). Asserting it amounts to saying:

"Judy Davis" is as tall as she is short.                                          (vii')

Again, we don't see anything wrong with this assertion. Our hypothetical actress is right in the middle, and a contradiction of a value 0.5 tells us exactly that. So, one could instead argue that, contrary to being unintelligible, there is additional information in true contradictions in fuzzy logic. Whereas in classical logic they all get the same truth value, in fuzzy logic their truth value tells us more [3, p. 31]. This logic is simply more expressive than its classical counterpart.

### *4.3.4  Vagueness Is Not Fuzziness*

In the preceding text, we have treated fuzzy logic based on fuzzy set theory as an answer to the problem of vagueness. This view has been prevalent in the philosophical literature. However, fuzziness can be seen as distinct from vagueness. Importantly, this is the view expressed by Zadeh himself. Dubois [5] further elaborates and expands on this view expressed in the following quote, showing that both epistemicism (vagueness as ignorance) and supervaluationism (there is a gap in truth value for borderline cases) are compatible with the notion of truth modeled by fuzzy sets. Zadeh argues:

> Although the terms *fuzzy* and *vague* are frequently used interchangeably in the literature, there is, in fact, a significant difference between them. Specifically, a proposition, *p*, is *fuzzy* if it contains words which are labels of fuzzy sets; and *p* is *vague* if it is both fuzzy and insufficiently specific for a particular purpose. For example, "Bob will be back in a few minutes" is fuzzy, while "Bob will be back sometime" is vague if it is insufficiently informative as a basis for a decision. Thus, the vagueness of a proposition is a decision-dependent characteristic whereas its fuzziness is not. [18, p. 396, n.]

Here we see that vagueness includes fuzziness, but there is another important characteristic of vague sentences—they don't offer enough information to be accounted for by fuzzy sets. With this distinction at hand, we can accommodate some theories about vague propositions. One can claim that there *are* truth value gaps, but they only concern vague propositions. Such propositions are in a way deficient, they are too underspecified to be ascribed a numerical truth value, be it classical of fuzzy, even type-*n* fuzzy. On the other hand, there is nothing underspecified in an exclusively

fuzzy description of a predicate. In Dubois' words: "While vagueness is a defect, gradualness is an enrichment of the Boolean representation" [5, p. 317].

Epistemicist theory of vagueness can also be incorporated to fit this distinction. Vagueness is still ignorance, not of just two possible truth values, but of the exact fuzzy truth value. Dubois calls this a "gradual epistemic view", according to which partially true propositions exist, but they appear vague or imprecise because of our (partial) ignorance. Similar point is elaborated by MacFarlane [10] in a view called "fuzzy epistemicism". Classical (bivalent) epistemicism claims that what distinguishes vague language from non-vague language has only to do with our knowledge, not with the underlying metaphysics of truth. However, "both uncertainty and partial truth are needed to understand our attitudes towards vague propositions" [10, p. 438].

This concerns some cases of higher-order vagueness. Firstly, if fuzzy epistemicism is correct, some first-order vagueness can actually be downgraded to fuzziness via amelioration of our epistemic position. And if there is still some vagueness about such fuzziness, it can again be a result of insufficient specificity. If so, it can then be alleviated with the corresponding type-$n$ fuzzy sets. It may take some conceptual analysis to come to know the "depth" of a (putatively) vague concept or proposition, but once we find that level, we can describe it mathematically.

## 4.4   Conclusion

Machine learning with deep neural networks is the prevailing paradigm of AI. However, the black-box property of deep learning algorithms may often propose a problem. This has recently sparked a movement called eXplainable Artificial Intelligence (XAI). Decisions made by AI should seek to become more transparent to humans.

Now, humans are the most accustomed to explanations in natural language. And it is exactly the insistence on natural language that is the hallmark of another approach to AI, Zadeh's fuzzy logic paradigm, which has been recognized as a viable approach toward XAI. This paradigm rests on "fuzzy logic" in the narrow sense, i.e., a logical calculus of partial truth.

However, it has been argued that fuzzy logic is not meaningful or acceptable (to humans) since some of its fundamental notions are mistaken or unintelligible. The aim of this text was to provide philosophical support for fuzzy logic. We first described the most common philosophical motivation for introducing this nonclassical logic—the sorites paradox. Then we addressed two common critiques. Fuzzy logic has been accused of harboring higher-order vagueness and allowing for true contradictions.

It is argued that such a nuanced view of a truth value as a number in the interval [0, 1] is itself vague since there is no transparent way of finding the exact value. We proposed two ways of alleviating higher-order vagueness. Firstly, it can be argued that finding the right (fuzzy) truth values for atomic propositions is not the domain of (fuzzy) logic. Secondly, even if in some cases the numbers are not unique, fuzzy

set theory can be expanded via type-$n$ fuzzy sets to mathematically describe this phenomenon.

Connectives in fuzzy logic are so defined as to allow some contradictions not to be fully false. This is often considered an undesirable feature which any correct theory should avoid. However, it is important to note that in fuzzy logic contradictions are at most half-true. We explored some true contradictions and argued that they can indeed be meaningful and even informative.

We also proposed arguments for distinguishing vagueness from fuzziness. In philosophy, fuzzy logic is often seen as just another theory of vagueness along with competing theories such as "epistemicism" and "supervaluationism". However, it can be argued that vagueness includes both fuzziness and an additional characteristic—lack of information. On this view, fuzzy logic doesn't compete with theories of vagueness—they can work in concert.

The notion of partial truth turns out not to be as counterintuitive as it first appears. This being the case, we think it is safe to assume that explanations provided by AI arrived at by using fuzzy logic can be understandable to humans, especially provided an accessible and coherent underlying philosophy of fuzziness.

# References

1. Alonso JM (2019) From Zadeh's computing with words towards eXplainable artificial intelligence. In: Fullér R, Giove S, Masulli F (eds) Fuzzy logic and applications. Springer, Cham, pp 244–248
2. Alonso JM, Castiello C, Mencar C (2018) A Bibliometric analysis of the explainable artificial intelligence research field. In: Medina J, Ojeda-Aciego M, Verdegay JL, Pelta DA, Cabrera IP, Bouchon-Meunier B, Yager RR (eds) Information processing and management of uncertainty in knowledge-based systems. Theory and foundations (Part I). Springer, Cham, pp 3–15
3. Belohlavek R, Klir GJ, Lewis III HW, Way EC (2009) Concepts and fuzzy sets: misunderstandings, misconceptions, and oversights. Int J Approx Reason 51(1):23–34
4. Brouwer LEJ (1949) Consciousness, philosophy and mathematics. In: Beth EW, Pos HJ, Hollak JHA (eds) Proceedings of the tenth international congress of philosophy. North-Holland Publishing Company, Amsterdam, pp 1235–1249
5. Dubois D (2011) Have fuzzy sets anything to do with vagueness? In: Cintula P, Fermüller CG, Godo L, Hájek P (eds) Understanding vagueness. Logical, philosophical and linguistic perspectives (Studies in logic). College Publications, London, pp 311–333
6. Gunning D (2017) Explainable artificial intelligence (XAI). Def Adv Res Proj Agency (DARPA) 2:1–36. https://www.darpa.mil/attachments/XAIProgramUpdate.pdf. Accessed 15 Aug 2019
7. Hagras H (2018) Toward human-understandable, explainable AI. Computer 51(9):28–36
8. Keefe R (2000) Theories of vagueness. Cambridge University Press, Cambridge
9. Łukasiewicz J (1970) Investigations into the sentential calculus. In: Borkowski L (ed) Jan Łukasiewicz: selected works. North-Holland Publishing Company, Amsterdam and London, pp 131–152
10. MacFarlane J (2009) Fuzzy epistemicism. In: Dietz R, Moruzzi S (eds) Cuts and clouds. Vagueness, its nature, and its logic. Oxford University Press, New York, pp 438–463
11. Mencar C (2019) Looking at the branches and roots. In: Fullér R, Giove S, Masulli F (eds) Fuzzy logic and applications. Springer, Cham, pp 249–252

12. Smith NJJ (2011) Fuzzy logic and higher-order vagueness. In: Cintula P, Fermüller CG, Godo L, Hájek P (eds) Understanding vagueness. Logical, philosophical and linguistic perspectives (Studies in logic). College Publications, London, pp 1–19
13. Smith NJJ (2017) Undead argument: the truth-functionality objection to fuzzy theories of vagueness. Synthese 194(10):3761–3787
14. Williamson T (1994) Vagueness. Routledge, London
15. Zadeh LA (1965) Fuzzy sets. Inf Control 8(3):338–353
16. Zadeh LA (1975) The concept of a linguistic variable and its application to approximate reasoning—I. Inf Sci 8(3):199–249
17. Zadeh LA (1975) Fuzzy logic and approximate reasoning. Synthese 30(3–4):407–428
18. Zadeh LA (1978) PRUF—a meaning representation language for natural languages. Int J Man-Mach Stud 10(4):395–460
19. Zadeh LA (1996) Fuzzy logic = computing with words. IEEE Trans Fuzzy Syst 4(2):103–111
20. Zadeh LA (1999) From computing with numbers to computing with words—from manipulation of measurements to manipulation of perceptions. IEEE Trans Circuits Syst I: Fundam Theory Appl 46(1):105–119
21. Zadeh LA (2008) Is there a need for fuzzy logic? Inf Sci 178(13):2751–2779
22. Zadeh LA (2008) Toward human level machine intelligence—is it achievable? The need for a paradigm shift. IEEE Comput Intell Mag 3(3):11–22