To claim that “Machines that make autonomous decisions are functionally equivalent to humans and other sentient beings” is one that is make assumptions about the internal constitution of beings, and whether or not the term “being” can be applied to all automata, finite state or otherwise.

The statement is qualified via the usage of the word *functionally*. In order to evaluate the statement, the ideas of functionalism and how they relate to an autonomous decision-making system must be discussed. Functionalism, specified in the broadest terms, is the idea that neural states are evaluated/identified based on the function that they serve for the system/being, and not what the neural states themselves are comprised of. This is to say that, the current state of the machine/being is a state that has been calculated using a previous state of the machine/being and its next state will also be the result of a calculation based on the current state of the system. For the rest of this discussion, we will refer to both machines and beings simply as systems, and we will diverge from this only when there is a distinction that must be made.

The Stanford Encyclopedia often uses the example of pain as a neural state. Pain, under the concept of functionalism, is a state that serves the function of removing the system from a negative environment. This neural state produces some sort of anxiety or fear, so as to remove the system from a potentially dangerous system or to make the system aware of some sort of external problem in some capacity. This concept itself already points to some type of internal composition, some internal recognition of a state and the imperative to remove the system from a state which it considers to be undesirable. This internal composition, and the complexity that it may or may not take in, is an important factor in the evaluation of the original statement.

A particular class of functionalism, machine state functionalism, references this internal composition as a part of the computation of one or many probable next states for the system to enter. Machine state functionalism posits that anything with a mind is an ideal Turing Machine whose operation can be fully specified by a machine table, a Turing machine being a machine that is sufficiently *responsive* to inputs so as to fool someone interacting with the machine into thinking that the machine is itself a human. This Turing machine is exactly implemented via the usage of what machine state functionalists refer to a machine table. For any input into the system, the machine table computes the next state of the machine (deterministic) or the probability of a machine entering any number of pre-defined states (probabilistic). Under this system, there must be some type of internal imperatives that the machine is attempting to meet in order to calculate the next state of the system. This internal composition could be one simple directive (minimize or maximize some condition), or a complex set of imperatives that are only locally defined for certain states. For example, a system would not simply wear gloves when handling something hazardous, it must have some kind of internal imperative to avoid harm in order to arrive at the decision to wear gloves.

Another foundational concept of functionalism is the portability of neural states, that is that while neural states may not be explicitly the same among many beings, they exist to serve some sort of higher-level role for the system. Using the pain example, pain is the related property of being in a certain state, and actions should be taken to remove oneself from this state if the internal composition of the system points to pain being an undesirable state (in the presence of whatever else the current neural state is comprised of). The portability of the

neural state is related to the higher-level role that the state serves. In the computation of a next state, we can see these higher-level roles as a part of the internal composition that then interacts with the current neural state to produce the next state of the machine. For most beings, pain would be a neural state that is avoided as a result of the dislike of pain that exists as a part of our internal composition. This is then related to environmental factors and an optimal next state is produced.

It is in the interaction of the internal composition and the current neural state where an important distinction is made between autonomous decision-making machines and fully realized beings. It would be a massive understatement to say that the internal composition of a being like a human is complex. In machine terms, there are static components to this internal composition (unshakable beliefs), locally defined portions (morality in the presence of different state), and potential hierarchies of system imperatives. Most importantly, the concept of malleability of this internal composition seems to be lacking in reference to autonomous decision-making systems. Machine state functionalism posits that machine tables should be runnable on different hardware, that is to say that a being is *only* defined by machine table it runs and not the system itself. Machines that make autonomous decisions *should* and *do* have their internal compositions ported to separate hardware; you can run a machine learning algorithm on many different types of computers in a manner that is nearly impossible to do with humans. The porting of the actual internal composition that interacts with the calculation of a next state is where humans and autonomous decision-making machines significantly differs in relation to functionalism. For most autonomous decision-making machines, there is some ground internal composition that must be predefined in order for the machine to start any actual autonomous decision making at all, this ground is typically unable to change in the presence of new neural states unknown to the machine.

Let us imagine a machine that has the ability of sight, in grayscale color. The machine, when exposed to colors, has no recognition of them, and cannot act upon these colors, its machine table unable to take note of them. If the machine were to somehow be upgraded to view these colors, or was ported to hardware that could, it would ignore them, as they are not an input in the machine table to produce an output. Sentient beings, on the other hand, would note the novelty of the new situation and change internal constitution because of it. If a child were to randomly gain the ability of hearing after not having it, they would take advantage of this and notice it, where as an autonomous decision making machine would not, as it is not a current part of the things it can understand.

This is all to say that, functionally, sentient beings and autonomous decision-making machines are not functionally the same, as their internal constitution differs in a manner that it is irreconcilable.