If someone claimed that modern machines were intelligent, how would we disprove them? We could think of tasks that humans can perform that machines cannot, but this is a two-way street: in some tasks, machines vastly outperform humans. Besides, examples alone can never get to the heart of the matter. We could use terms like "generalized" intelligence, but what exactly does that mean?

To be clear, my interest is not merely philosophical. I am interested in *mechanisms* that make complex, unstructured cognition possible. I believe that developments in machine learning bring us closer to answering these questions. Specifically, there is a development common to many recent innovations that I think touches on some of the most essential elements of intelligence: in contrast to traditional machine learning models, which take in data and output predictions, many of the most successful recent models incorporate mechanisms that take in ***state*** and output ***state***. This is actually the defining characteristic of a deep network: inner layers of a network operate on state, not directly on data or predictions.

In fact, the mechanisms by which deep networks manipulate their own state only seem to be advancing in complexity. I do not think this is simply a byproduct of models becoming generally more complex. I think that a model's ability to manipulate, mutate, and store state is one of the main determiners of intelligence. Discussion of a specific example will hopefully strengthen this claim:

When I ride a bicycle, my brain seems to upload what we might call a bicycle-riding "program" from memory. This "uploading" mechanism seems to take cues not from external sensory inputs but from some kind of internal signal (maybe something we could call *volition*?). Additionally, my brain draws on complex memories, such as the meaning of traffic signs, that are relevant to the activity. It also attends carefully to certain sensory signals like the movements of cars, while ignoring other signals that might be important in other circumstances, like expressions on faces. Given this complex system of conditioning mechanisms, my behavior on the bicycle from moment to moment requires little attention or effort.

In general, it seems to be true that in the moment of activity while performing a certain task, behavior is often (always?) automatic. Indeed, the progress of machine learning research suggests that current architectures are capable of mastering many moment-to-moment inference tasks. What distinguishes human intelligence is *not* the specific architecture for performing inference, but the *supporting* architectures that set the stage for inference. In computational terms, the significant element is not the CPU, but the programs that populate the computer memory and establish the state of the system before the computation is even performed.

The power of this meta capability becomes evident in applications that require rapid learning, for example, one-shot learning. In these applications, a model does not have time to slowly learn the inference function using gradient descent. Instead, models have to learn a more abstract mapping: from inputs to hidden states that then facilitate the rapid learning of the actual function. For example, when a person learns a new word, the brain's rapid assimilation of the word into vocabulary clearly suggests the presence of more a powerful *learned* program specifically responsible for vocabulary acquisition.

What happens when this idea is taken to its logical extreme? LSTMs operate on their own parameters, but imagine a model that operates on programs themselves. Humans seem to exhibit this capability: most ordinary adults with a sense of balance and a basic familiarity with wheeled vehicles can almost immediately learn to ride a scooter.

Somehow, the mind identifies related, learned programs and combines them to create new ones. Though slow, Hebbian learning might gradually improve the abilities of the scooter rider, the initial encounter with the scooter only benefits from learning related to other tasks. For most tasks, *learning derived from past experience seems to account for the bulk of overall learning*.