On the potential of model-based reinforcement learning

AI systems are well-suited to learn in situations where there is a clear cause and effect relationship between the agent's decisions and the utility of that decision. Swimming into a shark will cause a loss of life. Shooting an alien ship will increase the score. In these situations the agent must learn to associate the current situation and a decision with the direct consequences. However, in situations where the consequences of a decision are significantly delayed it is more difficult to learn this mapping. The most challenging and largely unsolved AI benchmark problems feature these delayed consequences: uninformative score signals, and open-ended worlds where the AI could engage in many tasks. Designing AI systems that can efficiently explore and learn in these challenging settings would (1) substantially improve the performance of our systems on current benchmark tasks, (2) broaden the industrial applicability of AI, and (3) potentially provide new insights into the fundamental principles of how young humans and animals learn and explore their world.

One way to deal with the problem of delayed consequences is for the AI to construct its own internal understanding of how the world works, usually called a model of the world. For example, a model might encode: when I am lined up with a shark and I decide to fire a torpedo, the shark will disappear. Given access to a model of this form, an AI can mentally simulate future situations that would result from behaving in particular ways. This mental simulation does not require time-consuming interaction with the world, and leverages prior learning. Just like a human can decide where they might end-up if they took a new path down to the river, without ever having taken that path before. Planning with learned models is the key to efficient learning.