In the context of artificial intelligence and particularly AGI (Artificial General Intelligence), naturalized induction is concerned with how an intelligent agent can form beliefs or make predictions about its environment while being a part of that very environment. This is a difficult problem because the agent's own actions and computations are events within the universe and can influence the phenomena it is trying to predict.There is a trade-off between the computational resources of a system and the quality of its predictions.

As an agent updates its world model or gains new information, it might undergo changes in the way it categorizes or understands the world. Naturalized induction must account for these shifts, especially when the agent itself can be a catalyst for such changes. An agent must maintain consistency between its beliefs and actions, even when those actions could change the environment or the agent's own future beliefs. Naturalized induction is closely related to the problem of embedded agency, where the agent is not an abstract entity but is situated within the world it is trying to understand.

If an AGI system uses naturalized induction, considerations about AI alignment, and ethical behavior become even more intricate, as the AGI is not just a tool but an actor within the ethical landscape.

Alternative phrasings

-

Related

https://docs.google.com/document/d/1ttLqTGNpbsT9ABY553GbQWth7Mz24jBUaeCAwVEZGml/edit