FONDAZIONE BRUNO KESSLER
MAGAZINE

The role of rationality in modern robotics

August 11, 2026

Leslie Pack Kaelbling is a Professor at MIT. Her research agenda is to make intelligent robots using methods including learning, planning, and reasoning about uncertainty. She was working on agentic AI way before it became fashionable. On 29 May 2026 we met her at FBK where she held a seminar co-organized with AIXIA, the Italian Association for Artificial Intelligence.

The classical approach to AI aimed to build systems that were rational at run-time: they had explicit representations of beliefs, goals, and plans and ran inference algorithms, online, to select actions.

The rational approach was criticized (by the behaviorists) and modified (by the probabilists) but persisted in some form.

In recent years, relatively unstructured data-driven end-to-end approaches have demonstrated great success in a wide variety of domains and began to seem like a plausible route to general-purpose intelligent robots. However, we are now starting to see the limits of pure behavior learning and many practitioners are re-integrating forms of search and explicit reasoning into their approaches.

Leslie Pack Kaelbling has revisited the rational-agent approach to the design of intelligent robots, from the perspectives of engineering effort, computational efficiency, cognitive modeling and understandability.

She also presented some of her current research, which focuses on understanding the role of learning in runtime-rational agents with the ultimate aim of constructing general-purpose human-level intelligent robots.

Giancarlo Sciascia (GS): You mentioned that many researchers are beginning to see the **limits of pure behavior learning**. In your view, what are the most critical signs that indicate when a purely data-driven approach is failing and requires the integration of search and **explicit reasoning**?

Leslie Pack Kealbing (LPK): If adding each new ability or scenario requires similar amounts of additional data, then we aren’t getting real generalization. Explicit reasoning can still be useful in a

largely data-driven approach, if what we're learning from the data is general-purpose knowledge that can be re-used and re-combined via reasoning to solve new problems.

GS: How do the modern forms of **search and reasoning** being re-integrated today differ from the inference algorithms used during the "classical period" of AI?

LPK: In many ways, they're similar: the classic techniques of search and even formal reasoning for theorem proving are still in use today. The difference is that they are operating on representations and/or using models that are learned from data or produced by LLMs.

GS: Regarding **agent architecture, i**n your approach, how do you balance learning with runtime computational efficiency?

LPK: Really, the question is how to balance explicit reasoning with runtime computational efficiency (because we can be reasoning from learned models.) We definitely have to find ways to make reasoning more efficient, and we can use learning to help! We can: "cache" the common cases by training smaller predictive models, learn reasoning heuristics/short-cuts, use decomposition techniques to solve several smaller reasoning problems rather than one big one.

GS: Specifically, which tasks should be delegated to learning and which should remain part of the agent's rational reasoning process?

LPK: (One more time I want to insist that reasoning can be performed on learned models, so contrasting learning with reasoning is not correct)
– Learning vs "built-in" knowledge: I think we should build in structural invariants that are true in our world and algorithmic/data-processing insights (e.g., ideas about processing images with convolution, using temporal hierarchical decomposition, maybe even the most basic ideas of physics and kinematics etc.) but learn (both "offline" before the robot is deployed and "online" at deployment time) detailed models/policies, particularly at the sensory-motor level.
– Representing policies vs representing world models: Typically, policies can be much more computationally efficient but harder to learn and larger to store; world models can be more general but slower to use. I think we should "compile" reasoning into policies that cover the common cases for routine behavior and fast reaction and use reasoning to maintain generality over all the unusual cases.

GS: You evaluate the rational-agent approach from the perspective of **understandability**. Do you believe that re-integrating explicit representations of beliefs and goals is the key to making robots safer and more explainable for humans?

LPK: I don't know if it's "the key", but it could certainly be helpful. Humans can understand the internal processing of a robot if it's structured this way, and individual sub-parts of the model have externally interpretable and verifiable semantics.

GS: The ultimate goal of your research is the construction of **general-purpose, human-level intelligent robots.** What do you consider the most difficult obstacle to overcome when moving from specialized robots in narrow domains to truly versatile systems?

LPK: Still and always it's generalization. Humans (and other animals) are incredibly robust in a very wide-ranging space of circumstances. And the robustness is at all levels: navigating difficult terrain, using objects in surprising ways as tools, solving cognitive problems. Each of these is difficult and

combining them smoothly is even harder!

GS: Given your interest in **reasoning under uncertainty**, how should future general-purpose agents handle entirely new situations for which they have no prior training data?

LPK: It depends on what we mean by “no prior training data”: if the learned representations have the correct combinatorial structure, and the solution to the problem can be assembled from *parts* that we have already learned, then we have generalized effectively to a new situation. But, of course, sometimes we won’t be familiar with enough parts, and will truly be uncertain. The key to robust behavior under uncertainty is explicit representation of the uncertainty: that enables the robot to (a) choose actions that are, for example, conservative in the case that the outcomes are not clear and (b) decide to choose other actions that will explicitly gain information to reduce uncertainty.

GS: Having focused on **“agentic AI”** long before it became a popular trend, what fundamental lessons from the past do you think the current generation of researchers is at risk of forgetting?

LPK: That was a little bit of a joke. But I do think it’s easy to forget that there are a lot of tools in our intellectual and algorithmic toolbox and just focus on the newest ones. But if the old ones are really important, the current generation will rediscover them!

GS: How does your background in **Philosophy** influence your vision of what constitutes a “rational agent,” and how is this reflected in the technical design of your robots?

LPK: My philosophy background helps me in framing problems formally (so we can solve them with algorithms or learning methods). A recent example is how to think about building a robot that is asked, for example, to retrieve an apple. How should it organize its search process when it doesn’t know about any apples yet? What should it do when it finds an apple but determines it is out of reach? It has to look for “another apple”, but what does that mean formally and how can we build a system that does this efficiently. Another example is reconsideration: imagine the robot has decided to make a particular kind of sandwich for lunch, and now is working on finding bread. Under what circumstances should it reconsider its higher level plan? When it finds the bread is stale? When there isn’t bread available? When it sees someone else has made a different, tastier lunch and left it in the fridge? We can’t reconsider all our choices all the time, but we clearly have to reconsider sometimes!

Leslie Pack Kaelbling, a Professor at MIT, has an undergraduate degree in Philosophy and a PhD in Computer Science from Stanford, and was previously a faculty member at Brown University. She was the founding editor-in-chief of the Journal of Machine Learning Research. Her research agenda is to make intelligent robots using methods including learning, planning, and reasoning about uncertainty. She was working on agentic AI long before it became fashionable.

PERMALINK

https://magazine.fbk.eu/en/news/the-role-of-rationality-in-modern-robotics/

TAGS

- #agentic ai
- #artificialintelligence
- #IA agentica
- #llm
- #philosophy
- #ragionamento
- #reasoning
- #robotica
- #robotics

RELATED VIDEOS

- https://www.youtube.com/watch?v=WmnF-NhKgzY

AUTHORS

- Giancarlo Sciascia