Welcome to Artificial Intelligence!

Ernest Davis

January 23, 2023

What is Artificial Intelligence?

What is Artificial Intelligence?

- Constructing an intelligent agent. (Russell and Norvig, earlier editions)
- The computational study of cognition (Charniak and McDermott, 1985)
- My definition: There are tasks that are easy – even automatic – for human but very hard to program computers to do:

  Vision, Manipulation and motion, Language, Some forms of learning, Some forms of reasoning

  AI is the study of how to get computers to do those.

Two major approaches to AI

- Knowledge-based AI (first half of this semester)
- Corpus-based machine learning (second half of this semester)

Knowledge-based AI

- Analyze the knowledge that is necessary to carry out intelligent tasks
- Determine how that knowledge can be expressed symbolically
- Build programs that use that knowledge.

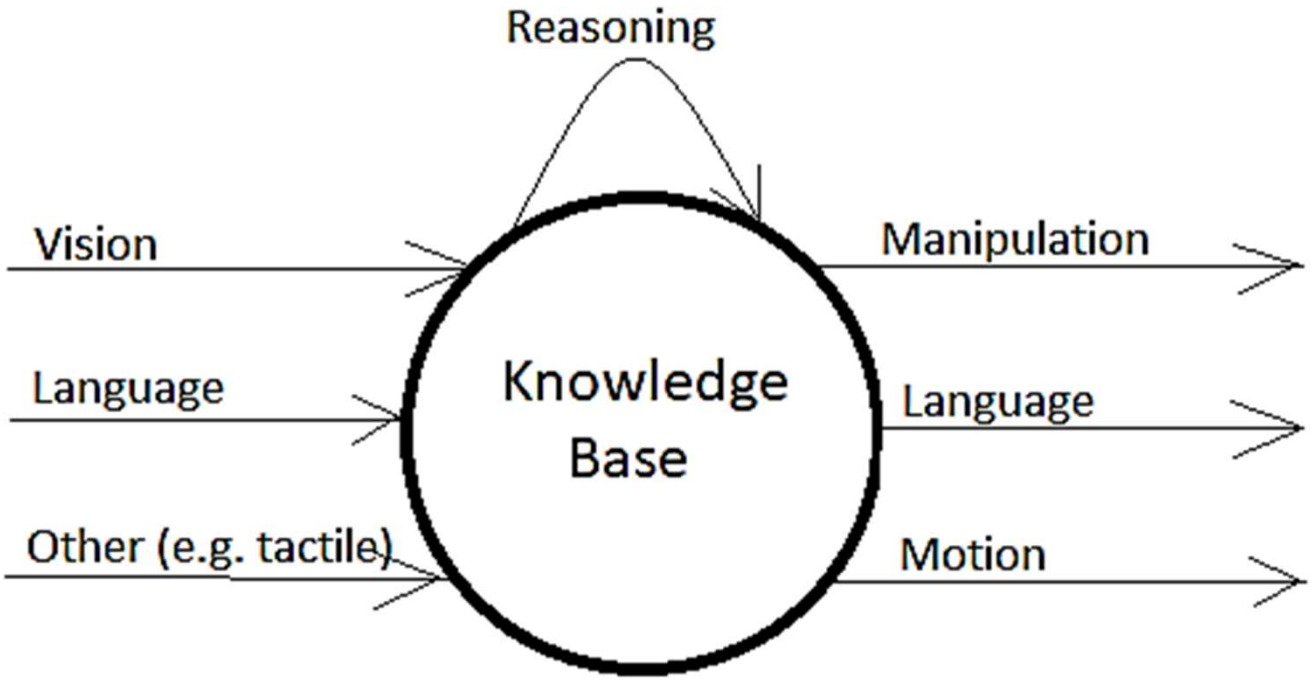
The representation of knowledge is

- Symbolic. DayOfWeek(January,23,2023) = Monday

  $\forall_{x,y}$ Parent(x,y) $\Rightarrow$ Older(x,y).

- Transparent. A human looking at a symbolic expression can interpret it.
- Modular. The knowledge base consists of propositions, each of which is meaningful by itself.

Architecture of Knowledge-Based AI

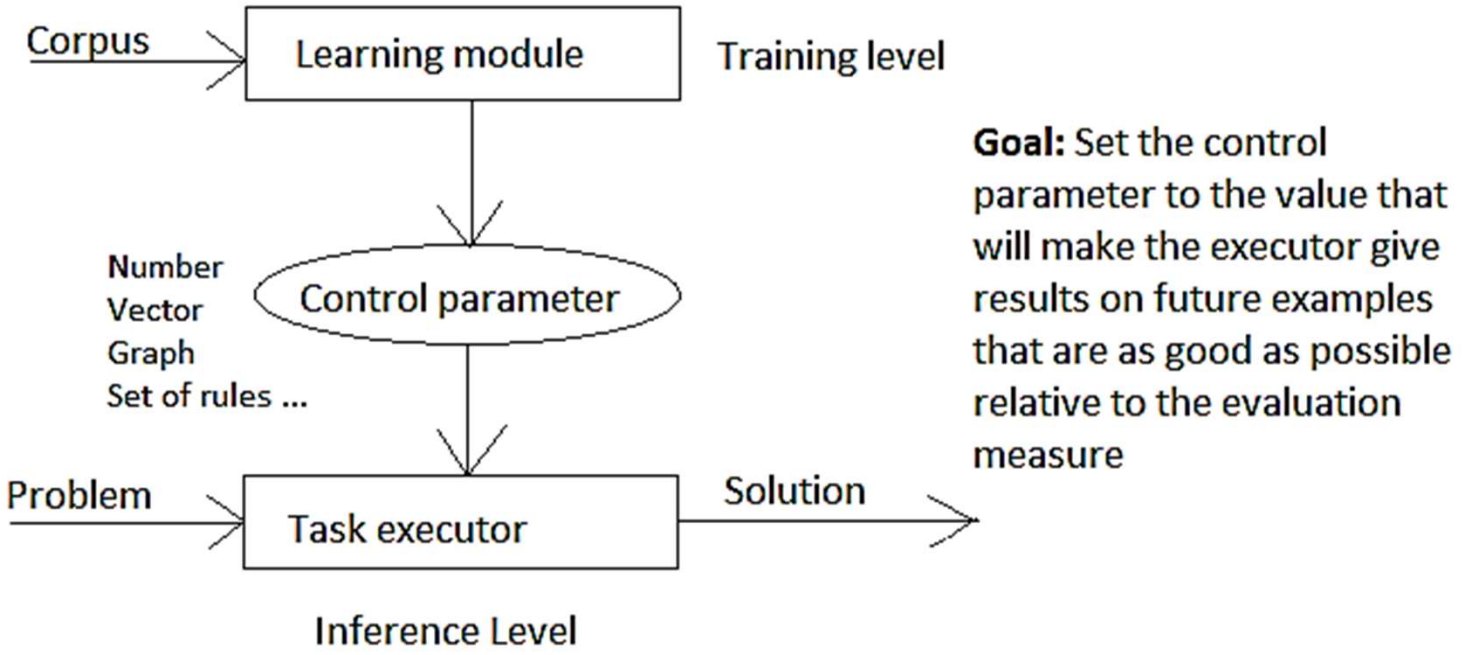


Corpus-based Machine Learning

- A specific task to carry out:
  - Tag photos with the category of image shown
  - Translate from Spanish to Urdu
  - Build a chatbot
- A corpus of relevant data

  (Either successful executions of the task or otherwise relevant)
  - Collection of labelled images
  - Collection of bitexts in Spanish and Urdu
  - A body of conversations
- An automated measure for evaluating success

ML Architecture



Characteristics of Machine Learning (particularly neural networks/deep learning)

- Opaque. It is difficult or impossible to understand the significance of the control parameter, or why it works, or why any particular answer is generated.
- Non-modular. It is sometimes possible to identify pieces of the architecture with meaningful characteristics, but not reliably.
- Undebuggable. If the system makes an error on a particular example, there is no way to find out why it went wrong and fix it. All you can do (hopefully) is improve the *overall* performance. An overall improvement may in fact cause the system to get wrong particular examples that it had previous gotten right.

Outline of the course

I. Search. Not actually AI, but a category of techniques widely used in AI (and also elsewhere in computer science).

II. Logic. A fundamental framework for the knowledge-based approach to AI.

III. Probability. Dealing with uncertainty. Important both in the knowledge-based approach and in many forms of machine learning.

IV. Machine learning.

V. Natural language processing:
   a. Knowledge-based approach
   b. ML-based approach