The 7 levels of AI

Warren B. Powell
Professor Emeritus, Princeton University

Revised September 6, 2024

Spanish translation: https://tinyurl.com/7levelsofAISpanish

There are now countless articles breathlessly talking about "AI," the vast majority of which are talking about a very specific form of AI which is some form of machine learning (mostly pattern recognition) being performed using neural networks. This is hardly the only form of artificial intelligence, and it is far from even being the most advanced form. In addition, there are breathless articles of advanced intelligence which I claim are pure science fiction.

As we observe the growing sophistication of computer-based tools, I felt it was time to distinguish the different levels of artificial intelligence. To begin, I think AI can be broadly organized into two categories:

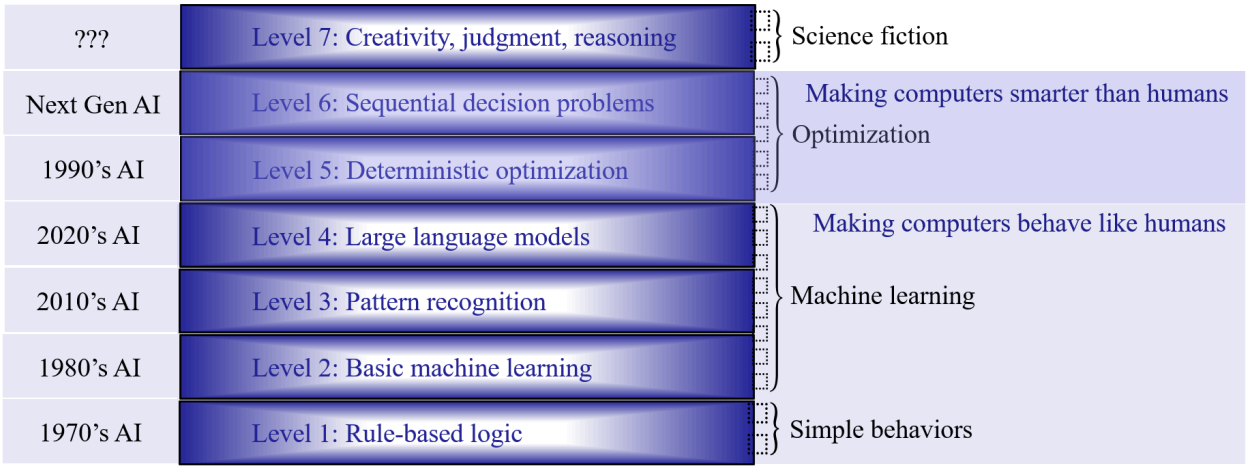
1) Making computers behave like humans.
2) Making computers smarter than humans.

Making a computer "behave like humans" has often been equated with a high level of intelligence since humans represent the most intelligent form of life. For major classes of unstructured problems, I think this is true. However, there are well-structured problems where computers can outperform humans.

Below I define seven levels of intelligence. Level 1 includes basic rule-based logic. Levels 2, 3 and 4 are all a form of supervised machine learning, which is what is typically meant twhen the press (today) refers to "artificial intelligence." Levels 1-3 are in widespread use, while level 4 (such as ChatGPT) seems to be enjoying very rapid adoption. Supervised machine learning requires a training dataset, and as such are technologies that try to mimic observed behavior.

Levels 5 and 6 are both forms of optimization problems, which means they focus on making decisions instead of just predicting or matching patterns. These are also widely used, although they are not always associated with "artificial intelligence." However, these are the only levels where a machine can outperform a human on well-structured problems.

Level 7 is where we reach the most advanced forms of human reasoning, creativity and judgment. There are specific, technical reasons why none of the existing "AI" technologies can perform at this level. The large language models such as ChatGPT are described as performing at this level, but this is only because it mimics word patterns which are produced by people. Levels 5 and 6 can only work on well-structured problems that make it possible to code all the physics of a problem into the computer.

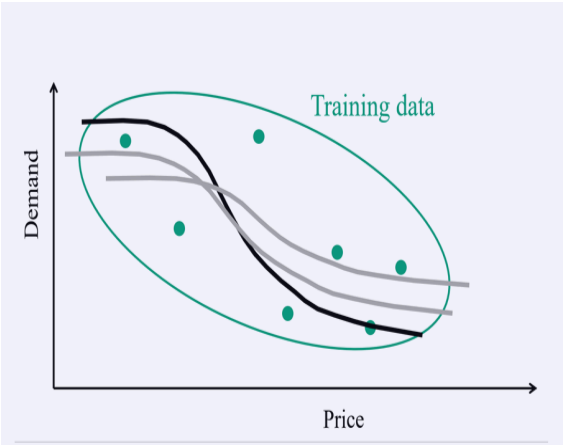


Level 1) Rule-based logic

Rule-based logic (alternatively, "symbolic rule-based reasoning") was the first form of "AI" that emerged in the 1960s and 1970s, and produced the initial surge of interest in what computers could do. Rule-based logic involves rules specified by a human, coded into logic. Simple examples include "if you are eating red meat, then drink red wine," but the number of cases quickly exploded as occurred in health: "a male patient, over 60, with high blood sugar, not a smoker, no parents with diabetes, …, should follow this [multidimensional] diet."

Rule-based logic did not come close to meeting the early expectations of AI, but far from being a complete failure (as it was once viewed), rule-based logic is used throughout modern machine intelligence. There are a number of commercial "rules engines" available today which are widely used.

Level 2) Basic machine learning



This covers well-known statistical models using lookup tables, parametric models (including linear or nonlinear models and neural networks) or nonparametric models. These methods first emerged in the early 1900s under the umbrella of statistics, but grew dramatically as computers became more powerful (computer scientists entered the field using the name “machine learning”). The most popular models were linear (in the parameters), capturing basic relationships between input variables (also known as explanatory or independent variables, or “covariates”) and a response. For example, we might use price with variations such as price-squared and log of price as inputs, to predict the response of the demand. These models matured with the introduction of nonlinear models (such as logistic regression and early neural networks), and nonparametric models (using local approximations). Neural networks have been popular since the 1970s, and widely used in many nonlinear (deterministic) estimation problems.

Level 3) Pattern recognition using deep neural networks

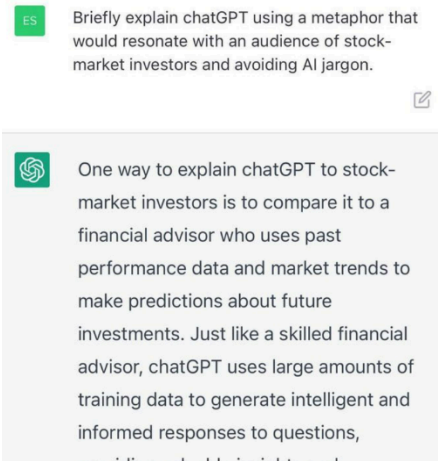


The research community has studied neural networks since the 1960s, but it was the use of deep neural networks, along with the availability of large datasets, starting around 2010 that produced the first true breakthroughs in pattern recognition. This was the foundation of the modern use of “AI” when it emerged in the 2010s for voice recognition, facial recognition, and image identification.

Pattern recognition using deep neural networks is just another form of nonlinear modeling, but it took machine learning into an entirely new class of applications, bringing a level of visibility to the field of machine learning far beyond what had been achieved with the work described under level 2.

Level 4) Large language models

Learning speech patterns using ultra-deep neural networks burst into the public imagination in 2023 with the introduction of ChatGPT building on research that has been evolving since the 2000s (but especially post 2010).



Although large language models can be viewed as merely an extension of deep neural networks for pattern recognition, we have reserved an entire level for this application given the dramatic increase in the complexity of language problems (LLMs require substantially more data preparation), and the sheer growth in the size of neural networks (and the training datasets to train them) used to support this problem setting. Unlike pattern recognition, where the "answer" to "what is this pattern" is deterministic, LLMs deal with much richer inputs, and can produce a range of outputs to a single query (which means it has a stochastic response). Some people equate its ability to create its own sequences of words and phrases, but it is nothing more than randomization among patterns found in the training dataset.

LLMs (such as ChatGPT) are often associated with a much broader category known as "artificial general intelligence" which includes more general forms of learning for unstructured problems. We put these capabilities in Level 7. We limit Level 4 to technologies that have to be trained, which implies it is still a form of supervised learning. This means that LLMs will always produce words and phrases that have been used in the training datasets. It is for this reason that companies supporting this technology are investing heavily not just in the training using massive datasets, but also in the active use of people to guide the behavior of the neural network.

There has been a lot of hype about the potential danger of LLMs. The only danger of an LLM is misinformation, which means any damage is still coming from people. There is, of course, no shortage of misinformation on the internet today, so we can hope that modern society has developed defenses to misinformation, but it is something that we have to be aware of.


ChatGPT is powered by these contractors making $15 an hour

PUBLISHED SUN, MAY 7 2023·11:22 PM EDT

NBC NEWS | David Ingram

Level 5) Deterministic optimization

Unlike machine learning models, deterministic optimization depends on an explicit model of a problem that includes both the physics of the problem as well as a performance metric. Inputs to the model include controllable parameters, also known as actions or decisions. Sophisticated algorithms are used to search over feasible regions to optimize the performance metrics.

Unlike the machine learning-based technologies (levels 2-4) which have to be trained using a training dataset, deterministic optimization does not involve any training. Instead, it uses a model of a problem that captures the physics of the underlying application, along with a performance metric that allows us to evaluate decisions. Then, algorithms are used to search for the best decision that is implementable, and which optimizes the specified performance metric. For example, in the 1990s tools emerged for scheduling airlines that were able to produce more efficient schedules. There are countless examples of achievements like this from the operations research community.

One class of deterministic optimization problems arises in machine learning. The problem is to find the best set of parameters that minimizes the difference between a parameterized function (often called a model) and a set of observations (the training data set). Using a neural network to solve a problem (which requires solving the optimization problem of fitting the neural network to the training dataset), versus using an optimization model to optimize some problem (model-based optimization) are often confused, but they are fundamentally different.

Model-based optimization, unlike training-based machine learning, is able to produce solutions that are better than what any human can achieve. The price of this performance is that we have to invest the time to create a model of the physical problem inside the computer. Note that a neural network does not directly capture the physics of a problem; instead, it tries to learn appropriate behaviors through the training dataset.

A major difference between a statistical model (such as a neural network) and an optimization model is the objective function. The objective when fitting a statistical model to a training dataset is always one of minimizing some distance metric (such as the sum of the squares of the difference between the model and the training dataset). By contrast, optimization models require that an objective function be specified by the analyst, in addition to the constraints that capture the properties of a problem.

Level 6) Sequential decision problems

Here we are solving a problem that consists of the sequence

decision, information, decision, information,…

where decisions have to be made given the uncertainty of information that has not yet arrived. The optimization problem is to find the best method for making decisions (known as a policy). Special cases of sequential decision problems include:

decision, information (this describes classical stochastic search problems)

and

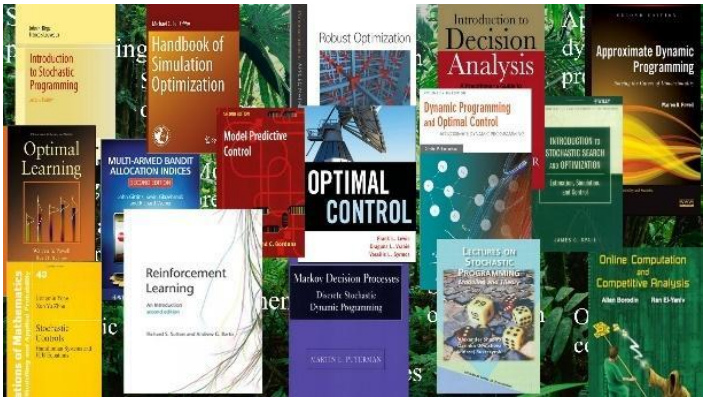
decision, information, decision (this describes two-stage stochastic programming problems).

More general problems have sequences of "*decision, information"* that extends over either a finite or infinite horizon.

While deterministic optimization problems search for the best decision x (which is typically a vector), sequential decision problems require searching for a function $X^{\pi}(S_t)$ known as a *policy* (designated "π"), which is a function that maps the information in a state variable to a decision. For sequential decision problems, a decision might be binary, discrete, scalar continuous, or any type of vector. The challenge is that the policy has to work well over time, according to some metric (typically an expectation, which means we are optimizing the performance on average).

$$\max_{\pi} \mathbb{E}\left\{\sum_{t=0}^{T} C_t(S_t, X^{\pi}(S_t)) \,|S_0\right\}$$

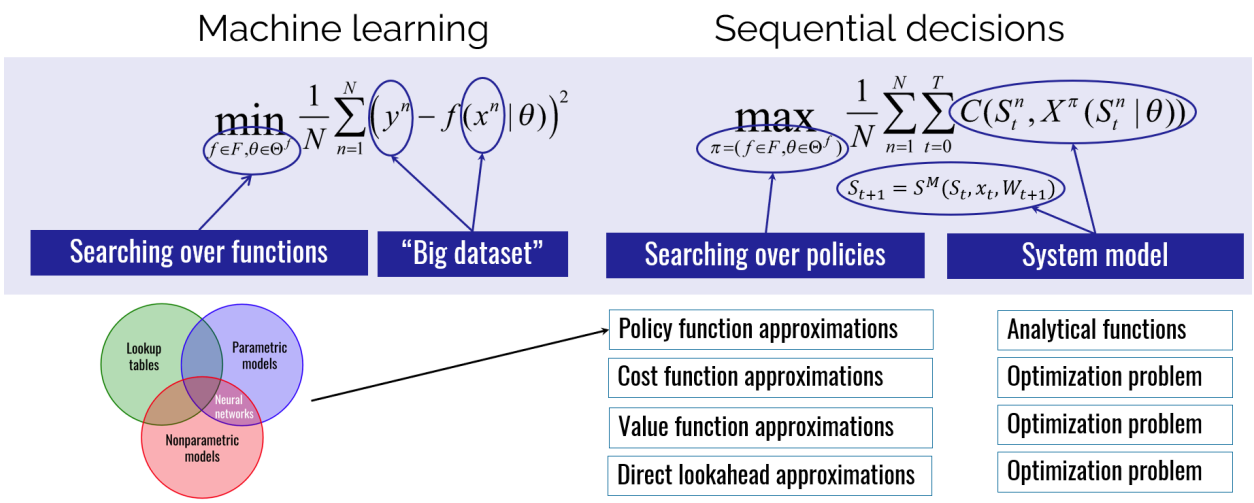
The AI community (in computer science) studies this problem under the banner of "reinforcement learning" but there are approximately 15 distinct fields, each with one or more books, using eight different notational systems. This diversity reflects the tremendous richness of sequential decision problems. A library of materials on "sequential decision analytics" is contained here.



Sequential decision problems represent a significant extension of deterministic optimization models. Like deterministic optimization, sequential decision problems require modeling the physics and dynamics of a problem and include a performance metric that allows us to identify the best decisions. Solving SDPs often uses machine learning, but like deterministic optimization, searching for policies does not require a training dataset. For this reason, SDPs can produce behaviors that outperform people. This is how computer science was able to create programs for playing chess and computer Go that would outperform trained experts. Other applications can be found in optimal control (such as landing SpaceX) and optimizing fleets of trucks.

Levels 5 and 6, which are both a form of optimization, can produce tools that outperform humans. Such problems may be quite complex, either due to size (such as planning an airline schedule), or because of the inherent complexity of sequential decision problems. However, these are for well-structured problems with properties that can be fully specified by a mathematical model.

The slide below compares machine learning (on the left) to sequential decision problems (on the right). Machine learning searches for the best function (the statistical model) to match a training dataset. Sequential decision problems also search over functions (policies) to optimize some performance metric, given a transition function that describes how the state variable evolves over time. While machine learning searches over three types of functions (lookup tables, parametric models and nonparametric models), sequential decision problems search over four classes of policies, where the first class includes any function that might be used for machine learning.



Level 7) Science fiction

We reserve level 7 for unstructured problems which require the highest levels of intelligence. I view this as pure science fiction - it is easy to speculate, but I just do not see the economic justification to develop such an advanced technology (just consider the cost of training an LLM, that does not possess any of these capabilities). This does not mean computer scientists will not try to claim these behaviors (as some have tried to attribute these capabilities to LLMs), but I just don't see it happening. Technology (especially expensive technology) does not happen unless there is a significant economic benefit to people.



Some examples of these highest levels of intelligence include:

- Knowledge representation reflects the ability to characterize a poorly defined problem. For example, we may have a robot that needs to find its way around an urban area, or a desert, or a room full of furniture. The robot needs to characterize the area that it is in and identify the questions that need to be answered that are suitable for the context. Then this information needs to be organized in a way that the robot can accomplish some objective.
- Creativity – Optimization involves finding the best of a well-defined set of decisions, where the set of decisions are clearly articulated, with a performance metric that can be used to evaluate different decisions. Creativity Is required when we have a performance metric but may not specify the decisions that are available. For example, we may wish to reduce the number of deaths from COVID, but do not have a well-defined set of actions (or decisions). In English, a new type of decision is called an "idea."
- Judgment – Decisions for simpler, well-structured problems have clear metrics (winning a game, maximizing profits, treating a medical condition), but there are problems that involve complex judgments such as choosing whether to swerve to avoid hitting a cyclist but sending the car into the path of another car, or whether to bomb a military target that will also kill civilians.
- Reasoning - This requires the ability to think through different steps to achieve a goal. While this is done in levels 5 and 6 in the context of well-structured problems, reasoning requires being able to think through steps that are less-structured or completely unstructured. This cannot be done using large language models since these are fitted with the singular goal of minimizing the difference between "predicted words" and "actual words", while reasoning requires the ability to incorporate other goals.

If computers ever make serious headway into unstructured problems, it is entirely possible that the skills may have to be divided into different levels of intelligence. In fact, we could probably say that is true of knowledge representation and creativity.

Notes:

1. We have three levels of artificial intelligence, 2, 3 and 4, that all involve supervised machine learning, and all three may include the use of neural networks. This raises the question of why these are at different levels. The reason is that the machine learning technologies, and most specifically the use of neural networks, are on completely different levels, from basic neural networks, to deep neural networks used in pattern recognition, to the ultra-deep networks used in large language models.

   Even more important is the dramatic difference between the problems that are being considered. Level 2 is the familiar set of estimation problems that have been studied since the early 1900s. By contrast, level 3 addresses a variety of pattern recognition problems that exhibit structure completely different than the problems addressed in level 2. In addition, it was with these pattern recognition problems where deep neural networks became prominent in the early 2000s which created what is currently meant when people say "AI."

   We created an entirely new level, level 4, for the large language models that only recently emerged on the public stage. We put this in a new level for two reasons: a) the neural networks are much larger and require substantially more training that the pattern recognition problems, and b) large language models as an application are fundamentally different than the pattern recognition problems in level 3. Word patterns are much more complex, and the response (often called a "label" in the machine learning literature), which consists of a set of words, is much more varied (we would say the response is "stochastic").

2. Optimization, which is represented in levels 5 and 6, is fundamentally different from machine learning since it does not use a training dataset. Instead, optimization requires a model of the problem being optimized, along with well-defined decisions.

   We put sequential decision problems into their own level because they require a fundamentally different way of thinking about optimization problems. Instead of determining a single vector x, we now have to search over functions (known as "policies"), just as we do in machine learning. However, policies for sequential decision problems fall in four fundamental classes:

   a) Policy function approximations (PFAs) – These are functions that map information in the state variable to a decision.
   b) Cost function approximations (CFAs) – These are parameterized versions of optimization models.
   c) Value function approximations (VFAs) – These are the policies based on approximations of Bellman's equation. This is what most people associate with reinforcement learning.
   d) Direct lookahead approximations (DLAs) – These policies are based on solving approximate lookahead models.

   The first class, PFAs, includes every function that might be used for machine learning, including lookup tables, linear or nonlinear models (including neural networks), and nonparametric

models. Hybrid policies can be created from combinations of two, three and even all four of the policies above.

3. Almost every use of “artificial intelligence” in the press today refers to the use of supervised machine learning where a training dataset is used to estimate a function so that it mimics the data. This might involve estimating demand as a function of price (level 2), identifying a picture as a flower (level 3), or “predicting” what someone might say based on a set of inputs (level 4). This raises a valid question: Can mimicry be true intelligence?

   We often associate true human intelligence with examples such as Einstein inventing general relativity or deriving E=mc^2. Or perhaps it might be the invention of the silicon-based transistor to replace vacuum tubes. Articles in the press continually stress over the potential for AI to make strategic military decisions which requires skills from level 7. Clearly these are examples of very high levels of intelligence.

   The case can be made that most day-to-day human behavior, which certainly requires intelligence, represents a form of mimicry. While universities aspire to teach higher level thinking, our educational system is founded on teaching students how to write well, how to solve well-structured problems, and what to do in certain situations. Medical schools and the medical literature are dominated by teaching “medical protocol” (that is, what to do in a specific situation). All of this is mimicry. From this background, a few students emerge who exhibit truly independent thinking.

4. There is considerable talk about more advanced forms of intelligence such as “artificial general intelligence.” There are endless books and articles in the press that discuss “artificial intelligence” as if machines have achieved what we are calling Level 7 intelligence. To the best of our knowledge, no computer has displayed true level 7 intelligence. All the tools I have seen in practice fall in the first six levels.