5 — Machine Learning

Fast forward, then, to the present—to deep learning and affiliated machine learning (ML) technologies associated with second-wave AI.[1] These systems have made definite progress on the first and second of GOFAI's failures (neurological and perceptual)—and have arguably begun to address, though they have by no means yet fully embraced, the third or fourth (ontological and epistemological).

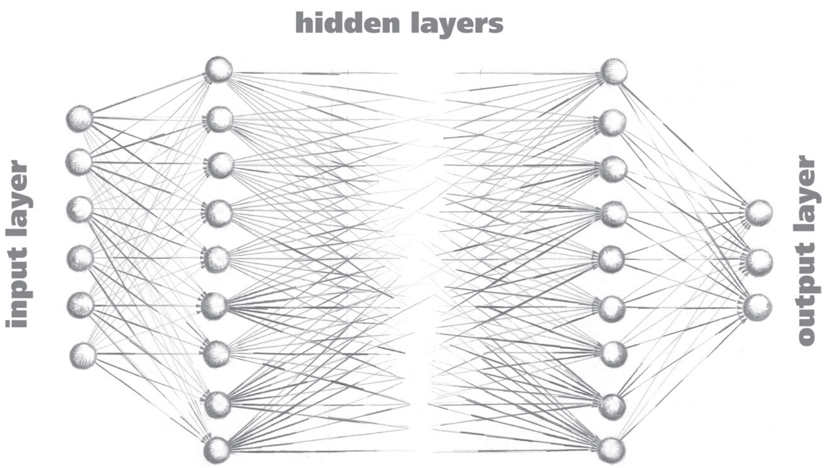
ML is essentially a suite of statistical techniques for:

1. the statistical classification and prediction of patterns
2. based on sample data (often quite a lot of it)
3. using an interconnected fabric of processors
4. arranged in multiple layers.

These techniques are implemented in architectures often known as "neural networks," because of their topological similarity to the way the brain is organized at the neural level. Figure 7 illustrates a way in which they are often depicted, but a better way to understand contemporary machine learning is in terms of the following four facts.

1. Though the phrase 'machine learning' was employed in era of first-wave AI, I will use it here, especially the 'ML' acronym, in its contemporary sense: to refer not only to deep learning algorithms but also to a variety of follow-on technologies, including deep reinforcement learning, convolutional neural networks, and other techniques involving statistical computations over complex graph configurations.

For deep learning in particular, see Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, "Deep Learning," *Nature* 521, no. 7553 (2015): 436–444.



Sandra Danilovic

Figure 7

D1 **Correlations**: As we have seen, first-wave (GOFAI) systems were built to entertain and explore the consequences of symbolically articulated discrete propositions implemented as formal symbols representing objects, properties, and relations in terms of a presumptively given formal ontology. Based on this model, rationality and intelligence[2] were taken to involve deep, many-step inference, conducted by a serial process, consisting of one or a few threads, using modest amounts of information, formulated in terms of a small number of strongly correlated variables (sidebar, next page). Standard logical connectives, such as negation (¬), conjunction (∧), disjunction (∨), implication (⊃), and the like, procedure and class definitions, and so on, can be understood as various forms of 100% positive and

2. Or thought, or cognition—as indicated earlier, no distinctions were being drawn at the time.

negative correlation. The model makes sense under the classical assumption of formal ontology, particularly under the grip of Descartes's desideratum of "clear and distinct ideas."

As indicated in the sidebar, below, contemporary machine learning is essentially the opposite. It consists of shallow (few-step) inference conducted by a massively parallel process using massive amounts of information, involving a huge number of weakly correlated variables. Moreover, rather than "exploring the consequences" of such correlations, its strength is to learn and reproduce *mappings* between inputs and outputs. Whether the mappings should be understood as relating causal patterns in the machine

GOFAI vs. Machine Learning

The most compact way to understand the difference between GOFAI and machine learning is in terms of their opposing positions on five conceptual axes.

GOFAI

1. *Deep* (many-step) inference
2. By a *serial* process, using
3. *Modest* amounts of information
4. Involving a *relatively small* number of
5. *Strongly* correlated variables

Machine Learning

1. *Shallow* (few-step) inference
2. By a *massively parallel* process, using
3. *Massive* amounts of information
4. Involving a *very large number* of
5. *Weakly* correlated variables

(i.e., as uninterpreted mechanical patternings) or complex representations of configurations of the world (i.e., as interpreted) is a question we need to examine. Most literature appears to discuss it in terms of mechanical configurations, though the critical probabilities are always understood in terms of what is represented.

What is called "face recognition" is widely touted as an ML success. But like many other terms uncritically applied to computational systems, the term "recognition" rather oversells what is going on. A better characterization is to say that ML systems learn mappings between (i) images of faces and (ii) names or other information associated with the people that the faces are faces of. We humans often know the referents of the names, recognize that the picture is a picture of the person they name, and so forth, and so the systems can be used *by us* to "recognize" who the pictures are pictures of.[3]

To be cautious, I will mark with corner quotes ("⌜" and "⌝") terms we standardly apply to computers that I believe rely on our interpretation of the semantics of the action or structures, rather than anything that the system itself can be credited with understanding or owning. Thus: image or face ⌜recognition⌝, algorithmic ⌜decision making⌝, and so on. (Perhaps we should even say ⌜computing⌝ the sum of 7 and 13, but that is for another time.)

3. If the capacity is built into a camera, one might argue that at least the camera is computing a mapping between *real people* and other information about them. But the question of whether what the system associates with the other (represented) information is the person in view, or their representation on the camera's digital sensor, is vexed. See the discussion of adversarial examples in chapter 6, note 5 (p. 57).

D2 **Learning**: Perhaps the most significant property of ML systems is that they can be *trained*. Using Bayesian and other forms of statistical inference, they are capable of what I will call ⌜learning⌝—a holy grail of AI, with respect to which the classic first-wave model provided neither insight nor capacity.

Several architectural facts are critical to the capacity of ML systems to be trained. The complexity of the relatively low-level but extremely rich search spaces, architecturally manifested in high-dimensional real-valued vectors, enable them—given sufficient computational horsepower (see D4, below)—to use optimization and search strategies (especially hill-climbing) that would be defeated in low-dimensional spaces.[4] Equally important, at the relevant level of abstraction the correlation spaces need not be discretely chunked—allowing steady incremental transitions between and among states, the epistemological opposite of ideas remaining "clear and distinct."[5]

Metaphorically, we can think of these processes

---

4. The higher the dimensionality of the search space—the greater the number of independent variables—the less likely it is that hill-climbing algorithms (strategies that move in the direction that, locally, has the steepest upwards slope) will encounter local maxima.

5. It was never clear how Cartesian models could accommodate the gradual shifting of beliefs or of concept meanings, except by the excessively blunt addition or removal of specific discrete facts. Estimable efforts were made within the GOFAI assumptions, including the tradition of non-monotonic reasoning and belief revision or maintenance. See, for example, Jon Doyle, "A Truth Maintenance System," *Artificial Intelligence* 12, no. 3 (1979): 231–272; and Peter Gärdenfors, ed., *Belief Revision* (Cambridge: Cambridge University Press, 2003). But it is fair to say that learning remained an Achilles' heel of first-wave AI.

as moving around continuously in the submarine topography depicted in chapter 3's figure 6 (p. 34), making much less mysterious their ability to "come above water" in terms of what we linguistic observers take to be "discrete conceptual islands." That is not to say that the conceptual/nonconceptual boundary is sharp. Whether an outcropping warrants being called an island—whether it reaches "conceptual" height—is unlikely to have a determinate answer. In traditional philosophy such questions would be called *vague*, but I believe that label is almost completely inappropriate. Reality—both in the world and in these high-dimensional representations of it—is vastly richer and more detailed than can be "effably" captured in the idealized world of clear and distinct ideas. (There is nothing vague about the submarine topology of figure 6; it merely transcends ready conceptual description.)

D3 **Big Data**: Once trained, machine learning systems can respond to inputs of limited complexity (though often still substantial; a single image from a good digital camera uses megabytes of data). Training these systems, however, at least given the present state of the art, requires vastly more data. This is why machine learning is at present a "post Big Data" development; training involves algorithms that sort, sift, and segment massive amounts of it, culling statistical regularities out of an overwhelming amount of detail.[6]

6. Humans may require massive initial training sets, too—the idea being that early childhood may be a long training sequence for infants, in order to set up the initial prior probabilities needed for subsequent recognition and processing.

D4 **Computational Power**: Training algorithms can require phenomenal amounts of computational power.[7] Some systems in current use employ the parallel processing capacities of banks of GPUs (video cards)—up to thousands at a time, each capable of processing thousands of parallel threads at gigahertz speeds.

The last two points are historically significant. As Geoffrey Hinton has remarked,[8] they reflect the substantial truth in the (in)famous 1973 Lighthill Report,[9] which threw cold water on the idea that first-wave AI could ever scale up to produce genuine intelligence. Given not only the ideas on which it was founded, but also the amount of computational power available at the time, first-wave AI was indeed doomed. The million-dollar, room-filling computers on which GOFAI was developed[10] had less than a millionth the processing power of contemporary cellphones; banks of current-day parallel processing video cards can extend that power by yet additional factors of hundreds or thousands.

But AI moves forward. Using different ideas, masses of collected data, and radically improved hardware, the results of machine learning are genuinely impressive. Recurrent networks, deep reinforcement networks, and other architectures are being developed to deal with time, to push feedback from later stages in a process back to earlier ones, and so on. New accomplishments are published

7. This is especially true of those in use at the time of this writing.
8. Geoffrey Hinton, personal communication, 2018.
9. James Lighthill, "Artificial Intelligence: A General Survey" in *Artificial Intelligence: A Paper Symposium*, Science Research Council, 1973.
10. Primarily Digital Equipment Corporation PDP6s and PDP10s.

almost daily—transforming machine ⌜translation⌝,[11] ⌜reading⌝ X-rays, filling in deleted portions of images, and such. Certainly AI researchers are more excited and optimistic than they have been in 50 years; it is not just the press that is heady. I too agree that the developments portend profound changes to the nature of society and our self-understanding.

Does that mean that we have figured out what it is to think? I think not.

11. Google translation is especially impressive when the languages are similar linguistically and capable of similar registrations; increasingly less so as these similarities fall away.