Phenomenology, Probability and Generative AI

by Ken Archer

Phenomenology has never been so important as it is today. Science continues to generate models, a growing number of which are AI models producing groundbreaking technologies, whose potential for benefits and for harms to humanity grows and seems to grow together. This interplay makes it challenging to orient the development of science and technology in a productive and responsible manner.

Husserl's phenomenology, particularly his final unpublished work, *The Crisis of European Sciences*, clarifies the internal structures orienting the advance of science and scientific modeling, and the interplay of benefits and harms that results. Phenomenological reflection on probabilistic modeling clarifies the nature of their enormous achievements, in particular those of Artificial Neural Networks (ANNs) underlying Generative AI, as well as the risks that probabilistic models have historically posed of objectivism and dogmatism. A clear understanding of the nature of probabilistic models, including ANNs, is critical to mindful and responsible development and application of such models.

1. Artificial Neural Networks and Probability

The uncanny achievements of Artificial Neural Networks (ANNs) in language translation and generation, and in image recognition, call for some kind of account. How are we to make sense of ANNs that can carry out a lucid conversation on advanced topics, and classify and generate images?

To make sense of ANNs, the central question that has been placed before cognitive science and philosophy is whether ANNs exhibit cognitive capacities, such as planning or causal inference. While this seems appropriate, given the concern of cognitive science and philosophy to understand cognitive capacities, this paper suggests that the achievement of ANNs is clarified and understood not only through interaction, relating to them from the outside, but through reflection on their internal architecture. This architecture manifests a conceptual achievement in the history of science that has gone largely unnoticed - the scientific modeling of learning through probabilistic modeling of the brain's plasticity.

ANNs that are used for image classification, for example, generate a probability distribution over possible classes of images. A probability distribution describes all the possible outcomes of a random variable, and the probabilities of each outcome. Mathematical probability is essentially the quantification of our uncertainty on a common scale, typically between 0 and 1. While a user interface on top of such an ANN can present the outcome with the highest probability as *the* classification, creating the appearance of a deterministic model, the model is in fact probabilistic. This probability is calculated by applying weights to sensory input, and it is these weights that are learned over time.

Similarly, language generation models that are used for generative text applications like ChatGPT are commonly characterized as predicting the next word. More specifically, language models generate a probability distribution across all words in a word corpus according to the probability of each word being the next word in a sequence of words, and select the word with the largest probability. And, again, it is weights applied to the sequence of words presented to the ANN that generate the probability distribution, and it is these weights that are learned when the language generation model is trained.

Furthermore, the central obstacles overcome throughout the history of ANN development have related to probabilistic modeling - specifically the structuring and updating of weights that generate probability distributions in the outer layer.

Why is it that mathematical probability is so central to the actual research driving ANNs, and yet is so rarely thematized in research seeking to understand AI? This dynamic - treating the modeling process itself as a technique, rather than as constitutive of the final model in ways that reveal its achievement and limitations - is not an oversight of commentators. It's a central aspect of the scientific process itself that phenomenological reflection upon scientific modeling helps to disclose. The growing distance of increasingly advanced scientific models from the lived experiences of scientists and non-scientists alike is central to their power for advancing science and technology as well as their tendency to objectification that motivates an unscientific dogmatism. Through reflection on the modeling process itself, these dual dynamics become evident.

To get a sense for how such a probabilistic model achieves the success that it does, it helps to start with a basic neural network, and then work our way up to contemporary computer vision and large language models.

The first functioning artificial neural network, called the perceptron, was designed by a psychologist named Frank Rosenblatt. It seemed a given to most psychologists that cognition approximates formal logic. The recent imaging available of neural activity seemed to even confirm this, given the functional similarity observed by a number of psychologists of neurons that fired in neural networks of the brain and the logic-gates of boolean logic.[1] However, Rosenblatt argued that symbolic logic could not account for the actual development of neural networks. In particular, recent discoveries of neural plasticity, particularly by psychologists such as Donald Hebb who studied the striking adaptability of damaged brains in returning WWII soldiers to reorganize and regain lost capacities[2], seemed resistant to explanation in terms of deterministic formal logic. The beginning of modern neural networks that are the basis for contemporary AI achievements essentially began with the following conceptual innovation from 1958.

> Unfortunately, the language of symbolic logic and Boolean algebra is less well suited for such investigations. The need for a suitable language for the

[1] McCulloch, W. S., & Pitts, W. (1943). A Logical Calculus of the Ideas Immanent in Nervous Activity. Bulletin of Mathematical Biophysics, 5(4), 115-133.
[2] Hebb, D. O. (1949). The Organization of Behavior: A Neuropsychological Theory. New York: Wiley.

mathematical analysis of events in systems where only the gross organization can be characterized, and the precise structure is unknown, has led the author to formulate the current model in terms of probability theory rather than symbolic logic.[3]

Rosenblatt's central innovation was the introduction of mathematical probability to model the stochastic development of brains. In particular, Rosenblatt modeled how weights he assigned to inputs, and which generated the probability distribution of outcomes, could be trained based on data. The perceptron was thus an idealized model of a biological neural network.

The early implementations of such neural networks modeled image recognition. To recognize an image, an idealized pixelation of the image generates sensory stimulus "intensity" values for each "cell" of the image. Each "cell" also has an associated weight, which indicates its importance in determining a particular classification. Each possible classification generates a probability by multiplying each incoming neuron's value and weight, and then summing these products, an operation known as the dot-product.

But how are the weights learned? This is the essence of any probabilistic model. First, the difference between the true outcome of a classification (1 or 0) and the probability of this classification is calculated. This difference - the error - is weighted for each incoming neuron according to its "intensity" value, in other words, according to its importance in determining the probability. Finally, the weighted error is added to the initial weight to generate an updated weight.

Through this training of weights associated with each neuronal connection, an ANN models the development of a neural network. However, this training algorithm would prove to be the central challenge for ANN development. The use of only two layers of neurons - an input and output layer - was immediately recognized as the chief limitation of the perceptron model. Obviously brains have numerous layers of neurons. It was not at all clear how to train the weights of hidden layers of neurons.

Once this challenge was overcome, through an application of differential calculus called backpropagation,[4] neural networks could include hidden layers of neurons. When neural networks include multiple hidden layers, they are known as deep learning, with the depth corresponding to the number of hidden layers. To train multilayer neural networks, a scientist uses trial-and-error to select a number of hidden layers, and number of neurons - which going forward we will call units so it's clear when we are speaking of physical neurons - in each hidden layer, until they find the network with the smallest predictive error.

The central customization of this model for computer vision is the structuring of the units in the form of maps to identify patterns across contiguous pixels of a visual field. Maps are grids

---

[3] Rosenblatt, F. (1958). "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain". Psychological Review, 65(6), 386-408.
[4] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). "Learning Representations by Back-Propagating Errors". Nature, 323(6088), 533-536.

of optical pixels, with weights associated to each pixel in the grid in a way that enables the detection of shapes, such as vertical edges, horizontal edges or curves. With multiple hidden layers, lower level shapes detected in lower unit layers can be combined into higher level shapes such as squares or triangles in higher unit layers, and ultimately shapes of objects at the output layer. This type of ANN is known as a Convolutional Neural Network (CNN).

Most well known today are the large language models (LLMs) that are used for machine translation like Google Translate, generative text like ChatGPT and other applications. Just like CNNs are ANNs customized to capture the gridlike contiguity relations among optical pixels for computer vision, LLMs use transformers which are ANNs customized to capture the relations between words in a sequence of words like a sentence or a paragraph. There are two essential elements of transformers that capture relations between words - word vectors and attention.

First, whereas CNNs structure input as grids of pixelated images, word vectors structure words as a vector of numbers that capture co-occurrence relations with other words. Before word vectors, ANNs treated words as arbitrary numbers. So, if a word corpus had 20,000 words, each word would be assigned to one of 20,000 input units, and as each word was fed sequentially into the ANN it would set that input unit to 1 while all others would be set to 0. Such models learned too slowly, because word co-occurrence patterns learned for one word would not generalize to other words with the same or similar meaning, as each word is just a meaningless number.[5]

In 2013 a group of Google researchers solved this problem by replacing arbitrary numbers with word vectors that capture their word co-occurrence relations.[6] A vector is essentially an array of numbers, rather than a single number, and refers to a line from the origin of a multidimensional space (0,0) to the point indicated in the vector. So, the vector [2,3] refers to a line from the origin to the point indicated by x=2 and y=3 on a 2-dimensional graph. This is similar to the way RGB values encode individual colors along 3 dimensions using 3-valued vectors. The idea is to capture the multiple dimensions of a word's meaning by identifying its patterns of co-occurrence with other words. In the words of linguist John Firth, "You shall know a word by the company it keeps."

These word vectors are learned by first creating a training data set of all word sequences - word pairs, or sequences of three or more words - from a large corpus of text. A neural network is then fed each word from each word sequence from the corpus, with arbitrary numbers to identify words as before, and trained based on their accuracy in predicting neighboring words. Once the network is trained, each word's vector is generated by feeding the word as input into the network and storing the weights input into the final hidden layer. Language models now always use word vectors as inputs, in essence using the outcome of one probabilistic model into another probabilistic model.

---

[5] Bengio, Y., Ducharme, R., Vincent, P., & Jauvin, C. (2003). "A Neural Probabilistic Language Model". Journal of Machine Learning Research, 3, 1137-1155.
[6] Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). "Efficient Estimation of Word Representations in Vector Space". Proceedings of Workshop at ICLR.

Word vectors enable us to use vector arithmetic to find pairs of words with similar relations. This is one of the most striking features of language models. Capitals of countries, gender associations, even analogies can be discovered through vector arithmetic.

The other central innovation that enables language generation applications is called attention, which enables words to take on different meanings in different contexts.[7] The input to a transformer ANN consists of all of the word vectors for all words in a sequence. Transformers add to typical ANNs an attention step to each word in the sequence, in which any other word from the sequence that is similar to the word (as determined with vector arithmetic) modified the word's vector (again, with vector arithmetic), thus accounting for relevant context.

These are the essential innovations behind contemporary achievements in AI. The tremendous accomplishments of computer vision, language translation and question and answer implementations of ANNs are provocative. They suggest that vision and speech are computational capacities, of which animal vision and human speech are only limited examples. This has prompted soul searching about the proper relation that we should strike with AI.

On the other hand, perhaps this is all an over-reaction. For many observers, language models are merely next word predictors, a big spreadsheet of word co-occurrences, that will hit a wall, on the other side of which is what is unique to human intelligence, like common sense or causal inference.

To provide some insight into the successes and failures of ANNs, as mentioned before, the cognitive science and philosophy communities have evaluated these systems based on their interactive capabilities, but what of their internal architecture? Surely that is relevant to the question of their capacities? The challenge, however, is to identify what feature of their internal architecture is relevant to the question.

Discussions of the architecture of generative AI typically focus on innovations from the decade prior to the first successful generative AI systems - word vectors and attention-based transformers. However, while these are the most immediate innovations that enabled generative AI, they are fundamentally implementations of the theoretical model first introduced by Rosenblatt. Like the grid-like maps of computer vision ANNs, word vectors and transformers are ways of structuring the weights that were the centerpiece of Rosenblatt's perceptron model, and which generate probability distributions. And perhaps the central innovation that enabled ANNs, backpropagation, also centered on weights, specifically updating weights during model training.

The central question that ANNs raise, from perceptrons to LLMs, is whether capacities like vision and speech can be modeled as input-output systems in which weights are applied to low-level inputs to generate probability distributions over meaningful outcomes such as image classifications and next words. Often in science the ability to experimentally validate a theory

[7] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., & Polosukhin, I. (2017). "Attention Is All You Need", Advances in Neural Information Processing Systems (pp. 30-48).

waits decades.[8] In the case of ANNs, data in sufficient volumes to validate the perceptron was not available until widespread use of the Internet and sufficient compute power was not available until the last two decades, and it was likely these two factors that spurred much of the engineering innovations needed to validate the fundamental model of ANNs.

The success and failures of ANNs, then, call for a reevaluation of their motivating theory, the probabilistic model of cognitive capacities. This was heavily contested within psychology and AI when initially proposed by Rosenblatt, but perhaps a reevaluation of this basic model sheds light on the successes and failures in its recent implementations.

2. A Reevaluation of Probabilistic Models

The role of probability in ANNs, while central to their initial formulation by Rosenblatt and to every major innovation in ANNs, is not central to the study of AI within cognitive science and philosophy. The distance of advanced probabilistic models from the pre-scientific experience of scientists and non-scientists alike contributes to their power, but also to their peripheral treatment in cultural understanding of probabilistic science and to risks of misinterpretation and misuse of this science. Here we seek to close this distance by starting with their meaning foundation in everyday experience, what Husserl termed the life-world.

It is often said that probability is a measure of uncertainty, so clarifying probability would seem to require clarity about uncertainty. In some interpretations, uncertainty simply is mathematical probability. But does this make sense?

Uncertainty, and the scientific structures that reduce uncertainty, are in fact both grounded in the contents of everyday intuition. It is only an abstraction that conceives of the contents of intuition as atoms of sensory input which are passively received by sensory organs. The sensory organs are always in motion, as is the sensible world of their surroundings, creating a flow of environmental intuition in which change is played off against permanence, variation is played off against invariants, and thus ambiguity is played off against structure.

It is an abstracted limit case to speak of purely passive hearing or seeing of purely inert input. We focus our ears on a speaker against a background of ambient noise, we direct our gaze on an object in an active scene, we move around to explore, all according to the mobile range of action of our bodies.

The stream of variation that results is given coherence through invariants that are picked up by our cognitive capacities. To render coherent the ambiguous flow of sensory experience, we always intend more than the present momentary profile of appearances, intending structures that persist through present and anticipated future profiles. Each sensory profile points beyond itself to a horizon of possible profiles, and invariants that structure this horizon.

[8] As Thomas Kuhn observes, "The first direct and unequivocal demonstrations of [Newton's] Second Law awaited the development of the Atwood machine, a subtly conceived piece of laboratory apparatus that was not invented until almost a century after the appearance of the *Principia*." Quoted in Sokolowski, Robert, *Pictures, Quotations, and Distinctions: Fourteen Essays in Phenomenology* (Notre Dame: Notre Dame Press, 1992), p. 150.

The most fundamental invariant is objects. For instance, while we perceive one side of a chair at a time, we intend the chair as a persistent object, as a chair, transcending the 2-dimensional profiles available to us through co-intending a variation of other profiles (e.g. the back of the chair) that are not intuitively given to one's perception at that moment. Object persistence is one of the fundamental invariants through which we structure our expectations of the varying profiles of flowing intuition.

Furthermore, we co-intend a variation of profiles potentially available to oneself and to others, as structures are constituted over against the flow of intuition through intersubjective recognition of such structures. Objectivity - the constitution of object structures - is only constituted over against the flow of sensory activity through intersubjectivity. Structured horizons thus develop intersubjectively, and reflect typical expectations of communities.

There are many other invariants through which the varying flow of appearances are stabilized into a structure of expectations. As a physical object, the chair exists in 3 dimensions, and thus has a front and a back. A primordial invariant through which animals structure their expectations of objects are the affordances provided by that object. A chair affords sitting. A branch affords digging or hitting.

Ambiguity and structure, uncertainty and certainty, variation and invariants, are thus reciprocal to each other in pre-scientific experience. Without one, we would lose the other, and simply have an undifferentiated throng of sensory experience that would be neither ambiguous nor structured. This structured ambiguity of the horizon of everyday, pre-scientific experience exerts a pull within consciousness towards greater coherence and structure, motivating the structures of science. We uncover further horizonal structures primarily through making distinctions between objects. A chair is non-living and thus will not die, as will living objects. Within chairs, we can distinguish between those made of material from dead objects, which will decay, and those made from artificial materials, which will not.

Such distinctions between kinds of objects are the basis for the classical science of formal causes that preceded the Scientific Revolution, and that continues to drive much of scientific discovery occupied with classification. The standard for such apodictic definitions through causes was Euclid's geometry, though natural philosophy, as inclusive of ambiguity, fell short of Euclidean validity.[9] The interplay between proto-scientific, everyday experience, and classical science is thus quite obvious, with each being closely related to the other, in a vital iterative relation. In fact, the more general sense of science prior to the exact sciences of the Scientific Revolution, in which any organized body of knowledge, including practical knowledge, was considered a science, reflects this closer connection of science with pre-scientific experience.

---

[9] "What is more, with Euclidean geometry had grown up the highly impressive idea of a systematically coherent deductive theory, aimed at a most broadly and highly conceived ideal goal, resting on 'axiomatic' fundamental concepts and principles, proceeding according to apodictic arguments". *Crisis*, p. 21

Classical science, despite its movement toward greater clarity and predictive accuracy, nonetheless possesses an inherent ambiguity that eventually motivates a more exact way to structure the varying profiles of intuitive experience. The major ways of knowing in science - classical science, exact science, probabilistic science - are distinguished in terms of how they bring structure to the flowing variation within intuitive experience.

Central to the transformation of science in the 17th century was the elimination of vagueness and uncertainty through the mathematization of phenomena. This wasn't just a matter of using math more in natural science. The exact sciences idealize dimensions of perceptual experience - perfectly flat surfaces, perfectly thin rays of light, an infinitely small rate of change - by ordering the pre-scientific experience of phenomena - flatter and flatter surfaces, light cut by a hand into smaller and smaller rays, smaller and smaller rates of change - and then imaginatively projecting a limit-shape of this ordering that does not actually exist in intuitive experience. Through imaginative limit-shapes the exact sciences constitute idealized indices against which perceptual observations are projected.

The idealization of the exact sciences brings with it great power for scientific advance, but also entails a trade-off. On the one hand, idealizing new dimensions of intuitive experience uncovers new dimensions of objects - mass, acceleration, rate of change, optical rays, electrical charge - that can be used to discover new functional relationships with greater predictive accuracy than in the non-exact sciences.

On the other hand, scientific structuring of our intuition in terms of imaginative structures not actually found in ordinary, pre-scientific intuition comes at a cost. We don't experience in pre-scientific intuition perfectly flat surfaces or perfectly straight lines; we experience objects that have approximately flat surfaces, or approximately straight lines. We don't experience light rays or energy, we experience color and heat. We don't experience motion only as inertia in accordance with Newton's laws, but rather we also experience self-initiated motion by whole beings, like a cat getting up from a mat, which seems contingent and uncertain before it has happened. We experience both structure *and* ambiguity.

In pre-scientific experience, we experience whole objects - the most fundamental structure - whose parts are approximate. The exact models of post-Galilean science shift the focus from wholes to the mathematized parts or dimensions in terms of which wholes are now to be understood scientifically. What is gained is a new way to understand wholes. What is lost, in the shift of attention towards mathematical objects, is ambiguity, the ability to scientifically enlighten uncertain cases or phenomena. Were the enlightening benefits of science available in domains that lack the knowable exactness of physics? This question would give rise to probabilistic models.

Probabilistic models arose in response to the earliest exact models of mathematical physics. Prior to the discovery of mathematical probability in late 17th century Europe, there is no historical evidence of any precursors of such a concept, which has raised the question among historians of why classical probability was discovered, or invented, at this place and time.[10]

The motivation becomes much clearer, however, once we consider that the central figure in the discovery of mathematical probability, Jacob Bernoulli, discovered many concepts of critical importance to the exact sciences.[11] Bernoulli sought to overcome the limitations of these exact sciences when it came to uncertain phenomena.[12]

The key idealization that overcame these limitations and enabled mathematized reasoning about uncertain affairs was randomness. Of course, perfect randomness does not exist in ordinary experience. In his *Art of Conjecturing*, Bernoulli began with an extensive discussion of the mathematics of lottery mechanisms, and then imagined that commercial and civic affairs could be understood *as if* their outcomes were determined by such a lottery mechanism. While randomness seems quite familiar to us, it was completely unfamiliar to Bernoulli and those before him, such that it required someone steeped in the idealizations of perfect determinism to first conceive of its inverse, perfect randomness.

For Bernoulli, if we isolate a collection of past cases that are similar to an uncertain present case, we can count the proportion of such similar cases with the outcome of interest, and then quantify our uncertainty as if we were drawing from a lottery with the same proportion of outcomes.

> *Nevertheless, another way is open to us by which we may obtain what is sought. What cannot be ascertained a priori, may at least be found out a posteriori from the results many times observed in similar situations, since it should be presumed that something can happen or not happen in the future in as many cases as it was observed to happen or not to happen in similar circumstances in the past. If, for example, there once existed 300 people of the same age and body type as Titius now has, and you observed that two hundred of them died before the end of a decade, while the rest lived longer, you could safely enough conclude that there are twice as many cases in which Titius also may die within a decade as there are cases in which he may live beyond a decade.*

What is going on here? Bernoulli here specifies for the first time the two key idealizations that constitute a probabilistic model, which are largely unchanged to the present-day.

---

[10] Hacking, I. (1990). *The Taming of Chance*. Cambridge University Press. Daston, Lorraine. *Classical Probability in the Enlightenment*, (Princeton University Press, 1988)

[11] Bernoulli's contributions to the exact sciences include those he made to the calculus and the discovery of the mathematical constant *e*.

[12] Uncertain reasoning at the time was the domain of practical reason. Psychology and social science, today considered central domains of probabilistic modeling, were still considered part of philosophy, and would themselves be reformed under the later expansion of probabilistic modeling in the 20th century, in part through the success of ANNs.

First, Bernoulli assumes there is a static, knowable collection of “similar circumstances in the past”, which he defines in the case of mortality prediction as “people of the same age and body type”. This ultimately became known as the reference class problem - how to define the reference class of similar cases from which to infer an unknown case - and has been the central challenge faced by statistics, machine learning, and any application of mathematical probability. The idea of a reference class imagines a “data generating mechanism” that is static and knowable. The reference class of similar cases reintroduces distinctions between kinds of objects (“people of the same age and body type”) that the exact sciences of physics, with their functional relations between idealized dimensions, had banished.

But, whereas the distinctions between kinds of objects made in classical science can be directly applied to individual objects, probabilistic models cannot be applied to individual cases, only collections of cases. That is due to the second key assumption made by such models, the assumption that all variation *not* accounted for by the reference class (in the case of mortality, variation not correlated with variation in age and body type) is due to random variation, of the type we observe in lottery mechanisms.

Whereas the exact sciences idealize individual dimensions, probabilistic science idealizes the entire manifold of intuition as a random manifold, generated by an imagined data generating mechanism. The concrete basis for this idealization is the random distribution of lottery tickets generated by the ticket generating mechanism of a lottery device.

If we look back, then, we can see more clearly the direction of modern science, moving further and further from the whole objects that are the central structure of pre-scientific experience. Whereas classical science makes distinctions between objects in terms of their observed parts that retain the transcendence of the objects, Galilean science defines objects completely in terms of their parts, and Bernoullean science defines classes of objects in terms of their parts. With each stage of abstraction, science grows in its predictive and explanatory power for society while also growing in remoteness from pre-scientific experience of objects.

This is the source of the power and the crisis of modern science of which Husserl writes. This is a real problem, because it stunts the interplay that is vital to the fulfillment of both the mathematical sciences and pre-scientific experience. Mathematical sciences, as we’ve seen, actually develop through the discovery of newly ordered dimensions and new reference classes within experience, not through applications of pure math. And mathematized models of science must always be ultimately validated against ordinary experience to ensure their continued progress.

In turn, denying the reality of pre-scientific experience alienates our pre-scientific lives from their yearning for coherence amidst the ambiguity of pre-scientific intuition, thus alienating society from science itself. Social alienation from science, often discussed in terms of a declining “trust in science”, is not coincidentally often associated with probabilistic models, such as pandemic response or climate change models. Probabilistic models in medicine and psychology, too, are often expected to replace clinical judgment.

But there is another aspect of this direction of modern science that phenomenology discloses - the turn of science from a concern for whole objects and the distinctions among them, as in classical science, to a concern for the subjective experience of objects. The idealizations of Galilean science, the imagined limit-shapes, carry with them reference to the open horizon of pre-scientific limit approximations, in a sense mathematizing this horizon itself.[13]

Bernoullian science goes further, carrying with its idealization reference to learning itself, to the overcoming of variation through structure. Uncertain phenomena are only meaningful in relation to certainty, a relation that refers us to a learning process that must then itself be modeled.[14] This gradual turn of mathematics to idealizations of consciousness' structuring of intuition itself discloses mathematics as more than geometry or arithmetics, but as the science of structure.

Just as Galilean models set the stage for other idealized, exact models, Bernoulliean models set the stage for the discovery of other probabilistic models of an outcome whose deterministic factors are not completely knowable, by idealizing the outcomes within a static, knowable reference class of similar cases to be randomly generated. This is obviously a powerful type of model for science, with widespread practical applicability, but one whose remoteness from everyday experience increases the risk of objectivism and dogmatism when probability is naturalized as a feature of the mind or objective reality.

This confusion is present in the philosophy of probability, whose central subject matter is the location of randomness, of probability, either in the mind or in objective reality. But talk of spatial location already presumes a naturalized world in which everything is lying around either outside the brain or inside of it. The idealizations of science, however, including the data generating mechanism that generates random data that constitutes probability, are not located anywhere. That doesn't make them false or fictions.[15] They are perhaps best thought of as methodological devices that enable us to organize the contents of intuition in insightful ways.

This is how 20th century economists Keynes and Knight conceived of probability. In probabilistic models, uncertainty is typically regarded as a quantifiable measure of doubt. However, there is another axis of uncertainty to quantifiable doubt, one not captured or acknowledged by much contemporary statistics, but thematized by Keynes and Knight and a constant presence in the probabilistic modeling process – non-quantifiable ambiguity within our pre-scientific experience.[16] When we are in a state of ambiguity, we are unable to elicit a

[13] "Hence we always have an open horizon of *conceivable* improvements to be further pursued." Husserl, E. (1970). *The Crisis of European Sciences and Transcendental Phenomenology: An Introduction to Phenomenological Philosophy* (D. Carr, Trans.). Northwestern University Press. p. 25.
[14] Hence, for Bernoulli, "[w]hat cannot be ascertained *a priori*, may at least be found out *a posteriori* from the results many times observed in similar situations."
[15] "Models are often understood in the context of an epistemology that separates the internal and the mental from the the external and the real….A model may carry more of our thoughtful activity, but it is still a transformation of something in the world." Sokolowski (1992), pp. 147-148
[16] Knight and Keynes in the early 20th century were notable for deepening our understanding of uncertainty as both doubt, or risk, and ambiguity. Frank Knight, Risk, *Uncertainty and Profit.* (1921,

quantitative measure of our uncertainty, as our understanding of the relevant features or dimensions of a situation are unclear. Only once we resolve our ambiguity around a situation are we able to abstract an idealization, a reference class within which to conceive a data generating mechanism, within which it makes sense to speak of quantifiable doubt.

3. The Achievements and Limitations of AI

The achievement and limitations of ANNs are clarified by recalling the idealizations through which they are constituted. Probabilistic models, unlike exact models of mathematical physics, model learning itself by idealizing the process of learning from data. This model requires the projection of a data generating mechanism that generates a distribution of probable outcomes. Assuming this distribution is static, it is possible to model this distribution.

All ANNs are premised on this idealization. For Rosenblatt, the plastic development of brains, such that two brains can acquire the same cognitive capacity through developing different neural networks, was conclusive evidence that cognition is about learning, not symbolic logic, and so must learn weights and biases that generate outcomes probabilistically.[17] Rosenblatt explicitly contrasted his approach with representationalist theories of cognition which are "forced to conclude that recognition of any stimulus involves the matching or systematic comparison of the contents of storage with incoming sensory patterns".[18] For Rosenblatt, such models cannot account for the process of learning a representation.

How to acquire weights that when applied to low-level input generate probability distributions of meaningful outcomes has motivated all major ANNs innovations, from backpropagation to update weights and gridlike maps for computer vision, to word vectors and attention-based transformers that enabled generative AI.

Rosenblatt was explicit about the assumptions that came along with the use of probability theory. Modeling intelligence as probabilistic, and not deterministic, while necessary to capture learning, meant he was shifting the focus of artificial intelligence from modeling individual brains to a data generating mechanism ("a defined environment") that generates classes of brains.[19]

Cambridge: The Riverside Press). John Maynard Keynes, *A Treatise on Probability*. (1921, London: Macmillan).

[17] "A perceptron, as distinct from some other types of brain models, or 'nerve nets', is usually characterized by the great freedom which is allowed in establishing its connections, and the reliance which is placed upon acquired biases, rather than built-in algorithms, as determinants of its behavior." Rosenblatt, Frank, *Principles of Neurodynamics* (Spartan Books: Washington, DC, 1962) p.5

[18] Rosenblatt (1958), p. 387

[19] A perceptron is never studied in isolation, but always as part of a closed experimental system, which includes the perceptron itself, a defined environment, and a control mechanism or experimenter capable of applying well-defined rules for the modification, or "reinforcement" of the perceptron's memory state. In most analyses, we are not concerned with a single perceptron, but rather with the properties of a class of perceptrons, whose topological organizations come from some statistical distribution. Rosenblatt (1962), p. 5

The central departure of ANNs from classical probability and statistics is fundamental to their success and challenges. In statistics, the reference class is defined in terms of distinctions between objects ("people of the same age and body type"). To refine this reference class, we can upweight or downweight the importance of a distinction. So, when we model probabilistically the likely outcome of an election, we define a reference class of likely voters.

ANNs learn these distinctions themselves, by applying weights to low-level inputs - optical pixels, words - and the relations learned between them, rather than to attributes of persons or cases that already relate to each other through these individual people or individual cases. What is lost is the top-down relations provided by objects, by contexts. What is gained is an expansion of our ability to model learning itself by allowing the model to explore relations freely.

Why does this work? The reason free exploration of relations between low-level inputs - optical pixels, words - picks up relations whose weights generate usually meaningful outcomes is precisely because of regularities in the training data, regularities which are only there because structuring of sensory contents is not bottoms-up. Let's look at vision models and then language models.

Bottom's up computer vision models work precisely because vision is not bottom's up. We interpret perceptual input through the selective registering of structure against change in one's perceptual field. Is this bottom's up or top down? It is bottom's up in the sense that it starts with intuition. It is top-down in the sense that it is not a passive reception of inert, pixelized intuition, but an active registering of structures that are meaningful within an intersubjective horizon of meaning.

To take an illustrative example, if vision training data is generated from a tropical region, snowflakes will have a certain optical structure that reflects a certain gaze, whereas an arctic region will generate vision training data in which many types of snowflakes will have distinct structures that reflect a different gaze. Vision isn't a matter of abstracting atomic relations outside of any larger context, but of picking up invariants that are registered as meaningful through the intersubjective, embodied motion of living.

What's notable about vision, then, is that each visual act is also an act of meaning-constitution. Vision and meaning-making are inseparable. A visual act is not a matter of processing pixels through a meaning filter, but of reconstituting visual fields into meaningful vision acts with every act of vision. This has two significant implications for ANNs.

First, the most fundamental structure through which perceptual fields are registered as coherent and meaningful is that of objects. As is well-known, object persistence is one of the central challenges of computer vision. This technical challenge of object persistence is the same as the neuroscientific challenge known as the binding problem - how to reassemble disparate sensory inputs into a single object. Object persistence is challenging for AI because it entails the detection of persistence when played off against change, and the intending of a persistent through a manifold of present and anticipated sensory profiles.

A well-known challenge of computer vision is recognition of scenes and visual context. It may be thought that this is a matter of relating bottom's up patterns to each other across spatially separated regions of a visual field, but this is not what composition of a scene entails. Composition of a scene is a composition of objects in relation to each other.

Second, to the extent that the meaning of a visual pattern is intersubjectively typical and normative, one can expect computer vision models registering the pattern to be more successful at inferring their meaning, but to the extent that the meaning of a pattern is created anew, then they will fail.

There is thus a symmetry between the behavior of ANNs, and of all probabilistic models, on the one hand, and of exact models of mathematical physics on the other, when it comes to limit cases. Exact models are valid for the limit case of physical phenomena unaffected by object wholes, such as living beings, which introduce uncertainty into physical phenomena. Probability mathematically models this uncertainty, but by imagining the limit case of a data generating mechanism of uncertain phenomena that is itself static.

In all probabilistic science there is a notion of distributions, and data that is in-distribution or out-of-distribution. Such concepts are bound up with the idealized fixed data generating mechanism, and are thus also idealizations, or, limit cases. Of course there are approximately stable patterns of life and meaning, but all acts are meaning-constituting acts with freedom to alter established meanings, such that the notion of a fixed data generating mechanism, or a fixed distribution in relation to which data is in or out, is a limit case.

Leibniz raised such an objection to Bernoulli - "the number of diseases does not always remain the same, but new diseases spring up daily". Bernoulli's response granted that this was a valid objection and, in line with his treatment of probability as a useful metaphorical device, limited the concerns of probability to processes like mortality and the weather that are stable for long periods.[20]

In language models, we see the role played by regularity of meaning even more clearly. The theoretical basis of language models is known as distributional semantics, the idea that the meaning of a word is reflected in its relations with nearby words. As with vision, this reflects a limit case in which meaning is static. And, most of the time, meaning is approximately static. But the structures through which we articulate a field of intuition - of objects and their parts, of the active and passive relations between objects - on the higher level of language, are rearticulated with every act of language. Such rearticulations may stay within the typical horizon of meaning, or may introduce new distinctions that alter this horizon.

---

[20] "We cannot deny that the number of diseases multiply with the passage of time, and anyone who wanted to draw an inference from today's observations to the antediluvian times of the Patriarchs would surely stray very far from the truth. But from this it only follows that new observations should be made in the meanwhile, just as with the tokens, if their numbers in the urn were assumed to change." Quoted in Weisberg, Herbert (2014). *Willful Ignorance: The Mismeasure of Uncertainty*. New Jersey: John Wiley & Sons. p. 144. From Bernoulli, Jacob (2006). *The Art of Conjecturing, together with Letter to a Friend on Sets in Court Tennis*. Edited by Edith Dudley Sylla. Baltimore: The Johns Hopkins University Press, pp. 329-330.

This is surely the reason why generative language models such as ChatGPT are most successful in closed semantic domains, most especially generation and debugging of software languages that are explicitly constructed and version controlled. However, software languages are the limit case of language, whereas natural language are always evolving as new distinctions are introduced through which we stabilize the ambiguity of experience.

The text generated by generative language models is thus less likely to fabricate meaningless text when it is within the stabilized, normative meaning horizon of an intersubjective community, or when it is within the distribution of its training data. Unlike humans, Generative AI models do not achieve language acts, which are articulations of intuitive structures on the propositional level of language, but they do learn the mathematical patterns of such articulations thanks to the availability of language.

ANNs thus reflect one of the great achievements of modern science. ANNs expand the domain of probability to encompass fundamental learning capacities and, in a mysterious way, extend the turn of mathematical science towards a modeling of subjective experience itself. We can imagine further idealizations that capture learning on an even deeper level, picking up persistents registered against change over time across multiple sensory modalities.

To gain a better understanding of the nature of this achievement, as well as its limitations, it is essential to clarify the mathematical idealizations through which such probabilistic ANN models are constituted. Mathematics is more than mere arithmetic. As a science of structure, mathematics enables us to idealize the invariants through which we bring structure to the variation of horizonal experience. To posit a wall of distinctly human concerns that are off-limits to science, as phenomenological hermeneutics often does, thus seems incongruous with the nature and historical achievements of mathematics.

Phenomenology, largely due to the work of Dreyfus, is sometimes associated with the skeptical stance towards AI as facing a wall of non-representational experience. Dreyfus appealed to a reading of classical phenomenologists, particularly Heidegger, according to which the horizonal structure of cognition is completely unconscious to us while we are mindlessly coping through daily life. Dreyfus pointed to experts in their field - expert swimmers or chess players - as exemplars of such mindless coping. It is only when our immersive engagement is interrupted by some breakdown of the functions played by our environmental tools that the horizon lights up for us, and we reflect on this horizon directly.

If this is how cognition works, then, as Dreyfus also claims, AI faces a wall in its progress towards more general, humanlike intelligence. That is because human expert cognition, and mathematical models, are incompatible. In fact, the latter threatens to crowd out the former. This hard distinction between mathematical and human domains of knowledge is reflected in much of post-Heideggerian phenomenological hermeneutics.

However, Dreyfus' interpretation of Heidegger has come under significant criticism within phenomenology.[21] For Heidegger, the horizonal network of references is always there in the background of our immersed being-in-the-world, otherwise there would be nothing to light up when it breaks down, and nothing to reflect on in phenomenology. Horizonal consciousness is like a light that shines from the back of our engagement in life, always there, though not always made an object of our reflection.

How, then, should one account for the horizonal content of cognition, without a representational theory of such content that ends up indistinguishable from a data processing view of cognition? The account provided here, of a proto-scientific, horizonal cognition, in which structured ambiguity motivates ecological distinctions that are non-representative and yet inform and motivate science, is inspired by the reflections of Husserl, Gibson and Sokolowski. These accounts contrast with accounts of cognition as an inscrutable domain of human concerns that are alien to mathematical science as one finds in Dreyfus.

While pre-scientific experience may motivate the idealizations of mathematical science, the further such idealized models extend from their pre-scientific basis in meaning, the greater the tendency to objectify such models as replacements of pre-scientific experience itself and render them incomprehensible to scientists and non-scientists alike. As probabilistic models in particular become more remote from the pre-scientific experience from which they arise, their persuasive power comes more from the technique of predicting and associated claims to supplant pre-scientific experience, and less from the clarity they provide to pre-scientific experience.

Just like objectivism casts idealized light rays and energy as more real than pre-scientific experience of color and heat which are subjective and illusory, so it casts human cognition as replaceable by artificial intelligence. We see this in an increasingly popular theory, originating with Helmholtz, known as predictive coding, active inference or the free energy principle, which casts vision and all capacities as hypothesis generation and confirmation based on probabilities.[22]

This is not to say that artificial intelligence is not an improvement upon human intelligence in significant ways. But one could make the same claim of all mathematized scientific models since Galileo, as they all *do* improve upon ordinary experience through math, not by replacing ordinary experience with pure math but by clarifying ordinary experience with mathematized idealizations of ordinary experience. Development and use of ANNs mindful of the idealizations through which they reconstitute pre-scientific experience too will benefit from AI and avoid the allure of replacing pre-scientific experience with these idealizations.

---

[21] Zahavi, Dan, "Mindedness, Mindlessness and First-Person Authority. In J. Schear (ed), *Mind, Reason and Being-in-the-World: The Dreyfus-McDowell Debate*, (London: Routledge, 2013), 320-343

[22] The flaw with this objectivism of probabilistic modeling is that it projects a radical neuro-representationalism according to which the brain stores multiple hypothetical representations explaining sensory input, and the probabilities of each, but exempts our apparently realiable knowledge of neuroscience from the radical skepticism that this entails. Zahavi, Dan, "Brain, Mind, World: Predictive coding, neo-Kantianism, and transcendental idealism", Husserl Studies.