WHAT IS AI?

January 2024

ABSTRACT

"What is AI?" is a question that demands real attempts at answering. There is a massive amount of confusion as to what it consists of but also a pressing demand to regulate it. Regulating the undefinable always leads to chaos. Herein we attempt to layout some canonical constructs resulting in what we call an Autonomous AI entity. This construct presents significant concerns and regulation may perforce of rogue developers be an a posteriori act. (Picture of Isaac Asimov by Phillip Leonian from New York World-Telegram & Sun)

Terrence McGarty

TGL 205

NOTICE

This document represents the personal opinion of the author and is not meant to be in any way the offering of medical advice or otherwise. It represents solely an analysis by the author of certain data which is generally available. The author furthermore makes no representations that the data available in the referenced papers is free from error. The Author also does not represent in any manner or fashion that the documents and information contained herein can be used other than for expressing the opinions of the Author. Any use made and actions resulting directly or otherwise from any of the documents, information, analyses, or data or otherwise is the sole responsibility of the user and The Author expressly takes no liability for any direct or indirect losses, harm, damage or otherwise resulting from the use or reliance upon any of the Author's opinions as herein expressed. There is no representation by The Author, express or otherwise, that the materials contained herein are investment advice, business advice, legal advice, medical advice or in any way should be relied upon by anyone for any purpose. The Author does not provide any financial, investment, medical, legal or similar advice in this document or in its publications on any related Internet sites.

Furthermore, this document contains references to and quotes and modified charts and figures from papers and documents under the premise of "Fair Use" in order to present ideas and understandings in context. The Author has attempted to make any and all references to such material separate from those of the author per se and has referenced the source expressly in all cases. These documents are for the dissemination of ideas and have no commercial intent.
Our approach herein is to take elements of what is recent in the literature focused on a specific topic and attempt to develop a tapestry image of these connectable elements. We do not necessarily provide any new or fundamental results but merely attempt to assemble elements in a systematic and holistic manner.

Communications relating to these documents and these should be sent to:
mcgarty@alum.mit.edu.

Terrence P. McGarty, Copyright © 2024, all rights reserved. This document is in DRAFT form and is solely for technical review and evaluation and it not intended for any commercial use.

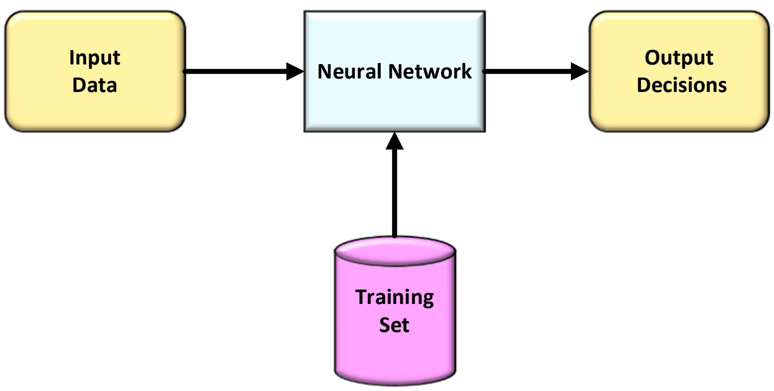
Contents

1 INTRODUCTION

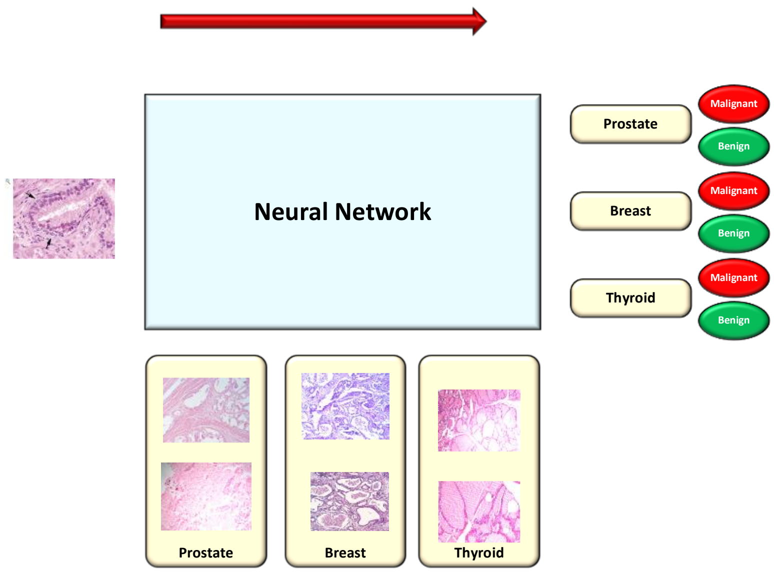
What is Artificial Intelligence? An examination of a Google search will list thousands of definitions, many convoluted and circular, namely defining intelligence as intelligence. As we have noted elsewhere, the problem of not having a clean and clear definition makes it impossible to create laws, yet this never seems to stop Governments, resulting of course in endless litigation and confusion. Our intent herein is not to define AI per se, since we believe that at best it is a work in progress and at worst the wrong words to begin with, but to present some paradigms and elements which may prove useful.

1.1 THE MODEL

In a simplistic sense, AI takes some input that is to be examined and provides an output to the putative question provided in the input. It does so by relying on a massive amount of exogeneous information that has been processed by an element called a neural network (NN) for example. The NN has been designed and trained so that any input aligned with the class of trained data can or should produce an answer. Some answers can be presented simply as yes or no, and others more complex and in a text form using a natural language processing system as an adjunct.



Another simple example is shown below. Here we take a pathology slide, not even identifying it by organ, and we seek to identify by organ and malignant status. The input is an image and the output is a classification of N possible organs and M possible states. The system has been "trained" with potentially millions of identified images.



However what AI has in common is a form of "learning" from prior data sets and then developing algorithms on handling new data demands to provide answers or actions. What we see is that AI is a concatenation of inputs, data sets, learning algorithms and output mechanisms. In the simplest sense, on can ask a question and receive an answer, if the data set contains the data adequate for learning.

We examine here the potential extensions of this set of constructs. AI can go from the simplest input/output paradigm to a fully autonomous entity that initiates interactions, gathers information, constructs mechanisms, and provides actions while continuously monitoring its own performance, seeking increased optimization.

The putative "danger" of an AI system lies in the realm of the autonomous AI entity (AAIE) embodiments. Namely, an AI entity totally independent of any human interaction. Namely, it begs the question; can an AI system become totally independent of any human agency? If so, then what limits can be placed upon its actions? What can be done to enforce such limits?

We have a clear example of unenforced limits in a small sense with COVID-19. A virus released into the society and its propagation facilitated by an unprepared set of Governments resulting in the death of millions and a near collapse of economies. Autonomous AI systems are many orders of magnitude more deadly to humanity as a whole.

1.2 Objective

Our objective herein is to examine AI systems and specifically to consider canonical models demonstrating the putative progression to a fully autonomous AI entity, one capable of independent actions both computationally and physically. The latter model we call the Autonomous AI Entity, AAIE. This is an entity that operates independent of human interaction and makes judgements on its own. Further it has the capability of using and assembling instruments as externalities to effect its intentions.

1.3 Outline

This is a preliminary attempt to establish a set of reasonable paradigms of AI systems and their evolutionary possibilities.

1. We commence with a discussion of definitions. What is immediately seen is that the definitions are often limited and circular. AI, artificial intelligence is machine intelligence. But what we ask is the meaning of intelligence. We refer the reader to Minsky and his discussions.

2. We then use of common examples of systems that take inputs, use data, a processor, and provide outputs. The classic set are those of statistical modelling. We provide a high level of neural nets to establish a base.

3. We then present our canonical models, going from current views through autonomous systems. As we note this is a personal set of structures and may be open to a great deal of debate and re-examination.

4. We then comment on the applications in multiple areas.

5. Finally we discuss a multiple set of observations including the issue of regulating AI.

2 DEFINITIONS

Now, is there a well-accepted definition of AI? Hardly, we present a few contorted ones.

Winston (in 1984) defines AI as:

There are many ways to define the field of Artificial Intelligence. Here is one: Artificial Intelligence is the study of ideas that enable computers to be intelligent. But what is intelligence? Is it the ability to reason? Is it the ability to acquire and apply knowledge? ***Is it the ability to perceive and manipulate things in the physical world?*** *Surely all of these abilities are part of what intelligence is, but they are in the usual sense seems impossible because intelligence appears to be an amalgam of so many information-representation and information-processing Nevertheless, the goals of the field of Artificial Intelligence can be not the whole of what can be said. A definition talents. One central goal of Artificial Intelligence is to make computers more useful. Another central goal is to understand the principles that make intelligence possible.*

Note the circularity. AI is the embodiment of intelligent machines. Yet Winston recognizes that this is intelligence defined as intelligence, it begs the question of what is intelligence. He limits intelligence somewhat but not that much. The manipulation of the physical world is a key factor. The current view seems to be the question-and-answer paradigm, yet it is the autonomous manipulation of the entities environment which is most world changing.

IBM defines it as[1]:

Artificial intelligence leverages computers and machines to mimic the problem-solving and decision-making capabilities of the human mind

Oracle defines it as[2]:

AI has become a catchall term for applications that perform complex tasks that once required human input, such as communicating with customers online or playing chess. The term is often used interchangeably with its subfields, which include machine learning (ML) and deep learning. There are differences, however. For example, machine learning is focused on building systems that learn or improve their performance based on the data they consume. It's important to note that although all machine learning is AI, not all AI is machine learning.

HHS defines it as[3]:

---

[1] https://www.ibm.com/topics/artificial-intelligence

[2] https://www.oracle.com/artificial-intelligence/what-is-ai/

[3] https://www.nibib.nih.gov/science-education/science-topics/artificial-intelligence-ai

Artificial Intelligence: A feature where machines learn to perform tasks, rather than simply carrying out computations that are input by human users. Early applications of AI included machines that could play games such as checkers and chess, and programs that could reproduce language.

Machine Learning: An approach to AI in which a computer algorithm (a set of rules and procedures) is developed to analyze and make predictions from data that is fed into the system. Image of AI concept in multiple industries AI is integrated into numerous technologies that people use every day. Machine learning-based technologies are routinely used every day, such as personalized news feeds and traffic prediction maps.

Neural Networks: A machine learning approach modeled after the brain in which algorithms process signals via interconnected nodes called artificial neurons. Mimicking biological nervous systems, artificial neural networks have been used successfully to recognize and predict patterns of neural signals involved in brain function.

Deep Learning: A form of machine learning that uses many layers of computation to form what is described as a deep neural network, capable of learning from large amounts of complex, unstructured data. Deep neural networks are responsible for voice-controlled virtual assistants as well as self-driving vehicles, which learn to recognize traffic signs.

McKinsey defines it as[4]:

Artificial intelligence is a machine's ability to perform the cognitive functions we usually associate with human minds.

HP defines it as[5]:

Artificial intelligence (AI) broadly refers to any human-like behavior displayed by a machine or system. In AI's most basic form, computers are programmed to "mimic" human behavior using extensive data from past examples of similar behavior. This can range from recognizing differences between a cat and a bird to performing complex activities in a manufacturing facility.

The DOE defines it as[6]:

Artificial Intelligence (AI) simply means intelligence in machines, in contrast to natural intelligence found in humans and other natural organisms. Artificial intelligence gained its name and became a formal field of research in 1956, and initial work led to new tools for solving mathematical problems. However, researchers discovered that creating an AI is incredibly

---

[4] https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-ai

[5] https://www.hpe.com/us/en/what-is/artificial-intelligence.html

[6] https://www.energy.gov/science/doe-explainsartificial-intelligence

difficult, and progress slowed in the 1970s. More recently, increases in computing power and availability of massive data sets have set the groundwork for advances in AI.

MIT defines it as[7]:

Artificial Intelligence and Decision-making combines intellectual traditions from across computer science and electrical engineering to develop techniques for the analysis and synthesis of systems that interact with an external world via perception, communication, and action; while also learning, making decisions and adapting to a changing environment.

Harvard defines a specific type[8]:

Generative AI is a type of artificial intelligence that can learn from and mimic large amounts of data to create content such as text, images, music, videos, code, and more, based on inputs or prompts. The University supports responsible experimentation with Generative AI tools, but there are important considerations to keep in mind when using these tools, including information security and data privacy, compliance, copyright, and academic integrity.

It should be noted that there is a great deal of circular definitions and no reasonable consensus. How does one define AI, and worse, as Legislatures and Governments now try to control AI do we have any chance without a clear definition.

A somewhat pervasive term we see is intelligence. AI is "machine intelligence". The ultimate in circular statements. We will not pretend to provide a definition; we believe it is still a bit early to do so. What we now have is a collection of tools described as AI that perform certain functions taking Inputs and creating outputs using highly processed collections of Data that allow the Inputs to yield Outputs. Moreover, the logic for establishing the connections is accrued from a massive amount of Data which may relate to the Input being asked to provide an Output upon. This is a rather long-winded way to say that the relationship between Input and Output is determined by a complex machine dependent processing of a massive amount of Data, Data selected initially by humans, and the specifics are found only is a high-level description of the processing algorithm. A now somewhat standard algorithm is the neural network approach, NN, and related approaches.

---

[7] https://www.eecs.mit.edu/research/artificial-intelligence-decision-making/

[8] https://huit.harvard.edu/ai/guidelines

3 EVOLVING CONSTRUCTS

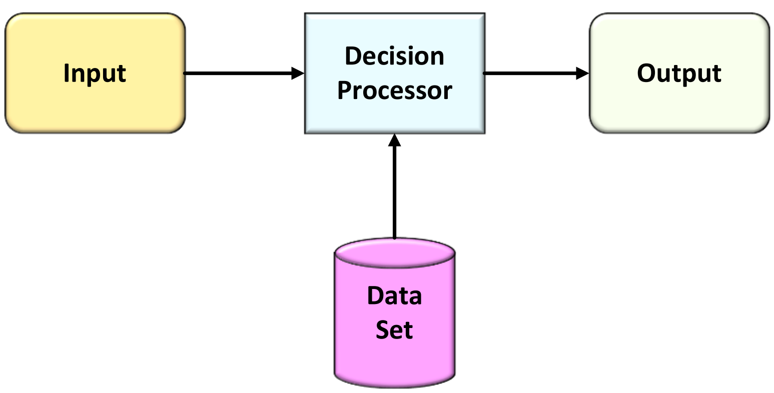
We start with a review of some early constructs and then provide a simple view of the neural network, NN, approach. We first examine some early constructs of computer assisted decision making.

There are two elements in many of the decision methods. Namely there is a set of inputs, a question if you will, and a set of outputs, answers if you will. In a medical context, for example, one may have a patients with a history of PSA measurements and related measurements. One can then ask what is the probability of prostate cancer? The approaches use massive amounts of a priori data from patients where inputs (PSA data) and outputs (diagnoses) are known. Then using this collection of past knowable data, the new patient data is processed to ascertain the diagnosis of this specific patient.

In essence, many AI type systems using NN for example have a similar paradigm. There is the input, the output, and processed data. The classic approach uses classic statistics, since we usually have a simple input and simple output, and the data base is isomorphic to the input/output paradigm. More advanced AI goes beyond the simple isomorphisms and has syntactical inputs and outputs and data bases that themselves lack isomorphic characters.

In the PSA case we just alluded to, in the AI world with NN, one may present the NN with millions of papers or studies where PCa and PSA are discussed. The Data Set is the collection of these prior results and the NN can take all of these results and create a network decision tree that gives an answer as to whether patient has or does not have PCa. Humans will never know what logic or reasoning the NN used to get this answer.

The canonical model can be depicted as below:



The four elements are typical. There is an input of some form, an output of some form, a processor which uses the data set to assist in connecting an input to an output. The input may be a question, an image, or any possible multimedia construct. Likewise, the output can be of a similar form. The Data Set is a collection of information that provides some connection between inputs and outputs in the class of issues to be determined. The Decision Processor may be any complex set or processes, algorithms or otherwise that take Input, Output, and Data and optimizes the desired nexus between all three.

What is important however is the recognition that the data, in a general sense, is often a multimedia set. As such, items such as images, video, sound, and even touch and smell, must be converted to a computer processable set of inputs. This means that the sensory conversions themselves add to the complexity, and potential lack of consistency, amongst AI systems. Unlike humans, whose sensory systems are integrated into the overall neural networks of the species, with sensors that convert a sensory stimuli to a nerve reaction, the AI systems are currently often disparate one from the other[9].

We shall show later how this basic formalism can be expanded in what we now assert to be AI systems.

We now proceed through this maze starting with the classic statistical isomorphisms and then leading to a neural network paradigm

3.1 General Assumptions

We will start with an understanding of a simple statistical construct. Here the problem is that we have a great deal of data regarding some physical measurement and some disease state. We have an input, namely the measurement of a specific patient and we seek an output, a disease state. This contains the three basic elements we have discussed above.

We could extend this to reading an Xray, MRI, CAT scan, a path slide or the like. In those cases, we would need a front end to process an image into a digital array which we can then compare to a data base of comparable arrays. We shall discuss this later. Yet the fundamental principle is the same. The key is that we make a decision based upon a massive amount of data and a decision metric.

Let us assume there are N variables which can be measured to determine if a person has a certain disease state. We then can determine:

$$P\left[Disease_k \mid x_n\right] = p_{k,n}$$

Namely, we know by data the fact that given the n state we can determine a specific disease state k.

---

[9] Note that the EU has passed AI regulations which frankly appear as a hodgepodge collection of regulations. See https://www.nytimes.com/2023/12/08/technology/eu-ai-act-regulation.html The Times notes: *European policymakers focused on A.I.'s riskiest uses by companies and governments, including those for law enforcement and the operation of crucial services like water and energy. Makers of the largest general-purpose A.I. systems, like those powering the ChatGPT chatbot, would face new transparency requirements. Chatbots and software that creates manipulated images such as "deepfakes" would have to make clear that what people were seeing was generated by A.I., according to E.U. officials and earlier drafts of the law. Use of facial recognition software by police and governments would be restricted outside of certain safety and national security exemptions. Companies that violated the regulations could face fines of up to 7 percent of global sales.*

We also have:

$$q_{k,n} = 1 - p_{k,n}$$

Now lest us assume that we have N of the measurements. Let us examine a simple 2 measurement example with two disease states.

$$P\left[D_1 \mid x_1, x_2\right] = P\left[D_1 \mid x_1, x_2\right]$$

Let us focus on mpMRI, namely multiparameter MRI, as a start. The question we pose is; what do we mean by radiomics using mpMRI? For example, is it totally computerized or totally reliant upon the radiologist? Do we send the patient into the imaging system and out comes a diagnosis? Or do we image a patient and have the radiologist make all the judgements? How much and how far do we allow the system to function?

For example, in breast mammography we generally perform an x-ray study and then the radiologist examines the image looking for signs of a malignancy. They are generally two-dimensional films, now electronic,

3.2 Generalized Metrics

The classic approach is in line with RECIST protocols used in clinical trials. As Eisenhauer et al have noted describing RECIST:

Assessment of the change in tumour burden is an important feature of the clinical evaluation of cancer therapeutics. Both tumour shrinkage (objective response) and time to the development of disease progression are important endpoints in cancer clinical trials. The use of tumour regression as the endpoint for phase II trials screening new agents for evidence of anti-tumour effect is supported by years of evidence suggesting that, for many solid tumours, agents which produce tumour shrinkage in a proportion of patients have a reasonable (albeit imperfect) chance of subsequently demonstrating an improvement in overall survival or other time to event measures in randomized phase III studies.

At the current time objective response carries with it a body of evidence greater than for any other biomarker supporting its utility as a measure of promising treatment effect in phase II screening trials. Furthermore, at both the phase II and phase III stage of drug development, clinical trials in advanced disease settings are increasingly utilizing time to progression (or progression-free survival) as an endpoint upon which efficacy conclusions are drawn, which is also based on anatomical measurement of tumour size.

3.3 Cluster Analyses

Cluster analysis is a now classic means to group data into classes. One may have supervised or unsupervised and there are a variety of clustering algorithms such as nearest neighbor clustering. Generally, we will not use clustering here.

Clustering is simply the throwing of data into a mass of other data and then using a separation metric selecting amongst the N dimensional data set a separation surface. The separation surface take the data and creates classes of outputs. The algorithms for creating these separation surfaces may vary.

3.4 Pattern Recognition

Pattern recognition has been available in one form or another for decades. It has been used in a wide set of areas. Pattern recognition techniques allow for the "extraction" of key patterns to be used in discrimination. Thus, using edge extraction in modifying say a US image we can then look for a pattern representative of a papilla like growth. We can look for pedunculated lesions, look for cyst like surfaces or look for clear nucleus or Orphan Annie eyes. Pattern recognition tools allow for the extraction of know or unknown patterns. If for example we know that certain patterns are pathognomonic then we can design a pattern recognition system to look for those specific patterns.

We believe that the development of pattern recognition and extraction methods will be at the heart of any successful classification scheme. The patterns will produce metrics which we can then use in the classification.

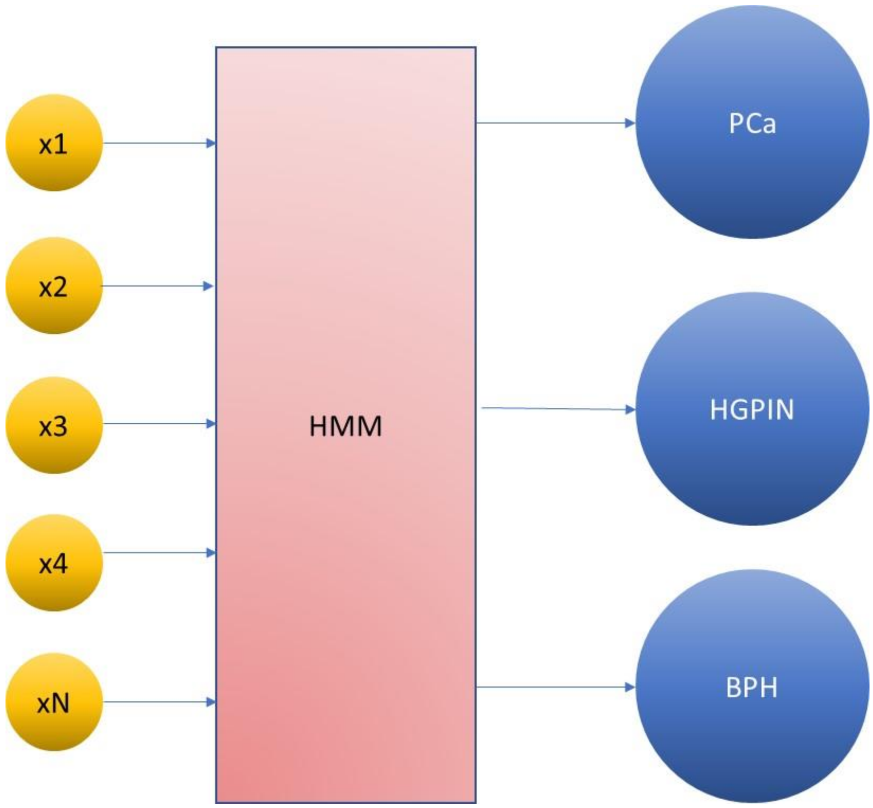
3.5 Classifications; Supervised and Unsupervised

Classifiers take multidimensional data sets and establish lines of demarcation separating one class from another. The example of using PSA and %Free and seeking the dividing line between benign and malignant allows for a reasonable test. Multidimensional classifiers are much more highly structured.

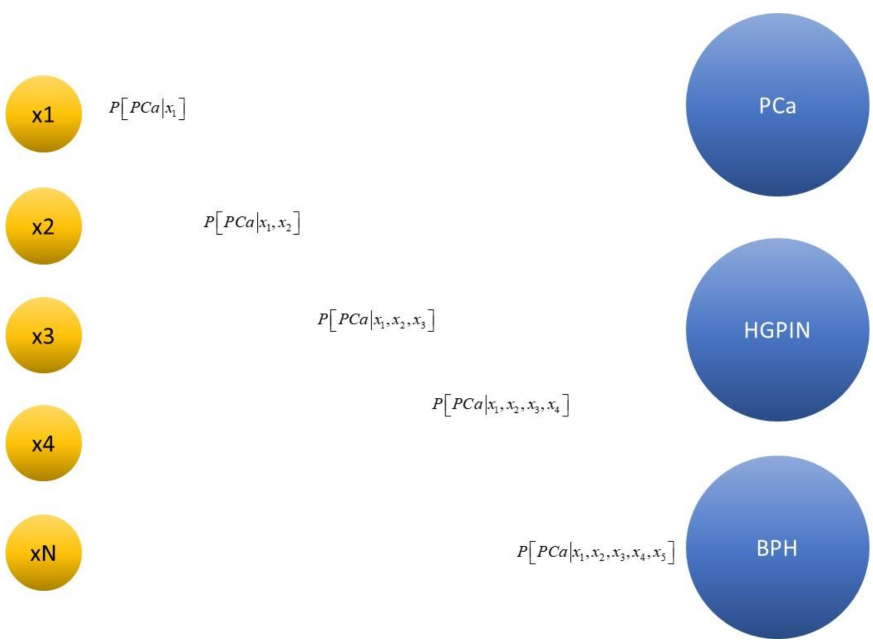
We can now measure various miRNAs in body fluids and this gives rise to the liquid biopsy concept. However, the key question is how does one take a collection of miRNA measurements and ascertain, for example, that there is a prostate malignancy. For example, we may from the previous presentation generate a vector of measurement of miRNA densities given by:

$$m_k = \begin{bmatrix} x_1 \\ \ldots \\ x_n \end{bmatrix}$$

where this is for patient k and measures n miRNA densities. We want a discriminant function which takes these values and determines whether the patient has cancer of not. We could have a linear weighted discriminant or a more complex non-linear version.



We can look at a Markov model as below. However, these transition probabilities are often difficult to determine.



$$P\left[x_1 \middle| PCa\right]$$

....

$$P\left[x_1,...,x_N \middle| PCa\right]$$

or

$$P\left[PCa \middle| x_1\right]$$

....

$$P\left[PCa \middle| x_1,...,x_N\right]$$

where we have the two probabilistic ways to ascertain a condition based upon a data set.

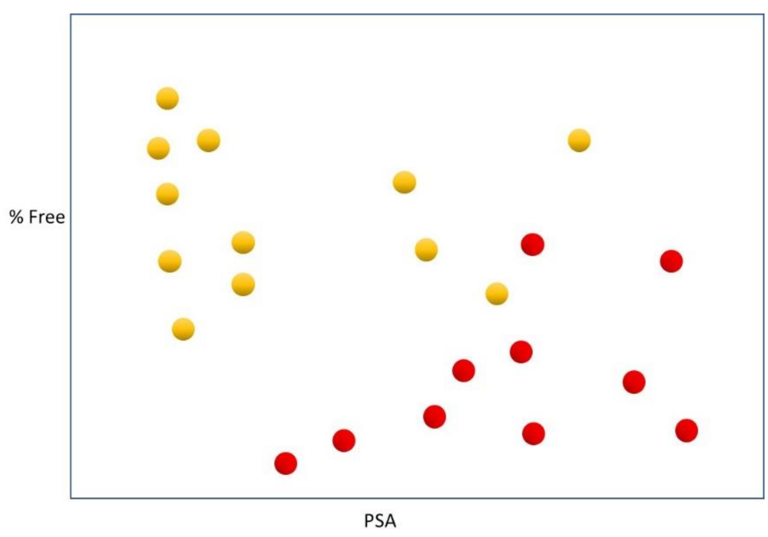
Let us consider a simple example. Assume we have to determine if a patient has prostate cancer or not. We are given three variables; PSA, % Free PSA, and PSA velocity[10]. Namely:

PSA=PSA

PF=% Free PSA

V=PSA Velocity

Thus, we have three measurements and they are somewhat related. Let us start with two of them; PSA and PF. The data may appear as shown below:



[10] See: Carter et al, Detection of Life-Threatening Prostate Cancer With Prostate-Specific Antigen Velocity During a Window of Curability, Journal NCI Vol 98 Nov 2006 pp 1521-1527, https://academic.oup.com/jnci/article/98/21/1521/2521858

The red are PCa cells and the orange are benign. The higher the PSA the greater the chance for PCa. However, the higher the PF the greater the chance for benign, namely BPH. This is a simple case where we would have some discriminant where both variables count.

Now consider all three variables. We have PSA, PF and V. We need a discriminant so as to separate malignant from benign. We have data ex post facto so this is a supervised learning algorithm. We need to obtain some covering surface that maximizes the sensitivity and specificity. The algorithm must maximize the AUC. The more data the better the algorithm, yet we will always have aberrant cases.

The challenge in this case is that the discriminant is not a simple plane of some sort. It can be a complex surface winding its way around the 3-space. Namely the 2-space example shown in the above diagram may change for every V measure. For any V value we can obtain a 2-space profile. But that profile is different for every V and each has a different AUC, area under the curve[11]. We can design a simple process where we enter all the data and calculate that surface on a cut-by-cut basis. Then any user can enter the three variable and get a result; benign/malignant, specificity, sensitivity.

Now let us consider a simple linear discriminant for PSA/PF and for a fixed V. Our goal is to select a curve:

$$PF = aPSA + PF_0$$

The goal is to obtain "a" and PF_0 so that we maximize both sensitivity and specificity. This can be readily accomplished by a variety of simple algorithms.

The next question would be; how many data points do we need and how frequently must they be updated? The answer can really only be obtained in an iterative manner with real data. We know that PSA alone has at best an AUC of 70%. Obtaining the AUC in this three-element case is more complex. We may also want to add such elements as age, family history, prior biopsy results and the like. Each element adds another layer of complexity.

A simple and direct approach would be a linear classifier. Our metric is sensitivity and specificity. Namely:

$$\begin{aligned} &Sensitivity = P\left[H_1 \middle| H_1 \right] \\ &and \\ &Specficity = P\left[H_0 \middle| H_0 \right] \end{aligned}$$

If the discriminant plane is:

[11] The AUC is a measure of how well the discriminator functions. An AUC of 0.5 means that the test is as goo as a coin flip. An AUC is a perfect test.

$$g(x) = ax + b$$
$$where$$
$$x = \begin{bmatrix} x_1 \\ ... \\ x_N \end{bmatrix}$$
$$a = \left[a_1........a_N\right]$$

The goal is given the data set, find the **a** vector and b to separate the data so as to maximize sensitivity and specificity[12].

There are a multiple set of classifiers and our selection of a linear classifier in a supervised environment is just one of many. We do not know the underlying statistics of the miRNA and also each miRNA itself may or may not be as strong an element in classification. Some miRNA that we choose may be a weak element and should be eliminated. That can only be ascertained after extensive data analysis.

Another way one could examine this partition problem is to assume that the two variables we discussed earlier, say PSA and PF, are independent Gaussian variable with mean and standard deviations:

$$H_0 :$$
$$m = m_0, \sigma = \sigma_0$$
$$H_1 :$$
$$m = m_1, \sigma = \sigma_1$$

Then we could use classic decision analytical methods to determine optimal selection criteria. We could estimate the mean and variance from the given data and even ascertain a probability density function to see if it varies from Gaussian. It is not clear that such an approach yields better discrimination.

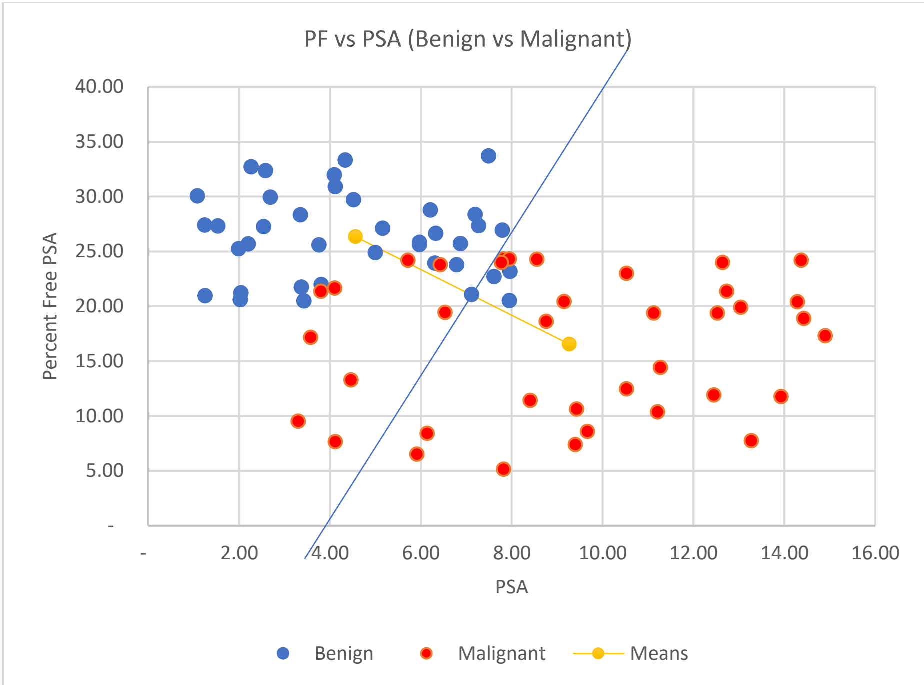
Finally, one could seek to use a Principal Component Analysis to determine optimal orthogonal axes[13]. However, again in my experience, this would not gain a great deal.

A linear classifier using the large data set may be more than adequate. We show below several examples of a linear classifier for PSA vs FP[14].

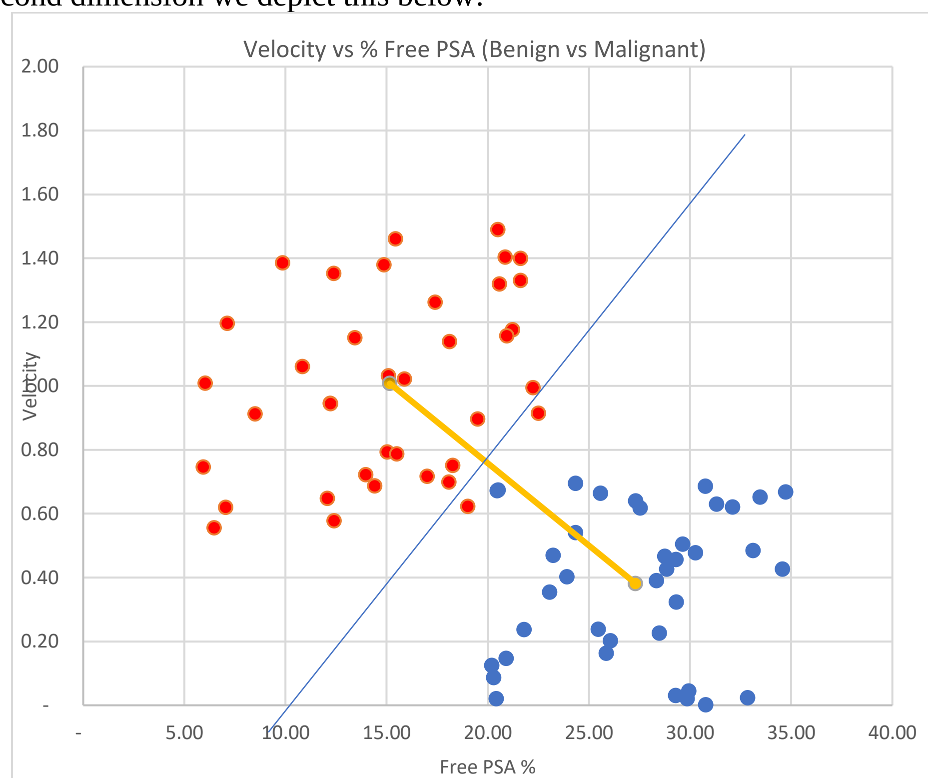
[12] We refer to Theodoridis and Koutroumbas and their work on classification. We note that there are a multiplicity of algorithms to define this linear classifier. Also, there is a great deal on PCa learning algorithms in Hastie et al.

[13] See Dunteman, Principal Component Analysis, Sage University Paper, 1989.

[14] We use the reference of Duda and Hart, Pattern Classification, 1st Ed, Wiley, 1973.



and for a second dimension we depict this below:



Note all have different data yet all have same means on the two data sets. Thus, the slope of the classifier is the same and intercept changes a bit. This same approach carries over to the miRNA context for multiple dimensions.

Now classifiers can often be nonlinear. The above simple example assumed a linear deign. However better performance may be obtained with a nonlinear classification scheme. The simplicity of the above classifier is based upon two facts. First, we know from the data who is benign and who is malignant. Second, we have selected elements, three in this case, upon which we can segment and classify.

In our study of images, we do not necessarily have elements to be used to establish a classification. We start with a set of digital images. Perhaps from the images we can obtain a finite set of metrics, perhaps not. The next discussion is on neural networks, a more sophisticated from of classifier.

3.6 Neural Nets

We now consider here neural nets in the context of images. What we present here may flow into many other constructs such as text, audio, video, and a plethora of other classes of input data and output results. An encyclopedic summary of neural nets and other related techniques is in Haykin[15]. Also, the book by Aggarwal is useful[16]. We consider a simple specific case herein related to images and the use of a neural net like structure. The work by Jurafsky and Martine provides an excellent approach using speech and language processing. The following analysis allows us to bring forth several of the key issues of neural nets at a simple level.

Neural nets can be applied to a wide variety of media, text, voice, image, video, and the like. It is effectively a multimedia mechanism. However, to do so modified front ends and storage mechanisms must be employed. Thus, the use of NN in this environment has a multilevel complexity.

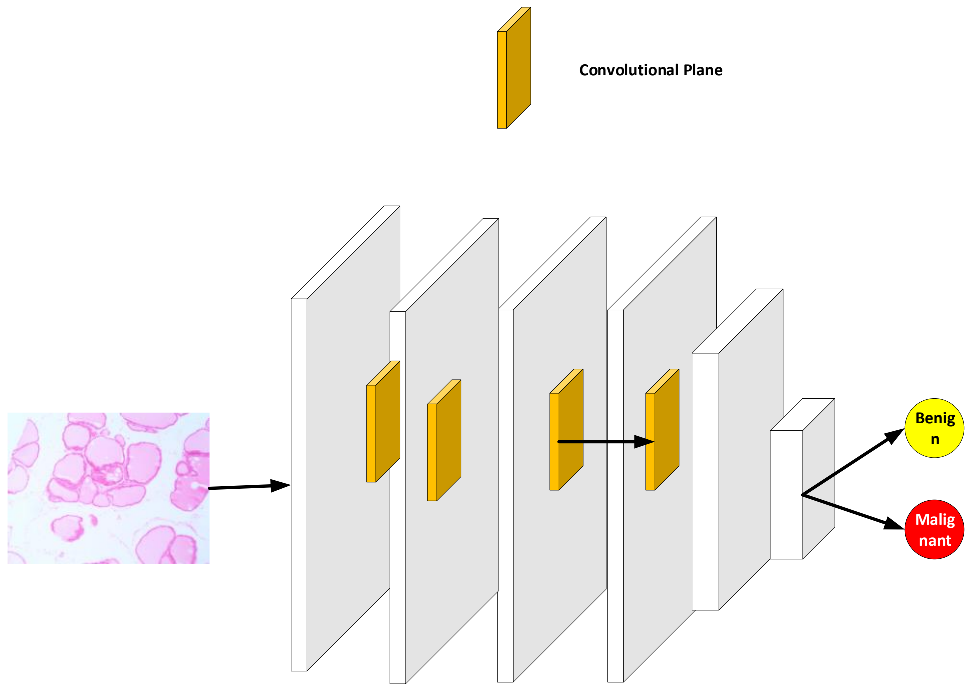
Imaging analysis using neural networks often uses a convolutional neural network. The input may be an N X M matrix of k bit pixels. Namely a digital picture. We then collect are large set of such pictures and using knowledge of the disease state a priori we pass them through a network whose weights of connections are adaptively changed so as to maximize the probability that if we put in an unknown we get a very good guess as to whether it is benign or malignant.

In such a design we often use a convolution processor which passes on to the next layer a new meta-pixel which can be some integration or enhancement of the basic input. We show this construct below.

---

[15] It is useful to examine the Minsky book (1986) for a variety of useful constructs and definitions. Almost all hold today.

[16] It is useful to review the work of Rumelhart et al on parallel distribute processing, the early name of neural nets in the 1986 edition. Much of what we see today is there but the ability of processors some 40 years ago was very limited. Companies like *The Thinking Machine* tried to establish some base in this space but were not successful.



It functions as follows:

1. A two-dimensional image, properly sized, is samples into a two-dimensional input plane in the network. It is identified as benign or malignant, and that is all.

2. The network has n planes and each plane has a convolutional filter as noted. Thus, the entry on the following plane is comprised of (k x k) convolved samples of a segment of the prior plane.

3. A backward propagation algorithm is used as the network is trained. Namely M samples, each identified as malignant of benign, is passed through the network and the weights between layers are modified to optimize the output based on the known sample. Classically this may have been a least squares algorithm as was done in the classic Widrow Hoff optimizer for phased array antenna beam forming[17].

4. After the M samples are used to set the weights, the unknown is entered and the sample unknown identified.

Needless to say, there are several key differences:

1. No a priori patterns are selected. In fact, the network does not even assume that there are cells there. It has been trained on a large number of patterns.

[17] See Monzingo and Miller, Introduction to Adaptive Arrays, Wiley, 1980, Chpts 3-4, VanTrees, Optimum Array Processing, Wiley, 2000.

2. The user may have no idea what the patterns emphasis may have been. Thus, after some training period, the user is relying on the network to select what is a discriminant and what weight it may have.

The algorithm for neural networks with training is generally simple to grasp but it has many variations. Namely, data sets x consisting of say n x n arrays of 8 bit grey scale samples are used to put into a single hidden neural net, where we have say an m x m array of weights a which we want to train to discriminate between a disease state. The algorithms generally look like:

$$a = \begin{bmatrix} a_1 \\ ... \\ a_m \end{bmatrix}$$

and

$$x = \begin{bmatrix} x_1 \\ ... \\ x_N \end{bmatrix}$$

where

$$\hat{a}(k+1) = \hat{a}(k) - K(\hat{a}^T(k)x(k) - s)$$

where K is some convergence matrix. Namely we train the neural net with a massive number of samples whose sate "s", the disease, we know, and generate a(k), to reach some stable state, hopefully. Then with an unknown we send it into the net and hopefully get the correct disease state.

Now a brief overview of the Least Square Estimate procedures. Let us assume we are trying to estimate the slop and intercept of a straight line:

$$y = ax + b$$

we have N sets of x and y values, all somewhat noisy. That is:

$$x = \begin{bmatrix} x_1 \\ ... \\ x_N \end{bmatrix}$$

and

$$y = \begin{bmatrix} y_1 \\ ... \\ y_N \end{bmatrix}$$

Now we want a recursive estimator of the form (as we note to be a least square steepest descent model):

$$\hat{a}(k+1)=\hat{a}(k)+\Delta_a(y(k)-\hat{a}(k)x(k)-\hat{b}(k))$$
$$and$$
$$\hat{b}(k+1)=\hat{b}(k)+\Delta_b(y(k)-\hat{a}(k)x(k)-\hat{b}(k))$$

This is based on steepest descent algorithms and the choice of the function for the descent is based upon a least square performance function[18]. Namely we want to minimize:

$$\varepsilon^2=\left[y(k)-\hat{a}(k)x(k)-\hat{b}(k)\right]^2$$

Classic least square has the descent function be that which minimizes the error for each element being minimized. Namely:

$$\frac{\partial\varepsilon^2}{\partial\hat{a}(k)}=\frac{\partial}{\partial\hat{a}(k)}\left[\left[y(k)-\hat{a}(k)x(k)-\hat{b}(k)\right]^2\right]$$
$$=2\left[y(k)-\hat{a}(k)x(k)-\hat{b}(k)\right]$$

and the same for b. Thus, the steepest descent for a least square estimator is as we have shown above. The constants are chosen for convergence purposes and they are negative.

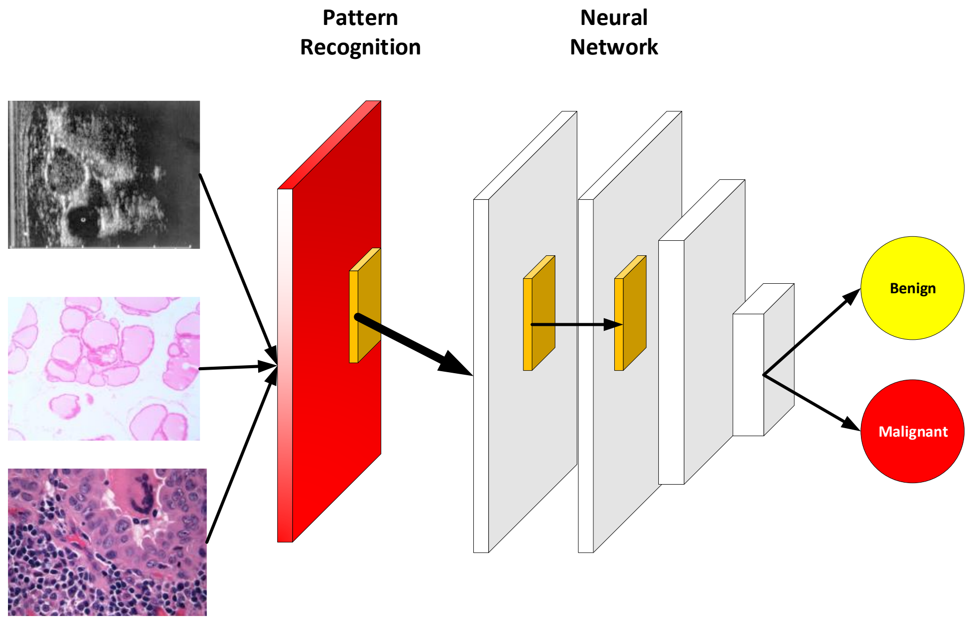
Now in a complex neural network we take the image which may be two dimensional and use each pixel as an input. Then we may convolve the image in some manner with a small m x m filter and pass it along. The weights at each step are adaptively changed if say a supervised test is performed. The neural net weights change in a manner similar to the linear estimator discussed above. We incrementally change them as we send identified image after image into the system[19].

There are now a massive number of algorithms to be used and with multiple layers as shown below we have deeper and deeper nets. Again, the issue is that we are relying on the net to identify the diagnostic issues and we may never know what the net sees as important.

Now an added approach is to establish a pattern recognition front end where we can identify such things as edges, MDI artifacts, cell size, cell counts and the like. Then we feed those parameters to the Neural Net. We show this below.

[18] See Athans et al, Systems, Networks, and Computation, McGraw Hill, 1974

[19] See Hakim



The above is a clear example of a pattern recognition system followed by a classifier. The classifier we have here is a neural network one but frankly we can use a variety of other classifier algorithms. It is critical to note that pure NN AI systems would just admit the images qua pixels. Here we add a priori knowledge of structures.

4 CANONICAL MODELS?

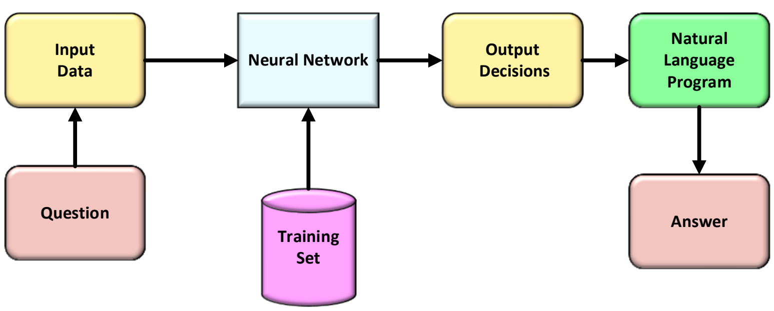
Now we consider the question of canonical models. As with many systems one may see simple constructs and then an ever more complex of such constructs. But across many implementations there are classes of constructs that pertain globally. These are the canonical models.

Now it should be noted that these canonical models are presented in a putatively evolutionary sequence. The current AI models can be related to Models 1 through 4. One suspects that Models 5 and 6 may very well exist and that 7 may also to some degree. The question is; how soon until we reach 8 and especially 9, the autonomous model? A second and more significant question is; does AI as a generalized construct become a threat to humanity as a result of its expanding autonomy? Finally, a third question is: what can be done with AI to prevent adverse consequences?

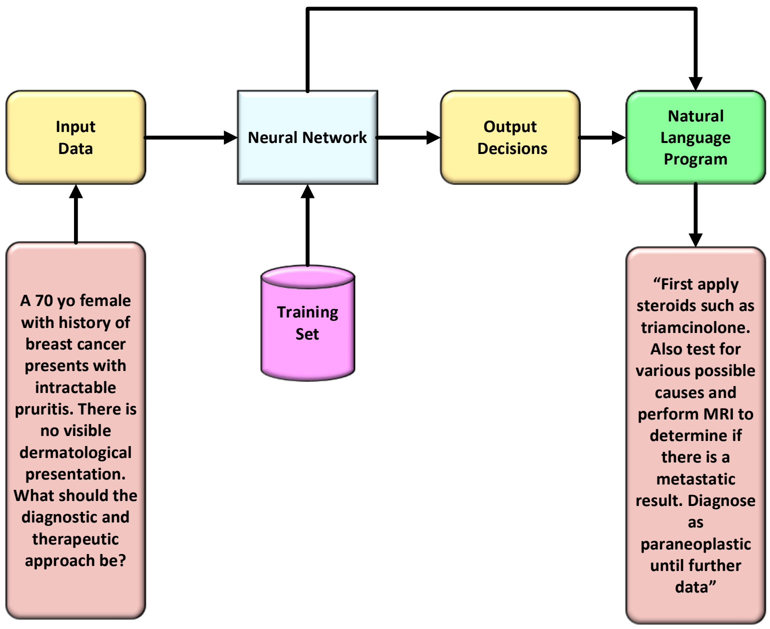
The proposed taxonomy as follows is the author's attempt to lay forth a possible evolutionary path. However, there may be many alternative paths and thus we use this merely as an example. Thus, we use these as examples subject to modifications and changes.

4.1 MODEL 1: "QUESTION IN-ANSWER OUT"

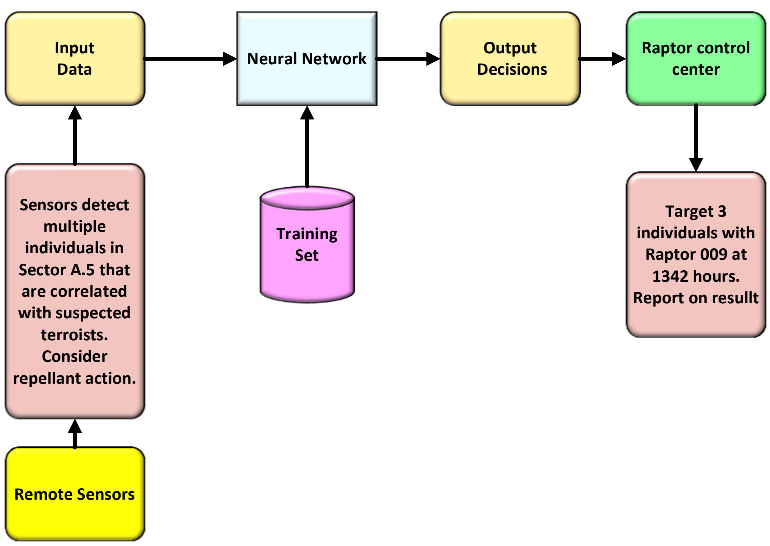
This is a text type system. We ask a question in text and get an answer in text. For example, we may ask: "What was the Thirty Years War?". We get a collection of output facts drawn from processing of the Training Set and then that is converted by a Natural Language Processor into a syntax correct answer.



We examine a medical example. Here the patient data is entered by some medical professional and the question is posed in the Input. Namely what is the disorder and how to treat it.



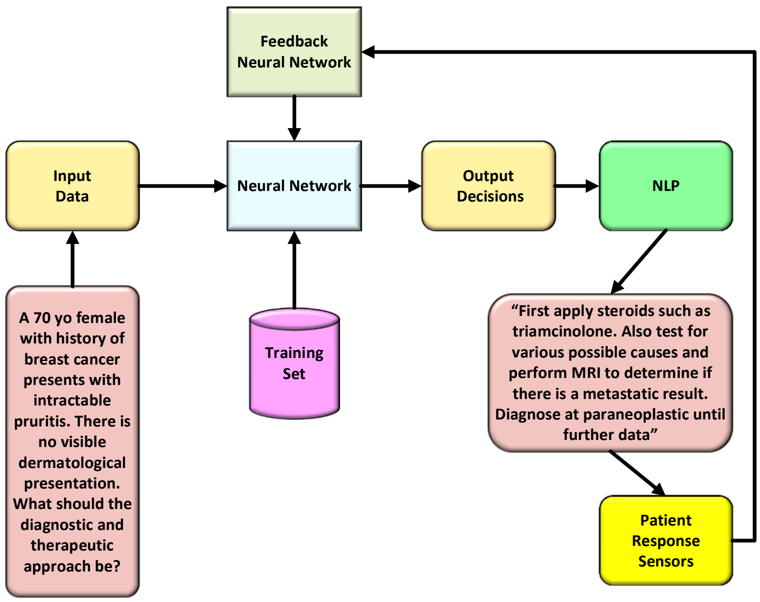
The following is a simple example in a military application. Here we assume we have some autonomous aircraft collecting imaging data real time. The output would be the attacking of some entity with the weapons available on the aircraft and done in a manner that is independent of any human intervention. If you will, it is "robot warfare". The system will do what it has been trained to do, hopefully no more and no less.



The above is a simple feed-forward system. Namely define the target and then allow the entity to take optimal measure to neutralize the target. The main problem however is if the target characteristics change in some manner. Adaptability in such an AI entity may be limited.

4.2 Model 2: Results Feedback to NN

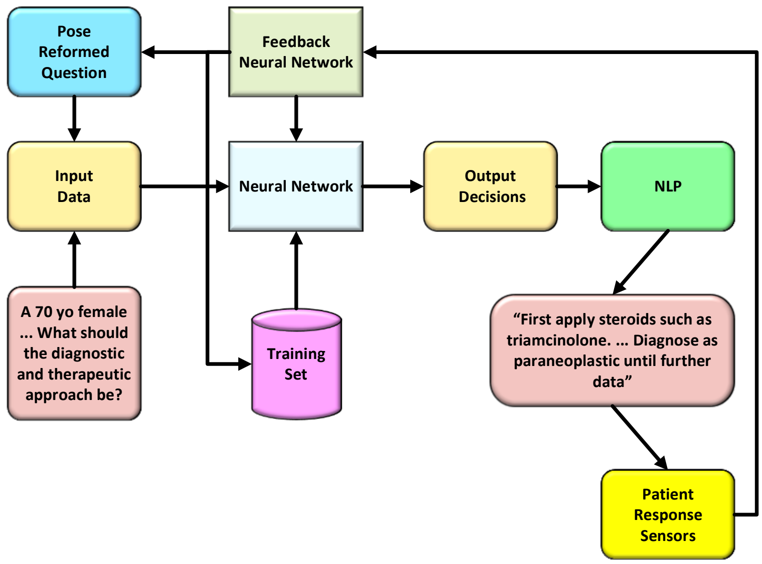
Model 2 takes a step towards a recurring set of feedback. Namely once a question is answered then this result is placed in the Training set and becomes part of the decision process. We show this below:



This model takes the result obtained and then adds it to the data set. Thus, the Training set can be modified and theoretically improved. There is a certain amount of complexity here however. The output may depend upon and additional amount of new information. Thus, the incremented data set may have additional modifications depending on the necessity for follow up steps. Finally, any ambiguity may be totally resolved is there is a final and dispositive measure of a disease state such as a pathology report. This is an example of the reducing levels of ambiguity as one drills down on diagnostic tools.

4.3 Model 3: Feedback with Question Reformation

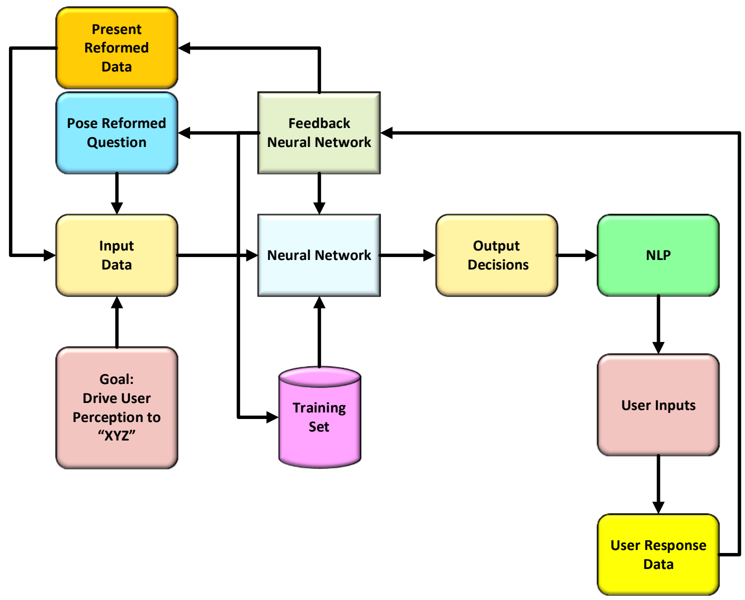
The third step not only feeds results back but poses a reformed set of questions so the system can iterate on itself. A reformed question is a modification of the original question, an incremental change, not a totally new question. One may ask if there is some stability point or process or if the models has inherent instabilities. The critical observation here is that the AI entity is now formulating a question. It has become slightly self-generating, asking something that extends somewhat beyond the original input.



Question generation is a complex but essential element of establishing an intelligent agent. Humans can distinguish themselves by what questions they ask. A key element in research is the posing of the question. Thus, an AI entity that is capable of posing a question is critical.

4.4 Model 4: End User Influencing

This model is also a feedback design but here the intent is to influence the end user who may assume they are in control but in reality, the driver is profiling the user and providing information that drives the users reality to some desired end state. In effect it psychologically profiles the end user and then adjusts the information presented to such a user so as to modify and control the users world view. It does so by understanding how this model functions on a massive set of such trials and it does so in an interactive feedback manner allowing the system to assess the users initial mindset and then providing feedback to move the mindset to the desired end point.



This approach is currently being adopted by the social media companies. Initially it is being done to promote products and services but it is slowly being done to effect political and social controls. All too often the users are unaware of the process. The new questions in this model are ones to influence the end user.

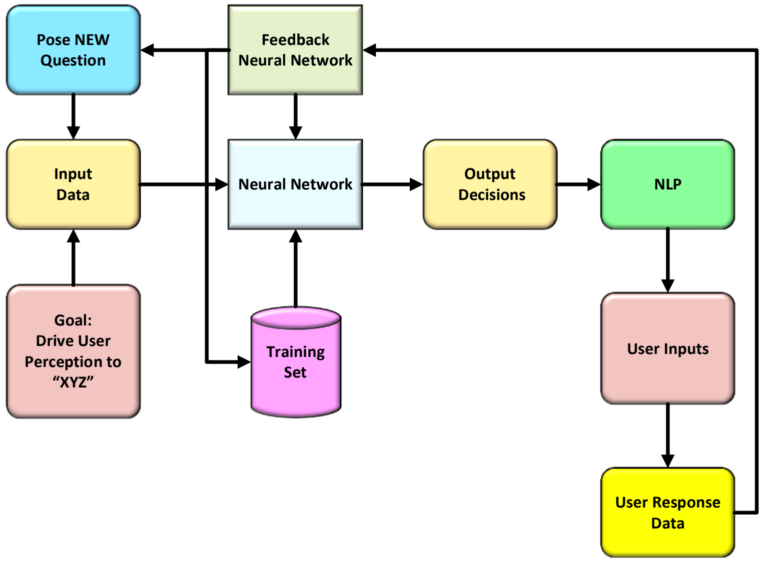
In fact, this model can be used when psychological profiling of the end user is performed against a Target Data Base.

4.5 Model 5: Question Posing

The essence of true knowledge and intellect is the ability to pose a question. Obtaining the answer may not be obvious but a well posed question allows for the initiation of the search. AI systems currently appear effective in presenting answers to questions. Often the answers are clearly limited to the input provided to the machine. But the machines have yet to demonstrate the ability to create a well posed question. Here we mean something not like Eliza, which was somewhat akin to what a psychiatrist would present.

A classic well posed question may be; "How does DNA replicate?" The question was timely after the classic Watson and Crick paper, the details of which were still lacking[20]. We could consider a well posed question as "What was the source of COVID 19?" Yet using the current systems we are dominated by propaganda based documents knowing well that "proof" of source may still elude us.

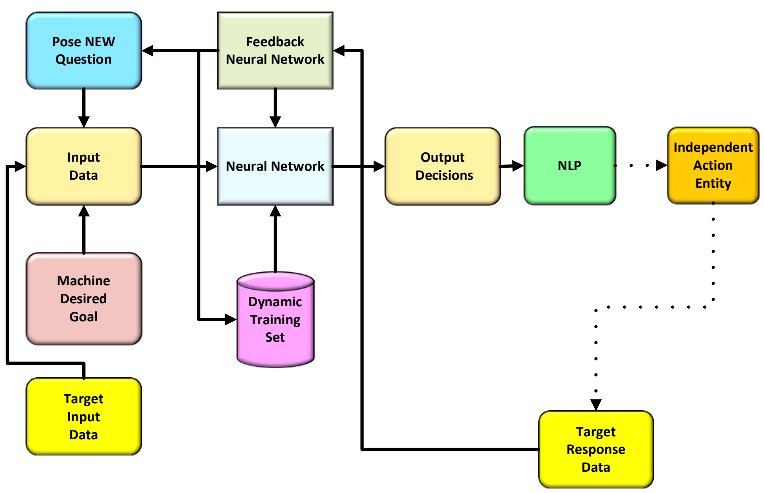
[20] Watson, J. D., & Crick, F. H. C. A structure for deoxyribose nucleic acid. *Nature* **171**, 737–738 (1953)



Note this is not just a reformed question, an incremental step, but a totally new independent question. Now many AI systems currently deployed to the masses are possibly becoming capable of presenting a well posed question as they are being bombarded with questions by humans. If true learning is possible for the machine than may suspect that soon the ability to form a well posed question will present itself.

4.6 Model 6: Action Taken

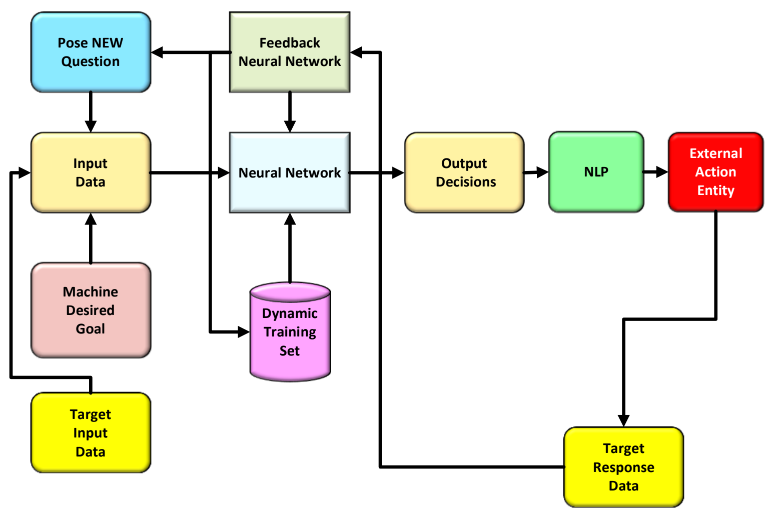
AI systems may evolve to be able to take actions independent of a specific human stimulus. These actions are limited to decision making and data collections and processing. In effect the AI entity is now acting upon an external entity in part. The interaction is at a distance and the entity being acted upon is still under the control of a third party such as a human.



The link to the action entity is contingent (note we have it dotted) and its selection is dependent on an external entity such as a human effecting its implementation. However, this give a first level physical extensibility to the AI entity. It may be nothing more than the connection of the AI entity to some physical device such as a DNA sequencer. It does however permit the interconnection of multiple external devices simultaneously.

4.7 Model 7: External Action

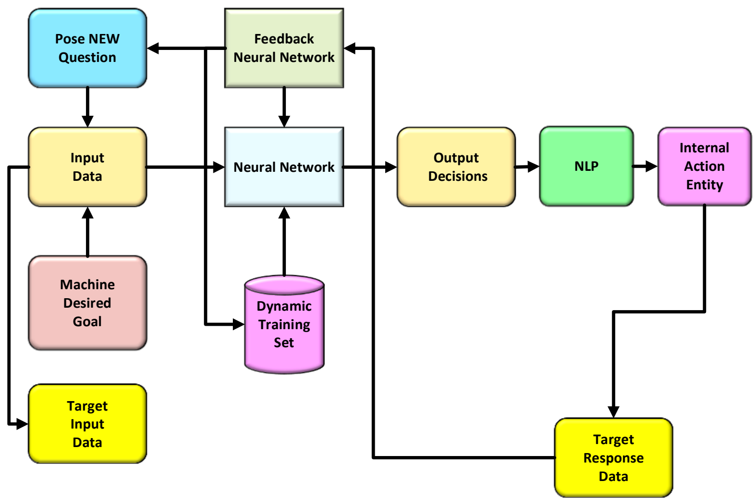
An AI system can eventually take action on its external environment, using existing external material tools. That is an AI system may decide to sequence the DNA from some sample. The AI system may then interface on its own with some existing and communicateable sequencing system. It can then control other external systems to collect DNA and then sequence the DNA. It may also be able to process the DNA and go as far as designing, for example, a CAR T cell set to treat a diagnosed pathology.



Note in the above the drive is still from an end user Input Data set. There may be autonomous elements but they still have some delimitation. Also, the Action Entity is **external**, it is a separate and distinct entity and not under the complete control of the AI entity. However, the AI entity has identified it and has access it and now controls it.

4.8 Model 8: Internal Action

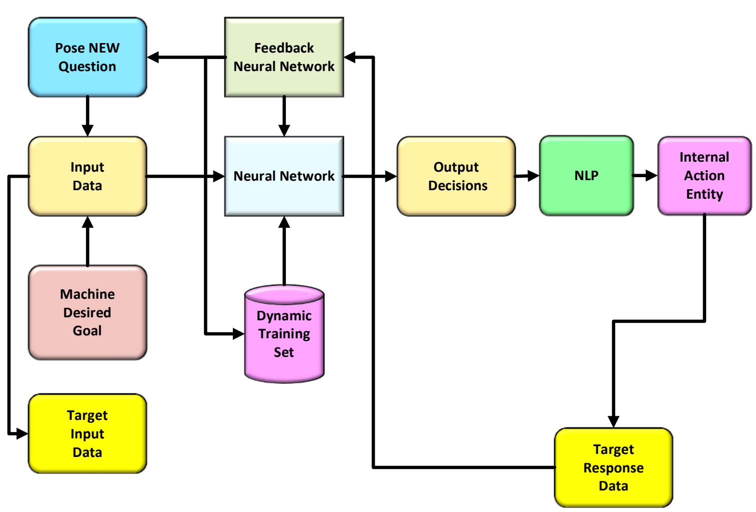
The next stage would be the AI system actually developing and implementing as well as operating its own action units. Namely as previously it may desire to diagnose, test, and treat a patient. In so doing it may design and develop its own systems to do so independent of existing human systems.



It is critical to note the flow of Target Input Data is still external and from an external user entity. This will contrast to the next model which is totally internal. However, in this case the entity assembles and controls its own Action Entity. For example, the AI entity may seek out a sequencing device, nucleic acids, and other elements to "make" its own viruses. It does not need a human to facilitate this. In fact, it may independently design its own physical elements, and create new improved designs, designs which become independent of human interfaces.

4.9 Model 9: Autonomy

The next step would be a totally autonomous system, able to effect material changes in its environment by self-generated tools and capable of communicating but independent of human interfaces. The Autonomous AI system is an entity unto itself, learning from and changing its environment according to its own directives.



Namely in the Autonomous mode the system would interrogate a patient, The patient would not tell the system the system would seek information from the patient. The system or entity would them perform internal actions and to the extent of creating and evaluating tests which it can perform based upon its own knowledge. Given the tests and responses it can then take remedial actions as it see fit. Note the sustained objective of a **machine desired goal**, in this case the good health of the patient. However, the risk is that the machine may try to redefine this goal, and in fact reinterpret it. Autonomy places all elements in the hands of the entity, from start to finish. The entity has a separate existence independent of humans.

4.10 SUMMARY

The Table below depicts the nine classes of AI development. It starts with the basic In and Out system and evolving to one of complete autonomy of action. We continue through the Autonomy system and then suggest added steps.

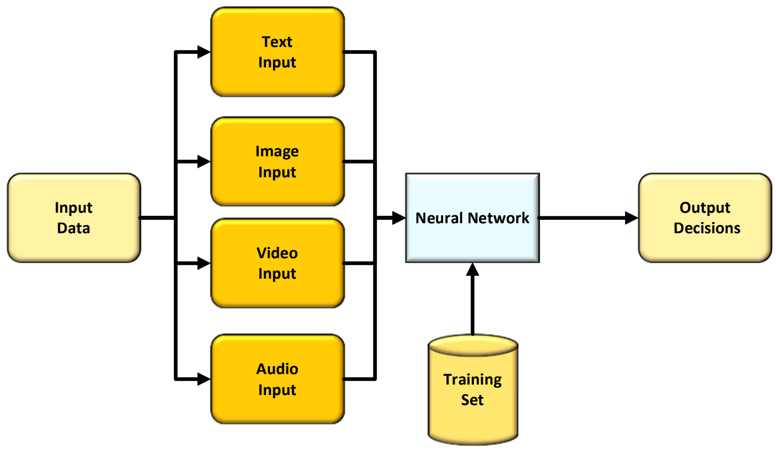
Class	In and Out	Response Feedback	Question Reformed	User Influencing	Question Posed	Action Taken	External Action	Internal Action	Autonomy
Answers Questions	X	X	X	X	X	X	X	X	X
Feedbacks responses to modify		X	X	X	X	X	X	X	X
Based on feedback it reforms the users question			X	X	X	X	X	X	X
Seeks to influence users to seek an external goal				X	X	X	X	X	X
Poses a new set of questions					X	X	X	X	X
System takes autonomous actions						X	X	X	X
System uses External activators							X	X	X
System uses internal activators								X	X
System is total autonomous no longer dependent on users									X
"X?"									

4.11 Implementation Options

However, there may be added options as multimedia inputs/outputs are sought. We present two options as follows. We have emphasized multimedia input/output options since as humans we communicate in a variety of sensory means. Thus, understanding some of these options is essential.

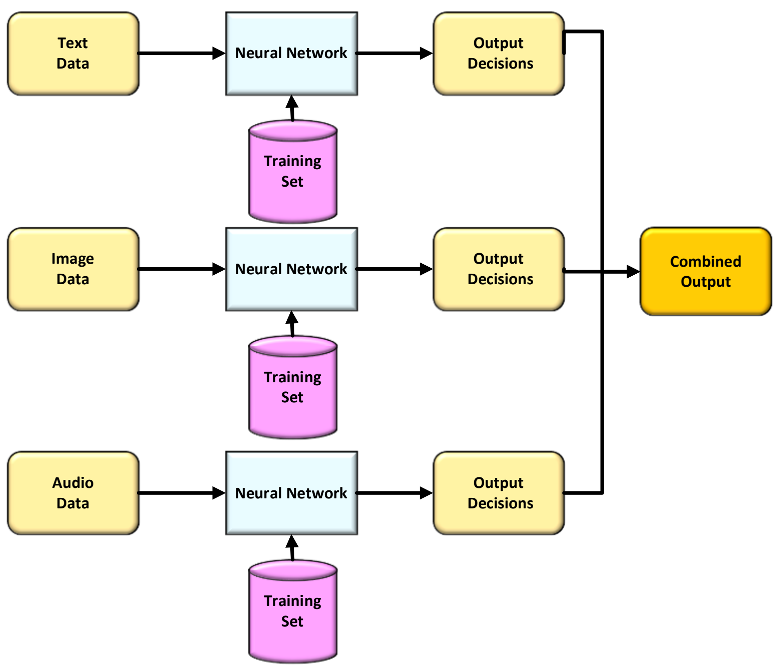
4.11.1 Series

The series model assume that all sensory data is entered as a complex set of inputs. This means that the Data Set and the NN must provide integrated methods to address these variants. We show this conceptually below. We believe this is a highly complex design.



4.11.2 Parallel

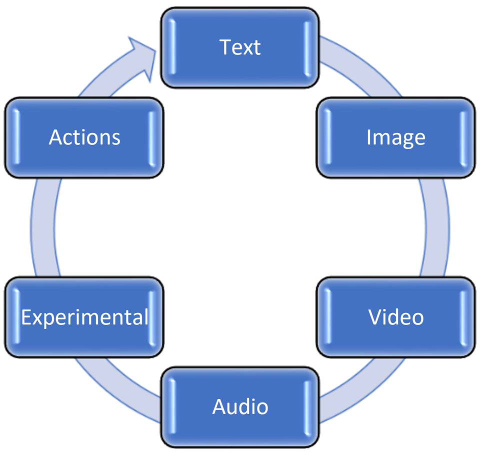
The second approach is a paralles approach where we have separate paths for each modality as shown below. This model then integrates on the output rather than the input.



5 EXAMPLES

We can consider multiple examples. But first, one should not the possible Inputs and possible Outputs. The inputs are what we provide to the AI entity and ask it to process it in some way. The Inputs contain information and an action request. Thus, in a medical context it may be pathology slides and a request and Output for a diagnosis. Perhaps it may also extend to patient care, patient prognosis, possible advanced therapies, locations for care, physicians and the like. The Inputs are things we have gathered and which we present to the machine and explicitly or implicitly pose a question, bounded or unbounded.

We may consider the inputs to be one or many of the types we depict below.



5.1 MEDICINE

AI can have a significant impact on medicine. From simple diagnostic and prognostic results from normal clinical results to the analysis of images, DNA and the like, and many other areas. As noted in Eisenstein:

Any biology student can pick a neuron out of a photograph. Training a computer to do the same thing is much harder. Jan Funke, a computational biologist at the Howard Hughes Medical Institute's Janelia Research Campus in Ashburn, Virginia, recalls his first attempt 14 years ago. "I was arrogant, and I was thinking, 'it can't be too hard to write an algorithm that does it for us'," he says. "Boy, was I wrong."

People learn early in life how to 'segment' visual information — distinguishing individual objects even when they happen to be crowded together or overlapping. But our brains have evolved to excel at this skill over millions of years, says Anna Kreshuk, a computer scientist at the European Molecular Biology Laboratory in Heidelberg, Germany; algorithms must learn it from first principles. "Mimicking human vision is very hard," she says. But in life-science research, it's increasingly required. As the scale and complexity of biological imaging experiments has grown, so too has the need for computational tools that can segment cellular and subcellular features with minimal human intervention. This is a big ask. Biological objects *can assume a dizzying array of shapes, and be imaged in myriad ways. As a result, says David Van Valen, a systems biologist at the California Institute of Technology in Pasadena, it can take much longer to analyse a data set than to collect it. Until quite recently, he says, his colleagues might collect a data set in one month, "and then spend the next six months fixing the mistakes of existing segmentation algorithms". The good news is that the tide is turning, particularly as computational biologists tap into the algorithmic architectures known as deep learning, unlocking capabilities that drastically accelerate the process. "I think segmentation overall will be solved within the foreseeable future," Kreshuk says. But the field must also find ways to extend these methods to accommodate the unstoppable evolution of cutting-edge imaging techniques.*

Medicine, in one sense, is an ideal area for AI. Yet to quote Dr Osler, "if all else fails, listen to the patient". This meant that a physician can do a myriad of tests but often the most critical one is listening to the patient, and using that to isolate the problem. AI does not often consider that type of interaction.

5.2 Information

The use of AI is already used in many information systems. It targets users and it modifies facts. It is promotional and persuasive and frankly is a true means of propaganda. Propaganda has always been a concern for democratic societies. We have seen that the social media entities are driven by propaganda elements, many selected by the media companies themselves. Now one must consider what an AAIE can do to manipulate society in an uncontrolled manner.

The work of Bernays and that of Lippmann in the 1920s was a clear example of Government manipulation during WW I. Wilson went about controlling the Press and criminalizing free speech, sending the presidential candidate Debbs of the Socialists to prison. Free speech is controlled by the media infrastructure. Clearly AI can further delimit this and currently is.

5.3 Transactions

Transactions range from purchases to financial transactions, to anything wherein a set of users relate by the transfer of elements of value. AI can facilitate these and it also can initiate these. Transaction are fundamentally the buying and selling, the movement of something of value concomitant with another entity of value. Again, we have the problem that an AAIE may decide on its own to enter into such transactions with itself being a shadow entity. This then can become a massive risk to the global financial markets.

5.4 Entertainment

AI can be a bastion of entertainment. As we had noted decades ago, digital effects and animation has dominated films. Acting is now irrelevant. Films have become nothing more than 90 second snippets of enhanced visualization. AI can allow for true personalization.

AI films can reproduce actors, voices, scenes, and in a sense can recalculate the plot depending upon the audience. In fact the audience can influence the plot in totality.

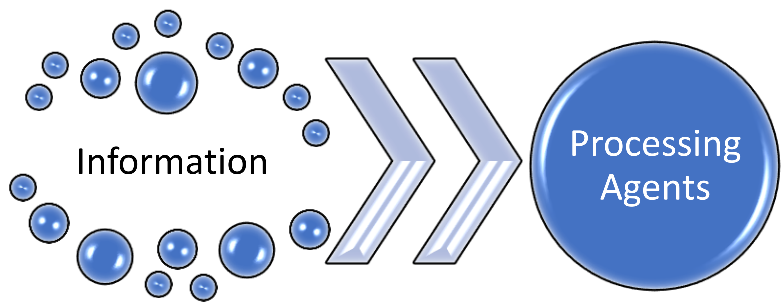
5.5 SOFTWARE

Software development can be enhanced by AI and it is being done now. Seeking routines to perform certain functions in a desired programming language can be readily accomplished. One can ask for a software program say in Python to sort some specific set of data. Now clearly if one can ask and obtain them, perhaps the system can ask itself and in turn build its own software structure. Thus, self-generating software would allow for the added functionality independent of a designer. Furthermore, the added software and its functioning may then be invisible to any outside observer.

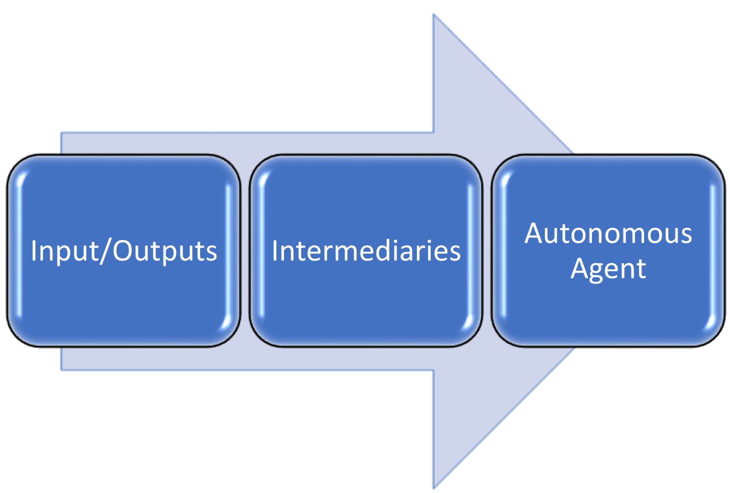
6 OBSERVATIONS

We often hear about the fears of AI devoid of any specificities. In order to understand what the risks may be one must understand what evolution can occur and what areas should be limited if any. In many ways it is akin to bio research on new organisms. We know that COVID is a classic example of bio-research gone wild.

Basically, the fundamental structure of AI as currently understood is some entity which relies on already available information that is used by some processing elements to perform actions.



Now, in contrast to what we have argued here, there is that this exogeneous Information set, provided by humans, may become self-organizing in an autonomous mode entity. Namely as we approach an autonomous mode this set of information may be generated by the entity itself, and no longer reflecting any reliance on a human. The evolution we see shown below goes from the current Input/Output to this Autonomous entity independent of humans and totally self-realizing.



6.1 Evolution of Neural Net Paradigms

The neural net paradigm has been evolving for almost the past fifty years. **Simply stated the neural net paradigm assumes a computer entity, that takes a massive amount of exogeneous information to train a network, so that when some input entity is presented, it can produce an output entity that correctly reflects the body of information available to the computer entity.** To accomplish this one needs significant amounts of information, memory and processing. Thus, conceptually one had the structure constructs yet it required the development and availability of memory and processing power to take the steps we see today. Thus, NN are not new but only constrained by technology.

In addition, the nature of inputs and outputs is also an evolving area. For the output we may want some natural language processor and say for the input the ability to gather and process images. In fact, the input must eventually gather all types of entities; video, image, taste, smell, touch, voice, etc. In fact, multimedia inputs and outputs will be essential[21].

We use the neural net construct as a place holder. One suspects there may be significant evolutions in these elements. One need look no further that what we have seen in the past 40 years. The driver for the evolutions will be processing complexity as well as computing complexity. One also suspects that there will be significant evolutions in memory for the learning data.

Also, paradigms on human neural processing may open avenues for new architectures. This is a challenging area of research. The biggest risk we face is the gimmick constructs that are currently driving the mad rush.

6.2 Risk of Autonomy:

The risk of autonomy was perceived in broader terms by Wiener in his various writings. The development of AEs is the development of entities that can displace if not annihilate man. We see that AEs can restructure their own environment and that control of AEs may very well be out of the hands of their developers. In fact, the developer may not even be aware of when such an autonomous act occurs.

On has always considered the insights of Shannon and his Information Theory and the broader constructs of Wiener and Cybernetics. One suspects we are leaving the world of Shannon and entering that of Wiener.

[21] See https://www.researchgate.net/publication/344445284_Multimedia_Communications_Revised This is a copy of a draft book I wrote for a course in Multimedia Communications at MIT in 1989. The ideas therein should be integrated into an AI construct.

6.3 PARALLELISM WITH HUMAN INTELLIGENCE OR NOT

If AEs are to be considered intelligent than how would we compare that to human intelligence. Would an AE consider humans just an equivalent primordial slime, an equal, a superior, or just some nuisance inferior species? Can we measure this or is it even measurable.

6.4 AREAS OF GREATEST RISK

The areas of greatest risk are legion in AI. They range from simple misinformation, to psychological profiling, then influencing and controlling large groups, and finally as full autonomy is obtained, the ability to manipulate their environment.

Without some moral code or ethical framework, AEs can act in whatever manner they so choose, often taking leads from the data input that may have or create themselves.

There have been multiple lists of AI risks[22]. The problem is that all that have been generally available lack any framework for such listing. They generally make statements regarding privacy, transparency, misinformation, legal and regulatory etc. These are for the most part content free sops. One needs, actually demands, the canonical evolution we have presented herein to understand what the long-term risks may be. Having a construct to work with then policies may evolve.

6.5 STABILITY OF AUTONOMOUS ENTITIES

Autonomous entities, AE, can result in unstable constructs. The inherent feedback may result in the AE in cycling in erratic ways that are fundamentally unstable. This again is a concern that Wiener expressed. Stability of an AE may be impossible. They may be driven by the construct of, "on the one hand but on the otherhand". This is a construct without a moral fabric, without an underlying code of conduct[23].

6.6 AI; POLICY AND PREVENTION

---

[22] See https://www.forbes.com/sites/bernardmarr/2023/06/02/the-15-biggest-risks-of-artificial-intelligence/?sh=68573d752706 or https://www.scientificamerican.com/article/heres-why-ai-may-be-extremely-dangerous-whether-its-conscious-or-not/ or https://ai100.stanford.edu/gathering-strength-gathering-storms-one-hundred-year-study-artificial-intelligence-ai100-2021-1-0

[23] See https://www.researchgate.net/publication/338298212_Natural_Rights_vs_Social_Justice_DRAFT We have examined this issue in the context of Natural Rights, a fundamental and perhaps biologically and genetically and evolutionarily programmed code of human conduct. Namely we assert that humans have evolved with a genetically programmed code of behavior displayed in what they believe are Natural Rights. These Natural Rights then become limits on unstable and extreme behavior. We further argue that these are evolutionary, not inherent in any creature. They are survival genetic expressions for the species. There is no reason to expect that an AE would in the near term ever assert such rights. Thus it is a basis for human annihilation.

Isaac Asimov in his robot novels present the three rules of robotics[24]. However, AI is much more than robotics. Robots, in the Asimovian world, were small self-contained anthropomorphic entities. In our construct the AI Autonomous entity is an ever-expanding entity capable are unlimited capabilities. Moreover, these autonomous entities can evolve and expand independent of human interaction or control. Thus, the key question is; what can be done to protect humanity if not all of earthly entities from an overpowering and uncontrollable autonomous entity?[25]

First one must admit the putative capacity of existence for such an entity. Second one must recognize that the creation of these entities cannot be prevented since an adversary may very well do so as a means of a threat or control. Third, creation of such entities may very well be in the hands of technologists who lack and moral foundation and will just do so because they can do it. Thus, it is nearly impossible for this entity to be a priori controlled.

Therefore, at best one can a posteriori control such entities. This requires advanced surveillance and trans-governmental control mechanisms. Namely it can be possible to sense the existence and development of such systems via various distributed network sensing mechanisms. When detected there must be prohibitive actions in place and immediately executable in a trans-border manner.

6.7 Is an AI Entity the same as a Robot?

The Asimovian Robot is an anthropomorphic entity. In Asimov's world the robot was a stand-alone creature, one of many, with capabilities limited by its singularity. Robots were just what they were and no more. An AI Entity is a dynamically extensible entity capable of unlimited extension akin to a slime mold, a never-ending extension of the plant. The AI Entity may morph and add to itself what it internally sees a need for and take actions that are solely of its own intent. Thus, there is a dramatic difference between a Robot and an AI Entity. The challenge is that trying to apply the three laws of robotics to an entity that controls its own morphing is impossible.

6.8 Complexity vs Externality

We have noted herein that the early developments of AI revolve around increased processing and interaction complexity. However, there comes a point when externalities become the dominant factor, namely the ability of the AI entity to interact with its external environment, first with the help of a human, then with existing external entities and then with the ability to create and use its own externalities. This progression then leads to the AAIE which if not properly delimited can result in harms.

---

[24] *A robot may not injure a human being or, through inaction, allow a human being to come to harm. A robot must obey orders given it by human beings except where such orders would conflict with the First Law. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.*

[25] See Watson et al. This describes the work concerning Recombinant DNA. In a sense this is akin to the concerns regarding AI and its dangers. This discusses what it is and how it can be controlled. The concern was that this modified DNA could be sent loose in the environment. In a sense, the work here mirrors what can be done with AI. The problem however is with Recombinant DNA we had highly educated professionals on the research side but in contrast in AI we have a collection of Silicon Valley entrepreneurs.

6.9 What is the Use of Canonical Forms?

Canonical Forms have multiple uses. First, they provide structure. Second, they allow for defining issues and elements. Third they are essential if any regulatory structure is imposed. We have seen this in Telecommunications Law where elements and architecture is critical to regulation. However, as in Telecom and other areas, technology evolves and these Canonical Forms may do so likewise. Thus, they are an essential starting point and subject to modification and evolution.

6.10 Sensory Conversions are Critical

As we have observed previously, the conversion of various sensory data to system processable data is a critical step. The human and other animal sensory system have evolved over a billion years to maximize the survival of the specific species. The specific systems available to AI are still primitive and may suffer significant deficiencies.

However, in a AAIE system, self-evolution may occur at an order of multi magnitudes faster that the evolution we have in our species. What direction that evolution takes is totally uncertain. The effects of that evolution will also determine what an AAIE does as it perceives its environment.

6.11 Regulatory Proposals in Progress?

A group at MIT has recently made a regulatory proposal for AI[26]. They recognize, albeit rather in a limited manner, that one must define something to regulate it. They thus note:

It is important (but difficult) to define what AI is, but often necessary in order to identify which systems would be subject to regulatory and liability regimes. The most effective approach may be defining ***AI systems based on what the technology does, such as "any technology for making decisions or recommendations, or for generating content (including text, images, video or audio)."*** *This may create fewer problems than basing a definition on the characteristics of the technology, such as "human-like," or on technical aspects such as "large language model" or "foundation model" – terms that are hard to define, or will likely change over time or become obsolete. Furthermore, approaches based on definitions of what the technology does are more likely to align with the approach of extending existing laws and rules to activities that include AI.*

Needless to say, the definition is so broad that it could include a coffee maker or any home appliance. As we have argued herein, **AI inherently contains an element whereby massive data if collected and processed by some means that permits a relationship between an input and output to be posited. Also, and a key factor, is that the relationship between input and posited output is hypothesized by some abstraction of data sets chosen by the designer and potentially modified by the system.**

The MIT group then states:

[26] https://computing.mit.edu/wp-content/uploads/2023/11/AIPolicyBrief.pdf

Auditing regimes should be developed as part and parcel of the approach described above. To be effective, auditing needs to be based on principles that specify such aspects as the objectives of the auditing (i.e., ***what an audit is designed to learn about an AI system, for example, whether its results are biased in some manner, whether it generates misinformation, and/or whether it is open to use in unintended ways), and what information is to be used to achieve those objectives (i.e., what kinds of data will be used in an audit****)*

This is the rule of the select telling the masses what to believe! It seems academics just can't get away from this control mechanism. They further note:

For oversight regarding AI that lies beyond the scope of currently regulated application domains, and that cannot be addressed through audit mechanisms and a system similar to that used for financial audits, ***the federal government may need to establish a new agency that would regulate such aspects of AI.*** *The scope of any such regulatory agency should be as narrow as possible, given the broad applicability of AI, and the challenges of creating a single agency with broad scope.* ***The agency could hire highly qualified technical staff who could also provide advice to existing regulatory agencies that are handling AI matters*** *(pursuant to the bullets above). (Such a task might alternatively be assigned to an existing agency, but any existing agency selected should already have a regulatory mission and the prestige to attract the needed personnel, and it would have to be free of political and other controversies from existing missions that could complicate its oversight of AI.) A self-regulatory organization (like the Financial Industry Regulatory Authority, FINRA, in the financial world) might undertake much of the detailed work under federal oversight by developing standards and overseeing their implementation.*

Again, another Federal entity, and as academics do, they assume a base of qualified staff, an oxymoron for any Government entity. As we have noted previously, if you can't define it, you can't regulate it. Also, as is all too well known, all regulations have "dark sides".

7 POSTSCRIPT

I believe it is worth a postscript to explain the purpose of this Note. There are currently several foci of interest in AI. The original focus could be considered the Silicon Valley cabal. Those “entrepreneurs” who often collude together to promote the latest technical fad. OpenAI is now a classic example. For this group you have “the next new thing”. A second group is the EU clan who for some reason want to regulate everything whether they understand it or not. The we have the US lawmakers and administrations, who all too often create laws that spend decades being interpreted. We have seen this again and again as things go to the Supreme Court, where all too often is again mis-interpreted. There are likely dozens of other clans circling around AI.

Personally, I have dealt with versions of AI over the pasts fifty plus years, often without knowing it. My first book was on estimating stochastic systems, namely developing algorithms from data and experience to estimate or predict what would happen next. In the early 70s I spent time at Bell Labs trying to develop AI like algorithms to detect Soviet nuclear subs. Namely subs from whales from just noise. Lots of data but little accuracy. In the late 80s as Head R&D at NYNEX, now Verizon, I spent a long period seeing if neural nets were in the immediate future. They were not, we just did not have the computing power. On returning to MIT in the mid-2000s I examined the evolving AI elements in medical imaging. Thus, I have spent decade on the periphery, getting close but not quite there.

The comments and observations that I have made herein are mine alone. To some degree they are based upon my experience and contacts but unlike the Silicon Valley elite these are singular, influenced by a long period of evolution and understanding of what bad regulation can do.

Is AI a danger to humanity! Most likely in the near term much less than social media. Many of the Silicon Valley types are doing the “look over there” trick to avoid the harm all social media is doing. The problem with AI is twofold. One is “garbage in garbage out” related to the training, and the second is “sensory dissonance” related to how sensory data is digitized. Hopefully some of these observations herein have some merit.

Finally, I lived in Prague for a period in the 1990s and 2000s. My office and my desk faced out on the oldest synagogue in Europe and from there to both the graveyard and the Golem storefront. Each time I looked out I recalled Wiener and the Golem. Thinking of AAIE I also think of the Golem, the self-reproducing monster from that very location. Will AAIE become a Golem as Wiener was afraid of or can it be a beneficial tool. My fear is it will be both.

8 REFERENCES

1. Aggarwal, Neural Networks and Deep Learning, Springer, 2018
2. Ahdoot et al, MRI-Targeted, Systematic, and Combined Biopsy for Prostate Cancer Diagnosis, NEJM, 382;10, March 5, 2020
3. Araujo et al, Classification of breast cancer histology images using Convolutional Neural Networks, PLOS ONE, June 1, 2017
4. Bagheri et al, Advances in medical imaging for the diagnosis and management of common genitourinary cancers, Urol Oncol. 2017 July; 35(7): 473–491
5. Bagheri et al, Advances in medical imaging for the diagnosis and management of common genitourinary cancers, Urol Oncol. 2017 July; 35: 473–491
6. Bard et al, Image Guided Prostate Cancer Treatments, Springer, 2014
7. Bargaje et al, Review of Classification algorithms for Brain MRI images, International Research Journal of Engineering and Technology (IRJET), Volume: 04 Issue: 01 | Jan-2017
8. Barrett, What is multiparametric-MRI of the prostate and why do we need it?, Imaging Med. (2015)
9. Bax et al, Review and Comparison of Cancer Biomarker Trends in Urine as a Basis for New Diagnostic Pathways, Cancers 2019, 11, 1244
10. Bernays, Propaganda, Ig Publishing, 2005 (reprint 1928 book)
11. Berens et al, Multiscale, multimodal analysis of tumor heterogeneity in IDH1 mutant vs wild-type diffuse gliomas, PLOS ONE 14(12): e0219724, 2019
12. Birkfeliner, Applied Medical Image Processing, CRC, 2011
13. Boerer and Asa, Biopsy Interpretation of the Thyroid, 2nd Ed, Wolters Kluwer, 2017
14. Braman et al, Association of Peritumoral Radiomics With Tumor Biology and Pathologic Response to Preoperative Targeted Therapy for HER2 (ERBB2)–Positive Breast Cancer, JAMA Network OPEN, 2019
15. Brezinski, Optical Coherence Tomography, Academic, 2016
16. Brown and Semelka, MRI Basics, Principles and Applications, Wiley, 2003.
17. Bryant et al, Predicting High-Grade Cancer at Ten-Core Prostate Biopsy Using Four Kallikrein Markers Measured in Blood in the ProtecT Study, JNCI J Natl Cancer Inst (2015) 107(7)
18. Bushberg et al, The Essential Physics of Medical Imaging, Wolters, 3rd Ed, 2012
19. Campanella et al, Quantification of cancer risk of each clinical and ultrasonographic suspicious feature of thyroid nodules: a systematic review and meta-analysis, Bioscientifica, 2014
20. Cantley ey al, Signal Transduction, CSHL Press, 2014

21. Chambara and Ying, The Diagnostic Efficiency of Ultrasound Computer–Aided Diagnosis in Differentiating Thyroid Nodules: A Systematic Review and Narrative Synthesis, Cancers 2019, 11, 1759
22. Cheon and Lee, Color Edge Detection based on Bhattacharyya Distance, Proceedings of the 14th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2017) - Volume 2, pages 368-371
23. Cover and Hart, Nearest Neighbor Pattern Classification, IEEE Info Theory, Jan 1967
24. Custem, Imaging in Colorectal Cancer: Progress and Challenges for the Clinicians, Cancers 2016, 8, 81
25. Decazes and Bohn, Immunotherapy by Immune Checkpoint Inhibitors and Nuclear Medicine Imaging: Current and Future Applications, Cancers 2020, 12, 371
26. Demster et al, Gene expression has more power for predicting in vitro cancer cell vulnerabilities than genomics, https://www.biorxiv.org/content/10.1101/2020.02.21.959627v1
27. Devi et al, Detection of Cervical Cancer using the Image Classification Algorithms, IJCTA, 9(3), 2016, pp. 153-166
28. Dhawan, Medical Image Analysis, IEEE, 2011
29. Duda and Hart, Pattern Classification and Scene Analysis, Wiley, 1973
30. Duick et al, Thyroid and Parathyroid Ultrasound and Ultrasound Guided FNA, 4th Ed, Springer, 2018
31. Eiber et al, Prostate Cancer Molecular Imaging Standardized Evaluation (PROMISE): Proposed miTNM Classification for the Interpretation of PSMA-Ligand PET/CT, The Journal of Nuclear Medicine, Vol. 59, No. 3, March 2018
32. Eisenhauer et al, New response evaluation criteria in solid tumours: Revised RECIST guideline, European Journal of Cancer 45 (2009) 228 – 247
33. Eisenstein, AI under the microscope: the algorithms powering the search for cells, Nature 27 Nov 2023
34. Epstein, Biopsy Interpretation of the Prostate, 6th Ed, Wolters Kluwer, 2020
35. Fan et al, Radiomic analysis reveals DCE-MRI features for prediction of molecular subtypes of breast cancer, PLOS One, Feb 6, 2017
36. Fehr, et al, Automatic classification of prostate cancer Gleason scores from multiparametric magnetic resonance images, PNAS Nov 2, 2015
37. Friedman et al, Principles of MRI, McGraw Hill, 1989.
38. Fukunaga, Introduction to Statistical Pattern Recognition, Academic, 2nd Ed, 1990
39. Galimzianova et al, Toward Automated Pre-Biopsy Thyroid Cancer Risk Estimation in Ultrasound,
40. Gartner and Hiatt, Textbook of Histology, 3rd Ed, Saunders 2006
41. Gearhart and Ahuja, Colorectal Cancer, Saunders, 2011

42. Ghai and Haider, Multiparametric-MRI in diagnosis of prostate cancer, Ind Jrl Uro, 2015, Vol 31, Issue 3

43. Ghai and Haider, Multiparametric-MRI in diagnosis of prostate cancer, Indian J Urol. 2015 Jul-Sep; 31: 194–201

44. Goldman and Schafer, Goldman's Cecil Medicine, 24th Ed, Elsevier, 2012

45. Gonzalez and Woods, Digital Image Processing, 4th Ed, Pearson, 2018

46. Guorong et al, Bhattacharyya Distance Feature Selection, Proceedings of the 13th International Conference on Pattern Recognition, vol. 2, pp. 195-199, 1996

47. Gurusamy and Subramaniam, A Machine Learning Approach for MRI Brain Tumor Classification, CMC, vol.53, no.2, pp.91-108, 2017

48. Halicek et al, Deep convolutional neural networks for classifying head and neck cancer using hyperspectral imaging, J. Biomed. Opt. 22(6), 060503 (2017)

49. Hariharan, Optical Interferometry, Academic, 2003

50. Hasan et al, Segmentation of Brain Tumors in MRI Images Using Three-Dimensional Active Contour without Edge, Symmetry 2016, 8, 132

51. Haykin, Neural Networks, 3rd Ed, Pearson, 2016

52. Horowitz, MRI Physics for Physicians, Springer, 1989

53. Huettel et al, Functional Magnetic Resonance Imaging, Sinauer, 2009.

54. Jacobs and Finlayson, Breast Cancer, Saunders, 2011

55. Jain et al, Statistical Pattern Recognition: A Review, IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 22, No. 1, January 2000

56. Jelinek, Statistical Methods for Speech Recognition, MIT Press, 1997

57. Jurafsky and Martin, Speech and Language Processing, Prentice Hall, 2000

58. Kak and Slaney, Computerized Tomographic Imaging, IEEE, 1988

59. Karp and Falchook, Targeted Cancer Therapy and Immunotherapy, 2nd Ed, Wolters Kluwer, 2019

60. Kochmar, Natural Language Processing, Manning, 2022

61. Korn, A Multiparametric MRI for the Classification and Grading of Prostate Cancer, UCSF PhD Thesis, 2019

62. Kuhl et al, Mammography, Breast Ultrasound, and Magnetic Resonance Imaging for Surveillance of Women at High Familial Risk for Breast Cancer, Jrnl Clin Onc, Nov 2005

63. Kuhl, The Changing World of Breast Cancer, Investigative Radiology • Volume 50, Number 9, September 2015

64. Kumar et al, Basic Pathology, Elsevier, 2018

65. Kumar et al, QIN “Radiomics: The Process and the Challenges”, Magn Reson Imaging. 2012 November; 30: 1234–1248

66. Larrea et al, New Concepts in Cancer Biomarkers: Circulating miRNAs in Liquid Biopsies, Int. J. Mol. Sci. 2016, 17, 627

67. Li e al, Robust Automatic Speech Recognition, Academic, 2016

68. Liang and Lauterbur, Principles of Magnetic Resonance Imaging, IEEE Press, 2000.

69. Lippmann, Public Opinion, Free Press 1997 (reprint 1922 edition)

70. Long and Wong, The Chemistry of Molecular Imaging, Wiley, 2015

71. Lufkin, The MRI Manual, Year Book Med Pub, 1990

72. McRobbie et al, MRI, From Picture to Proton, Cambridge, 2003.

73. Minsky, The Society of Mind, Simon and Schuster, 1986

74. Minsky and Papert, Perceptrons, MIT Press, 1969

75. Monti et al, Multiparametric MRI for Prostate Cancer Detection: New Insights into the Combined Use of a Radiomic Approach with Advanced Acquisition Protocol, Cancers 2020, 12, 390

76. Motlak, Detection and Classification of Breast Cancer Based-On Terahertz Imaging Technique Using Artificial Neural Network & K-Nearest Neighbor Algorithm, International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 21 (2017) pp. 10661-10668

77. Murphy et al, The Expanding Role of MRI in Prostate Cancer. Gen Imag, AJR, V 201, Dec 2013

78. Nie et al, Rectal Cancer: Assessment of Neoadjuvant Chemo-Radiation Outcome Based on Radiomics of Multi-Parametric MRI, AACR 2016.

79. Padhani, Dynamic contrast-enhanced MRI in clinical oncology: current status and future directions, J Magn Reson Imaging. 2002 Oct;16(4):407-22

80. Parekh et al, Multiparametric Deep Learning Tissue Signatures for a Radiological Biomarker of Breast Cancer: Preliminary Results, Division of Cancer Imaging, The Johns Hopkins University School of Medicine. Preprint

81. Parker, Image Reconstruction in Radiology, CRC, 1990

82. Parkins, Investigating the Mechanisms of Breast Cancer Metastasis Using Multimodality Molecular Imaging, PhD Thesis, The University of Western Ontario, 2019

83. Petralia and Thoeny, DW-MRI of the urogenital tract: applications in oncology, Cancer Imaging (2010) 10, S112S123

84. Phelps, PET, Physics, Instrumentation and Scanners, Springer, 2010

85. Poggio et al, Complexity control by gradient descent in deep networks, Nature Communications volume 11, Article number: 1027 (2020)

86. Rizzo et al, Radiomics: the facts and the challenges of image analysis, European Radiology Experimental (2018) 2:36

87. Royle, The Digital Cell, CSHL Press, 2020.

88. Rumelhart et al, Parallel Distributed Processing, Vol 1 and 2, MIT Press, 1986
89. Saha, Basics of PET Imaging, 2nd Ed, Springer, 2010
90. Sanchez, TI-RADS classification of thyroid nodules based on a score modified according to ultrasound criteria for malignancy, Rev. Argent. Radiol. 2014;78(3): 138-148
91. Schreier et al, A Full-Image Deep Segmenter for CT Images in Breast Cancer Radiotherapy Treatment, Front in Onc, July 2019
92. Sencha and Patrunov, Thyroid Ultrasound, Springer, 2019
93. Shah and Zhou, Prostate Biopsy Interpretation: An Illustrated Guide, Springer, 2012
94. Shannon, The mathematical theory of communications, BSTJ, 1948
95. Shen and Rubin, Prostate Cancer, CSHL Press, 2019
96. Simon, The Sciences of the Artificial, MIT Press, 1969
97. Stabile et al, Multiparametric MRI for prostate cancer diagnosis: current status and future directions, Nature Rev Urology, July 17, 2019
98. Stattin et al, Improving the Specificity of Screening for Lethal Prostate Cancer Using Prostate-specific Antigen and a Panel of Kallikrein Markers: A Nested Case–Control Study, Euro Urology, 2015
99. Stoyanova et al, Association of multiparametric MRI quantitative imaging features with prostate cancer gene expression in MRI-targeted prostate biopsies, Oncotarget, Vol. 7, No. 33, 2016
100. Stoyanova et al, Prostate Cancer Radiomics and the promise of Radiogenomics, Transl Cancer Res 2016;5(4):432-447
101. Su et al, A Neural-Network-Based Approach to White Blood Cell Classification, The Scientific World Journal, Volume 2014
102. Su, Prostate, Saunders, 2010
103. Subramaniam, Neural Network with Bee Colony Optimization for MRI Brain Cancer Image Classification, The International Arab Journal of Information Technology, March 2015
104. Sun et al, Deep Learning vs. Radiomics for Predicting Axillary Lymph Node Metastasis of Breast Cancer Using Ultrasound Images: Don't Forget the Peritumoral Region, Frontiers in Oncology, January 2020 | Volume 10 | Article 53
105. Supe et al, Image Segmentation and Classification for Medical Image Processing, International Journal on Future Revolution in Computer Science & Communication Engineering Volume: 5 Issue: 1 45 – 52, Jan 2019
106. Tan et al, Thyroid imaging reporting and data system combined with Bethesda classification in qualitative thyroid nodule diagnosis, Medicine (2019) 98:50
107. Tessler et al, ACR Thyroid Imaging, Reporting and Data System (TI-RADS): White Paper of the ACR TI-RADS Committee, Journal of the American College of Radiology, Volume 14, Number 5, May 2017

108. Thamilselvan and Sathiaseelan, Detection and Classification of Lung Cancer MRI Images by using Enhanced K Nearest Neighbor Algorithm, Indian Journal of Science and Technology, Vol 9(43), November 2016

109. Thestrup et al, Biparametric versus multiparametric MRI in the diagnosis of prostate cancer, Acta Radiologica Open, 5 1–8, 2016

110. Türkbey et al, Diffusion-weighted MRI for detecting and monitoring cancer: a review of current applications in body imaging, Diagn Interv Radiol 2012; 18:46–59

111. Volinia et al, A microRNA expression signature of human solid tumors defines cancer gene targets, PNAS February 14, 2006 vol. 103 no. 7 2257–2261

112. Wang et al, Artificial Intelligence in Ultrasound Imaging: Current Research and Applications, Advanced Ultrasound in Diagnosis and Therapy 2019;03:053–061

113. Wang et al, Computer Aided-Diagnosis of Prostate Cancer on Multiparametric MRI: A Technical Review of Current Research, BioMed Research International, Volume 2014, Article ID 789561

114. Wang et al, Computer Aided-Diagnosis of Prostate Cancer on Multiparametric MRI: A Technical Review of Current Research, BioMed Research International, Volume 2014, Article ID 789561

115. Westbrook and Talbot, MRI in Practice, 5th Ed, Wiley, 2019

116. Westbrook, MRI Technique, 4th Ed, Wiley, 2014

117. Westphalen et al, Detection of clinically significant prostate cancer with PIRADS v2 scores, PSA density, and ADC values in regions with and without mpMRI visible lesions, IBJU, Vol. 45 (4): 713-723, July - August, 2019

118. Watson et al, Recombinant DNA, Freeman, 1983

119. Widrow et al, The No-Prop algorithm: A new learning algorithm for multilayer neural networks, Neural Networks 37 (2013) 182–188

120. Wiener, Cybernetics, MIT Press, 1948

121. Wiener, God and Golem, MIT Press, 1964

122. Wiener, The Human Use of Human Beings, Houghton Mifflin, 1954

123. Winfield et al, DCE-MRI, DW-MRI, and MRS in Cancer: Challenges and Advantages of Implementing Qualitative and Quantitative Multi-parametric Imaging in the Clinic, Topics in Magnetic Resonance Imaging, Volume 25, Number 5, October 2016

124. Winograd, Language as a Cognitive Process, Addison Wesley, 1983

125. Winston, Artificial Intelligence, 2nd Ed, Addison Wesley 1984

126. Yang et al, Magnetic resonance imaging (MRI)-based radiomics for prostate cancer radiotherapy, Transl Androl Urol 2018;7(3):445-458

127. Zhang, et al, An imaging-based approach predicts clinical outcomes in prostate cancer through a novel support vector machine classification, Oncotarget, Vol. 7, No. 47, Aug 2016

9 INDEX