The Periodic Table of AI

Intelligence

The question of what constitutes intelligence has always been problematic for us.

As we look at our place in the world, we want to believe that human intelligence is unique and separates us from the animals. But the moment we try to define what makes us unique, we confront examples of birds catching food to plan for the future, dogs that find toys by reasoning from exclusion,
and apes using sign language to communicate with us. There isn't one universal thing that sets us apart.

As we try to find the thread of intelligence that binds us, we have to deal with the difference between book learning and street smarts, the war between emotional and cognitive intelligence, and the difference between visual and verbal learners. And though we see ourselves as creatures of reason, the entire field of behavioral economics and a good chunk of cognitive psychology argue that we survive using a loose collection of decent heuristics and a complete misunderstanding of statistics.

Given our history, it is not surprising that we are now wrestling with the question of what constitutes machine intelligence. And given our special relationship with intelligence as a defining characteristic, it is equally unsurprising that this debate gets emotional. It is hard to think clearly about a technology that we assume is going to outthink us, take our jobs and ultimately wipe us off the planet.

Nonetheless, we must study the nature of machine intelligence to understand functionality and trustworthiness and to get a handle on its role in our world. In particular, dissecting the components of these systems and exploring them enables us to make informed decisions.

This is an attempt at that exploration, which aims to start a conversation around the functional elements of intelligence. The goal here is to work through the hype around AI, consider what we really have at hand, examine the tests that get us closer to a true definition of AI, and finally deconstruct machine intelligence into its functional components.

Intelligence clarified

The need for clarity in the AI space is particularly pressing given the explosion of AI technologies over the past five years. We have seen massive achievements and the pace of innovation continues to accelerate. As we deploy intelligent systems, it is important that we understand the breadth of their abilities and shortcomings. In particular, it is imperative that decision-makers understand what is currently achievable, the limits of those technologies and enough about how they work to recognize possible shortcomings. To gain this clarity, we should probably look at the factors that are getting in the way.

Denying the <u>potential</u> of intelligent systems

A couple of years ago I was working with a United Nations team considering policy issues related to autonomous lethal devices, aka *killer robots*. We were considering devices with long-term deployment plans and one of the participants commented that the responsibility for actions taken by a device that had been at sea for six months was still solely in the hands of the individual who deployed it.

When the idea that these systems would be making decisions on their own came up, and that we needed to consider the scope of their autonomy and thus responsibility, the team was shocked. It wasn't just the idea of machines making life and death decisions that was disturbing. They were shocked by the more fundamental idea that machines would be capable of making decisions at all.

This resistance to the idea of intelligent machines as decision makers points to a very real problem. Tautological as it seems, if we don't understand the capabilities of the systems we are using, we will end up deploying systems whose actions we cannot understand. Few things are as scary as the idea of implementing intelligent systems that are making decisions if we are in denial of the idea that they are even capable of doing so. Denying capabilities doesn't make them go away.

Listening too hard to the <u>marketing</u> of intelligent systems

We face another problem related to the marketing of intelligent systems. Recently, the CTO of a technology company that provides a cyber intrusion alert product stated, "We don't use rules, it's all machine learning." Of course, he meant that the rules used by the system were learned from historical data and automatically updated based on new data rather than built and updated manually. But framing the distinction as one between "learning" and "rules" in service of the marketing gets in the way of people understanding the technology.

The point is that we live in a mixed world of science, technology and marketing. We have to understand and address the possibility of conflicting goals. Both technologists and marketers are struggling to differentiate ideas and products and this goal to be different sometimes results in framing distinctions that only serve to confuse.

Because of this, we face some confusion around ideas such as *machine learning*, *cognitive computing* and even *speech recognition*. This plays out in the confusion between learning systems that acquire knowledge of the world and the execution systems that apply it. Similarly, marketing terms such as "cognitive computing" or "smart systems" are sometimes confused for indicating technical distinctions that often don't exist. And because of the focus of popular press on the most visible applications, there is an ongoing confusion between a system's ability to *recognize* words in an audio stream versus its ability to *understand* what they mean in context.

The core technologies of machine learning, cognitive computing and speech recognition are genuinely brilliant. But we have to see them for what they are so that we can have the appropriate expectations (and make the appropriate decisions) about what they can and cannot do for us. Even more importantly, we must understand what needs to happen next.

Glossing over how intelligent systems work

Very few people need to understand the details of every piece of technology in their lives. However, we are entering an era in which decision makers need to know more about the core technologies of AI simply because they need to figure out how to work with them, manage them and understand their shortcomings.

A good example of this is the public's attitude towards Microsoft's Tay chatbot. Tay was a Twitter-based system designed to learn how to interact using examples of what people said to it. In a failure that was both very public and very dramatic, Tay was taught to be a "racist, misogynistic jerk" in less than 24 hours by a focused group of 4chan users who were looking to have some fun.

There are two issues here. First, Tay did not become misogynistic. It simply learned phrases that one would associate with being a misogynist, but had no understanding of what the phrases actually meant. Second, Tay was not a failure from a technological point of view. It was, and remains, an incredibly effective call-and-response engine. Tay's failure was the result of the community that raised it. It was a good technology that fell in with a bad crowd.

Of course, because human behavior is the only model we have for understanding these sorts of behaviors, we interpret them in the same way we interpret those of the people around us. But as we unpack systems like Tay, we can determine the areas of strengths and weaknesses. And we can use that understanding to evaluate when and where we can apply them.

Another example is Tay's older sister, XiaoIce, that Microsoft deployed in China over a year ago. It is incredibly successful and popular and never developed the problems that Tay exhibited simply because it was trained in an environment that is more restrained and polite than what you would find on 4chan.

As people look at different AI and Cognitive Computing technologies, this kind of misattribution plays out over and over again. Our knowledge of complex human comprehension influences how we understand the capabilities of various systems. We look at Watson, and because it provides answers in the form of text, it must understand what the text means. Because deep learning techniques cluster images of cats together, we assume that it must also understand what cats are. And we think that Siri must actually understand what we are saying because it is difficult to even imagine recognizing words without having ideas attached to them.

As we look at intelligent systems, we have to understand that we are often looking at a piece of the puzzle, and that the puzzle is not yet complete. Our experience with human thought sometimes misleads us by allowing us to attribute what we see as being broader and deeper than it may be. The point is that although Watson, Siri and deep learning are all genuinely brilliant technologies, none of them are even close to the complete story of intelligence.

Lessons from Intelligence Tested

There have been many definitions of Artificial Intelligence over time and many approaches to evaluating it.

Alan Turing famously avoided defining AI by proposing an empirical approach to testing it that took the form of playing a variant of the Imitation Game. In the original game, a man and woman hide in respective rooms and answer questions written on pieces of paper slid under a door. The challenge of the players outside the door was to figure out which is which. In the variant proposed by Turing, the computer plays the role of one of the hidden players, and the challenge becomes one of figuring out which is the machine and which the human.

Setting aside the possibility that the test Turing proposed was as much a reflection of his relationship with other people as it was a way to evaluate intelligent systems, it is valuable as it views AI from the perspective of functionality rather than technique. That, plus the fact that it is easy to implement, is what accounts for its sustained popularity over time and the spawning of multiple variants.

If the strength of the Turing Test is its focus on functionality rather than technology, its weakness is its relatively narrow focus. Though Turing was arguing that we should evaluate intelligent machines with regard to how well their behaviors align with humans, fairly specific terms defined the test itself. Unfortunately, it tends to be interpreted in terms of the latter rather than what it was designed to reveal.

There have been a variety of responses to the Turing test. On the one hand, there are those that dislike the behavioral approach in general and tend to argue for evaluations that focus on the mechanisms or theoretical basis rather than the behaviors. On the other, there are those who see the test as simply incomplete and have suggested modifications and extensions to the behavioral model. Though different responses, a model of what constitutes intelligence in the first place is the driving factor behind each of them.

For example, Hector Levesque has proposed the *Winograd Schema Challenge*. The Challenge aims to evaluate whether or not a system can apply general knowledge of the world to tasks such as language understanding. It centers on sentences such as

"The trophy would not fit in the brown suitcase because it was too big <small>" and "*The town councilors refused to give the demonstrators a permit because they feared <advocated> violence.*" These sentences are used to determine if systems can apply knowledge of concepts like size, politics and human goals to the understanding process. The assumption in this proposal is that long-term knowledge of the world is going to be a necessary component of any intelligent system.

Gary Marcus has proposed a different approach, called the *Marcus Test*, with a similar theoretical thrust. His take on intelligence is that it requires the ability to take in information, synthesize it and then use the resulting knowledge. He envisions a test in which systems watch television and then answer questions based on what they have seen and understood. As with the Winograd Challenge, the assumption is that common sense reasoning is a core requirement of intelligence and has to be part of any evaluation.

In fact, most of the Turing alternatives, such as the Visual Turing Test or the Lovelace Test, are aimed at some specific aspect of intelligence. The Visual Turing Test and the Construction Challenge are designed around integration of visual recognition and spatial reasoning with interpretation and plan execution while the Lovelace Test is designed to test creativity. The point being that each has its own focus and point of view on what constitutes intelligence.

That we have multiple metrics for evaluating AI is excellent. That we try to apply these metrics to any and all systems that call themselves AI seems shortsighted. Visual acuity and spatial reasoning are important, but not to a system that is providing you with financial advice. Speech recognition is necessary for conversational systems unless designed around typed input. And the ability to form abstractions and reason analogically is part of the human experience, but unnecessary for a vehicle that is just trying to keep in its own lane.

Different systems have different functions and different goals. It makes sense that we apply different tests and metrics when we evaluate them.

Even when we are considering the concept of *general AI*—systems that are designed to be complete models of all human capabilities—it is important that we tease apart these different functionalities so that we can focus on both developing and evaluating them. It is equally important that we think about how the individual components work together to create that complete system along with how they support and augment each other.

Similarly, even if we believe that there is a single master algorithm or that all elements of an AI system derive from the use of Deep Learning, we need to understand what those elements are. In fact, understanding the nature of these elements—their use and the information they require to function—is crucial if we are going to build learning systems that produce them. The techniques at the core of Deep Learning are only meaningful in the context of the idea that the networks that result from them can perform tasks such as recognition, categorization and transformation. Hinton didn't design a learning system and then wonder how to use it. He saw the networks as a way to support reasoning based on a model that was also learnable.

Productizing Intelligence

Another way to start deconstructing intelligent systems is to look at them from a product perspective. If we see intelligence as a product that has to be developed and managed, we can talk about *what* these systems are doing rather than *how* to implement them. We can evaluate them from the perspective of their utility, scope and scale rather than whether they adhere to the dictates of a particular theory. Aside from allowing us to avoid turf battles, it gives us the tools to evaluate not only the systems but also the theories that drive them.

If we take a product point of view, we need to begin with the functionality of an overall system rather than the individual features. We need to start with a model of a *Complete Artificial Intelligence* system that has all of the end-to-end functionality to take information in, process it and then take actions in response. Even if we stay agnostic with regard to the question of *Narrow* versus *General AI,* we still need to consider what is required for a system to function end-to-end.

So if we view Complete AI as a product, what does it need to do for us? At the most abstract level, it is clear that it needs to be able to do three things: assess, infer and respond.

- Systems need to be able to assess inputs and form characterizations of what is happening based on some connection to the external world. Speech and image recognition, sensor integration and situation assessment all fall under this functionality.

- They need to be able to infer new information from earlier assessments and inference. Language understanding, predictive analysis, goal tracking and reasoning about causes and outcomes tend to fall under this category.

- They then need to be able to respond to their understanding of the world based on some sort of goal, impulse or intention. Response includes forming plans of action, providing advice or even taking physical actions.

These three aspects of intelligence support the ability to know what is happening, know what it means and then know what to do. And while this is a painfully simple characterization, even here there is nuance.

Real-time systems are cycling through this loop on a constant basis. Some systems might include their own responses as part of the next round of assessment. Some may have multiple loops running at different time scales and dealing with different types of input, reasoning and response. And of course, many, if not most systems, will be based on work that only fits into one of these elements.

A system doing entity, location and topic extraction and aggregating this work is functioning at the level of assessment and inference, but might very well have no response component. A complete system like Google's Alpha Go has a learned network to do initial assessment of a board and a decision tree to manage the follow-on characterization and choice of move. Likewise, IBM Watson uses natural language techniques to form an assessment of the problem from its input, pattern-based search and aggregation to draw the right inferences, and then uses selection through ranking to select the response. Deconstruction of intelligence into functional components enables us to characterize the roles that the components play in an overall system.

Of course, these initial three areas are not enough. We need to dive deeper to get a full sense of the ecosystem and unpack each of these aspects.

Unpacking Assessment

At its most basic, assessment is recognition and identification, often from a sensory layer such as audio or visual inputs.

Speech recognition systems that map audio signals to words are the canonical example of assessment. The input is a raw audio signal and the output is a set of words that can then be used to determine meaning. The recognition process provides the assessment of spoken words while other processes will provide an interpretation of what those words mean.

This first step is crucial; however, knowing that someone has said that they want pizza is very different than understanding how to interpret the utterance in the context of looking for a place to eat, putting together a shopping list or talking with a waitress in a small Italian restaurant.

Even in the audio space, speech recognition is only the beginning. Beyond words, there is the possibility of recognizing intonation and the voices of specific speakers.

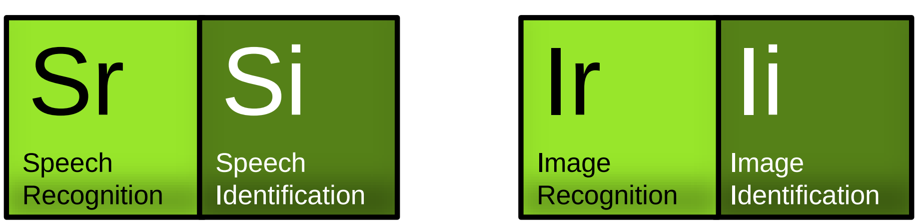
Stepping beyond voice, audio signals can be used to recognize shifts in tone that indicate a piece of music or the subtle changes in the sonic landscape of an engine that potentially linked to diagnostics. And audio signals can be used to identify the rasp in breathing indicative of the location of an infection. While the focus in the audio space has been on speech, for the obvious reason that we want systems that can understand the words we say, the range of capabilities that we can develop is enormous.

In the visual space, we see a similar dynamic. Faces and specific classes of objects have been at the center of the work in this space, but this is clearly just the beginning. The range of possibilities in this space around recognition of objects in general and specific instances, in particular, is enormous. From the identification of quasars from telescopic data to the recognition of meaningful configurations on a Go board, we see the intense value that comes from this early stage assessment that maps signal sets onto features that have clearer semantic value.

Of course, the possibilities in terms of these signal sets scope well beyond the limitations of human senses. From laser range finders to RFID to the entire electromagnetic spectrum, we will be building systems to do assessment, recognition and identification for the entire suite of inputs provided by the emerging Internet of Things. The hallmark is that the functionality is one of mapping data from low semantic spaces (e.g., pixels, audio signals, and sensor data) onto higher order spaces (e.g., faces, words, object proximity). That mapping may take the form of recognition of identifying either general characteristics (e.g., a voice saying "My fellow Americans"), or specific instances (e.g., the president is speaking).

For each of these, we can see parallel systems. For every recognition system, there may be similar identification functionalities. In speech, for example, it would be the distinction between recognizing the words and identifying the speaker. With images, it is the difference between recognizing that the image is of a building versus identifying that it is the Empire State building.

If we were to map these functions onto elements, we could distinguish between these different situations visually:

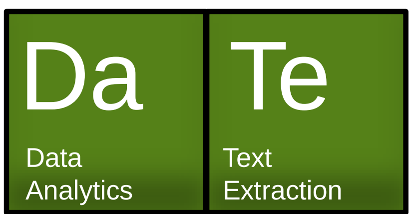


While an argument can be made that assessment stops there, it is useful to consider the layer of simple relationships synthesis as part of the process. Noting the increase in the volume or shift in tone of someone's speech, or the fact that there is a small dog next to a larger dog in an image are examples of this type of assessment. But this local juxtaposition is different than the next level of interpretation that someone is getting angry, or perhaps the small dog should be concerned.

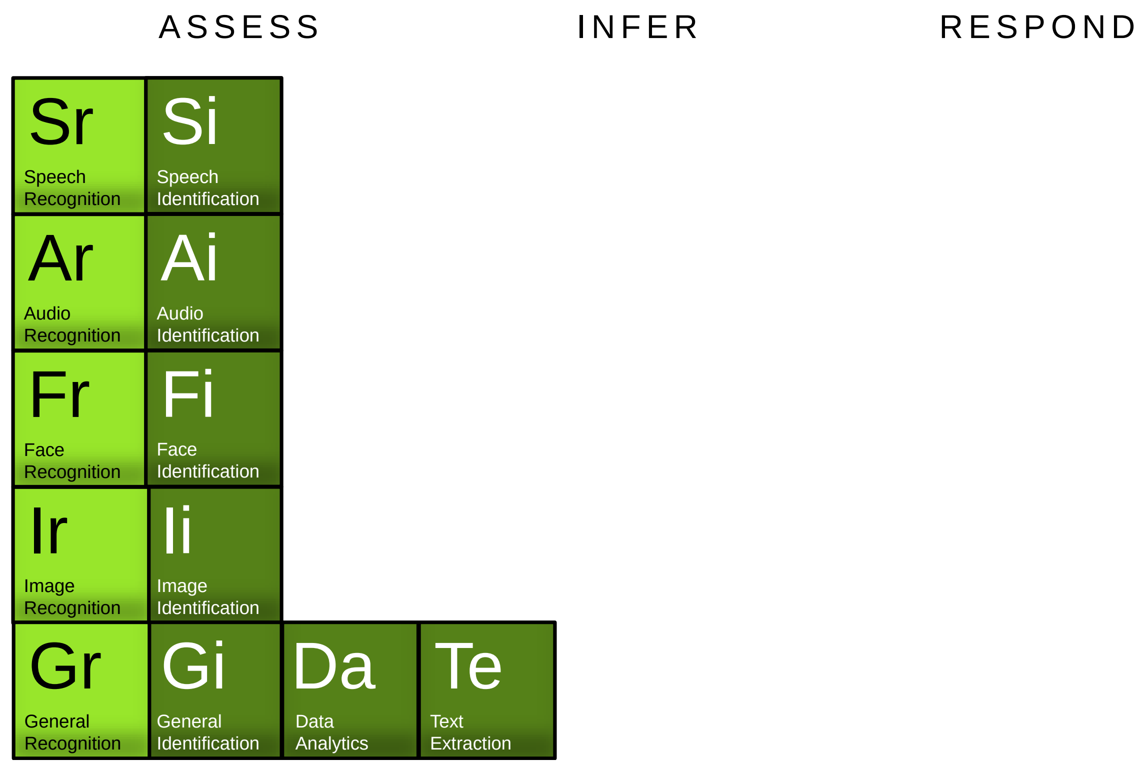
As with the early stage recognition layer, this type of interpretation is also driven by the input itself rather than features external to the input, aka other contexts.

Examples of this interpretive aspect of assessment appear outside of the scope of sensory inputs. Systems that extract names, locations and even topics from text are performing the same sort of input-centric assessments. They map terms (words and phrases) to entities and concepts but do not develop a complete picture of how the elements work together to create meaning. The crucial point is that they are extracting decontextualized elements from the text, not engaging in a language-understanding process that results in meaning.

Likewise, systems that process data to extract facts are performing similar assessments that result in a baseline description of the world through the lens of that data. As with the other types of input, the goal is the recognition of certain core facts based on the input and some analysis of it. Text extraction and data analysis are new elements for our growing periodic table.



While they may use radically different techniques, these are all examples of systems that start with inputs associated with the world, and make assessments as to what is happening based on those inputs. The central point and differentiator is that they use very little in the way of what we tend to consider as context in order to interpret beyond the input. Some may use learned networks; some may depend on lists of terms while some may be doing direct analysis of the data. But all of them are aimed at extracting a core set of data or facts related to what is happening in the world based on an assessment of some initial input.



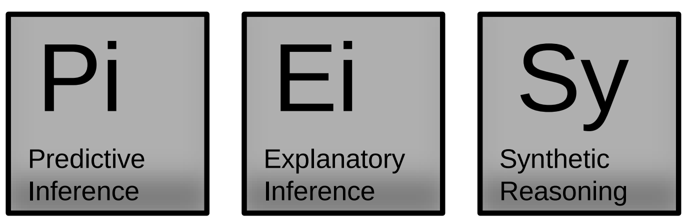
Speech and image recognition and identification, the more general sensor assessment, text extraction and data analysis give us the beginning of our periodic table of AI functionality. It provides us with a first pass at the assess stage and gives us a baseline off of which we can consider the next stage.

Examining Inference

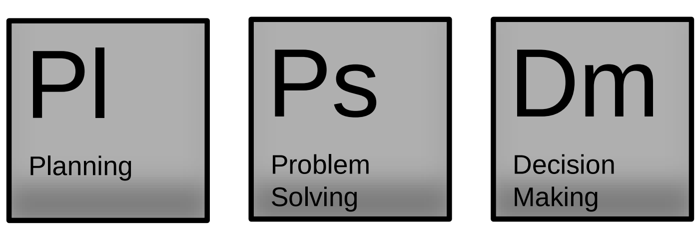
The notion of inference has been at the core of Artificial Intelligence since the beginning, often characterized by the application of logic or rules that are applied to create chains of reasoning. In the current wave of AI development, this notion of rule-based reasoning has fallen to the wayside, replaced by approaches that depend on machine learning. As we examine inference here, however, the focus is not on the mechanism but the behavior. From both scientific and development points of view, there may be arguments and disagreements about how things are implemented (and what theories they follow), but that is very different than understanding what core functionalities have to be built or accounted for in an intelligent system.

The need for inference seems clear. The ability to anticipate the future, characterize the present and explain the past is essential to any intelligent system. All three of these core inferential capabilities—prediction, explanation and synthesis—are necessary for any intelligent behavior that goes beyond the purely reactive.

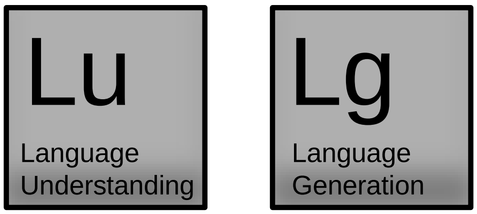
This does not imply that there needs to be an explicit representation of different states or even rules that provide the transitions between them, only that the behaviors have to be available. Given that, we can add these three new functions to the mix:



If explanations, characterizations and predictions are reflections of reasoning about the past, present and future, it is clear that we need to also consider the world of the hypothetical. In particular, we need to consider reasoning that supports problem-solving and planning. There is a gray area between these two functionalities, but one feature that can be used to differentiate them is that problem-solving tends to combine the reasoning with the execution (think of math or puzzles) while planning has an execution aspect to it. In the same view, decision-making, the selection between courses of action or solutions, seems to be part of this mix.



There are some capabilities that lay in gray areas but are clearly part of human inferential skills and need to be part of machine intelligence as well: language understanding. At some level, we could consider language understanding as part of general inferential skills, but it is so central to the core intelligence capability of communication that it seems that we should attend to that status. Similarly, language generation, in the sense of communication driven by content, could be seen through the lens of planning, but its central role in interaction argues that it should define itself as a separate element.



In doing this, we should also attend to the difference between the processing of text that results in the identification and extraction of entities, or even sentiment, and the understanding of an utterance. The former results in a list of people, places and things while the latter results in a characterization of how they relate to each other.

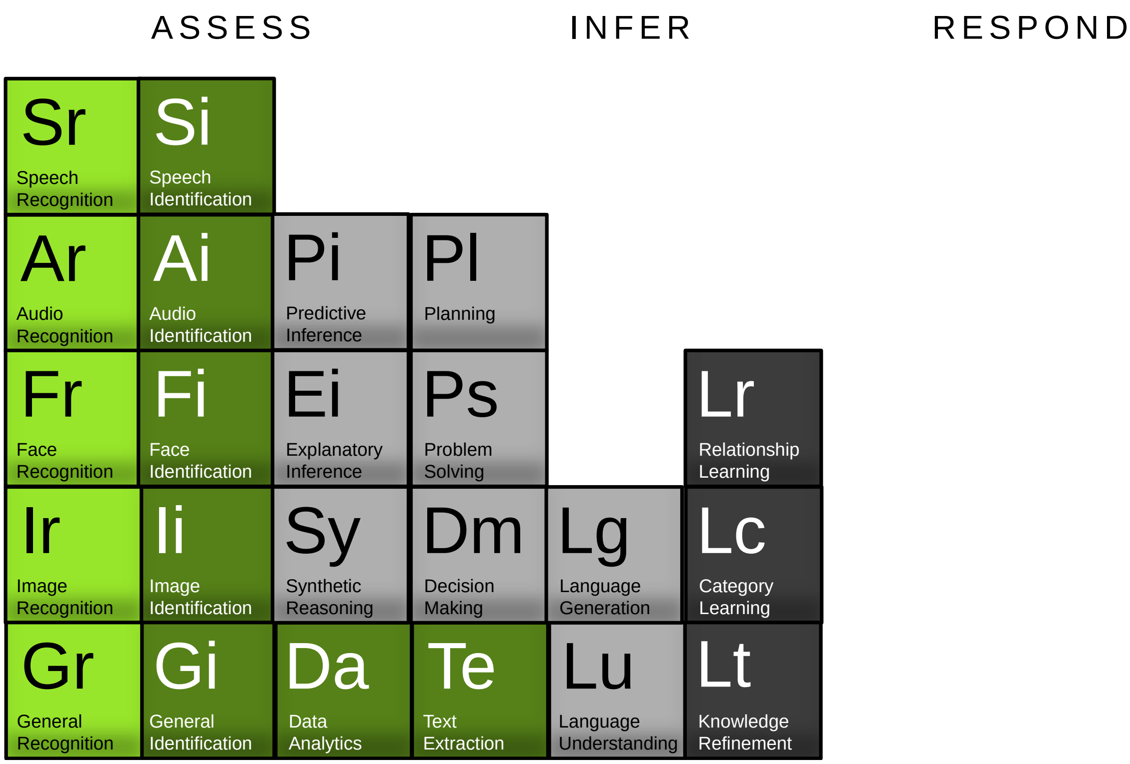
The final element in this category is probably the most visible and is at the core of the latest wave of Artificial Intelligence research and development: learning. Fundamentally, learning is an inference process. Regardless of the mechanism, learning is the inference of long-term characterizations of the world from data. There are different dynamics based on the flow of the data and the target information type, but the nature of the behavior is always the same.

Again, stepping away from the mechanisms, we can see three basic classes of learning: high-data-volume learning in the mode of Google's image clustering and categorization, learning focused on finding correlations within data sets and incremental learning of task refinements based on feedback during execution. There are further refinements, but this level provides us with a useful categorization.

Our functional mode under inference ends up comprising four overall types of elements: inferences related to states of the world past, present and future; inferences related to impact change through planning and decision-making; inferential capabilities related to language and the ability to infer structure from examples and data flow that supports learning.

There is a crucial point here. Although the language around these functional elements includes words such as inference, plan and knowledge, this is a reflection of how we tend to think of these elements from the outside. It is not a commitment to their method of implementation or what drives work associated with them. There are some researchers dedicated to the idea that functionalities such as prediction and explanation require explicit semantics representations and rule-based approaches. There are others who see the same processes as supported by activations, moving through networks in which a single fact only exists in the form of weights on links between nodes in that network.

I say this as a reminder that this framework is designed to be agnostic with regard to how things are implemented, and is aimed instead at what components and what elements seem to be necessary. That said, the components associated with inference can now be joined with those from assessment, bringing us closer to a complete set.



Response

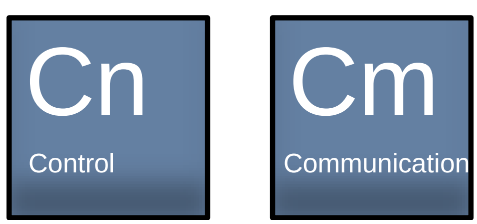
The final collection of functional elements falls under the category of response. These are functionalities that bring intelligence back into the world, real or virtual. Just as the elements of assessment draw information in from outside of the system, the elements of response provide mechanisms for impacting that world. As we step through the different response elements, it is important to recall that what we are defining are functional components but not necessarily how they interact with each other. There is no implied cycle from input to processing to output implied here. These are components that can be fit to various architectures. These are elements that can be combined to make a wide variety of compounds.

On the response side, there is a strong break between the physical and the virtual. While the underlying planning and problem solving may be shared between systems in these different realms, there are differences.

On the physical side, there are two major functionalities necessary to include: mobility and manipulation. Within mobility, there is the further breakdown into large- and small-scale devices. In essence, we can differentiate between devices that carry us and those that exist side by side with us, or, more colloquially, between cars and robots. Scale is less of an issue with manipulation, so we end up with three more elements.

The elements on the virtual side also scope across the physical, but have somewhat fewer constraints in that engineering issues of the physical world simply do not apply. These final elements are control and communication.

The first of these, control, includes work in the area of intelligent systems that are making decisions that play out as interactions with or control of other systems. The result of their reasoning has external impact but only through the trading of a stock, the adjustment of a valve or the ordering of new supplies. There may be eventual physical impact, but not at the connection point of the system.



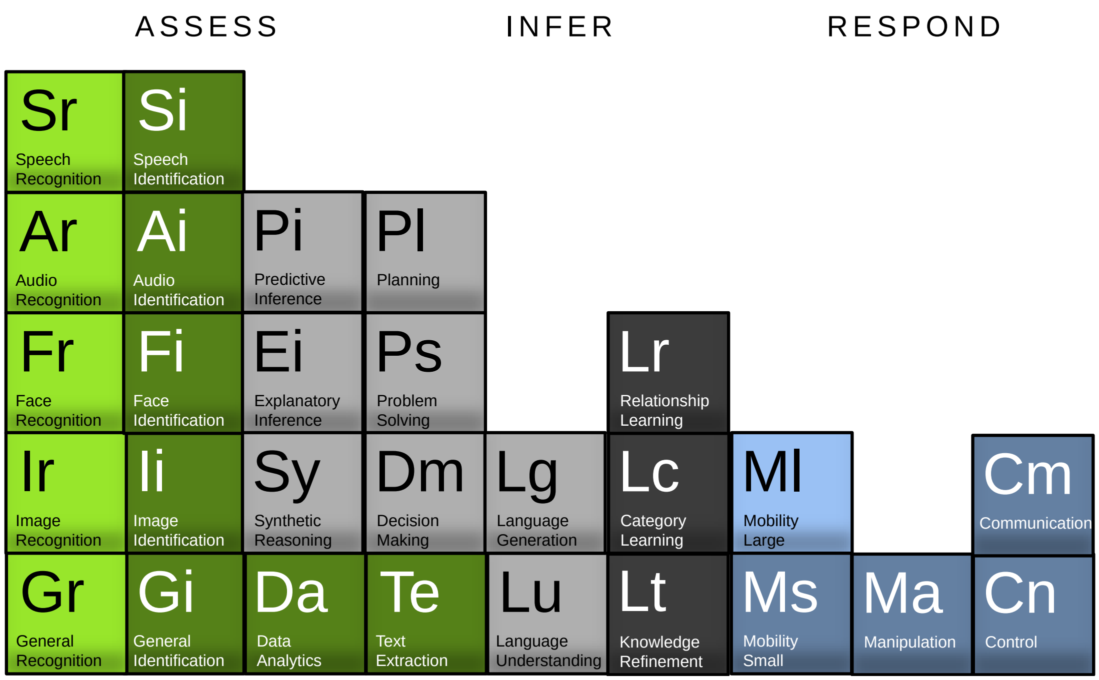
The second is communication. As discussed earlier, the planning and reasoning behind language generation is one aspect of inference, but the last mile of communication is in the realm of response. It is not just connection with the world but more so to the humans that inhabit it.

The Table

The result of this exploration is a table of elements with each element corresponding to a functional part of a complete AI system. Now that we have it, we need to consider two things: how to use it and how to expend it.

There are three main uses for this table: categorization, evaluation and development.

As we discussed earlier, there are many things that get in the way of our understanding of the different technologies including our own denial of capabilities, misattribution and the occasional over marketing of technologies. This organized table of elements provides us with at least a first pass at a language for asking questions about the capabilities of the different technologies that we encounter. It gives us a set of questions to uncover what different technologies actually do. The elements themselves can drive questions about how different systems assess, infer and respond to the world and allow us to identify not only capabilities but also gaps.



This deconstruction into elements also gives us a way to organize how we think about evaluating AI systems. By teasing out the different functionalities and identifying what various systems do or don’t do, we can do a better job of evaluating them. Seeing that a system is focused entirely on evidence-based or synthetic reasoning and has no assessment capabilities, allows us to evaluate it as such. Likewise, the breakdown of functionality that the elements provide gives us a framework for comparing systems at a component level rather than on monolithic devices.

Finally, it provides developers and researchers with a framework for organizing their work and considering what they need to build independently and what they can adopt and apply from others. A team working on managing the reasoning associated with a communication layer can explicitly decide to not focus on speech recognition or even generation. Likewise, someone developing a language understanding system need not worry about functionalities associated with language generation. In essence, it provides a framework for organizing and segmenting efforts.

On the expansion side, there is nothing but opportunity. This is the first step in building out the functional breakdown of AI into a set of building blocks that we can use to drive thought and discussion. In being a first step, it also includes an invitation for others to expand and extend this model as needed.

The goal is to work towards a way to view and evaluate AI systems that allows us to compare different systems along the lines of the functionality they were designed to provide. To give us all a framework for considering the value of different pieces of work without worrying about those elements that they may not include and to be able to ask the question of how these elements might work together. The goal is to collaborate.

This is the beginning of the conversation, not the end.