Introduction

The science is really clear. Millions of tonnes of plastics polluting just the Atlantic, soil toxicity, air pollution, climate change, pandemics, mass depletion of natural resources, peak antibiotics. Humans have been so successful we've almost used up the gifts of nature.

Wildfires, floods, droughts, cyclones, tornadoes, polar melting, ocean temperature change, species' extinction, all that and more. The cataclysm has already begun. Yet our short-term self interest and innate tribalism counteract collective action. It seems we're simple not emotionally equipped with a drive to put the future first. Hence, we urgently need machines that will. But to act in our place, they'll need artificial general intelligence (AGI).

Happily, the AI project to create AGI began soon after WWII using an electronic machine developed during the conflict. Yet in the 70 years since, despite success in limited domains, the general case has proved extremely challenging.

Quantum physicist David Deutsch concludes the project lacks the concepts or conceptual consistency needed to understand generality:

> ...today in 2012, no one is any better at programming an AGI than [project founder] Turing himself would have been ... The lack of progress in AGI is due to a severe log jam of misconceptions. I cannot think of any other significant field of knowledge where the prevailing wisdom, not only in society at large but among experts, is so beset with entrenched, overlapping, fundamental errors.

According to historian of science, Thomas S. Kuhn (1962), science progresses via changes of conceptual framework:

> 'normal science' presupposes a conceptual and instrumental framework or paradigm accepted by an entire scientific community ... [T]he resulting mode of scientific practice inevitably invokes 'crises' which cannot be resolved within this framework; and that science returns to normal only when the community accepts a new conceptual structure...

The AI project seems to easily fit within Kuhn's definition of normal science, and its failure to make significant progress towards its original goal of generality is clearly a crisis.

Yet one explanation of AI's failure to achieve generality is that its concepts are adequate but the needed machine configuration hasn't yet been found. Another is that key concepts are faulty or missing. Alternatively, AI's machine might simply lack the parts or basic operations needed for a mind.

But we know machine intelligence is possible. We are organic machines. And there are reasons to think AI's electronic device does have suitable parts and simple operations. So why, after 70 years' effort at the best universities are initial AGI milestones yet to be reached? It seems AI lacks the needed concepts to understand how to realize generality in its machine.

At this point one should probably say something about generality. General intelligence includes idiomatic conversation, common-sense knowledge, and the avoidance of combinatorial explosion. Combinatorial explosion happens while running through consequences of potential courses of action in response to a current situation in a hostile, complex and changing world. The quantity of possibilities soon exponentially explodes, rapidly consuming processing power and increasing processing time. The AI system fails to respond quickly, and "dies".

For instance, today, "narrow AI" systems control self-driving vehicles. These systems detect stationary objects forward of the vehicle. However, the processing time needed to determine whether the objects are

benign (such as parked cars, trees and lamp posts outside of the lane; and overhead bridges, walkways and signs) is so great that the processing needed for moving objects such as oncoming cars is severely compromised. The engineers turn off processing for stationary objects. Self-driving vehicles at full highway speeds hit stationary objects, such as police cars and fire trucks attending roadside accidents.

Yet identifying benign verses dangerous objects is something humans do almost immediately and without, it seems, even thinking. This stark difference can be put down to serial verses parallel, but is that correct? The machine might be capable of quick and accurate assessment, but AI might be using inadequate concepts to understand the machine.

AI might reply that its understanding is science. And that as scientific fact, it's not in dispute. Yet is it scientific fact? Today, two different conceptual frameworks are used to understand the machine: AI's programming-level conception for sequencing its simple operations, and the scientific framework of the basic chemistry and physics of the electronics.

AI's framework rests on concepts of the inner symbol and computational symbol manipulation. The scientific framework contains concepts of chemical doping, gallium arsenide, transistors, electrical bias, diodes, electrons, copper, semiconductor switch states, electromagnetic radiation, magnetic orientations of iron oxide, quantization, rare earth metals, quartz crystals, and many more.

So in fact AI's understanding is not science. It's a framework developed during and shortly after WWII for easily explaining human use of the device (that use being programming it then using it once programmed). It's a framework of the inner symbol and its computational manipulation.

This conception might be good for explaining human use of the device, but is it suitable for knowing how the machine could intrinsically acquire general intelligence?

Well, yes it would be suitable, if in fact the device manipulated inner symbols. But the only place symbols exist is on exposed surfaces of attachments: keyboard, screen, printed paper (besides ones printed on components such as capacitors and circuit boards). Put there for only one reason – so humans can see them. The central unit by its design is incapable of manipulating inner symbols, because by its design it is incapable of *having* inner symbols. When AI talks of manipulating symbols it means symbols inside the machine, by virtue of which manipulation the machine operates, or runs.

So while the symbolic understanding of the device is suitable for one purpose, is it suitable for knowing how the machine could intrinsically acquire genuine intelligence? Or is the fountain of all AI's problems the belief of the programer that they are dealing with a device that manipulates symbols?

The present text

It's hoped to try to answer some of these and also related questions. That is, to work out, or I suppose realistically speaking to try to make some contribution to working out, why, since its inception in the late 1940s, the Artificial Intelligence project has failed to make significant progress towards its original goal of artificial general intelligence. Early hopes were so high.

The present text focuses mainly on the machine. It came first. Then its inventor (said to have designed the machine in abstract form) promoted his belief that it could be configured to have a mind. That was the birth of the AI project. And the machine is AI's only platform. Thus, its conception of mind is essentially an expression of its understanding of the stored-program digital computer.

One problem with configuring the machine to have a mind is that the principles of general intelligence are unknown (one will be suggested). But we do understand the computer. But as noted, with two different conceptual frameworks: AI's programming-level conception for sequencing the simple operations, and the scientific framework of the basic chemistry and physics of the electronics.

An early well-argued philosophical attack on AI theory and practice, the 1965 paper, "Alchemy and Artificial Intelligence", was written by a philosopher at perhaps the center of AI research at the time, the Massachusetts Institute of Technology. Though his claim that AI is a modern alchemy was used as a strong criticism, I want to argue that some alchemical concepts can nevertheless make a positive contribution to solving some of AI's difficulties.

Basic alchemical concepts are radically different from those AI uses to understand its electronic equipment. Why not try to explain the computer using basic concepts of the alchemical framework? The goal would be to force AI to try to explain why the alchemical understanding is wrong and why AI's very different understanding is right, and through this, reveal a more fundamental understanding of the machine. Perhaps a way more suited to realizing generality.

AI is founded on the ideas of the inner symbol and computational symbol manipulation. Theoretical alchemical concepts include those of spiritually endowed metals and the ratio of the quantities of component elements as determinants of properties. A key practical concept of alchemy is repetition.

The Alchemists' Guide to AI tries to explains AI theory and practice to a virtual Medieval adept using concepts of alchemical theory and practice. AI theory includes ideas about why a mind could be realized in a computer. AI practice means its understanding and use of what it calls the computer.

Foundational errors are found. A replacement programming-level framework is developed for AI. This is based on a proposed principle about how sensors, causal interfaces from outer to inner, create knowledge; and on the form this knowledge takes in the data streams transmitted from sensors to the central system.

It seems that existing concepts can't explain how instances of the form contain semantic content, a necessity of knowledge. New concepts are developed to explain this, for example the idea of temporal structure. It's argued that these new concepts explain how to realize the principle in AI's machine, and how the resulting structure can have intentionality in Searle's sense.

As noted, key to AI's framework are the concepts of the symbol and computational symbol manipulation. The proposed framework contains neither AI's concept of the symbol nor AI's concept of computation. It also lacks the concept of information.

This represents a radical departure from the idea of the computer as a symbol manipulating device or as an information processing system. It offers a very different programming-level conception, one derived from the proposed principle of the creation of and then embodiment of knowledge in transit. But also one consistent with the science.

Finally, I argue that the new framework, developed with help of certain alchemical ideas along with the proposed principle, can be used to overcome the most well-known theoretical attack on AI, American philosopher John Searle's 1980 Chinese room argument, and various other objections including the problem of design (1843), problem of machine translation (1959), problem of common-sense knowledge (1965), frame problem (1969), problem of combinatorial explosion (1973), symbol grounding problem (1990), the related problem of encodingism (1995), and the problem of knowledge quantity (1996).

Chapter 1 is an outline. Subsequent chapters progressively add detail. The springboard is the proposed principle that states the form knowledge takes in the streams emitted by senors.

Explaining the form has been very challenging. Stating the principle is easy enough. But how could it possibly be right? It's an outlier. It doesn't mesh with existing conceptual frameworks. Yet no other principle seems viable. What else exists in a stream other than instances of the relation of temporal contiguity?

After many unsuccessful attempts to explain the principle using existing frameworks, I thought new concepts must be needed (and the new concepts in their turn would need to be explained and justified). It seemed a daunting task, but I imagined adequate concepts could be developed.

With independent income I had spare time. I had such a great love of research, and decided to try to develop the concepts. In any case, realizing AGI is important. That was in the mid-1990s. Now under covid-19 I've had the freedom of solitude to consolidate and reflect on the research.

Parts

The present text:

- tries to understand AI theory and practice using the conceptual framework of alchemical theory and practice, and through this, force AI to reveal and justify its foundations;
- assesses a proposed principle of general intelligence in the light of what was learned in the first step just above, then develops a new conceptual framework for understanding the computer, so called, and how the principle might be realized inside it;
- uses the proposed framework to addresses the most well-known attack on AI theory and practice, American philosopher John Searle's Chinese room argument, and various others.

The alchemical framework

The metallurgical theme of Western Medieval alchemy is chosen because of the relatively plentiful surviving material indicating its concepts. Alchemy itself is selected because of the procedural similarities between the Hermetic art and the AI project.

Both strongly seek to realize theory in equipment by following recipes (in AI's case, called programs). Both projects enthusiastically endorse, believe and follow their theory, but it fails to predict the successful recipe. As a result, both projects include extensive laboratory testing. And in both cases, failure of tests is easily explained as failure to discover the right recipe, and the theory itself is rarely challenged.

Principle of generality

It's assumed senses create knowledge. And that knowledge is a necessary element of general intelligence. The senses then transmit their knowledge in streams to the central system where it is extracted and stored. But the mode of embodiment in the streams (and in storage) is unknown.

A mode is suggested. A principle of perception is proposed that says all knowledge gained through sense experience is reducible to instances of the relation of temporal contiguity.

This relationship is examined then concepts developed to explain how instances of it can contain knowledge and how a computer can extracted then store it. Simple extraction algorithms in Intel assembler and C are given, then the resulting storage structure illustrated and explained.

Objections to computationalism

American philosopher John Searle accepts AI's symbolic framework, explaining that linguistic symbols have shapes: are formal, or syntactic, objects. That AI's machine by its fundamental design reacts only to that form: is a purely syntactic device. That the meanings of the shapes, the semantics, don't arrive with the symbol. That all the computer gets is the symbol. And hence the machine is forever a prisoner in a universe of mere shape. It could never understand the world including language.

I reason that Searle's adoption of AI's conception of the computer as a symbol-manipulating device is a fundamental mistake. He rightly rejects AI's computational theory of mind. He should have also rejected AI's understanding of its machine.

If this is true, it seems great news for AI. It means AI can accept the main conclusion of the Chinese room argument but keep its machine – because the argument isn't about AI's machine. Rather, it's about some different imaginary one.

Removing AI's symbolic understanding of the computer, Searle's argument produces a very different conclusion. It doesn't show that computers could never understand the meanings of the shapes of the things they receive and internally manipulate. Rather, it concludes that the conceptual framework AI uses to understand the inner machine is fundamentally wrong.

And the argument, so construed, indicates the locus of error: AI needs to abandon the concept of the inner symbol (and hence of inner computation). Searle should have abandoned Turing completely, not just his computational explanation of the mind but also his computational understanding of the electronic device he called a computer.

Thus, it seems progress towards AGI presupposes abandoning Turing. This seems a very unhappy prescription. But what else could be the starting point of a paradigm shift? So is abandonment really unfortunate? After 70 years bereft of demonstrable progress? The absence of progress – that's what's unfortunate, given the increasing environmental tragedies.

Kuhn says "*'normal science' presupposes a conceptual and instrumental framework or paradigm*". In a sense, via the appropriate paradigm shift, AI would be fulfilling its destiny. It would need to jettison its conceptual framework, but it would keep its current instrumentation. Alchemy's theoretical foundations were abandoned, but its equipment founded the science of chemistry. Perhaps to signal the AI project's change to a new paradigm (whatever that new framework might be) and also to honor the birthplace of the computer (so called), the project could rename itself with Donald Michie's original British term, Machine Intelligence. After all, what sort of intelligence would want to be called artificial?

The AI founder's inspirational 1950 paper explaining, justifying and promoting the computational framework concludes, "*We can only see a short distance ahead, but we can see plenty there that needs to be done*". For 70 years, very much work has been done. Yet with no demonstrable progress to AGI. Instead: the identification of serious theoretical difficulties. Rather than reaching demonstrable incremental milestones, fundamental problems have been uncovered. The opposite of early scientific expectation.

Regarding the design problem, machine translation problem, frame problem, problem of common-sense knowledge, symbol grounding problem, problem of encodingism, problem of knowledge quantity, and the problem of combinatorial explosion, I try to show that these, like the Chinese room argument, are problems not with the machine *per se*, but rather with the symbolic computational framework used to explain it. They don't arise (I argue) with the proposed new framework. Though any new conception might bring its own problems. But that possibility shouldn't forestall examining new frameworks.

