Introduction

The Alchemists' Guide to AI seeks to understand the computer with radically different concepts from those used by AI. Through attempts to reconcile the two understandings, insights are gained into recognized problems with AI's computational conception of the device.

Next, aided by basic alchemical concepts, a principle of general intelligence is proposed and examined. Respective algorithms and data structures are detailed. Finally, it's argued that the new understanding of the machine developed with help from basic alchemical concepts reveals that the most well-known attack on AI, the Chinese room argument, is both unsound and crucially important to appreciating how to realize general intelligence in a computer.

Chapter 1 is an outline. Subsequent chapters progressively add detail.

The motivation for writing *The Guide* is the strong belief that AI's biggest challenge is understanding the principles of general intelligence, but a fresh perspective is needed. One way to achieve a fresh perspective is to seek to understand AI's subject matter with radically difference concepts and through these gain insights into AI's foundations.

General intelligence includes idiomatic conversation, common-sense knowledge and avoidance of combinatorial explosion[1]. It was the project's original goal. In the 70 years since, though achieving strong success in narrow domains, the general case has proved extremely challenging. The project now divides into narrow and general (AGI), most resources directed into narrow.

Noted theoretical physicist David Deutsch concludes AI lacks the concepts or conceptual consistency needed to understand general intelligence[2]. The present text accepts this as a starting position. Hence the initial discussion of the problem of adequately understanding general intelligence is largely an analysis of concepts.

Thomas S. Kuhn (1962) says science progresses through 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...

1 Combinatorial explosion, first highlighted by Sir James Lighthill in his 1973 report to the British Government on AI, and sometimes considered the core component of the frame problem, is the exponentially increasing number of possibilities encountered when evaluating potential courses of action in a complex world. The recent "deep learning" statistical method, which is the old artificial neural network method used with internally faster machines with more internal memory, has achieved commercial success in narrow domains such as face recognition and vehicle control systems, but fails generality because of its necessarily narrow application and fragility both in "edge" cases and under adversarial attack. An edge case is a case that is unusual. Adversarial attack makes small changes that result in radical misidentification by deep learning networks, such as identifying a turtle as a rifle, a cat as guacamole, or a stop sign as not a stop sign.

2 "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. Yet it has been one of the most self-confident fields in prophesying that it will soon achieve the ultimate breakthrough. In 1950, Alan Turing expected that by the year 2000, 'one will be able to speak of machines thinking without expecting to be contradicted' ... yet today in 2012, no one is any better at programming an AGI [a computer to have human-like general intelligence] that Turing himself would have been. ... The lack of progress in AGI is due to a severe log jam of misconceptions."

The Alchemists' Guide to AI is an attempt to identify then apply principles of general intelligence through a three-step process:

(a) Try to understand AI theory and practice using a very different conceptual framework from that of AI. And through this, force AI to reveal and justify its foundations.

(b) Assess a proposed principle of general intelligence in the light of what was learned in the first step, and develop a new conceptual framework for understanding the computer and how the principle might be realized in it.

(c) Address the strongest attack on AI theory and practice, American philosopher John Searle's Chinese room argument, and explain firstly, why it is unsound, secondly, why it so important to AI, and thirdly, how a computer could be an intentional device.

(a) The first step adopts the framework of the metallurgical theme of Western Medieval alchemy. Western Medieval alchemy, because of the relatively plentiful surviving material explaining its concepts. Alchemy itself, because of the procedural similarities between the Hermetic art and AI.

Both are strong research projects that seek to realize theory in equipment by following recipes. Both theories fail to predict the successful recipe, or program. This results, in both cases, in extensive testing. Both also provide an easy explanation of experimental failure as due to not yet testing the right recipe. Reasonably, in both cases theory need not be challenged.

Both theories are not scientifically testable in the sense the theories of phlogiston, the speed of light, the blood as the food of the body, and relativity are testable. Nevertheless, though not scientifically testable in the usual sense, both theories are compelling and inspiring, and to the research communities of the time seemed obviously correct.

AI's theory of mind comes from its understanding of its machine. In the present text, an attempt is made to understand AI practice, its machine, with alchemical concepts in the hope of forcing AI to reveal its foundations. But not as AI explains them using its own computational framework, but rather using alien and more fundamental concepts of metallurgical alchemy.

The first substantive critique of AI after American information scientist Mortimer Taube's 1961 book, *Computers and Common Sense: The Myth of Thinking Machines*, MIT philosopher Hubert L. Dreyfus' 95-page paper "Alchemy and Artificial Intelligence" uses the allegation AI is a modern alchemy as a deep criticism. We now say alchemical theory is drastically in error. But rudimentary alchemical concepts can, it's argued, shed a positive and valuable light on the foundations of the research project of Artificial Intelligence.

(b) The second step considers perception, the front-end of knowledge, and says understanding the principles of perception is essential to understanding the principles of knowledge, and understanding the principles of knowledge is essential to understanding the principles of general intelligence. Step (b) concludes that since all knowledge originates from senses, knowledge must be embodied in the streams of units of substance emitted by sensors and transmitted, or progressively propagated, to the central system, but that the mode of embodiment of this knowledge in these streams is unknown.

A mode of embodiment 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 discussed and explained and the principle critically examined. Then data structures and algorithms are outlined that may realize the principle in the electronic machine AI, using two concepts of its present conceptual framework, calls a digital computer.

(c) Step Three considers American philosopher John Searle's renowned Chinese room argument against AI. Searle argues that the conceptual framework AI uses to understand the computer contains an essential concept that contradicts what is known about the mind. That is, contradicts what is known about the mind as understood with the accepted conceptual framework of mental content.

Searle accepts the two frameworks: that used to understand mental content, and the one AI uses to understand the device it calls a computer. He assumes AI's understanding of the computer is accurate, or true to the hardware. And he assumes that today's relevant understanding of mental content is accurate and true to the mind.

Then he argues, very persuasively given his two assumptions, that the understanding of the computer and the understanding of the mind are fundamentally at odds. An essential concept used to explain the computer contradicts an essential concept used to explain the mind. The argument's conclusion: computers could never be or contain a mind. They could never think, they could never have general intelligence.

The Guide accepts Searle's understanding of the mind. But not of the machine. There are actually two conceptual frameworks currently available that may be used to understand the machine. The one AI uses, but also and independently, there is the scientific framework.

The scientific framework comprises concepts of the chemistry and sub-microscopic electronics of semiconductors. Thusly understood, semiconductors internally realize and manipulate binary difference. AI's framework, by contrast, is that of computation, the internal manipulation of symbol tokens according to rules about the meanings of their shapes. The inspired and inspiring British mathematician who founded the AI project used this computational framework to explain both the mind and the electronic device AI calls the computer.

A symbols is token whose shape has been assigned a meaning by an observer. A token is a particular instance. For example, the following: A , is a token, an instance, of the shape "A". The quotation marks highlight the shape of the object between the marks, as opposed to the meaning of the shape.

In order to properly argue against AI, Searle must adopt the second of these two frameworks, that of computation as used by AI. But that computational understanding is not true to the hardware. Not true to the science. The computational understanding is founded on the idea of the symbol, or more broadly, syntactic formality. But the scientific explanation lacks the concept of the symbol. AI's understanding has proved very useful, but what it actually explains is an early human use of the machine, the one that gave the device its name, not the essential machine itself.

The device contains parts with names that imply inner computation, parts such as the arithmetic-logic unit of the von Neumann and other architectures (such as the Harvard architecture). Many operating system or programming language functions can be considered computational, such as the trig functions (functions eked out of binary difference often with considerable difficulty).

But even though a part or function assists a certain use of a machine, it doesn't follow that the use describes the fundamental nature of the device. The science says the machine internally operates by realizing and reacting to inner binary difference. On this understanding, there is no inner reaction to symbols, tokens whose shapes have meanings assigned by observers. Rather, the computational understanding is used to explain a human use of the machine.

This is seen from the fact that, apart from symbols printed on inner components such as capacitors, symbols exist only as encrustations on exposed surfaces of machine attachments (keyboard, display, paper in a printer, ...). Put there for just one reason – so human users can see their shapes. In fact, by its design, the electronic machine is incapable of internally reacting to shape. It lacks the causal power. And in the sense symbols have meanings, the things the device does internally manipulate have no meanings.

The Chinese room argument adopts AI's understanding of the machine, as it must in order to oppose AI's position. But AI's understanding of the hardware internals is fundamentally wrong. And hence so is the Chinese room argument's understanding. In the argument, AI's errors take the form of false premises. And if a premiss is false then the argument is unsound. The Chinese room argument as presented by Searle is unsound.

But when correctly understood, the argument is extremely valuable to AI. What it really shows is not that computers will never think, but rather that the conceptual framework AI uses to understand the internals of the electronic machine, that of computation, is fundamentally in error. Further, the argument indicates where AI's conceptual errors originate: the concept of the linguistic symbol. Searle's argument shows that AI needs to abandon the concept of the inner symbol (and hence of inner computation) and use a more scientifically accurate framework.

But a framework at the programing level rather than semiconductor level, for AI needs to easily alter the causality of the device in its testing program of different recipes. However, to be able to give effect to the full potentiality of the machine, the programing-level understanding needs to be true to the science of the device. A suggested programming-level framework true to the basic electronics is outlined in the next chapter of the present text.

Of course, a computer, correctly understood, might still perform a computation in the sense humans do – by manipulating external symbols on sheets of paper.

One of AI's biggest mistakes is thinking the Turing machine is the fundamental design of the electronic computer. Turing machines do actually manipulate inner symbols. By believing Turing, who said the electronic machine is a practical realization of his Turing machine design, AI fundamentally misunderstands its machine.

A Turing machine can operate on binary difference. But the Turing machine is understood as the internalization of external symbol manipulation, of the human activity of computing with pen and paper. That's how AI's founder explains the Turing machine: "*we may now construct a machine to do the work of this [human] computer*". As a result, the electronic device is understood as the internalization of external manipulation of linguistic symbols. And that's how Searle understands it.

But the concept of the Turing machine as internalized human computation with external symbols has given AI a conceptual framework that prevents it understanding how binary electronics could become intelligent. As the adept's alchemical understanding of his equipment prevented him realizing the full scientific potential of the equipment, so AI's computational understanding of AI's equipment prevents AI realizing the full scientific potential of AI's equipment.

Searle rightly concludes that, on AI's computational understanding of its machine, the device could never be or contain a mind. But AI's program-level understanding is fundamentally inaccurate.

Deutsch: "*The lack of progress in AGI is due to a severe log jam of misconceptions*." For AGI, the initial 1950s clarity and hope, after 70 years of diligent and largely well-funded effort, has become uncertainty and confusion.

AI's founder in 1950 concluded, "We can only see a short distance ahead but we can see plenty there that needs to be done". In the seven decades since, plenty *has* been done. But with no demonstrable progress towards general intelligence. *The Guide* concludes that in those very early days of 1950, the mathematician, in squinting down the computational path, was wrong about what lay ahead. Intelligence is not computational. But happily, neither is the machine AI calls a computer.

