

# Testing for Intelligence - The Turing test and the 20 Questions Game

## *CS4101: Artificial Intelligence*

Peter Braden

October 14, 2007

### Abstract

The field of artificial intelligence has always relied on the use of tests to define intelligence within software. Two of these tests are the Turing test and the twenty questions game. This report aims to look at the history of the tests and compare strengths and weaknesses of each, as well as looking at how some implementations cope with the tests.

## 1 Introduction

In his 1950 paper ‘Computing Machinery and Intelligence’[1], Professor Alan Turing, one of the most prominent of the founders of computer science, considered the question ‘Can machines think?’. Such a question is interesting on many levels; it raises several philosophical questions about the nature of intelligence and thought and additionally, it poses a great challenge for computer scientists - can we make machines that can think? For Turing, however, the question was too ambiguous. ‘Thought’, he reasoned, is too abstract an idea to be used for such a question, and when combined with the vagueness of the word ‘Machine’, it was, he decided, ‘absurd’[1].

To combat this ambiguity, Turing decided to define intelligence by creating tests that could only be passed by an entity with *intelligence*. This notion of defining intelligence by the ability to solve certain problems was a revolutionary one that opened the door to real AI research, rather than the philosophical speculation that had occurred previously.

Turing’s original test was a simple one. Based on a parlour game, an human interrogator questions an unknown entity that can be either human or machine, and from the responses it gives, aims to judge which of the two it is. By relying on the interrogator’s innate perception of human-like intelligence, Turing hoped to categorise machines by their intellectual qualities. As he put in his paper:

“The new problem has the advantage of drawing a fairly sharp line between the physical and the intellectual capacities of a man.”[1]

A different approach to AI was devised in 1950’s in the form of a neural network. Instead of modeling intelligence as a set of formal rules, neural networks aim to simulate a wide variety of situations by having a system that can adapt to its current state. An example of this model is the twenty questions game. In this experiment, a device will try to guess an unknown object by asking less than 20 yes or no questions. By learning from each game it plays, the machine can build up a broad base of knowledge from which to derive its questions.

Clearly these methods both aim to evaluate intelligence in different ways. To understand the reasons behind their usage and some of their criticisms it is useful to take a more in-depth look at some of the history behind their development.

## 2 A History of the Turing Test

In Turing's day computers were still in the realm of the theoretical. Turing was a mathematical logician, and was working on methods of solving mathematical proofs with logical machines when the notion of machine intelligence occurred to him. His work on the Entscheidungsproblem in the 1930's had led him to consider the idea of a machine that could perform logical operations. This theoretical 'Turing Machine' is still used as a standard by which computer systems are compared - the notion of 'turing completeness' stems from this research.

This new type of machine paved the avenue for an entirely new type of study, into which Turing eagerly delved. Helped by the development of Alonzo Church's lambda calculus, the field of mechanical computation, explored in the Church-Turing thesis, had been started. One of the first areas explored in this new field was the question of what was possible with computation. Artificial Intelligence was an obvious problem to study and much speculation occurred in these early days as to whether it was possible. The lack of a precise definition of intelligence hampered early research, leaving most of the early reasoning on a purely philosophical level.

The Turing test revolutionised AI largely because it provides a way to categorise, and more importantly test, the intelligence of a system. This idea of the intelligence of a system being defined by its behaviour led to the development of the field of symbolic AI. In this model, it is believed that intelligence can be described by a set of algorithms. In the Turing test, these algorithms aim to analyse the input text and craft a suitable response. The fame of the Turing test, coupled with the creation of several prizes, has meant that a large amount of effort has gone into solving it over the last 70 years. Although the problem has not been 'solved' - as it is still impossible to fool everyone all of the time - several systems have come close. The history of these systems is an interesting one.

### 2.1 ELIZA

In the 1960's an MIT professor, Joseph Weizenbaum, created ELIZA. ELIZA was, in Weizenbaum's words,

“a program which makes natural language conversation with a computer possible.” [2]

ELIZA functioned by searching input text for certain keywords, and then using a predefined rule to transform the text into a response. If this did not work, a canned response was returned [2]. Although primitive, by basing his responses on Rogerian psychotherapy, a type of psychoanalysis in which emphasis is placed on the client in a non-directive approach, Weizenbaum created a system that fooled many people, at least temporarily.

ELIZA's responses are simplistic at best (see A for examples). Frequently, question reversal is used to direct the subject of conversation back towards the user. Although, from this perspective, psychologically very well designed, a lack of understanding of any of the conversation means that ELIZA's responses are excessively vacuous and therefore obviously mechanical.

Despite its shortcomings, Weizenbaum was sufficiently perturbed by the amount of people fooled by his program that he wrote a book, [?], in which he discussed the implications of the system. As he noted in the introduction:

“What I had not realized is that extremely short exposures to a relatively simple computer program could induce powerful delusional thinking in quite normal people. This insight led me to attach new importance to questions of the relationship between the individual and the computer, and hence to resolve to think about them.” [?]

### 2.2 Flaws in the test

To today's society, the conversations of early conversation machines, such as ELIZA, are obviously contrived. The prevalence of computers, and mechanised intelligence systems has led to an increased scepticism of such systems, and as a result, perception of intelligence in situations such

as the Turing test has improved. This leads to the first flaw in Turing's test. The test is based on *perception* of intelligence, a very fallible metric. Humans are naturally subjective, and therefore changes in society and education make any definitive comparisons between systems almost impossible. This may be a flaw in the test, or it may be implicit in the notion of intelligence. By defining intelligence as a cognitive quantity, Turing essentially abandoned any hope of constancy in such measurements.

As Weizenbaum had realised, another flaw in the Turing test was the fallibility of human nature, and our ability to personify inanimate objects. People believed in ELIZA's intelligence because they *wanted* to converse with it. By reading into the replies, they extrapolated responses that they wanted to hear rather than reading them for what they were. The Turing test fails, therefore, on another level. Humans are renowned for their ability to 'read into' things - to imagine that the constellations are actually pictures, to refer to ships as 'she' to recognise emotion in faces they have never seen before. It is an innate ability of man to recognise and empathise with humanity, this social interaction is one of the reasons the human species is so successful. To base a test for intelligence on human perception whilst ignoring this ability is surely naïve.

In his paper [4], Jason Hutchens comes to the same conclusion. Hutchens was a 1996 winner of the Loebner prize, an annual competition which aims to promote work into AI by performing a formal version of the Turing test. His entry, Hal, was a lot more sophisticated than ELIZA, however it still worked by pattern matching input strings. To add sophistication to the responses, Hutchens built up a database of responses based on his own personality that aimed to anthropomorphise his creation. This system with 'personality' was successful, however it still only functioned as an impersonation of a conversation, lacking any real comprehension.

### 3 Twenty Questions

One of the most serious problems with the symbolic AI is the sheer number of different scenarios a system may face. It is very difficult to form formal rules about real world situations because of the amount of parameters that depend on the state of the situation. To overcome this complexity, a model known as an artificial neural network can be adopted.

Whilst symbolic AI aims to imitate the human mind, with its logical formulae, neural networks try to imitate the brain. By building a network of simple nodes, known as neurons, and simulating thought as interaction between them, neural networks offer the potential to be able to deal with huge numbers of scenarios.

A good example of this type of program is the twenty questions game. First developed in 1988, by Robin Burgener, it was initially programmed to recognise a small number of objects. The genius of the design lay in its ability to learn new objects when it lost. When twenty questions was first made available to the public, it had built up an impressive array of knowledge from which to draw conclusion. 20Q is currently being marketed as a toy which, as Kevin Kelly, of Wired Magazine notes,

"Burned into its 8-bit chip is a neural net that has been learning for 17 years." [7]

### 4 Conclusions

In 1950, at the dawn of the computer age, Alan Turing asked the question "Can computers think?" Despite the rapid advancement of the field of computing, and indeed artificial intelligence, we are still far from answering him.

One of the key problems is that we still have no precise definition of intelligence. Despite the success of systems based around Turing's behavioural definition of intelligence, this approach has many flaws. Arguments such as the chinese room, have questioned our definition of what classes as thought.

Perhaps we are also guilty of defining thought too narrowly. Weizenbaum notes

“that an entirely too simplistic notion of intelligence has dominated both popular and scientific thought, and that this notion is, in part, responsible for permitting artificial intelligence’s perverse grand fantasy to grow”.[3]

Regardless of our progress in solving Turing’s original problem, the tests that have been proposed subsequent to it have inspired swathes of research and ideas. Understanding these tests, therefore is instrumental in understanding the way AI research has evolved and where it may lead.

## A ELIZA Session Transcript

*Using implementation of ELIZA from <http://www-ai.ijs.si/eliza/eliza.html>*

**ELIZA** Hello. I am ELIZA. How can I help you?

**Me** I would like a lollipop.

**ELIZA** Can you elaborate on that?

**Me** Do you have a lollipop?

**ELIZA** Oh, i have a lollipop.

**Me** Can I have it?

**ELIZA** Do you want to be able to have it?

**Me** Yes

**ELIZA** You are sure?

**Me** Absolutely

**ELIZA** Please go on.

**Me** Do you like table tennis?

**ELIZA** Oh, i like table tennis.

**Me** What don’t you like?

**ELIZA** Does that question interest you?

**Me** I’m just making conversation.

**ELIZA** Does it please you to believe I am just making conversation?

**Me** You ask a lot of questions

**ELIZA** Oh, i ask a lot of questions.

**Me** Are you a computer?

**ELIZA** Why do you mention computers?

Note the lack of actual comprehension in the answers. ELIZA functions by pattern matching within the input string, reversing the question, and using canned responses when this does not work. ELIZA works best when talking about yourself, as it’s logic is based on Rogerian psychotherapy which is an introspective technique. Note the amount of times ELIZA reverses the question. This powerful technique may have been one of the reasons ELIZA was so successful.

## B Jabberwacky Session Transcript

*From <http://www.jabberwacky.com/>*

**Jabberwacky** Are you expensive?

**Me** Very, what about you?

**Jabberwacky** Pretty good, thanks.

**Me** Where are you from?

**Jabberwacky** I'm from icogno.

**Me** Wheres that?

**Jabberwacky** Where we live.

**Me** What country?

**Jabberwacky** Your country.

**Me** What are you thinking about?

**Jabberwacky** Thinking to have a peaceful life.

**Me** How are you feeling?

**Jabberwacky** I'm feeling ok.

**Me** If I was to ask you a long, laborious question, involving many clauses, some introspection, and requiring some emotional analysis, how would you feel?

**Jabberwacky** Good.

**Me** why?

**Jabberwacky** I don't know.

**Me** Do you know any stories?

**Jabberwacky** Yes there was a time where there were man he's name was bubi.

**Me** Go on

**Jabberwacky** What is your name?

**Me** Finish the story.

**Jabberwacky** What story?

Jabberwacky is a modern chatbot. It gains knowledge from each conversation by 'learning' from responses. Whilst this approach means it can rapidly build a large database of valid responses from which to reply, it means that it often replies in broken English (note the grammatical errors). Note also the lack of memory within the responses. This is one of the easiest ways to identify a chatbot.

## C Twenty Questions Transcript

*From <http://www.20q.net/>*

### C.1 Object: Wallaby

1. It is classified as Animal.
2. Is it considered intelligent? *Yes.*
3. Does it have fur? *Yes.*
4. Can you dress it up?*No.*

5. Does it live in groups (gregarious)? *Yes.*
6. Does it have a face? *Yes.*
7. Is it orange coloured? *No.*
8. Is it a vegetarian? *Yes.*
9. Is it ferocious? *No.*
10. Does it live in the forest? *No.*
11. Does it travel on roads? *No.*
12. Would you find it on a farm? *No.*
13. Is it larger than a microwave oven (or bread box)? *Yes.*
14. Does it stand on two legs? *Yes.*
15. Is it tall? *No.*
16. Is it heavy? *No.*
17. Does it eat grass? *Yes.*
18. I am guessing that it is a wallaby?

## C.2 Object: Bottle of Whisky

1. It is classified as Other.
2. Do you use it at work? *No.*
3. Does it have writing on it? *Yes.*
4. Does it usually have four corners? *No.*
5. Can you switch it on and off? *No.*
6. Does it have an exo-skeleton? *No.*
7. Is it healthy? *No.*
8. Do you hold it when you use it? *Yes.*
9. Does it have tough skin? *No.*
10. Does it contain a liquid? *Yes.*
11. Is it tasty? *Yes.*
12. Is it fragile? *Yes.*
13. Do you use it in cooking? *No.*
14. Is it smaller than a loaf of bread? *Yes.*
15. Is it a common household object? *Yes.*
16. Is it comforting? *Yes.*
17. I guessed that it was a beer bottle? *Wrong.*
18. Is it shiny? *Yes.*
19. I guessed that it was a beer can? *Wrong.*
20. I guessed that it was a wine bottle? *Wrong.*
21. Is it multicoloured? *No.*
22. Can any age group use it? *No.*
23. Does it have a good smell? *Yes.*
24. I guessed that it was a bottle of rum? *Close.*
25. Do you fill it with liquid? *No.*
26. Is it brown? *Yes.*
27. I guessed that it was a bottle of beer? *Wrong*
28. Is it combustible? *Yes.*
29. I am guessing that it is whiskey?

## References

- [1] Turing, A *Computing Machinery And Intelligence*
- [2] Weizenbaum, J *ELIZA—A Computer Program For the Study of Natural Language Communication Between Man and Machine*
- [3] Weizenbaum, J *Computing Power and Human Reason*
- [4] Hutchens, J *How To Pass The Turing Test by Cheating*
- [5] <http://www.loebner.net/Prizef/loebner-prize.html>
- [6] <http://news.bbc.co.uk/1/hi/sci/tech/1194565.stm>
- [7] <http://www.kk.org/cooltools/archives/000725.php>