How far along is the Artificial Intelligence domain in creating their ultimate goal?

Building an intelligent robot that matches/exceeds human intelligence
(examples of Siri, Watson, pokerbots and robots in elderly homes)

Frank Mulkens, ANR 384287

Department of Information Management
Tilburg University
Tilburg, 5000 LE, Netherlands

ABSTRACT
The goal of Artificial Intelligence (AI) is to create an intelligent robot or computer program that matches or exceeds the level of intelligence of humans. After 25 years of research (1950-1975) it became clear that this could not be realized in the near future. As a consequence, AI fragmented in multiple disciplines.
In this research three components (question and answer, voice recognition and learning AI) are selected to define how far AI is developed. By conducting a literature review, this study found what is state of the art in each particular subfield, how it works and which future developments are needed to create a robot with human intelligence.

Keywords: Artificial Intelligence, voice recognition, learning, question and answer.
1. INTRODUCTION

According to Russell & Norvig (2003) Artificial Intelligence (AI) strives to build ‘intelligent’ entities and understand them. The term intelligent is in quotes, because not every scientist agrees that computers have intelligence. Their argument is: humans wrote their code and/or computers don’t have a mind or understanding. These systems only simulate intelligence, called weak AI. The opposite, and focus of this research, is strong AI. Strong AI is intended to produce machines with an intelligence level that matches (or exceeds) that of human beings (Brooks 1991).

From the beginning of AI development (1950s) and in the next 25 years scientists tried to reach the goal of building a system with a human-like intelligence level, but only accomplished simulating isolated aspects of intelligence. From that moment the domain of AI fragmented in different subfields. Examples range from general areas like perception and neural networks to videogames and customer support. AI is diverse and it is hard to differentiate between all the fields where it can be used. Nowadays scientists concentrate on the potential commercial aspects of intelligent assistants.

In this research three components are identified which are important to create an intelligent robot or computer system. These aspects are also mentioned by Negnevitsky (2005), Bishop (2006) and O’Leary (1998). The components also contain a specific sub-question:

Component 1: AI contains knowledge
With this knowledge a program or system can answer a question. Before it can answer a question its database needs structured and labeled content. However this content usually unstructured, so the challenge is to structure the content. This is done in the natural language processing field. A recent project, Watson, is discussed in which natural language processing is important. In this section the following sub-question is answered:

*What is nowadays the smartest program that can answer a broad range of questions?*

Component 2: AI eases the way people live
One way to improve multiple products is to change how to use it. In recent years the use of touchscreens in the telecom industry changed the way we handle a mobile phone. Presently scientists have developed voice recognition to a reasonable level so that it can be used in commercial products. So how we use mobile phones might change again when they can be controlled trough voice recognition. Siri is one of the few voice recognition systems that can be used on Apple’s IPhone. This is the corresponding sub-question that will be examined:

*Which commercial system is using only voice recognition as input/output to make their system user-friendlier?*

Component 3: Learning AI
AI can learn by itself, not by entering more and complex rules by a programmer, but by making use of statistical data. By giving more data to the system the algorithm gets better each time. With a lot of data or large pruned trees AI systems can beat the human championship level in a couple of games like chess and checkers. As a result of the next question pokerbots will be examined. In this section the following question will be answered:

*How and in what game, that isn’t fully dominated by programs, can AI learn?*
This paper will show how far AI is developed in these three different subfields of AI. These subfields are chosen because of two reasons:

1. It is too complicated to create an intelligent robot that matches the human intelligence level. Instead there are all kinds of small subfields where intelligence is created in isolated experiments.
2. Each subfield touches one of the three important components for highly advanced AI.

The three examples (Watson, Siri and Pokerbots) are elaborated in detail. It is explained how the example works, what holds back the development and which future potential it contains. This is needed to combine these technologies in section 6. With all the information on how it works, which problems there are and future potentials of these three examples, a practical example is added. In this example serving robots in elderly care homes are examined. These robots make use of all the technologies in the three previous sections. With the information of the three sub-questions, the main question can be answered. The main question is:

*How far is AI with creating their ultimate goal?*
(Building an intelligent robot that matches/exceeds human intelligence)

This research is intended for people interested in Artificial Intelligence. It is fascinating to see that there could be intelligent robots that can learn by themselves, to ease the life of humans in the near future. So it is interesting to consider why these robots aren’t here yet and what are the problems that are encountered.

Section 2 starts with the methodology that is used to conduct this research. Section 3 elaborates on a question and answer system (Watson). Non-human contestant Watson can answer questions in a knowledge quiz and outperform the best human players. Section 4 discusses a computer-based personal assistant on Apple’s Iphone named Siri. Siri can help its user by performing multiple tasks on their phone. This program makes use of voice recognition. Section 5 is about AI in games. In 1997 computer machine ‘Deep Blue’ defeated world champion Garry Kasparov in chess. In multiple games the computer matches or exceeds human championship level. In this section poker bots are evaluated. These bots recognize a situation, select the best line to play, perform this action, and also learn from the past to adapt in the future. Section 6 is a practical example with serving robots. These robots make use of learning AI, voice recognition and natural language processing. The serving robots are active in an elderly care home. Section 7 contains the discussion and the last section is the conclusion.

2. METHODOLOGY

To answer the main question and sub-questions a literature review will be conducted. This type of research depends on existing literature. To find relevant literature, available sources of Tilburg University, complemented by other Internet-databases like Google Scholar and Scirus will be used.

No new data will be gathered, except for the security measures of poker bots. This is needed to determine how providers of online poker protect their players from pokerbots. This information will be gathered by contacting worlds largest poker client: Pokerstars.
3. **AI QUESTION AND ANSWER**

In the introduction there is a brief explanation about the term AI. In this section we take a closer look at AI and especially to the problems that are encountered when trying to build a program/machine with advanced AI. Therefore this section is about a question and answer system (Q&A system). The task of Q&A systems is to obtain appropriate answers, for a given domain, from independent questions, written in natural language, from a large document collection (Fukumoto 2001). This system is needed for every advanced and complex AI system as we can see in section 4, 5 and 6. However, as we can see in this section, it is also difficult to make a good Q&A system due to the natural language problem. In section 3.1 we will explain the problems of making to make a Q&A system and in 3.2 and project Watson is taken as a state of the art Q&A system. This system is more developed than other Q&A systems because it can answer difficult questions and can answer questions from multiple categories.

3.1 **Q&A difficulties**

The main problem for a Q&A system is the natural language processing (NLP) problem. According to (Warschauer 1998) NLP is the process of a computer extracting meaningful information from natural language input and/or producing natural language output. So a computer has to understand the meaning of a sequence of words, because a single word can have a totally different meaning if the context changes. Besides that, new words are frequently introduced and old words are consistently used with new meanings. So the context of a word is very important in NLP. This makes it difficult for the programmers to create a computer that ‘understands’ the context. To understand a sequence of words, multiple steps have to been taken in a general NLP system. This is shown on a high level in figure 1.

![Figure 1: A general NLP system (Kumar 2011)](image)

The input for an NLP system is usually given in natural language. This goes to a parser which generalizes a syntactic (1) structure of the content. Then a semantic interpreter captures its semantic (2) details and creates a deeper structure of the content. The conversation rules take this deep structure and changes it, then the database handler can store the data. This is the way to process natural language in a NLP system. The other way around, the generation of natural language, is also possible and ends with a generator which provides the output. On a lower level Kumar describes there are actually six categories of knowledge to process natural language, which happens in the parser/semantic interpreter phase. Beside syntax and semantic knowledge the following is also needed:
With knowledge of these six categories inside a NLP system there is a fair chance that the computer understands the words and the context.

In the next paragraph IBM’s Watson is taken as an example to examine a Q&A system. Watson is selected because of its state of the art abilities to understand the context and find an answer in his (offline) database. This system is more developed then other Q&A systems because it can answer more difficult questions and answer questions from multiple categories.

(1) Syntax: the composition and structure of sentences. It is needless to say that the positions of words create a whole new context. For example dog bit boy or boy bit dog.

(2) Semantic: the meaning of words.

### 3.2 Watson

Scientist of IBM believed that highly advanced Q&A systems should be able to understand complex questions and answer it precise, with confidence, and fast. These Q&A systems could support employees with critical and timely decisions in multiple business areas like business intelligence, healthcare and customer support. For almost three years 20 employees of IBM worked on a Q&A system (Ferruci et al. 2010). This QA system has been named ‘Watson’, after the 1st President of IBM (Watson 2003). Their challenge was to compete, real time, at the human championship level, in a quiz show (Jeopardy!). Ferruci and his team of IBM employees had to make an automatic contestant who could outperform the best two human players in the game Jeopardy!(1). In a special episode of Jeopardy!, Watson played versus Ken Jennings and Brad Rutter. Both human players won more than three million dollars on the show, but they couldn’t defeat Watson.

(1) Jeopardy! Is an American television quiz show. The show was first broadcasted in 1964. In this show a contestant needs knowledge of various topics like history, literature, art, sports etc. The show became popular through the unique answer-question format. Contestants are giving clues in the form of answers and must phrase their responses in a question form. The show was also popular because contestants could win high prizes. Ken Jennings won in total over 3 million dollars in 74 TV appearances in this show. He also played the special episode versus ‘Watson’.
3.3 Metrics

There are three critical metrics that are important for Watson to be successful in Jeopardy!. Three players are playing versus each other and the fastest candidate can answer first. When the fastest candidate doesn’t give the right answer he loses money, otherwise he gains money. So speed is important to be the first contestant to answer a question. Precision is a metric because it is important to answer correctly. The 3th metric is confidence, because there are some special events in the show were the player can wager their own points. If the player is really confident to know the answer he/she can wager more money and increase their chance to win the game. After a study of 2000 games of Jeopardy! the human championship level of this game was to answer roughly 70% of the questions with 80% precision in 3 seconds or less. So in the end Watson had to reach this level. To compare, Ken Jennings won 74 games, on average he answered 62% of the questions and answered with 92% precision.

3.4 How does Watson work?

In figure 2 is explained how Watson can answer a question. The text below will clarify the illustration.

![Figure 2: Explanation of Watson (Ferruci et al. 2010)](image)

3.4.1 Question Analysis

When Watson receives a question it has to process the question analysis. In this analysis it tries to understand what the question is and analyses how the question will be processed in the other stages. Besides different parses, logical forms and relationships between words, this step focuses on detecting LATs (lexical answer types). A LAT is a word in the clue that indicates a type of answer. The input for Watson looks like this:

**Category:** 1990s  
**Clue:** this woman had an affair with the US President in 1995  
**Answer:** Monica Lewinsky  
**LAT:** woman

Other examples of LATs, which had a high frequency, are ‘he/she’, country, city or film. With these LATs a boundary can be set on the candidate answer (possible answers to the question).
3.4.2 Query Decomposition
Sometimes a clue can be composed in multiple clues or sub clues. Watson has to recognize and decompose this. 3 examples are given:

**Category:** “Rap” Sheet
**Clue:** this archaic term for a mischievous or annoying child can also mean a rogue or scamp
**Subclue 1:** this archaic term for a mischievous or annoying child
**Subclue 2:** this term can also mean a rogue or scamp
**Answer:** rapscallion

Multiple answers can be given on subclue 1 and subclue 2. When Watson found answers to both clues it has to compare them and select the answer that can be answered on subclue 1 and subclue 2.

**Category:** Diplomatic Relations
**Clue:** of the four countries in the world that the United States does not have diplomatic relations with, the one that’s farthest north
**Inner subclue:** the four countries in the world that the United States does not have diplomatic relations with (Bhutan, Cuba, Iran, and North Korea)
**Outer subclue:** Bhutan, Cuba, Iran and North Korea; the one that’s farthest north
**Answer:** North Korea

From this clue Watson has to make two questions. Question 1 (inner subclue) has to be answered and with the four candidates. These four candidates are the input for question 2 (outer subclue).

The most difficult ones to program are the puzzles in some questions.

**Category:** Rhyme Time
**Clue:** it’s where Pele stores his ball
**Subclue 1:** Pele and ball (soccer)
**Subclue 2:** where store (cabinet, drawer, locker, and so on)
**Answer:** soccer locker

The system needs to detect 2 subclues, answer them, but also make sure these two answers rhyme with each other. This was very difficult to program because the clues give little information about the answer. So Watson had difficulties to answer most of the puzzles.

3.4.3 Hypothesis Generation
In this step the system is searching for candidate answers in his database. Its database contained a wide range of encyclopedias, dictionaries, newswire articles, literary works etc. This will result in several hundreds of candidates.
3.4.4 Soft filtering
In this stage a lightweight (less resource intensive) scoring algorithms are used to prune down
the list of initial candidates. When this is done only the remaining candidates go through the
more intensive scoring components. The soft filtering scoring model is based on machine
learning through training data. The candidates that don’t pass the lightweight scoring models go
directly to the final merging stage. The result of the soft filtering stage will be around 100
candidates.

3.4.5 Hypothesis and Evidence Scoring
The remaining 100 candidates will go through a new evaluation process. In this process the
confidence level of each candidate/hypothesis will be measured through a wide variety of deep
scoring measures. An example of a scoring algorithm: Consider the clue “He was presidentially
pardoned on September 8, 1974 in the US”. One of the candidate answers is “Nixon” because
Watson found “Ford pardoned Nixon on Sept. 8, 1974”. Then the machine uses the ‘one
passenger scorer’. This scorer checks in the document that Watson found, if Ford, Nixon and
1974 are common used words in that document. If that is true, the confidence level of this
answer will be higher.

Another algorithm that is used is the sequence-matching algorithm. This algorithm measures the
lengths of the longest similar subsequences between the question and passage. If the total
question can be found in a piece of literature, the answer in that document will have a higher
confidence level. Many other scoring techniques are used, some very advanced and some quite
simple. For example, how reliable is the source, how many times this candidate’s answer is found
in the literature or if the answer must be an American President.

3.4.6 Final Merging and Ranking
The last step is to merge the answers and rank all candidates. Merging is very important because
different answers can boost the confidence level of one answer. For example Watson could find
the answers, Barack Obama, President Obama or the current president of America. It has to
merge these answers and combine their score and confidence level. After merging the answers it
selects three answers with the highest confidence level. It pronounces the best answer in the TV-
show and shows the top three answers and the comparing confidence level on his display (see
figure 3). Note that the top three answers can exceed 100% because Watson gives his confidence
level for each of the answers. In this case he was 93% confident that ‘Michael Phelps’ is the
correct answer.

![Figure 3: A fragment of Jeopardy, broadcasted 14th of February 2011](image-url)
3.5 Remarks on Watson winning the show

Right after Watson won the show many people didn’t think it was a fair race between humans and a computer. The biggest complaint was that a computer could ‘buzz in’ faster. When a contestant ‘buzzes in’ he presses a button. When the contestant presses the button first he is allowed to answer the question. Because Watson is a computer, they believed that he could buzz in way faster than the human opponents. This isn’t true because Watson could only buzz in when he was confident enough to answer a question. A possible answer had to exceed a 50% confidence level. Human players are even advantaged to buzz in faster. They can ‘gamble’ if they think they know the answer already. For example when they read a word they know a lot about (for example: “this card game…”’) they can buzz in and they have 5 seconds to read the question and answer it. Watson had to be confident that he knows the answer and doesn’t use this ‘gamble’ technique. Buzz-in very fast isn’t the only way to win this game. When the player buzz-in and give a wrong answer the contestant will lose an amount of money. Watson also wasn’t super fast because he had to search in a huge database in which thousands of terabytes were stored (it wasn’t allowed to use the internet because human players can’t use internet either). In the beginning Watson needed two hours to answer a question. In the end this was reduced to 1 to 6 seconds (depending how confident it was of a certain answer).

Watson couldn’t listen or see. When it was Watson’s turn he always answered the best option it could find. Unfortunately, in one of the questions on TV, Watson gave the same wrong answer as the previous player. Normally in Jeopardy! there are also clips that are used instead of a question. Watson couldn’t see, so they didn’t ask such questions.

3.6 Future/Further use of Watson

For the game Jeopardy the database of Watson contained information of many different kinds of fields. Imagine Watson containing only information of a certain company or one industry. Most companies suffer from an information overload. Key is to take the intelligence out of the data. For example in the stock market there is a lot of data and numbers. These numbers are influenced by all kinds of activities like people that interact with companies, financial blogs and newspapers that write about a certain brand or industry. All this information is in natural language. Watson could digest this information to check and understand the interdependencies between companies and shares. In the end Watson could make financial decisions because his database is so rich with this information. Not only in the financial industry would Watson have its value, but also in the medical industry. Imagine Watson containing only medical information. The patient sums up the symptoms and Watson returns a diagnosis and a confidence level for each disease. In fact, there are some computers designed to do this kind of work. However Watson’s response time is faster and he can interpret a patient better. For example when a patient says he has difficulty swallowing, Watson can immediately link this with the medical term ‘dysphagia’. Also the health care sector changes so rapidly that it is hard for doctors to keep up with the latest trends. Watson can be used as a support tool to store all the trends in his database. A similar technique for diagnosing a patient can be used in a completely different industry. For example customer support. Whenever a client has a problem with his computer, it tells the problems to Watson and it can return a solution for the problem. When Watson won the knowledge quiz it is not the end of the project, more likely it is the beginning of Watson to fulfill useful tasks to manage large data collections written in natural language and return information in all kinds of industries.
To conclude, in section three Q&A systems are discussed. These systems try to obtain appropriate answers, for a given domain, with questions given or answered in natural language. The knowledge to answer these questions comes from a large document collection. To process natural language, to store (or capture) the right answers in a database, multiple steps have to been taken into account. This is difficult because not only words are important but also the context (or sequence of words) is necessary to get to a correct answer. IBM thought it was an interesting challenge to build a Q&A system that could compete on the human championship level in a knowledge quiz. This project (Watson) had to follow up their chess computer (Deep blue) from 1995. For Watson to compete successfully in this knowledge quiz there are three important metrics: Speed, precision and confidence. Speed is needed to answer the question first, but therefore it has to be confident that the precise answer is correct. This information is gathered from his own database (humans can’t use internet so Watson cannot either). With use of multiple algorithms, Watson could break down the question, ‘understand’ the context, search for possible answers, rank the answers and choose the best option. In a couple of seconds (dependent on the confidence level) Watson could answer the question. In the end it won the knowledge quiz against the best players of the show. This Q&A system can be used in other industries. Like in a medical sector to assist a doctor. Therefore the database has to contain different information.

4. VOICE RECOGNITION

The last section contained a question and answer system. This section discusses voice recognition because voice recognition is a good example of AI that eases the way people live. Instead of all kind of controllers or keyboards the users can use their voice as input. Voice recognition is another subfield of AI and came up around the 1960s. With the rise of the Internet and mobile devices speech recognition is growing fast and might take over all normal keyboards and touch screens markets. In this section Siri (a personal assistant on Apple’s smartphone) will be evaluated. Siri is taken as an example because it is pretty new to the market (2010-2011), works well, and the owner (Apple) won’t tell others how it works to prevent competitors from copying it. Furthermore Siri is one of the first commercial products that is widely used that uses voice recognition after roughly 50 years of international research. What Siri does, how it works, its future potential, and competitors of Siri will be discussed in section 4.2, 4.3, 4.4 and 4.5.

4.1 History of voice recognition

This section summarizes the history of voice recognition systems. With this information it is clear which aspects of voice recognition are known and which future developments are needed to gain more products that use voice recognition.

According to Juang et al (Juang 2004) in the 1960’s small vocabularies (between 10-100 words) of isolated words could be recognized by systems. An important technology that had its breakthrough was the filter-bank analyses. This filter bank could separate multiple input signals and create a single frequency sub band of the original signal.
In the 1970s the vocabulary of isolated words rose to 100-1000 words. Simple template-based and pattern recognition methods were used to do this. With template matching a word has to be spoken a couple of times by the user. An average sound of the word is calculated and when the user pronounces a word that looks like an average sound of the word this word will be used. The positive side is that accents will be recognized, the downside is that every new word that has to be recognized must be recorded a couple of times by the system. Background noise can also affect the average sound of a word or give problems when the user is in a noisy environment. A different method is pattern recognition. With this method each input gets a label witch classifies the input. In voice recognition the input is a voice and the label or classification can be a word (or letters). This input receives a label according to the database win which all the letters/words are stored.

In the 1980’s speech recognition problems are tackled by statistical methods (hidden Markov Model and stochastic language model) with a wide range of networks for handling language structures. The goal of the hidden Markov Model is to determine hidden parameters from observerable parameters. Also learning from a test and training set became common. With these two methods it was easier to handle virtually continuous speech recognition with a higher efficiency rate and performance.

In the 1990’s the vocabulary could contain infinite words and wasn’t sector related. The key technology is the finite state transducer. Before that there was a finite state machine, which is a mathematical model used to design computer programs and digital logic circuits. It defines a ‘state’ of a system and can only work in that state if the input and output ‘language’ is the same. For example a radio could be in the state ‘radio’ or ‘cd’. The action ‘next’ performed in the state ‘radio’ will take the radio to the next radio station while the same action in the cd state plays the next song. So the input and output relate to each other in one state. The finite state transducer, developed in the 90’s, has an input and an output tape. An input is recognized by the system and matches that to an output, even if the output is in another state. So it translates the input to the correct output (Jurafsky et al 2000).

From 2000 until the present text-to-speech (instead of speech-to-text) is developed significantly and multi-modal inputs (keyboards, mice, touch screen) can be used. The technology of voice recognition is now so far developed that it can compete with keyboards and touch screens. So some companies are now trying to use voice recognition in commercial products. Apple is trying this with their new product Siri.

4.2 What is Siri?

Siri stands for Speech Interpretation and Recognition Interface. It is a personal assistant which operates on Apple’s IPhone 4S. Siri is an application that uses a natural language user interface to answer the users question. The application allows a user to ask Siri:
• for a reminder
• to send a text message (SMS)
• about the weather
• for information (from Yelp, Wolfram|Alpha, or Wikipedia)
• to set a meeting
• to send an email
• for a number
• to set an alarm
• for directions (from Map apps)
• about stocks
• to set the timer
• Siri about Siri (Shows which commands can Siri perform)

Once Siri is installed the home button needs to be pressed to give Siri a command. It will process the question within three seconds and gives (if possible) an audible answer. The user can ask, for example, to send a text message to person X. Siri makes this text message for person X and only wants one confirmation from the user to send the message. This way the user doesn’t have to press various buttons to make the text message itself. So it is easier and faster to use Siri for a couple of commands. Also blind people can make good use of this application.

4.3 How does Siri work?

How Siri works is illustrated in figure 4. Siri will be explained in six steps.

figure 4: Siri, from input to output
1. The sound of the users’ speech is encoded in a digital form.
2. This form is send via a cell tower to the Internet Service Provider (ISP). Your ISP communicates with a server in the cloud. This server contains all sorts of models to decode your speech.
3. Simultaneously with step 2, your speech is evaluated locally on your smartphone. Your smartphone communicates with the cloud server whether your command needs access to the internet (where is the nearest restaurant?) or that it can handle the command locally (Play ‘Das Veilchen’ by Amadeus Mozart).
4. The server compares your speech (the digital form of the user) against a statistical method. It evaluates through the sounds and order you spoke, which letters your message contains. At the same time, the local recognizer compares your speech to an abridged version of that statistical model. The highest probability estimates go to step 5.
5. The best candidates of step 4 will be ranked and are run through a language model. This model estimates the words that your speech is comprised of. The computer creates, based on his confidence level, the best candidate.
6. If the confidence level of the best candidate is high enough, the smartphone executes this command. When there isn’t enough confidence (when your voice is too ambiguous) it gives the user a follow up question. “Did you mean Frank Mulkens or Frank Mulder?”

The total time needed for these six steps is around three seconds.

---

4.4 Competitors of Siri

A possible goal for Apple is to gain the technological advantage of being the first firm with a successful voice recognition program. There are of course multiple companies with voice recognition programs but they don’t have the huge commercial success to develop their systems rapidly. Using Siri it is possible to ensure widespread distribution of the technology in smartphones and in the future to other devices. In this section there is a closer look at nine factors that influence whether a company can gain the technological advantage. Because when there is a technological advantage created in a firm, this doesn’t necessarily result in an economic benefit for this firm.

According to Teece (1998) there are two factors for a firm to cope with their benefit:

1. The ability to translate technological advantage into commercially viable products or processes
2. The firms capacity to defend their technological advantage against imitators

This results in nine factors that influence the firm’s capacity to benefit commercially from its technology:
(I)  **Secrecy: Prevent competitors knowing how it works**
Apple didn’t tell others how Siri works. Because many companies are interested in this application and many tools to reengineer exist, the French app-developer Applidium found out how Siri worked and posted a blog post with all the technical details. This had been done in less than a month after the release of Siri in October 2011.

(II)  **Accumulated tacit knowledge: When the knowledge of a certain innovation is only in people’s heads, it is hard for competitors to imitate**
To imitate Siri there is little or no tacit knowledge needed. Because Siri is an application, Applidium could monitor the tcp traffic (the audio streams), which goes to the servers of Siri. They made an audio stream that looked like Siri, this went to the Siri servers and there they could spoof their servers and knew how Siri works.

(III)  **Lead times and after-sales service: When there are large lead times to produce a product/service it will take a while before competitors can imitate it**
Because the precursor of Siri already existed for a while, other competitors like Microsoft could build a similar application. Because this application was not so well known they started as soon as possible to recreate it. This will result in approximately a 6-month lead-time for Apple.

(IV)  **The learning curve: When a firm produces more, the costs for production will lower which can result in cheaper products (advantage over imitators) or larger margins**
Siri doesn’t need to be produced over and over again, however the more commands Siri has, the easier it gets to program new commands. So, the more useful Siri becomes.

(V)  **Complementary assets: Extra assets to increase the value of the product**
In fact Siri can be seen as a complementary asset for the IPhone. There are no complementary assets for Siri yet.

(VI)  **Product complexity: When it is difficult to produce a product, the more advantage a first mover (Apple) has**
Natural language is already quite difficult and also to get from speech to text. So there can be said Apple has a big advantage here over other competitors. In the first place the makers of Siri (Siri Inc.) announced that Siri would be available on BlackBerry and Android phones. After acquisition by Apple these plans were cancelled. So those companies have to make their own product, if they are interested. Unfortunately for Apple, Microsoft already had some small products in which they used the speech to text technique from an idea in Word (speaking against ‘Word’). As soon as they discovered Siri they started to imitate it. So Apple’s advantage isn’t that big when it is compared with Microsoft but compared with others it raised a barrier.

(VII)  **Standards: When the product has a widespread acceptance it widens its own market and raises barriers against competitors**
An example is the QWERTY-keyboard. This keyboard had been copied from the typewriter, because people were used to this keyboard they didn’t buy other keyboards. Siri could be the standard if it is allowed to be used in other products or industries. So far this isn’t the case and they even try to patent their technology (see also IX).
(VIII) Pioneering radical new products
It is not necessarily a great advantage to be a technological leader in the early stages of the development of radically new products. Because the performance characteristics, the features valued by users or producers, are not always clear. For example the keyboard or touch screen has to be used because Siri can’t be used to control all functionalities on the smartphone. So it can’t replace the keyboard yet.

(IX) Strength of patent protection
It is hard to patent Siri. First of all, patents are more effective in chemical-related sectors, because it is more difficult to ‘invent round’ a clearly specified chemical formula than round other forms of invention. (Tidd et al 2005) Furthermore patents primarily served to help companies protect their intellectual property. Nowadays, especially in the mobile phone industry, they try to sue each other with their patents. For example Oracle is suing Google over the use of Java in Android. Apple is suing HTC, claiming that its phones copy the iPhone. The result is that every big company in the mobile industry is now patenting the most small and sometimes ridiculous things. For example the ‘sweep’ to switch to another page, or the method to unlock a smartphone. With their portfolio of patents, sometimes traded between each other, they try to sue their competitors, so they can’t sell their products in certain countries or make extra money when a firm has to pay for using their patent. For example Microsoft has to pay around $400 million for licensing agreements to Android. So it will be pretty hard to patent Siri because a couple of their competitors will have a similar patent in their portfolio.

In general Apple didn’t do a good job to stop competitors from imitating. The only advantage they have is that their product is quite complex and they have a lead-time of around nine months. So it is no surprise that AT&T and Microsoft announced their Siri-like products. Microsoft announced that their ASR called Iris will come out in fall 2012. AT&T is releasing their application ‘Watson’ this summer (totally different from the Watson Jeopardy champ discussed before). The difference (apart from handling more commands) is their strategy. They announced that everyone can buy their Automated Speech Recognition (ASR) and use it in other apps for $99 each. This can release a revolution in the app store if everyone, relatively cheap, uses a sophisticated ASR in their apps.

4.5 Future of Siri and comparable AI
Siri, or a similar product, will be common in the future. When other companies sell speech recognition software (e.g. apps) for a low price, this sector will grow very fast. Otherwise every big company has to make (or imitate) their own voice recognition software and the development will slow down. The vocabulary of voice recognition systems can contain unlimited words. So the next years other roadblocks have to be removed. These roadblocks are slang, dialects, accents and background noise. When these problems are solved a system can give a good indication about the user’s input. Voice recognition can be handy for a lot of products which need controllers or keyboards. With a small cell phone, there is a heavy tax on traditional keyboard based information entry. I think it is easier and faster to communicate by voice, not only for a cell phone but also for other products like TV, radio, IPads and other computer devices. In fact Apple is working on their own Apple TV’s. This is a smart TV with Internet access, there are also rumors that this TV can be controlled by using Siri.
The next step in the future of voice recognition is intent understanding. Intent understanding is about identifying the action a user wants a computer to take or the information she/he would like to obtain, conveyed in a spoken utterance or a text query. In section 4.2 there was a brief history of voice recognition, from 2000 till now, the biggest change was that there will be more commercial products. So in the next 10 years I expect that many input devices (controllers, keyboards, mouse, etc.) will change to voice recognition inputs.

To conclude this section, voice recognition is an important domain of AI. This eases the life for their users because they don’t need a keyboard or controller to provide input, instead they can use their voice. The beginning of voice recognition started in the 1960’s and it took around 30 years to come up with a system that could recognize many words from different sectors. A couple of models like Markov, hidden layers and training and test data contributed to this development. Since the 1990’s the use of voice recognition in commercial products has risen gradually. In 2011 smartphone assistant Siri became popular despite the small amount of commands that Siri could handle on the smartphone. However Siri is faster and easier to use than a touch screen smartphone. Voice recognition could be used in all kinds of other products that currently use controllers or keyboard. For example, TV, radio, computer (programs), IPads, etc. Other companies besides the owner of Siri (Apple) were also interested and started to make imitations of Siri. Apple didn’t tell others how Siri worked, but because a French company found out how Siri worked, and placed this information on the Internet. Other companies like Microsoft and AT&T can easily create an imitation. Although their imitations aren’t on the market yet, AT&T announced that their voice recognition system can be used in other applications for low prices. Therefore AT&T has a chance to become the new leaders in voice recognition because their product could be the new standard.

5. AI LEARNING

It is known that AI can contain and extract knowledge and how voice recognition works but to make these systems smart another method is needed. There are two ways to improve AI, machine-wise or hand-wise (manually). This will be further explained is section 5.1 in which it is explained how AI can learn or improve itself. After this, section 5.2 will elaborate how poker bots work. Poker bots are smart programs that learn from statistical data from their opponents. Poker is taken as an example because this game isn’t dominated by computer systems (yet) while statistical data is very useful to exploit the game. In section 5.3 the future of automatic learning AI is discussed.

5.1 AI learning

In essence there are two ways for AI to learn, machine-wise or hand-wise. With the hand-wise procedure there are programmers who insert additional rules to the program. This way the more rules and the more complex it becomes the more intelligent the system gets. This is a very time consuming and difficult approach because programmers have to work in code which can contain thousands (or much more) lines to create an ‘intelligent’ system.
As Charniak (2001) already said (and many other scientists agree) machine learning is in most cases a better solution because the learning process automatically focuses on the most common cases. Furthermore automatic learning procedures make use of statistical interference algorithms to produce models that are robust to unfamiliar input (containing words or structures that have not been seen before) and to erroneous input (misspelled words). Also machine learning systems will be accurate just by providing more data, while hand-made can only be made more accurate by increasing the complexity of the rules, which is a much more difficult task.

With machine learning the system makes use of statistical data to improve itself. This can be done in three different ways: supervised, unsupervised and reinforcement learning.

**Supervised learning:** There is training data available; this training data contains an input and a desired output. An appropriate algorithm analyses the training data and can predict the correct output value for every input value. This way of learning there are examples with input and the desired output needed.

**Unsupervised learning:** With unsupervised learning there is tried to find hidden structures in unlabeled data. So there are no examples with the correct input or output. It is possible (for example with k-means) to cluster a set of data, so the input gets clustered but there still isn’t a desired output of a cluster. This way is mostly not preferable because it is hard to learn something from this data.

**Reinforcement learning:** In this learning way the system takes an action in a certain environment to maximize the reward. It learns from previous actions in the same environment and checks what the best/desired action is. For example a taxi driver wants to maximize his tip. When a customer goes home at 2 AM from a bar, it is more likely that this person tips if the taxi driver has a funny conversation with his customer and tells him a couple of jokes along the whole ride. When the taxi driver tries this same behavior with a businessman who goes to work at 6AM the tip will be lower on average. So to optimize his tip (reward), the taxi driver has a different, or no, conversation (action) with different customers (environment).

In the next section we take poker bots as an example. These bots are systems and take an action based on statistical data. The question remains, also for other systems, how can you make sure that you have the right algorithm based on the data?

Any hypothesis that is seriously wrong will almost certainly be ‘found out’ with a higher probability after a small number of examples, because it will make an incorrect prediction. Thus, any hypothesis that is consistent with a sufficiently large set of training examples is unlikely to be seriously wrong: it must be probably approximately correct. (Russel and Norvig 2003)

So with additional data the algorithm gets better each time. Thus the larger the database for a poker bot the better it will take its decisions. In the next section it will be explained how more data results in a better decision.
5.2 Poker

Since 2003 the poker industry is growing exponential, since an amateur player (Chris Moneymaker) won the World Series of Poker (WSOP) for 2.5 million dollars and qualified himself for only 40$ (Grohman 2006). These days the big poker clients have over 50.000 people playing online every night (GMT). There are also lots of poker shows on TV (the big game, poker after dark, etc.). However there is a big threat for the poker industry and that is the use of software among players. There are different kinds of software that players can use. When these software programs are used, these players have an advantage over other players who don’t use this software. (Johansson et al. 2006). In general all software is legal as long as the system doesn’t make a direct decision for the player. This results in a threat. What if a human creates a poker bot that makes all the decisions and is better in poker than humans? In fact there are bots created and they might win money in some poker disciplines. So the poker clients have to detect and delete these bots in order to provide a save environment for human players. These human players don’t want to play against bots and especially not when the bots win their money. The question remains, how good are these poker bots? In the next section will be discussed how these poker bots work and what the strengths and weaknesses are of these bots.

5.2.1 Pokerbots and AI

This section discusses the most common and one of the most strategic variants of poker namely Texas Hold’em. There are 4 reasons why poker is an excellent domain for AI:

*Imperfect information:* The player can only see their own hand and the cards on the table. They don’t know what the other players have in their hands.

*Stochastic outcomes:* The player has to select the most profitable action in an uncertain environment, however when a new card is dealt his winning position can turn in a losing position and vice versa.

*Exploitation is important:* In poker there isn’t an optimal game strategy. To maximize the long-term winning of a player he/she needs the best counter-strategy against each opponent. Therefore the player have to exploit your opponent’s strategy.

*Partially observable information:* Players have to construct a model of their opponent even though they won’t have all the information about that player.
Especially all the parts with imperfect information are a challenge in the AI domain. Almost every game with perfect information (chess, checkers, etc.) there is a system that can beat the human championship level (Billings et al. 2002). To win these perfect information games there is a need to do a deep look-ahead search of the game. If a decision tree is performed with all the possible steps that is at least one step ahead of your opponent and an algorithm that searches for the best step, there is a big chance you will win the game. However with imperfect games you can’t make the best decision because the next action (of your opponent, or for example a new card on the table) can change your position. Furthermore theoretically correct plays in an imperfect information game requires probabilistic mixed strategies, where different moves are chosen some fraction of the time in identical circumstances. This differs from perfect information games where there is a single best solution in identical circumstances. Therefore popular imperfect games, like poker, are now a target for the AI domain.

5.2.2. Inside a pokerbot
First of all a pokerbot contains a set of rules to play its ‘standard’ game. This standard game has to reach an optimal game solution (Nash equilibria) which doesn’t take into account the players playing style. This solution provides a randomized mixed strategy, which suggests how to play in each possible situation. With this strategy the bot will gain the maximum game-theoretic value of each game, regardless of the opponent’s strategy. Unfortunately this Nash equilibria solution only derives a very small problem, it is all about retrieving information about the possible cards of your opponent and his playing style.

Now we take a step into reading your opponents cards. Imagine a 1vs1 poker game versus an average player. Your opponent can have 1326 different starting hands \((52 \times 51)/2 = 1,326\) from a standard 52-card deck. However many hands have the same value, for example \(A\heartsuit J\heartsuit\) and \(A\spadesuit J\spadesuit\) are identical in value, because each is a hand consisting of an ace and a jack of the same suit. If you leave out the identical value hands, 169 combinations remain. (See figure 5) The red part (s) means cards from the same suit and the blue part (o) means two cards with different suits.

![Figure 5: All possible different starting hands.](image1)

![Figure 6: Opponent calls preflop.](image2)
When you receive your cards and decide to raise (put in more money), a portion of his range (possible card combinations) will fold (pass). Another portion will call (to match a bet) and a portion will re-raise. The portion that an average player will call looks like Figure 6 (the yellow part). This is around 40% of all starting hands.

After the preflop game there comes the flop. Both players can see these three cards. An example of a flop could be:

If we would bet this flop an opponent with nothing (for example a 7 and a 6) would be likely to fold. Let’s assume every opponent would fold except when you have a pair with King or a 9 or better. We could now narrow the range of our opponent to only 16% of all possible starting hands.

So the player started with 40% of all hands to the flop and remains after one bet with only 16% of his hands. A poker bot can now calculate if we have to bet or not. This involves four parameters:
(1) Compute hand strength: the probability that a given hand is better than that of an opponent.
(2) Positive Potential: Chance that a hand that is not currently the best improves and ends up winning.
(3) Negative Potential: The chance that a currently leading hand ends up losing.
(4) Effective hand strength: How good are our two cards compared to an opponent potential cards.

When this sum retrieves a positive value we bet this flop. This happens when we have, on average, a better hand than our opponent and/or we have a worse hand but we can make him fold his cards.

However now comes the tricky part, we cannot take for granted that a player only continuous with a King or 9. So we also have to take into account that a player might have a 2 maybe not even a pair. This data can only be extracted when playing many games against the same player. Then the bot could calculate what the optimal strategy (so far as is known) is against this player. For example, one opponent may bluff (betting with weak hands) too much, the other too little. Against the first opponent the bot would call his raise more often and versus the second he folds more often (or bluffs more often when the second player doesn’t bet). So each player has a different range of hands. The range of hands from someone who (almost) never bluffs is a lot smaller then the range of an average opponent. So that makes it easier to calculate which hands he probably has and which action (bet or check) is better. When the bot would only call with the optimal (average) frequency he would decline an extra opportunity versus these players.

So there is an absolute ultimate playing style which can be approximated by a bot. However it is more important to play on the weaknesses/errors of your opponent. To see the errors of the opponent you have to track some statistical numbers and let the bot adjust accordingly to that. Poker professionals can adjust very quickly and have more background information (experience) about the most common errors in poker. So when a bot plays versus the human Championship level (Billings 2006) the bot looses a little in the beginning but when more hands are played (and the bot has more data about a player) it is a hard opponent for the world class. In 2008 a competition is held with two very good poker players (Phil Laak and Ali Eslami). In this competition pokerbot Polaris defeated these players in a 1vs1 game over a large sample. This was in a limit game (the bet that is made is a fixed amount every time), this variant is in favor of a bot because the decisions to make are easier (the decision tree is way smaller). So the more popular variant ‘no limit’ is way harder to program. Besides that an optimal solution versus one player is easier to find than versus six or more players. So far the best bots can only win versus the human championship level in a 1 versus 1 match with fixed betting sizes.
5.2.3 Security of Poker bots

In section 5.1 is already said why poker bots are a threat versus human players. To prevent bots to play on a poker client multiple measurements can be taken. The researcher asked the three largest poker sites how they make sure that bots aren’t playing on their site. Unfortunately the 2nd and 3th largest site (Ipoker network and Party Poker) wouldn’t cooperate with this research because it is inside information and they wish to keep it secret. Pokerstars however agreed to give some basic functionalities how they prevent bots for playing on their site. They only gave the basic functionalities because for their bot detection and mitigation techniques to remain effective, they must remain secret.

To detect a bot Pokerstars has a large database with all kinds of data from each player. When a player does something ‘strange’ this will be recorded in the database and an investigation will start. Some ‘strange’ behaviors are:

- Playing much more games simultaneously then normally
- Playing games on a higher/lower stake then normally
- Playing individual hands totally different than normally

These behaviors are very abstract. In reality this will be way more specified because these three rules are no good indicators for bot detection. Every player will have some days they play more hands than normally, play a different stake or messes up an individual hand (for example by clicking on the wrong button by mistake). Although these behaviors are on a high level it shows where the employee is going to. When a player start using a bot there will be a difference in speed, stakes and playing style. The speed will increase dramatically. The speed of the bots are so fast they can be easily play 50 games simultaneously. This will trigger the database because it is near impossible to play more than 30 games simultaneously. Only the best and fastest player can make decisions so fast and these players have developed their skill with months/years of practicing.

Normally players who use a bot will start on a lower stake than he normally plays. Just to check if the bot can make money on that stake. This will cost the player less money if the bot can’t win versus human players. The fast speed of the bot can also be tested to set a larger sample size. The bigger the sample size, the higher the probability the bot is profitable at a certain stake.

When a player makes his own bot it is very hard to copy the exact game play from that player. Because there are so many factors that influence a decision. This will result in a different playing style when the bot and player are compared to each other. This playing style will differ even more when a player uses a bot of someone else.

Unlike the amount of games and stake played, a difference in playing style is harder to measure. Probably this would be only measured when a player is suspected to be a bot. Other players can report a player. This button is mainly used when players are using offensive language but it can also be used when a player thinks someone is cheating.
When the database thinks the player might be a bot, it sends the player a captcha in his game. The player must enter this captcha within 100 seconds. When a captcha isn’t (correctly) answered in 100 seconds the player is likely to receive a captcha in the near future or get an investigation of his playing profile. When the outcome is that a player used a bot, his account can be frozen and the remaining money on his account will be given to charity. The player will be banned from the Pokerstars Network.

5.3 Future of learning bots

With the previous information on how a bot works it is shown that a bot can win versus the human championship level in a fixed, one versus one game. It is just a matter of time to create a bot that can win in other (more popular) variants in poker. So far the biggest concern is that there is not enough statistical data to play against the same player(s). For example in a session where a bot plays one hour on one table with 5 other opponents he plays around 90 hands (1). When taken into account that an average winning strategy plays around 25% of their hands from a certain position there is a sample of 15 hands from every position from 5 opponents. This sample is very small so it is very hard to estimate (like there is done in 5.2) which range of hands a player plays. So a bot won’t be able to optimize his strategy in an hour. Most players won’t play longer than an hour in a session and large poker clients have 300,000 unique players every day. So the chance is very slim that a bot plays against the same player in the near future. After a while this data isn’t accurate because a player can change his playing style over time. So the data of a certain player will be useless. Also when a bot plays against 5 or 8 other opponents the decision tree of possible actions from other players will be very large and difficult to handle for a bot.

Nevertheless I am convinced AI will win in every game. The decision tree games (checkers, chess, go, etc.) are already dominated by computers. After this the games with more random factors are dominated by bots. This is for example Backgammon (where the player throws a dice) and a couple of card games like Hearts (Tesauro 1994). And the last years there are even competitions to build AI that can compete in videogames (Mario Kart and Unreal Tournament) (Shaker et al 2011). So in the future AI will win in every game where decision trees, calculations and statistical data is involved, which is basically every game. But self learning bots that can play (video) games aren’t that interesting for human society. It would be a lot more interesting if AI can do something useful. So in the next section we take a closer look to a couple of real life examples of advanced AI in the elderly sector.

(1) This comes from my own experience, I played around 1000 hours of poker. This data is captured in an advanced database (Hold’em Manager)
To conclude in this section we saw AI that can learn. This is not by entering more and complex rules by a programmer but to make use of statistical data. With this data the system can learn in three ways, supervised, unsupervised and reinforcement learning. By giving more data to the system the algorithm gets better each time. With many data or large pruned trees AI systems can win from the human championship level in a couple of games like chess and checkers. In this section a 1 versus 1, fixed betting poker game is elaborated. A bot can win in this match if he complies with three rules. The first one is a basic game which approximates the optimal way to play (Nash equilibrium). The 2nd rule is to find out which range of hands his opponent is playing each game. The last rule is to find out which strategy his opponent is using (e.g. bluffing or trapping) and match this to a counter strategy. The faster a bot can adapt to rule two and three, the better the bot will be. Therefore he needs to gather as much as possible statistical data from that specific player, so he has to play against the same opponent for a long time. Due to complex rules and very large decision trees (without enough data) poker bots can only win in 1vs1 poker games and not versus 5 or 8 opponents. This however concerns poker clients because no human player want to play versus computers, especially not when these computers can win money from this player. So poker clients ban every bot they can detect. These bots can be recognized because a player can suddenly plays more games simultaneously, plays on a higher/lower stake then normally or plays individual games totally different. Other (and probably better) methods aren’t mentioned by Pokerstars. But these three methods aren’t ‘water proof’, a clever player will notice these methods and try to let the bot play the same style, same amount and same stakes. In the future other AI systems will likely to beat humans in every game. This isn’t very useful but other learning systems can be valuable in the society. The next section will take a closer look to useful AI.

6. SERVING ROBOTS

In the previous sections we saw that AI eases the way people live, AI contains knowledge and that AI can learn from itself. In this section we combine these three abilities and look for real life, useful examples. Therefore I selected the elderly care, this is a hot topic in the Netherlands. Because of the high number of births between 1946 and 1970 in the Netherlands the group of elderly will grow from 2,4 million in 2008 to 4,5 million in 2040 (van Duin 2009). Because of the relatively high life expectancy and a shrinking labor force there is a need for AI to support the elderly. In the next section a couple of robots are shown which support the elderly with every day routines.

6.1 Functions of Mobiserv and Companionable

This trend of a growing population of elderly is not only in the Netherlands detected but a worldwide trend. According to Pollack (2005) the proportion of people over 60 years will have doubled in 2050. The highest increase will be the amount of people over 85 years of age. Therefore AI can support older people with cognitive impairment in a couple of ways. (1) By providing assurance that the elder is safe and performing daily activities. (2) To assist the elder with his/her impairment (3) and/or to train the brain of the elder. There are a couple of robots who assist the elder like MobiServ and Companionable (shown in figure 8). These two robots (and most similar robots) have three components:
(1) Daily Activity Plan (DAP): What is the planning for the user today?
(2) Client Modeler: What is the user currently doing?
(3) The Intelligent Reminder: Which actions are need to be taken to fulfill the DAP?

1. **Daily Activity Plan**

   In a daily activity plan is stored what the user has to do today. This can be all kind of daily activities like eating, drinking, taking medicines, bathing etc. This system also works as an agenda; the user (or other people) can insert activities like ‘visit from daughter at 3PM till 4PM’. The user or caretaker can check which activities are on the agenda and which activities are done.

   ![figure 8: Companionable in action in an elderly facility](image)

2. **Client Modeler**

   To identify what the user is doing multiple sensors are placed in the residence. This can be pressure sensors in couches or chairs or RFID (Radio-Frequency Identification) tags in household objects. The system can recognize where the person is and what he/she is doing. For example if the person is walking from the couch to the kitchen and uses the teakettle and sugar, there is a fair chance that the user is making tea. With extra use of biosensors it can be measured what happens in the body. This can be a (small) increase in body temperature by drinking hot tea. So in the DAP it is now tracked that the user drank something. Other uses of the biosensors are the detection of the user’s physical movement and what their heart rate, blood pressure and other medical conditions are. When something goes wrong the system gives an alarm to the caretaker. This person can make contact with the user, the robot knows where the user is so it drives automatically to the user and then a screen pops up where the caretaker is shown (webcam) and they can interact with each other by sound and video.

3. **Intelligent Reminder**

   The smartest part of both robots is the intelligent reminder. Its basic function is to remind the user of a certain activity. For example to have lunch at 1PM. But this reminder isn’t static. Suppose that a client has to take her medicine an hour after breakfast. Less advanced systems would alarm the user to have breakfast at 8 AM and alarm at 9 AM to take medicines. These systems make use of the DAP. If there is listed in the DAP that the user usually has breakfast between 7AM and 8AM and the client is lying in bed at 8.15AM (recognized by the pressure sensors) it will drive to the client and asks the user if it wants to have breakfast. If the user then has breakfast at 8.30 this data is stored in the DAP and asks at 9.30 if the client took her medicines. The intelligent reminder also stores how long an activity takes place (bathing 25 till 35 minutes) and thinks about future activities during the day. When it sees that bathing might take 35 minutes and that the elder usually watches a TV show on 11 AM it can warn her at 10.25 to take a bath or suggest her to take a bath after the TV show. Furthermore it also suggests using items at certain activities. Elderly with dementia typically forget which items belong to which activities (Mihailidis et al 2001). These robots help with this impairment. When the user is leaving the apartment to visit her daughter it asks if the user needs items such as keys or a wallet. With the RFID tags on these items the robot can also tell where these items are (other robots can
also bring these items to the user). Besides reminding the user, also warning the caretaker, is an important function of these systems. Warning signs can be triggered at multiple occasions, for example with the biosensors when the heart rate changes rapidly or when the user doesn’t respond to the robot.

6.2 Serving robots and AI

With these robots all components of the previous sections are discussed. These systems have a voice recognition component. This component can define what the user wants based on his/her voice input. With this input the robot can support the elder with all kind of tasks or answering a question. This question and answer component was discussed in section 3. Watson could answer a question based on his large database with all kind of knowledge from Wikipedia and researches. ‘Companionable’ and ‘Mobiserv’ can answer the questions by using their database with input from the DAP and learning database. This learning database is discussed in section 5 with poker bots. These robots also make use of statistical data. The input is the time when a user is doing an activity. When the system doesn’t know what the user is doing it can ask it directly to the user. Also when the user is taking more time with a certain activity in the last few days (e.g. an afternoon nap) it takes that into account with other activities in the DAP. The statistical data can even protect the elderly for example when a person with dementia is cooking but went for +5 minutes to the couch in the living room. The robot can remind this person that he/she is cooking, or alarm the caretaker that the kitchen is unattended while someone is cooking. Other learning components in these robots are in the voice recognition system. The model (to compare input with a language model) changes when someone is speaking multiple times with the system. In this way also elderly that have a slurred speech can properly communicate with the system.

6.3 Future use of robots in elderly care

The use of robots in the elderly care is already a success. The clients are positive about the robots but there is also space for improvement. The elderly say that the robot is more useful if it can support with other activities like calling or controlling the climate control. I agree that adding more tasks on the robot will increase the value of the robot. Also the serving robots can be used as external control. With these serving bots the caretaker needs to supervise the elderly. But a couple of tasks can be supervised or automated by the bot. For instance when it is known that the elderly has heart problems and all of a sudden the heartbeat drops, the robot could already call the emergency number and the employee takes over the call. This will decrease the amount of activities of an employee and speed up the efficiency of other activities for the user. However, I think, the biggest improvement that these serving bots can make, is to link all the robots of a certain place together (e.g. elderly home). When this is done people in different rooms could communicate with each other (they don’t need to call or walk to each other). Furthermore they could make an appointment with each other. If the other one agrees this will pop up in both of their DAP (e.g. lunching together at 1PM). The caretaker (or the robot itself) could also promote activities. This can be an invitation for a special event for example Queens day. Or the robot can check in its statistics that the person skipped two gym-activities the last week while the user normally participates in. So with a connection between serving robots the functionalities and quality can increase.
To conclude, in this section we discussed robots in elderly care. These robots can support the elderly with managing, remembering and scheduling activities on a daily basis. Not only the elderly benefit from these robots but also care takers because the robot can alarm the caretaker when something goes wrong. Mobiserv and Companionable are a good match in this research because they combine the components of AI, which are discussed in earlier sections. Both robots make use of voice recognition to communicate with the user, they can question the user or answer the user’s question and they learn from statistical data.

7. DISCUSSION

As been said in the introduction there is a distinction between strong and weak AI. Strong AI is intended to produce machines with an intelligence that matches (or exceeds) that of human beings, and weak AI is only simulates intelligence. Searle says that strong AI doesn’t exist by using his ‘Chinese room’ experiment.

Chinese room
The Chinese room is a thought experiment by John Searle. Suppose that there is a program that gives a computer the ability to carry on an intelligent conversation in written Chinese. If we give the program to someone who speaks only English to execute the instructions of the program by hand, then, in theory, the English speaker would also be able to carry on a conversation in written Chinese. However, the English speaker would not be able to understand the conversation. Similarly, Searle concludes, a computer executing the program would not understand the conversation either.

Another (old but popular) way to test intelligence of a computer is the Turing Test. When a computer passes the Turing Test a human questioner can’t tell whether the person is a computer or human based on chatting with this other person or computer. However to this day, no computer passed this test. There are however a couple of chat bots that give some realistic answers, for example Cleverbot. This bot learns from the input of other chatters. When it discovers that some people give the same answers when it asks the same question it stores this answer in its database. This way it is in fact copying sentences from other chatters. In this way it gives sometimes very surprisingly good answers about hot topics.

The question remains how AI can get intelligent. This question is difficult, perhaps impossible, to answer because we don’t know much about (human) intelligence. In general it is accepted that intelligence is an attribute of the brain. Some researchers claim intelligence is not on one spot of the brain but the whole brain together (Haier 2009). Other researches claim that intelligence is created when different parts of the brain integrate with each other (Thomson et al. 2001).
Without the knowledge how intelligence is created in the human body, I think it is impossible to create true intelligence in AI. So I agree with Searle that AI does not really understand the situation and only gives the correct answers. On the other side most of the time when a systems gives the correct answer it is all we want to know. When I was searching for sources on this discussion I found some people that where referring that (strong) AI was even dangerous, especially self learning AI. I think these statements are funny because we don’t know how intelligence works so it would be highly unlikely that AI can discover this suddenly. Everything AI ‘knows’ is programmed by humans, with use of (lots) of statistical data AI can provide a better algorithm. To conclude, scientist need more information on how the human brain works in order to create ‘more intelligent’ AI.

8. CONCLUSION

With all the information of the previous sections we can now answer the sub questions and finally the main question. Below is a recall of the components and main/sub questions.

<table>
<thead>
<tr>
<th>Component:</th>
<th>Subfield of AI:</th>
<th>Example:</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI contains knowledge.</td>
<td>Natural Language Processing</td>
<td>Watson</td>
</tr>
<tr>
<td>AI eases the way people live.</td>
<td>Voice recognition</td>
<td>Siri</td>
</tr>
<tr>
<td>AI can learn themselves.</td>
<td>Learning AI</td>
<td>Pokerbots</td>
</tr>
</tbody>
</table>

Sub question 1: What is nowadays the smartest program that can answer a broad range of questions?

AI that contains knowledge and with this knowledge Q&A systems can be developed. Q&A systems try to obtain appropriate answers, for a given domain, with questions given or answered in natural language. The knowledge to answer these questions comes from a large document collection. To process natural language, to store (or capture) the right answers in a database, multiple steps have to been taken into account. This is difficult because not only words are important but also the context (or sequence of words) is necessarily to get to a right answer. IBM builds a Q&A system that could compete on the human championship level in a knowledge quiz. For Watson to compete successfully in this knowledge quiz there are three metrics important: Speed, precision and confidence. This information (to answer the questions) is gathered from his own database filled with researches and Wikipedia pages. With use of multiple algorithms, Watson could break down the question, ‘understand’ the context, search for possible answers, rank the answers and choose the best option. Watson is currently the best program because of his large database and unique natural language processing algorithms. With these algorithms it is capable to understand the context which is necessary to answer the questions. This Q&A system can be used in other industries. Therefore the database has to contain different information.
Sub question 2: Which commercial system is using only voice recognition as input/output to make their system more user friendly?

The beginning of voice recognition started in the 1960’s and it took around 30 years to come up with a system that could recognize many words from different sectors. A couple of models like Markov, hidden layers and training / test data contributed to this development. From the 1990’s there slowly became some commercial products which uses voice recognition. Voice recognition could be used in all kind of other products that currently use controllers or keyboard which will ease the life of their users. For example, TV, radio, computer (programs), Ipads etc could use voice recognition. Siri (smartphone assistant) makes use of Q&A algorithms and voice recognition. Other companies besides the owner of Siri (Apple) were also interested and start to make imitations of Siri. Other companies like Microsoft and AT&T have it fairly easy to create an imitation. Although their imitations aren’t on the market yet, AT&T announced that their voice recognition system can be used in other applications for low prices. Therefore AT&T has a chance they can be the new leaders in voice recognition because their product can be the new standard.

Sub question 3: How and in which game, that isn’t fully dominated by programs, can AI learn?

In the section 5 we saw that AI can learn. This is not by entering more and complex rules by a programmer (like IBM’s project ‘Watson) but to make use of statistical data. By giving more data to the system the algorithm gets better each time. With many data / large pruned trees AI systems can win from the human championship level in a couple of games like chess and checkers. In this section a 1versus1, fixed betting poker game is elaborated. A bot can win in this match if he complies three rules. The first one is a basic game which approximates an optimal way to play (Nash equilibrium). The 2nd rule is to find out which range of hands his opponent is playing each game. The last rule is to find out which strategy his opponent is using (e.g. bluffing or trapping) and match this to a counter strategy. The faster a bot can adapt to these three rules, the better the bot will be. Therefore he needs to gather as much as possible statistical data from that specific player, so he has to play against the same opponent for a long time.

Combining the previous three subfields:

The serving robots combine voice recognition, Q&A and self-learning systems. All these components are a little less impressive in itself but the strength is their combination of all the three the techniques. It is not surprisingly that these robots are made with many help from different authorities like elderly homes, universities, speech recognition companies etc. These robots are also selected to give this research a more practical note, the other sections contain a high level about one of the three components. Besides the high level of AI the most applications are not commercialized. So they don’t have much usefulness in a community. But these robots can support the elderly with managing, remembering and scheduling activities on a daily basis. Not only the elderly benefit from these robots but also care takers because the robot can alarm the caretaker when something goes wrong.
Main question: How far is AI with creating their ultimate goal? (Building an intelligent robot that matches/exceeds human intelligence)

There is no clear answer to this question because AI is fragmented in all kinds of subfields. There is not a single project that tries to reach this goal at once. That project would need too many subfields of AI combined to create a human-like robot. One of the problems is that the subfields often fail to communicate with each other (McCorduck 2004). So at most projects handle one isolated aspect of AI. Some subfields like natural language programming and systems that learn on statistical data are becoming quite advanced and produce commercial products. Others didn’t reach this level yet because there are technical issues or no specific (urgent) problems to solve. Another problem, and maybe the biggest challenge, might not even be to program difficult algorithms but to discover how our own brain works. Without the knowledge how our own brain works it is difficult (perhaps impossible) to let a programmer (re)create the knowledge of a human brain into a program. The accomplishments in the field of AI and neuroscience are only the beginning of creating a system with intelligence, and it will take many years before a machine can pass the Turing test, namely achieving performance that matches or exceeds that of a human.
REFERENCES


