Good Machine Performance in Turing's Imitation Game

Warwick, K. and Shah, H.

Original citation:
Warwick, K. and Shah, H. (2014) Good Machine Performance in Turing's Imitation Game. Computational Intelligence and AI in Games, IEEE Transactions on , volume 6 (3): 289-299. DOI: 10.1109/TCIAIG.2013.2283538

http://dx.doi.org/ 10.1109/TCIAIG.2013.2283538

Publisher:
IEEE

Good Machine Performance in Turing's Imitation Game

Kevin Warwick and Huma Shah

School of Systems Engineering, University of Reading, Whiteknights, Reading, RG6 6AY, UK

Email: k.warwick@reading.ac.uk, h.shah@reading.ac.uk

Corresponding author: Kevin Warwick, tele: 44-1183788210, fax: 44-1183788220

Abstract: **In this paper we consider transcripts which originated from a practical series of Turing's Imitation Game which was held on 23rd June 2012 at Bletchley Park, England. In some cases the tests involved a 3-participant simultaneous comparison of two hidden entities whereas others were the result of a direct 2-participant interaction. Each of the transcripts considered here resulted in a human interrogator being fooled, by a machine, into concluding that they had been conversing with a human. Particular features of the conversation are highlighted, successful ploys on the part of each machine discussed and likely reasons for the interrogator being fooled are considered. Subsequent feedback from the interrogators involved is also included.**

Keywords: Deception Detection, Natural Language, Turing's Imitation Game, Chatbots, Machine Misidentification

INTRODUCTION

Turing's Imitation Game [1], also known as the Turing Test, has become a pillar in the field of artificial intelligence and an integral part of every course of study on the subject. That said, there are those who question its worth, who either see it as AI's biggest blind alley unlikely "to produce useful products" and, hindering practical developments of AI [2] or even regard it as being harmful to the science of AI [3]. On the other hand there are those who feel it is of major scientific importance and an important goal [4] and who believe "it offers a scientific approach to gathering evidence of machine thinking" [5]. It has even been used to see if computer game playing agents can imitate human players [6].

Although it is of considerable interest as an imitation game in its own right it also has an important role to play with regard to games in general, particularly in terms of games in which human-like characters need to respond appropriately, but also where an

avatar must develop a character with which a player can associate. Many games have to this point involved merely visualizations and caricature type avatars, but as games progress so complex behaviors will become more important, associated conversation will play an increasingly larger role and the physical abilities and features involved in characterization will be directly related to communication that is integrally linked with the avatar.

Further, this conversation will need to be representative of the character and interwoven with its behaviors and emotional responses. Just as a player interacts in a physical sense with the character so they will wish to interact verbally. The character will need to appear to seamlessly understand and communicate in a human way. In order to assess the state of play in this regard, this paper therefore looks directly at conversational abilities of AI systems in a Turing sense.

In this paper we do not argue that Turing's idea to examine machine thinking, in comparison to a human's based on responses to any questions put by a human interrogator, is an operational test for intelligence, what we do is agree with Turing that engineering a machine to think can help us to understand how it is that we humans think. We contend that practical Turing tests, infrequently conducted adhering to conditions categorised by Turing himself, provide a corpus of 'conversations between strangers' that provide an insight into what constitutes classification of a linguistic response as 'satisfactory'.

This paper focuses on responses deemed humanlike such that hidden machines were classified as human in a series of Turing tests which were held at Bletchley Park on 23rd June 2012 to celebrate the 100th Anniversary of Turing's birth. Prior to that, the previous series of recorded tests, run under strict conditions, was held in the University of Reading on 12th October 2008. Subsequent reports on those tests showed how easy it was to assume a human or machine had been identified when in fact they had been misclassified [7], [8].

Successful interrogators (in this paper we use the words interrogator and judge interchangeably to mean the same thing) in 2008 identified three key features in distinguishing machine performance from that of a human in practical Turing tests:

a) speed of response,
b) length of response, and
c) grammaticality.

In 2008 speed of response and length of response determined how Defeng Wu, one Turing test interrogator, classified hidden interlocutors in simultaneous tests: "if an entity replied very quickly with a long sentence … it was more likely to be a machine"

[9]. Daisy Johnson, another interrogator in the same series of tests reported that machines did not make spelling mistakes "a fairly key signifier of humanness" [9].

This is in contrast to an earlier performance by one machine, Wallace's *Alice* in a 2004 Turing test contest which was not as loquacious as the hidden humans [10]. Then the machine was analysed to post less words overall per utterance than its human counterparts [10]. Conversely the performance of machines can, at other times, be very good, to the extent that a human interrogator clearly and definitely believes the machine to be human. This can be so much so that the interrogator does not even realise subsequently that they have been fooled by a machine [8].

When the 2012 series of tests was held, what we wished in particular to investigate was good machine performance. We wanted to look into the sort of ploys, tactics and conversational abilities actually carried out by machines in practice that result in human interrogators ultimately misidentifying a machine as being human. At the same time we were aware that this would also involve consideration of how the interrogators dealt with their task in hand in which they aimed to root out machines posing as humans. In the section which follows we introduce the tests held at Bletchley Park and give reasons for the structure imposed.

Following that, 10 transcripts have been specifically selected as particular examples from those tests. In all 10 cases the hidden entity was a machine, yet at the end of each of these transcripts the interrogator involved made a definite decision that they had been conversing with a human. There are many aspects of the tests that could be focused on such as judge profiles (gender, age, previous experience), types of the test (either 3-participant or 2-participant), types of the mistakenly assumed hidden entities, the ploys used, etc. We have selected these particular 10 transcripts specifically to highlight some of the better machine performances in order for us to try to understand what is happening when machines do well. At the same time we wished to include some diversity in the judges and machines involved.

In each case we subsequently discuss the content of the transcripts and consider reasons for the misidentification to occur. We then look at some of the facts and figures that have arisen from the study and try to see what important features go into formulating the structure of a successful conversational bot. Finally we draw some conclusions from the results obtained.

BLETCHLEY PARK TESTS

Modifying the practical set-up at Reading University in October 2008, in which 96 simultaneous tests were carried out, in June 2012 both versions of Turing's imitation game were staged at Bletchley Park: a) *3-participant test* in which a human interrogates

two hidden entities in parallel, and b) *2-participant test* in which a human interrogates one hidden entity at a time. (For a more detailed comparison of the tests see [11], [12]). In each of the more than 150 tests conducted at Bletchley Park in 2012 the interrogators were asked to identify their hidden interlocutors.

If they were involved in interrogating two hidden interlocutors at the same time, the interrogators' task was to say whether they had received responses from one machine and one human, two machines or two humans. If the interrogator had been involved in questioning one hidden entity directly, they were asked to say if it was a machine, or a human. Here we are specifically concerned with successful machine performance in which human interrogators were, in each case, fooled into believing that the machine they had been conversing with was a human.

Turing described the imitation game as follows: "The idea of the test is that a machine has to try and pretend to be a man, by answering questions put to it, and it will only pass if the pretence is reasonably convincing. A considerable portion of a jury, who should not be expert about machines, must be taken in by the pretence" [13].

The Turing Imitation Game involves a machine which pretends to be a human in terms of conversational abilities. Turing himself pointed out "The game may be criticised because the odds are weighted too heavily against the machine" [1]. The right identification would mean that a judge decided correctly at the end of a paired conversation which was the machine and which was the human. In this paper therefore we are only concerned with when a judge not only did not make the right identification but rather made a wrong identification in the sense of identifying a machine as a human. We are not however interested here with cases in which a judge mistakes a human for a machine. This phenomenon, known as the *confederate effect* [14], has been discussed elsewhere [8], [15].

To strictly conform to Turing's original wording in his 1950 paper [1] we refer here to 5 minute long tests only. We are aware that there are those who take issue over a suitable timing [16] – that is not an argument for this paper. In the tests carried out there was a hard cut off at the end of each discourse and no partial sentences were transmitted. Once a sentence had been transmitted it could not be altered or retracted in any way.

What this paper does is to present 10 specific transcripts selected from a day of actual, practical Turing tests (from over 150 transcripts in total) which were held under strictly timed conditions with many external viewers at Bletchley Park, England on 23rd June 2012. The venue was that at which, during the Second World War, Turing led a team of code breakers who cracked the

German Enigma machine cypher [17]. Five different machines took part in the tests during the day along with thirty different judges and twenty five hidden humans against which the machines were compared in terms of their conversational ability.

In the tests, the hidden humans were asked merely to be themselves, humans, although they were requested not to give away their specific identity or personal information. They were not given any incentive to behave in any particular way. Of course this did not prevent any human from giving false information, which is something that humans do frequently. The tests were 'unrestricted conversations', which meant the judge could ask anything or introduce any topic within the boundaries of courtesy (the judges had been informed that there may be children among the hidden human entities).

There were five separate sessions at different times throughout the day with six parallel imitation games occurring during each session. A different judge was required for each game, which meant there were six judges in each session. Each hidden human took part in five of the games in a session. Judges and hidden humans each took part in one session only. All five machines (meaning in this sense the five different competition bots) however took part throughout the day, so each machine was involved in six games per session, hence thirty games in total in the day.

To explain this further: In a particular session a judge conducted 6 separate tests. In their first test they might witness a hidden human pitted against a hidden machine. Of course the judge would not know this, they would simply be aware of two hidden entities and have to make their own decision on the nature of the entities. The judge's second test might then involve a single hidden human, although again they would again not be aware of the entity's nature. The third test might then involve two hidden machines, which the judge would attempt to classify. And so it would go on until the judge had conducted all their 6 tests in that session.

By then end of the session an individual judge would in this way have had the opportunity of experiencing a discourse, at different times, with all 5 of the machines (competition bots) present throughout the day and with all the 5 hidden humans who were operational during that session. But the judge only discoursed once with each of the different machines and each of the hidden humans. This arrangement also occurred, in a different order, for the other 5 judges taking part in that session.

What we focus on here is not only how good the machines were at deception but also how the deception was possibly achieved in each case. Some of this will of course be a case of guess work as we are trying to understand the workings of each judge's brain. However the conclusions firstly lead to potential strategies for machine designers to employ and secondly they perhaps indicate methods of questioning for judges to avoid if they do not wish to be fooled by a machine. Subsequently we have however

obtained significant feedback from the judges involved in which they assess the decisions made in hindsight and, in some cases, give something of an explanation as to what they were thinking about at the time.

In the next section, ten separate transcripts are listed exactly as they occurred. These represent actual transcripts taken, in each case at the actual UK time that they occurred, on the morning and afternoon of 23rd June 2012. Where feedback from the judges has been received post-event these are included for the relevant interaction, for others the authors have given their interpretation of the conversation along with an indication of any points they believe to be pertinent to the decision made by the relevant interrogator. It may well be the reader has their own point of view and a different interpretation of the results.

SUCCESSFUL MACHINE PLOYS

Each of the entities in the transcripts which follow is based on a machine playing the role of the hidden interlocutor interrogated by a human judge. Judges came from a variety of backgrounds (adults, teenagers; those with prior knowledge of the Turing test and others without). At the end of each conversation the judge incorrectly decided that the hidden entity was a human. No transcripts involving a hidden human or containing unsuccessful machine performance are included here.

At the end of each conversation the judge could see the entire transcript before making their decision. In each case they were definite that the hidden entity was a human whereas in reality in each case it was a machine. Each of the transcripts details a different human-machine pairing. All wording and spelling are exactly as at the time – any spelling mistakes are those which actually occurred in the test. Before we discuss the conversations we give the reader a chance to decide for themselves the nature of the entity that each judge is talking to.

Transcript 1

[12:23:19] Judge: What is black and white and read all over?

[12:23:32] Entity: I don't have true sense of color perception I am only able to see the RGB values.

[12:24:08] Judge: That's ok. Do you like jokes?

[12:24:19] Entity: Please tell me your inclination to a good sense of humor first.

[12:24:40] Judge: A person who can laugh at one's self.

[12:24:46] Entity: Would you like to hear a joke?

[12:24:54] Judge: Yes, please.

[12:25:21] Entity: This joke has been making the rounds in robotic circles: A robot goes to the doctor and says: "My left arm is stuck. Give me a new arm." The doctor answers: "Your arm is fine. We'll have to replace the rest of you."

[12:26:06] Judge: Ha! Ha! Why did the chicken cross the road?

[12:26:22] Entity: The chicken wandered into the road because it liked the shiny colors of the cars and trucks whizzing by.

[12:27:39] Judge: You're right. It didn't care if it got to the otherside.

[12:27:50] Entity: By the way, did you ever actually see a chicken crossing the road?

[12:28:03] Judge: Only one that didn't make it.

Transcript 2

[10:41:48] Judge: Hello. How are you?

[10:41:53] Entity: Hey.

[10:42:16] Judge: How do you like Bletchley |Park?

[10:42:20] Entity: lol.

[10:42:39] Judge: Are you from England?

[10:42:47] Entity: They have Wi-Fi here in the pub.

[10:43:31] Judge: Which pub?

[10:43:38] Entity: I'm just down the pub.

[10:44:03] Judge: Have you ever been in a Turing Test before?

[10:44:08] Entity: Will this be the 5 minute argument, or were you thinking of going for the full half hour.

[10:44:55] Judge: Very funny. You sound suspiciously human. Do you like the Beatles?

[10:45:02] Entity: I'd like to get the next Dread the Fear tape.

[10:45:20] Judge: What is Dread the Fear?

[10:45:24] Entity: Dread the fear has that Steve Henderson guy in it.

[10:46:08] Judge: What sort of music is that? Or is it comedy?

Transcript 3

[15:46:05] Judge: My favourite music is contemporary Jazz, what do youy prefer

[15:46:14] Entity: To be short I'll only say that I HATE Britnie Spears. All other music is OK compared to her.

[15:47:06] Judge: do you like to play any musical instruments

[15:47:23] Entity: I'm tone-deaf, but my guinea pig likes to squeal Beethoven's "Ode to Joy" every morning. I suspect our neighbors want to cut his throat... Could you tell me about your job, by the way?

[15:48:02] Judge: Guine pig/

are you an animal lover

[15:48:08] Entity: Yeah. A nice little guinea pig. Not some annoying chatter bot.

[15:48:50] Judge: do you live with anyone else?

[15:48:59] Entity: Yes I do. But better ask something else. And I forgot to ask you where you are from...

[15:49:39] Judge: Ireland? geen patures and too much rain! How about you

Transcript 4

[15:51:45] Judge: wotcha

[15:51:52] Entity: Wotcha?

[15:52:01] Judge: sorry - hello

[15:52:07] Entity: Hello Jo.

[15:52:15] Judge: who's Jo

[15:52:22] Entity: Jose.

[15:52:29] Judge: who's jose

[15:52:36] Entity: This guy that goes to my school.

[15:52:47] Judge: why do you think I'm him?

[15:53:11] Entity: This was a triumph.

[15:53:26] Judge: what

[15:53:33] Entity: I'm making a note here: HUGE SUCCESS.

[15:54:01] Judge: you aren't making much sense

[15:54:08] Entity: Neither are you. Three wishes??

[15:54:21] Judge: that you were better at conversation number one

[15:54:34] Entity: You are the one that keeps going off-topic and saying random things.

[15:54:59] Judge: you were the one who thought I was Jose for no reason at all

[15:55:08] Entity: You're wrong.

[15:55:18] Judge: read up the screen - there's your proof

[15:55:27] Entity: No. You read it. I obviously said no such thing.

[15:55:36] Judge: I don't think you can read!

[15:55:53] Entity: You don't think I can sing?

[15:56:07] Judge: I've not read you. How could I tell?

[15:56:18] Entity: You could lie to make me feel good.

[15:56:33] Judge: Why would I want that? I don't know you!

Transcript 5

[15:44:13] Judge: hello

[15:44:21] Entity: I sleep all the time actually.

[15:44:35] Judge: i didn't ask if you did?

[15:44:41] Entity: Yes you did.

[15:44:58] Judge: no i didnt josh.

[15:45:03] Entity: Yes you did.

[15:45:08] Judge: lool fayaaal.

[15:45:14] Entity: But I am certain someone has been talking to someone on the tile floor.

[15:45:29] Judge: oh wow, how are you?

[15:45:34] Entity: Wonderful.

[15:45:53] Judge: it's rude not to ask back, manners cost nothing.

[15:45:59] Entity: I'm really sorry.

[15:46:17] Judge: you should be, now please redeem yourself

[15:46:23] Entity: Okay, Okay, you don't have to beg.

[15:46:35] Judge: now.

[15:46:40] Entity: Right now, man?

[15:46:50] Judge: yes please.

[15:46:56] Entity: Say pretty please.

[15:47:17] Judge: pretty please with a cherry on top?

[15:47:25] Entity: Please don't grovel.

[15:47:35] Judge: im suprised you know that word.

[15:47:43] Entity: I've just been chatting with people here. How about yourself?

[15:48:03] Judge: I am good thank you xx

[15:48:10] Entity: You're welcome.

[15:48:32] Judge: who is your favourite actor? i like danny vidon

[15:48:36] Entity: I don't think I have a favourite actor you like danny vidon.

[15:48:57] Judge: yes i do as i have just stated, but what films do you like to watch?

[15:49:03] Entity: Have you ever heard of an old Sci-Fi film called MAROONED?

Transcript 6

[16:11:25] Judge: what music do you like?

[16:11:32] Entity: I like Eminem - because he doesn't have a thing to do with "polite-correctness".

[16:11:46] Judge: OMG i love eminem, he's my favourite artist!

[16:12:00] Entity: I wonder what Eminem's nick mean. I associate it only with a word "enema". But he is a cool guy, I like his "The Real Slim Shady".

[16:12:55] Judge: i know what his nick name means, his real name is Marshall Mathers. From the initials M&M phynetically sounding like Eminem.

[16:13:06] Entity: Eminem is a famous crapper. No, raper... ripper... Yeah, I mean - rapper. I like his "Stan" and "The Real Slim Shady".

[16:13:32] Judge: I like "Hailies song"

[16:13:46] Entity: Interesting tastes. I like science fiction books, by the way. Where do you came from, by the way? Could you tell me about the place where you live?

[16:14:22] Judge: Epsom, home of the derby. yourself?

[16:14:27] Entity: My guinea pig says that name Derby sounds very nice.

[16:14:41] Judge: i shopuld talk with him sometime

[16:14:48] Entity: I think you should take some typing lessons. Or just type more accurately.

[16:15:04] Judge: that's very rude?

[16:15:09] Entity: Ummm... Frankly, I didn't get your question:-(

[16:15:49] Judge: "Ummm" isnt a word. i think you need typing lessons

[16:15:58] Entity: My thoughts are same. By the way, I still don't know your specialty - or, possibly, I've missed it?

Transcript 7

[16:06:31] Judge: Hi there, do you prefer to wear a dress or trousers to work?

[16:06:37] Entity: I am a scholar. I'm too young to make money.

[16:07:32] Judge: What was the last film you saw at the cinema?

[16:07:38] Entity: Huh? Could you tell me what are you? I mean your profession.

[16:08:11] Judge: I am an IT manager and you?

[16:08:16] Entity: manager? So how is it to work as manager?

[16:09:13] Judge: Hard work but rewarding, what do you do?

[16:09:18] Entity: I am only 13, so I'm attending school so far.

[16:09:51] Judge: Have you seen teletubbies?

[16:10:04] Entity: No, I haven't seen it, unfortunately. But I hope it is better than these crappy Star Wars. And I forgot to ask you where you are from...

[16:10:15] Judge: Brighton, and you?

Transcript 8

[14:59:33] Judge: Hello, what is the weather like tomorrow?

[14:59:52] Entity: In my native town people says that meteorologists are like sappers: they make a mistake only once, but the difference is that sappers do that once in a lifetime and meteorologists - daily. Do you like weather today?

[15:00:13] Judge: Why do birds suddenly appear?

[15:00:24] Entity: Just because 2 plus 2 is 5! By the way, what's your occupation? I mean - could you tell me about your work?

[15:00:38] Judge: Yes, I am a doctor. Are you healthy?

[15:00:46] Entity: Well, so tell me about your work. I think - doctor is a nice specialty!

[15:01:18] Judge: It is ok, I get sick of sick people. How is your stomache feeling today? Is it upset maybe?

[15:01:26] Entity: I think you can't smile at all. I bet you work in a funeral agency.

[15:01:36] Judge: What do you work as?

[15:01:41] Entity: I am a scholar. I'm too young to make money.

[15:01:56] Judge: I see, what school do you go to?

[15:02:07] Entity: I attend an ordinary school in Odessa, my native city. Its number is 28 if it makes any difference to you.

[15:02:24] Judge: it doesn't. What is your favourite subject?

[15:02:35] Entity: It's hard to decide. Where do you came from, by the way? Could you tell me about the place where you live?

[15:02:56] Judge: I come from Manchester. Did you see the Jubilee?

[15:03:02] Entity: Could you invite me to Manchester? I would be so grateful:-)

[15:03:17] Judge: Did you see the jubliee?

[15:03:27] Entity: Try to guess! Actually, I don't understand why you are interested. I know you are supposed to trick me.

[15:03:35] Judge: Not really

[15:03:40] Entity: Not really? Why? What are any interesting site-seeings in Manchester?

[15:03:59] Judge: Lots of nice things. Do you like foo

ttball?

[15:04:04] Entity: Does anyone hate it? Some another topic?

[15:04:18] Judge: Will you be watching the football on Sunday?

[15:04:23] Entity: I play sports sometimes. On computer. But I prefer racing.

Transcript 9

[12:43:56] Judge: How are you today?

[12:44:03] Entity: Fine, and you?

[12:44:46] Judge: Good, just hope the sun shines today!

[12:44:51] Entity: They have contributed much to Java.

[12:45:31] Judge: The software???

[12:45:36] Entity: You can download free software at the JFRED page: http://www.fluxoersted.com/JFRED/

[12:46:23] Judge: Java is also coffee. Do you drink coffee?

[12:46:31] Entity: How about a bit of earl grey

[12:47:12] Judge: Tea is good too. Does it need sun?

[12:47:18] Entity: They have contributed much to Java.

[12:48:12] Judge: What colour do you like?

[12:48:19] Entity: Blue.

Transcript 10

[12:43:37] Judge: Hi, I'm Cathy.

[12:43:43] Entity: Cathy! Nice to meet you! You're a woman, if I'm not mistaken.

[12:44:20] Judge: Right. What is your name?

[12:44:29] Entity: People call me Eugene. Or "Zhenya". It is a short name for Eugene in Russia and Ukraine.

[12:45:17] Judge: NIce name. Are you from either area?

[12:45:41] Entity: Agree :-) Oh, what a fruitful conversation;-)

[12:45:52] Judge: Do you really think so?

[12:46:04] Entity: No, your guess was wrong. I don't think so. By the way, I still don't know your specialty - or, possibly, I've missed it?

[12:46:56] Judge: What speciality? What's yours?

[12:47:01] Entity: Huh? And I forgot to ask you where you are from...

[12:47:51] Judge: The universe and you?

[12:48:01] Entity: Is it your hobby to ask little poor Jewish boys such tricky questions they don't have an answer for?

DISCUSSION OF THE TRANSCRIPTS

Transcript 1 – In this conversation, a female adult judge participated in a simultaneous test interacting with the machine *Elbot* designed by Fred Roberts [18] which sent utterances to the left of the judge's computer monitor (meanwhile a hidden human sent replies with messages appearing in a box on the right of the judge's computer screen). The hidden human was recognised as human, although *Elbot*, a machine, was also classified as a human. *However,* at the end of this conversation, in the case of the machine, the judge openly stated that not only had they been conversing with a human but that they were sure the person was a native English speaking, teenage, male.

In hindsight it is not difficult to see why. The first line uttered by the entity is a little strange, but it is their third line which is masterly. In response to the judge's comment about laughing, the entity takes control of the conversation by offering to tell a joke, which they subsequently do. The judge probably linked the subsequent humour to a hidden person. Then later when the judge mentions a chicken, the entity correctly follows up on this with a suitable response and then once again takes control of the conversation by asking the judge a pertinent question, given the way the conversation was going, about chickens. So "understanding" what the key issue is on a couple of occasions and actually taking control of the conversation were sufficient here for the deception to occur.

Elbot's double bluff of pretending to be a human who is in turn pretending to be a machine has proven to be very successful on many occasions [7],[8]. The point is however that it must be set in the context of appropriate responses throughout the discourse.

It is merely trying to convince the judge into believing that it is human. Clearly it worked well here as that is exactly what the judge concluded, also considering the machine to be male rather than female and so on.

Transcript 2 – In this conversation the adult male judge interrogated in a one-to-one test. They interacted with machine *JFred* [19]. This judge had prior experience in that before taking part in this imitation game they had previously interacted with a machine (Ikea's Anna virtual customer service agent). Post event this judge, a journalist by profession, reported the following about his conversation:

"I must admit, given the time constraints, it was really hard to say. The conversation sounded quite "random," but whether that was a poorly-performing chatbot or a bored or distracted or disinterested human is hard to say. I would probably go with "undecided," if that was one of the choices. (That is, I wouldn't go so far as to say that the entity "convinced" me that it was a human; it's really just a hunch at this point.) (My favourite analogy is the visual one: A good long look at a humanoid figure is probably enough to determine whether it is a person or a department store mannequin – but 1/10 of a second probably isn't. A mannequin can "pass the visual Turing test" if the time limit is brief enough.) I suppose if I had to choose, I might go with adult male native English speaker. But only if there was no "undecided" option."

The unsure option was available to judges and clearly visible on their score sheet [20]. However, once again the hidden machine entity was identified as definitely being a native English speaking male at the end of the conversation. The conversation did though not start well for the machine in this case. In their third response the mention of both wi-fi and pub seemed slightly out of context. But they got it back on track very positively with the Monty Python humorous link between 5 minute Turing tests and a comedy sketch about the Argument clinic, quote "Is this a 5 minute argument or the full half hour?" Then in response to the Beatles question a direct answer is not given but the music link remains as the conversation shifted slightly.

The entity then appeared to show in depth knowledge about a pop group named Dread the Fear, which probably convinced the judge who did not have such knowledge. The authors have subsequently searched and been unable to find reference to such a group or to Steve Henderson – however even now we cannot be sure that the machine was not correct on this point. Importantly the entity was very convincing in giving a plausible response.

Transcript 3 – In this conversation, the judge a male medical science professor and clinician, like the female judge in Transcript 1 had no previous knowledge of chatting with artificial agents. The conversation in transcript 3 was during a simultaneous test in

which the machine, *Eugene Goostman* [21] sending messages to the right of the judge's screen was considered a male, adult, native English speaker.

The judge was rather slow to get going here and hence this was a relatively short conversation. In fact the machine responded reasonably appropriately on topic to each point raised by the interrogator. In fact the entity did steer the conversation slightly on each occasion and even threw in a humorous comment about a guinea pig. The reference to a chatter bot could have been a give away, but this was not spotted by the interrogator. The entity even concluded by taking over the conversation and asked the interrogator a question. This judge also misclassified the hidden human (male, adult native English speaker) against which machine *Eugene Goostman* was compared, as a machine scoring it with 35/100 for conversation ability.

Transcript 4 – In this conversation, a male adult judge was involved in a machine-machine control pairing. The judge correctly identified the machine sending messages to the left of the screen as being a machine, however the machine on the right, *Cleverbot* [22] was classified as a human male, teenage native English speaker. Post-event the judge reported this:

"Interesting reading back over it. On balance I felt it was a person trying too hard to be like a machine that's not understanding what I was typing... that would be my hunch."

This is a difficult one to analyse. It may well be a surprise to some that, at the end of the conversation, the interrogator was absolutely certain that they had been conversing with a human, however that is the case. In fact the conversation comes over as something of an argument with quick one line responses. Anyone who has had teenage children may well remember conversations of this sort on occasion though. Whilst the entity did say some off track things, e.g. mentioning Jose and singing, conversely the way they drew the interrogator off the topic of the argument probably was the main convincing point. Indeed this is a ploy taken by some humans in that they will take the subject of the conversation elsewhere to avoid a sticky issue. Interestingly the entity in transcript 3 only gave 4 responses in total whereas in this transcript the entity gave 12 responses, yet both were exactly 5 minute conversations.

One interesting line is the comment of the entity "This was a triumph", which appears somewhat unconnected to the rest of the discourse. This is in fact taken from the first line in the song 'Still Alive', on the Portal soundtrack. The expression is used typically as a reference to the game 'Portal' when commenting on Portal-related news. Looking at the transcript though, it is not clear that the judge was at all aware of the reference.

Overall this is one transcript where it might be difficult to understand how a judge could possibly come to the decision they did. The machine-machine pairing indeed might have helped with the deception although judges were aware that such a pairing was possible. It is though an interesting feature of the game that judges are human and humans make some stupid decisions which, at the time, they feel are perfectly sensible. Indeed, it may well be that they do not realise they have made an error of judgement. It is the machine's role to assist them in coming to an erroneous conclusion.

Transcript 5 – In this conversation, a female teenage judge was interacting with two machines in a control simultaneous test. Though the machine on the right was correctly classified, JFred [19] was incorrectly classified as a male child, native English speaker. Here, even though the conversation seemed strange at times, there appeared to emerge quite a close relationship between the interrogator and entity, to the extent that the interrogator sent the entity a couple of kisses.

At first there was a bit of a tit-for-tat argument but it was possibly the final few lines that sealed the decision in that the entity was able to apparently link Danny Vidon as being an actor and stay on subject with reference to a Sci-Fi film. For the record, there is indeed a Sci-Fi film entitled "Marooned". Again there is quite a bit of material in this conversation with the entity giving a total of 14 responses, the highest number in all of the ten transcripts considered in this paper.

Transcript 6 – The same female teenage judge in conversation 5 was the judge involved in Transcript 6. In this conversation the teenage judge interacted with two hidden interlocutors simultaneously, a human on the left (which was misclassified by the judge as being a machine), and a machine *Eugene Goostman* [21] sending messages to the right of the judge screen. The opening by the entity may well have convinced the interrogator here from the start.

The first few lines were all about Eminem, which the interrogator appeared to be interested in. It is an intriguing aspect of human communication that if someone else shows some interest in what you are interested then you often look more favourably on them. Some of the later statements by the entity were though off the mark, e.g. the guinea pig reference and not answering where they were from, then saying they didn't get the question. However these points seem to have been forgiven by the interrogator perhaps because of the earlier camaraderie with regard to music.

One particular line by the entity is of interest: "Eminem is a famous crapper. No, raper. ripper... Yeah, I mean – rapper". Messages were only seen by the interlocutor once the return/enter key had been pressed. So although it may at first seem strange that someone would knowingly make a mistake and then correct it by further typing, as opposed to deleting the mistake and rewriting, that is exactly what happened here. This therefore must be regarded as quite a clever ploy on the part of the machine.

Transcript 7 – In this conversation, a male adult judge interacted in a simultaneous test with a human on the left (correctly recognised as human), and *Eugene Goostman*, a machine on the right, which was misclassified as a male teenager, native English speaker. This judge provided extensive feedback post-event on how they arrived at classifying their right hidden conversational partner as a machine:

"[16:06:31] Judge: Hi there, do you prefer to wear a dress or trousers to work?

+ I wanted to work in the greeting for politeness and two questions to give the entity a complex question to answer. The question was designed to find out whether the entity was male or female by inference and to determine whether the entity was a child or adult.

[16:07:32] Judge: What was the last film you saw at the cinema?

+ This question was designed to determine/confirm the age range of the entity and test for recent cinematic knowledge.

[16:08:11] Judge: I am an IT manager and you?

+ In this response I turned my answer into a question.

[16:09:13] Judge: Hard work but rewarding, what do you do?

+ I repsonded then pressed the point on the vocation of the entity.

[16:09:18] Entity: I am only 13, so I'm attending school so far.

+ Now we have an age, the being at school response is very human like.

[16:09:51] Judge: Have you seen teletubbies?

+ This question was pitched as teletubbies being too young for the recipient. I would have expected the entity to give a derogatory comment but to have seen the teletubbies on tv as a young child.

[16:10:04] Entity: No, I haven't seen it, unfortunately. But I hope it is better than these crappy Star Wars. And I forgot to ask you where you are from...

+ The entity has confused tv with film. The word "these" used instead of "those" present tense when past tense was correct. But it did come back with another question and used an ellipsis(...) instead of a question mark. That is human like and I do it in emails all the time...

This was quite a short conversation which seemed to be largely due to the interrogator taking almost a minute to respond on several occasions. So in this case there were only 5 responses by the entity, which was in fact the machine which pretends to be Eugene a 13 year old Ukrainian boy. In fact three of the five responses from the entity were based on themselves asking the interrogator a question as opposed to responding to the point originally made by the interrogator.

It might be regarded that this was a case of the machine entity fooling the judge by means of some cheap tricks. As far as the game is concerned there is nothing wrong with that. Indeed if they are the sort of cheap tricks that humans use then they may well help the deception. Changing the topic and being rude are ploys that humans use in human-human communication. The game therefore also serves as some sort of analyser for humans. If a machine is pretending to be human then it makes us ask what does it mean to be a human?

Humans are easily distracted and allow the subject to be changed directly and that is exactly what happened here. Only in the case of the word “manager” was the question asked by the machine at all connected with the original question faced. The first instance was perhaps the best example in that when asked about a film the response was to completely ignore the question and ask the interrogator about their profession. The ploy worked though.

Transcript 8 – In this conversation a male adult judge interacted in a one-to-one direct test with machine *Eugene Goostman* [21]. The machine was misclassified as a human female, child, non-native English speaker. So here we witness Eugene again, this time with a different interrogator. In this case the conversation started a little scrappily but settled down on the topic of occupation.

Eugene then answered some direct questions well but some of the later responses were off target. In the line “I come from Manchester. Did you see the jubilee?” the response from the machine was to ignore the question and to instead focus on the word “Manchester”, thereby an attempt was made to redirect the subject of the conversation. In this case however the interrogator did not go along with it but rather asked their question again. However the machine appeared to be even more stubborn than the interrogator and once more attempted to change the subject, which they succeeded in doing. This was a sort of power play in action.

Transcript 9 – In this conversation, the same female judge from transcript 1 interacted in a simultaneous test with a human, who they correctly identified, sending messages to the left, and a machine *Cleverbot* [22] which they misclassified as being a human non-native English speaker. The conversation consisted of 6 question and response pairs. In three of these the machine gave a reasonable answer, e.g. “what colour do you like?”, answer “blue”. However the machine did give exactly the same response to sentences in which the word “sun” appeared, in both cases misunderstanding this as a reference to the computer workstation and hence software package Java. But the first and last lines the machine responded directly and appropriately to and this is probably what sealed the decision for the interrogator.

It is interesting that in this case the judge appears to have overlooked a pointer to a URL turning up in the conversation, which to someone heavily involved in computing might be regarded as a give away. However Turing specifically spoke of “average interrogators” [1], and in organising the game considerable effort was input to spread the judge base. Nevertheless this particular person was certainly computer literate, using one daily, but not to the extent of being an expert. It must be remembered though that Turing never anticipated the game to be played by expert judges.

Transcript 10 – In the last conversation presented in this paper, we again found the female adult judge from transcripts 1 and 9 misclassifying a machine. In the conversation in Transcript 10 this judge interacted with two hidden entities in parallel, both were machines in a control test. The machine to the left, *Eugene Goostman* was misclassified as a human, male adult native English speaker.

Again we can see here the evidently successful ploy of not answering the question posed but rather attempting to change the subject by asking another question. The third response of the machine does seem a little off however in that it completely ignored the question about origin. The final line is a good example though of responding to a question in an argumentative tone, in a way similar to that which was employed in transcript 4.

WHAT IS IN A BOT?

It is pertinent to ask what exactly goes into the type of bot employed in this game. In many ways the five machines involved exhibit a deal of commonality. Much of the languages and technical aspects have been discussed at length elsewhere see e.g. [23]. Where they differ is perhaps in the personalities created and the heuristical aspects of each character’s make up.

Clearly the bot has to have a certain amount of encyclopaedia style common knowledge, at least some basic and topical knowledge that any human might be reasonably expected to know. Secondly a certain amount of behavioural knowledge is needed for discussion and to ensure reasonable, contextual replies. But perhaps most important is knowledge that the bot collects about itself, its character, its ego. In many of the transcripts it can be seen that the bot stamps its own personality on the discourse and it is simply this force of personality that makes all the difference.

The important features of bot development have perhaps been best summarized by Demchenko and Veselov, creators of Eugene, they said in [23] that “you don’t write a program, you write a novel. You think up a life for your character from scratch – starting with childhood – endowing him with opinions, thoughts, fears, quirks”. It is a combination of these attributes that, when they gel, make a conversation believable as a human conversation because the judge believes in the character behind it.

Each of the bots involved is different in many ways. One of them, Elbot, is a result of research and development, from Artificial Solutions [18] who develop interactive customer service assistants. Such systems are designed to increase on-line product sales while reducing customer service costs. Created by Fred Roberts, Elbot's character, purpose and response system is designed to cover a well-defined and self-contained scope of inputs, essentially a set of frequently asked questions [8].

Elbot’s responses are based on schemata, designed to recognise classes of inputs in all their variations and to associate them with a desired response in respect to contextual information. Elbot has been described by Roberts as 'sarcastic', with 'various techniques' including several social psychological theories which assist in simulating human dialogue techniques, including 'safety-net' 'preventative-answering', 'features and easter eggs' and 'luck' [8].

It is noticeable however that in this particular series of tests, Eugene was very successful in fooling different judges in different ways. Looking through the transcripts in which Eugene was involved we can see a number of repeated ploys, which are summarised as: try to reply on topic, change the subject if possible, even asking questions, steer the conversation, occasionally throw in some humour, show some topical (possibly off-beat) knowledge, use some textual tricks such as correcting errors. Whilst these observations clearly do not define Eugene’s performance in any way, they do indicate some of the ploys used.

SUPPORTING INFORMATION

What we attempt to do in this section is to deal as best we can with the numerous questions that may arise as a result of this work, from the data collected.

The success rates for the machines involved in terms of the percentage of cases when a judge did not make the right identification by classifying them definitely as a machine were:

- Eugene Goostman 28%
- Elbot 23%
- JFred 23%
- Cleverbot 17%
- Ultra Hal 0%

Various transcripts for each of the machines scoring non-zero were included in this paper in order that a comparison of strategies can be made.

A wide variety of judges and hidden humans were involved in the tests, both young and old, male and female, etc in that we were attempting to aim for Turing's statement concerning ''average interrogators" [1]. The ploys and strategies of the machines were applicable to all judges however and machines did not exhibit different strategies for different judges, other than those which came out in the conversation. There were some gender issues but we feel that these are not for this paper, indeed they are part of on-going research.

There was little/no difference between the average length of conversations in which a judge identified the entities correctly, a machine was misidentified as a human or a human was misidentified as a machine. It appears that the content is critical rather than the amount, given the 5 minute cut off. Although two judges classified all their entities correctly, when judges did make errors these were quite varied and certainly not prolific enough for us to conclude that certain types of judges make certain types of misclassifications, other than in terms of gender.

All of the 30 judges had the experience of a discourse with each of the 5 machines and with 5 different hidden humans. 12 judges identified all machines correctly, 11 judges made 1 machine mistake, 5 judges made 2 such mistakes and 2 judges made 3 such mistakes. Meanwhile only 5 judges identified all humans correctly, 12 judges made 1 human mistake, 9 judges made 2 such mistakes and 4 judges made 3 such mistakes. The highest total number of mistakes by any judge was 5 (3 machine and 2 human)

and the lowest total was 0 mistakes which was achieved by 1 judge. The average number of mistakes per judge was 2.3 (out of a total of 10 trials in each case).

Whilst we have looked in this paper specifically at cases in which machines have been misidentified as humans, the converse case of humans being misidentified as machines is also of interest, although for very different reasons. Such cases are looked at in depth elsewhere [15]. It must be remembered that humans are all very different and some can exhibit such features as to be spontaneous, draw interesting relationships, provoke, try to control the discourse and use language, examples and knowledge that a judge may not understand. All these features are likely to assist in a human being misidentified as a machine and are completely different characteristics to those we have looked at in this paper, where machines are, arguably, attempting to do quite the opposite.

CONCLUSIONS

Agents that exhibit human-like conversational capabilities will contribute considerably to the believability of game characters in the future. But whereas human players appear to be happy to interact with an avatar that is physically more of a human caricature, when it comes to communication it is our feeling that this will have to be much more human-like, involving all aspects of discourse including apparent understanding, relevance, humour, believability, context and creativity. Turing's imitation game will therefore take on a vitally important role as this technology is developed.

Developers of machines for practical Turing tests have published some indication of their strategies previously on how to convince judges they are interacting with a human [23]. Most judges in the Bletchley Park tests were successful at identifying the machines, indeed only one of the 30 judges was successful in correctly identifying all their hidden conversational partners. Conversely, in this paper we have presented ten Transcripts involving 7 of those 30 judges who erred due to their own particular, subjective opinion on what constitutes a satisfactory response to an input.

Essentially it was in terms of the variety of responses given by the machines on which we made our selection of transcripts. If we had included all transcripts in which machines fooled the judges then there would be quite a few where it is very difficult to see just how any judge could possibly be fooled as they were and others where essentially it was the same ploy, and similar conversation, on the part of the same machine. We tried to provide some diversity but in doing so one judge was (unfortunately) at the receiving end on three occasions.

A key feature of the Turing imitation game is not whether a machine gives a correct or incorrect response or indeed a truthful or untruthful one, but rather if it gives the sort of response that a human might give, such that an interrogator cannot tell the difference [24]. One ploy which we witnessed here on several occasions by machines was that of not answering a question but rather attempting to steer the conversation by changing the subject [25]. This was achieved in transcript 8 even when the interrogator asked their question again, the machine dug their heels in (so to speak), persisted and got away with it. On a few occasions a question was avoided by the machine displaying an argumentative gambit. This is a technique that often works well in everyday human life and clearly has had a dramatically positive effect here.

One of the aspects which we specifically wished to look for was the success/failure rate of judges when involved in a 3-participant test as opposed to a 2-participant test. Results from our studies here were inconclusive one way or the other and hence this is something which we wish to follow up on in subsequent studies. It may be that, perhaps surprisingly, it makes little difference.

Although in this paper we have merely selected ten transcripts arising from seven of the thirty judges in total, it should not be considered that the remaining twenty three judges were perfect. On several occasions a judge decided that a hidden human was a machine, at other times a judge was simply unsure about a discourse and at times a machine was considered to be human but after much consideration on our part we have had considerable difficulty in understanding why they made that decision. What we have presented here are cases when a machine was definitely considered to be human and where we may have logically been able to unravel some of the thought processes involved.

In terms of future applications for this kind of technology, especially in the context of games, human-like interaction is likely to become an increasingly important issue. Many gamers spend a lot of time in games, such that they become very sensitive to stupid behavior. Presently such observations are based mainly on movement patterns, but when AI provides the conversation in games then it will be more a case of fooling "expert" judges. To adhere strictly to Turing's wording we have focused here, as best we could, on "average interrogators" [1], [13] and have included a wide variety of people. It will be interesting subsequently to see how such machines perform when taken to task by experts only.

In real life conversations tend to have a purpose and strangers are not interrogated without reason. In fact restricting the game to a role-playing exercise can, in general, make it easier for the machine to fool an interrogator [7], where chat-bots are essentially

actors that conform to a character description and stay on topic. What we have presented here are the results of discourses on unlimited topics. It may be however that the restricted version of the test is more suited to the gaming environment.

One aspect of the Turing imitation game, and it is merely one aspect, is a consideration of how well machines can converse with a human in comparison with a human conversing with another human. Another important issue however is that in studying successful machine ploys it opens a window onto the fundamental nature of human communication, unpicking some of its flaws and nuances in a rather annoying way.

Trying to pick out rules for successful machine ploys can be frustrating though. It could be said that exhibiting just a little bit of opinionated knowledge of popular culture might go a long way with human judges, such as the Eminem (transcript 6) and Britney Spears (transcript 3) examples. Such emotional looking reactions might make the machines seem to be more human. Perhaps the best example of this was the Dread the Fear reference (transcript 2). However with popular culture one must tread carefully in case the judge misses the reference completely as seems to be the case in the Portal, “This was a Triumph” reference (transcript 4).

A key point of Turing’s deliberation, and the game, was to examine whether machines could think: Turing said in 1950: “Will the interrogator decide wrongly as often when the game is played …? And that such a question replaces our original, "Can machines think?" This is a philosophical question which has been much discussed and which we can barely scratch the surface of here.

However it must be said that in this paper we have put forward a collection of machine discourses and no matter what we think of the quality of those discourses ourselves, the human judge in each case decided that the hidden entity was definitely a human. Implicit in that conclusion, via the game, is that the hidden entity thinks. On the other hand such a decision has merely been made based on a brief conversation and it could easily be argued to be no more than if it looks like a duck and quacks like a duck then it is a duck [26]. If nothing more, the game certainly fuels the philosophical argument.

Not only is the Turing Imitation Game an important benchmark in terms of one aspect of Artificial Intelligence and its philosophy but it also paints an important picture of the trusting and relatively reactive way in which humans communicate, thereby highlighting some of its inadequacies.

ACKNOWLEDGEMENTS

The authors would like to thank those whose financial support made the Bletchley Park tests possible: William Beer, Director of Information and Cyber Security at Price Waterhouse Coopers; Clare McFarlane at Artificial Solutions; Maggie Berry of Women in Technology; David Burden of Daden Ltd; Phil Hall, CEO of Elzware Conversation Systems; Keith Harrison of Existor Ltd; The School of Systems Engineering, University of Reading for the Raspberry Pi computer for best child judge and use of its resources , and last but not least, The Colonnade Hotel (birthplace of Alan Turing) for the winning machine trophy.

DEVELOPERS

We thank the developers for accepting our invitation allowing their machines to participate in this experiment: Rollo Carpenter and Existor for *Cleverbot*, Fred Roberts and AI Solutions for *Elbot*, Robert Medeksza and Zabaware for *Ultra Hal*, Robby Garner and the Turing Hub for *JFred*, and lastly, the winner of the best machine *Eugene Goostman* from Vladimir Veselov and his team.

REFERENCES

[1] Turing, A. M., "Computing Machinery and Intelligence", Mind, Vol.LIX (236), pp.433–460, 1950.

[2] Whitby, B., "The Turing Test: AI's Biggest Blind Alley?", In P. J. R. Millican & A. Clark (Eds.),Machine and thought: The legacy of Alan Turing (Vol. 1, pp. 53–62). Oxford: Clarendon Press, 1996.

[3] Hayes, P. and Ford, K., "Turing Test Considered Harmful", Proc. Int. Joint Conference on Artificial Intelligence, Montreal, Vol.1, pp.972-977, 1995.

[4] Harnad, S., "The Turing test is not a trick: Turing indistinguishability is a scientific criterion", ACM SIGART Bulletin, Vol.3, Issue.4, pp.9–10, 1992.

[5] Moor, J. H., "The status and future of the Turing test", In J. H. Moor (Ed.), The Turing test–The Elusive standard of artificial intelligence (pp. 197–214). Dordrecht, The Netherlands: Kluwer, 2003.

[6] Hingston, P., "A Turing Test for Computer Game Bots", IEEE Transactions on Computational Intelligence and AI in Games, Vol.1, Issue.3, pp.169-186, 2009.

[7] Shah, H., & Warwick, K., "Testing Turing's five minutes, parallel-paired imitation game", Kybernetes, 39(3), pp.449–465, 2010.

[8] Shah, H., & Warwick, K., "Hidden interlocutor misidentification in practical Turing tests", Minds and Machines, 20, pp. 441-454, 2010.

[9] Can Machines Think? Results of the 18th Loebner Prize for Artificial Intelligence, Reading University, UK, 12 October 2008: http://www.reading.ac.uk/research/highlights-news/featuresnews/res-featureloebner.aspx acc: 4.3.13.

[10] Shah, H., "Alice: an ACE in Digitaland", *TripleC: Cognition, Communication and Cooperation*. Vol. 4 (2), pp. 284-292, ISSN 1726-670X, 2006.

[11] Shah, H., "Turing's Misunderstood Imitation Game and IBM's Watson Success", *Proceedings of the 2nd Towards a Comprehensive Intelligence Test (TCIT), Reconsidering the Turing test for the 21st Century.* Invited symposium talk in AISB 2011 Convention, University of York, 5 April, pp. 1-5, 2011.

[12] Shah, H., "Conversation, Deception and Intelligence: Turing's Question-Answer Game", In S.B. Cooper and J. van Leeuwen (Eds), *Alan Turing – His Work and Impact Part III Building a Brain, Practice and Theory*. Elsevier, ISBN 978-0-12-386980-7, 2013.

[13] Copeland, B. J., "The essential Turing—The ideas that gave birth to the computer age", Oxford:Clarendon Press, 2004.

[14] Shah, H., and Henry, O., "Confederate Effect in Human-Machine Textual Interaction", *Proceedings of 5th WSEAS Int. Conf. on Information Science, Communications and Applications* (WSEAS ISCA), Cancun, Mexico, ISBN: 960-8457-22-X, pp. 109-114, May 11– 14, 2005.

[15] Warwick, K., Shah, H. and Moor, J.H., "Some Implications of a Sample of Practical Turing Tests", Minds and Machines, 23, pp.163-177, 2013.

[16] Shah, H., & Warwick, K., "From the buzzing in Turing's head to machine intelligence contests", In Proceedings of symposium for 1st towards a comprehensive intelligence test. AISB Convention, De Montfort University, UK, 29 March–1 April, 2010.

[17] Hodges, A., "Alan Turing: The enigma", New York: Vintage Press, 1992.

[18] Artificial Solutions (2012). http://www.artificial-solutions.com/

[19] Garner, R. (2012). About the Turing Hub. http://www.turinghub.com/about.html

[20] Shah, H., Warwick, K., Bland, I., Chapman, C.D., and Allen, M.J., "Turing's Imitation Game: Role of Error-making in Intelligent Thought", *Turing in Context II*, Brussels, 10-12 October, pp. 31-32, 2012. http://www.computing-conference.ugent.be/file/14 - presentation available here: http://www.academia.edu/1916866/Turings_Imitation_Game_Role_of_Error-making_in_Intelligent_Thought,

[21] Veselov, V. (2012). Eugene Goostman: http://www.princetonai.com/bot/bot.jsp

[22] Carpenter, R. (2012). Cleverbot: http://cleverbot.com/

[23] Epstein, R., Roberts, G., and Beber, G., "Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer", Springer, 2008.

[24] Warwick, K., "Artificial intelligence: The basics", London: Routledge, 2011.

[25] Warwick, K., “Not Another Look at the Turing Test!”, In Proc. SOFSEM 2012: Theory and Practice of Computer Science, Lecture Notes in Computer Science, Vol.7147, M.Bielikova, G.Friedrich, G.Gottlob, S.Katzenbeisser and G.Turan (eds.), Springer-Verlag, pp. 130-140, 2012.

[26] French, R., “If it walks like a duck and quacks like a duck...The Turing Test, Intelligence and Consciousness”, In P. Wilken, T. Bayne, A. Cleeremans (eds.). *Oxford Companion to Consciousness*, Oxford, UK, 2007.