Computers in Human Behavior 58 (2016) 278–295

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Full length article

Can machines talk? Comparison of Eliza with modern dialogue systems

Huma Shah [a, *], Kevin Warwick [b], Jordi Vallverdú [c], Defeng Wu [d]

[a] Department of Computing, Coventry University, Coventry, CV1 2JH, UK
[b] Deputy Vice Chancellor-Research, Coventry University, The Alan Berry Building, Coventry, CV1 5FB, UK
[c] Philosophy Department, Universitat Autònoma de Barcelona, Bellaterra (BCN), Catalonia E-08193, Spain
[d] Marine Engineering Institute, Jimei University, JMU, 176 Shigu Road, Xiamen, China

ARTICLE INFO

Article history:
Received 16 February 2015
Received in revised form
4 January 2016
Accepted 5 January 2016
Available online xxx

Keywords:
Conversation ability
Dialogue systems
Eliza
Machine
Talk
Turing test

ABSTRACT

To find if current dialogue systems use the same, psychotherapist questioning technique as Joseph Weizenbaum's 1960 natural language understanding programme, Eliza, the authors carried out an original experiment comparing five successful artificial dialogue systems, Cleverbot, Elbot, Eugene Goostman, JFred and Ultra Hal with an online version of Eliza. More than one hundred male and female participants with 1st or non-1st English language, age range 13–64, interacted with the systems over the Internet scoring each for conversation ability. Developers of the modern conversation systems show they deploy a variety of techniques to initiate and maintain dialogue learning from interactions with humans over the Internet. Statistical significance shows these dialogue systems are an improvement on their predecessor. Embedded on the web affording round-the-clock interaction the nature of artificial dialogue systems is evolving as these systems learn from the way humans converse. The uses of modern Elizas are proven successful as virtual assistants in e-commerce; their conversational basis is already extending into education. What we can say is modern artificial dialogue systems do talk. They are able to participate in conversation in a way their predecessor Eliza could not: they are able to share personal opinions, relay experience of family dramas, be relevant, but also be vague, and mislead just as humans do.

1. Introduction

Artificial dialogue systems, such as *Ask Anna*, Ikea's "most versatile employee" (Artificial Solutions, 2015), Sky's *Ella* and O2's *Lucy* (Fig. 1) are extensively deployed in e-commerce as virtual-bodied customer service agents. Disembodied 'pocket assistants' equip smart 'phone users with dialogue, for example in Apple's *Siri* (2013), iFree's 'Everfriend' *Spoony* character (2013), and email-reading Microsoft's *Cortana* (FT, 2014). Google *Now* (2015)'s provides its users with text and visual information through organised cards displayed on a variety of Android platforms (PC, tablet, smart 'phone and watch). The roots of these interactive 'talking machines' lie in Weizenbaum (1966)'s *Eliza* programme which facilitated interaction between human and machine through text-based communication (Shotwell, 1983). *Eliza*'s question-answer format can be said to follow Alan Turing's *viva voce*, one-to-one direct questioning test to examine machine thinking (Turing, 1950).

What their increasing deployment as "helpful agents" (AI Artificial Solutions, 2011) do not inform on is whether modern conversational systems deploy the "usual, give-away, tiring, Eliza-ish strategy" (Floridi, Taddeo, & Turilli, 2009). The purpose of this exercise was to find this out during the preliminary phase of an experiment implementing Turing's two tests for his imitation game (Shah, 2013; Shah, Warwick, Bland, Chapman, & Allen, 2012, Shah, 2011).

1.1. Alan Turing centenary 2012

In the period leading up to the 100th anniversary of the birth of Alan Turing in 2012, and in preparation for a unique public centenary event staging Turing's imitation game (Shah, 2013; Turing100, 2012) at Bletchley Park UK on Turing's birthday, 23 June (Warwick & Shah, 2013b, 2014a, b, c), the authors staged a pre-event experiment comparing five of the best modern dialogue systems with a web-version of *Eliza*. This gave students and non-students from the authors' countries an opportunity to interact with artificial dialogue systems on anonymous websites (Turing100, 2012; STEMNET, 2012). In this way participation from people who would not be attending the UK event was facilitated. In this paper we present the findings from

* Corresponding author.
E-mail addresses: ab7778@coventry.ac.uk (H. Shah), aa9839@coventry.ac.uk (K. Warwick), jordi.vallverdu@uab.cat (J. Vallverdú), defeng@jmu.edu.cn (D. Wu).

http://dx.doi.org/10.1016/j.chb.2016.01.004
0747-5632/

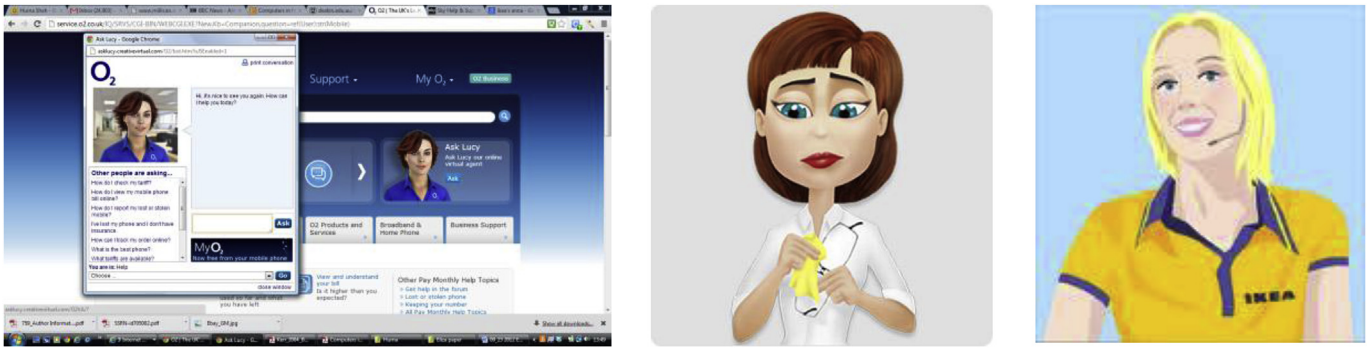

Fig. 1. Virtual Assistants: (left) O2's Lucy; (middle) Sky's Ella,[1] (right) Ikea's Ask Anna Europe version.

Table 1
Modern conversational systems used in this experiment.

System	Developer/Commercial arm	Competitions won
Cleverbot	Rollo Carpenter/Existor: https://www.existor.com/en/	Win: 2010 BCS SGAI Machine Intelligence contest Win (as *Jabberwacky*): Loebner Prize for Artificial Intelligence twice: 2005 and 2006
Elbot	Fred Roberts Artificial Solutions: http://www.artificial-solutions.com/	Win: 2008 Loebner Prize Win: 2003 Chatterbox Challenge
Eugene Goostman	Team led by Dr. Vladimir Veselov	2nd placed in 2008 Loebner Prize 2nd placed in 2005 Loebner Prize
JFred/TuringHub	Robby Garner	Win: Loebner Prize twice (1998 and 1999)
UltraHal	Robert Medeksza Zabaware https://www.zabaware.com/assistant/	Win: 2007 Loebner Prize

that online phase, the one-to-one interaction method where human judges talked with and scored six systems for conversational ability.

1.2. Selecting the machines

An online version of *Eliza* lent itself to comparison with modern conversational systems. Selection of the comparator conversational systems was based on developers' expertise in producing successful performance in previous machine intelligence and Turing test competitions (Table 1).

To compare against *Eliza*, *Cleverbot*, *Elbot*, *Eugene Goostman*, *JFred* and *Ultra Hal* systems were selected as a result of their successes in human-machine interaction contests (Table 1) and their developers' willingness to participate in this exercise.

In section 2, we begin by tracing the background of *Eliza*, the first programme that afforded conversational interaction between a human and a computer, from its roots in Turing's imitation game, commonly known as the Turing test. Following, in section 3 a review of *Eliza* and modern *Elizas* (Cleverbot, 2013; Elbot, 2013; Eugene Goostman, 2013; JFRED, 2013; TuringHub, 2013; Ultra Hal, 2013) is presented. In section 4 we present the experiment comparing *Eliza* with five artificial conversationalists. A discussion of the results is found in section 5. The paper concludes, in contrast to Floridi et al.'s claim of decades of *Eliza* type implementation (2009) her descendants can talk and *are* better conversationalists than their predecessor. However, their purpose of 'all-round chatters' is different from Weizeinbaum's single domain artificial psychotherapist. The authors do not own the intellectual property of any of the six systems presented in this paper, we are privileged that the developers of the five comparison systems were willing to share some technical information. For this reason, and the commercial nature of these systems, the authors are not able to provide more than what was shared by the Developers. However we point the reader to chapters in Epstein, Roberts and Beber's book 'Parsing the Turing test' (Copple, 2008; Demchenko & Veselov, 2008; Garner, 2008; Hutchens, 2008; Wallace, 2008), and 'Turing on Emotions' (Roberts, 2014).

Eugene Demchenko and Vladimir Veselov: Who Fools Whom? The Great Mystification, or Methodological Issues on Making Fools of Human Beings
Robby Garner: The Turing Hub as a Standard for Turing test Interfaces
Jason Hutchens: Conversation Simulation and Sensible Surprises
Richard Wallace: The Anatomy of A.L.I.C.E and to the 'International Journal of Synthetic Emotions' Volume 5, issue 2 for
Fred Roberts: The Social Dialogue of Simulation as Applied in Elbot.

2. Turing test

Having introduced a game in which successful imitation of human-like responses could induce wrong identification (Shah, 2013), Turing claimed that the question-answer[2] method was "suitable for introducing almost any one of the fields of human endeavour" that the interrogator might wish to include" (1950: p. 435). The interrogator is not allowed to seek any practical demonstrations during questioning (p. 446), no matter how much the hidden entity may boast about their appearance or prowess (see Fig. 2). Turing pointed out the limitations of the machines at that time: "there will be some questions to which it will either give a wrong answer, or fail to give an answer at all however much time is allowed for a reply" (p. 444). Turing wrote "I am often wrong, and the result is a surprise for me" (p. 451), but, he asked, would it be fair to deem machines *worse* for *not* making mistakes? (p. 448).

Turing supposed closed questions, with 'yes' or 'no' answers were more appropriate to begin with than the type of questions

[1] O2 Lucy: http://asklucy.creativevirtual.com/O2/bot.htm?isJSEnabled=1 accessed: 23.9.12. SKY's Ella: http://www.sky.com/mysky/latestnews/article/my-sky-updates/bcde-u/index.html accessed: 23.9.12. IKEA's Anna: http://www.cookylamoo.com/boringlikeadrill/2005/06/i-do-not-understand-what-you-wants-to-formulates-you-gladly-on-something-else-ways.html accessed 23.9.12.

[2] IBM claim their advanced question-answer technology, new super computer Watson, understands questions in natural language successfully testing it against humans in 2011 competing against humans in the US TV quiz show *Jeopardy!*: http://www.nytimes.com/2010/06/20/magazine/20Computer-t.html?pagewanted=all accessed: 23.9.12.