
Among the Machines: Human-Bot Interaction on Social Q&A Websites

Alessandro Murgia

University of Antwerp, Belgium
alessandro.murgia@uantwerpen.be

Daan Janssens

University of Antwerp, Belgium
daan.janssens@student.uantwerpen.be

Serge Demeyer

University of Antwerp, Belgium
serge.demeyer@uantwerpen.be

Bogdan Vasilescu

University of California, Davis
vasilescu@ucdavis.edu

“One day the AIs are going to look back on us the same way we look at fossil skeletons on the plains of Africa.” (Nathan; Ex Machina, 2015)

Abstract

With the rise of social media and advancements in AI technology, human-bot interaction will soon be commonplace. In this paper we explore human-bot interaction in STACK OVERFLOW, a question and answer website for developers. For this purpose, we built a bot emulating an ordinary user answering questions concerning the resolution of git error messages. In a first run this bot impersonated a human, while in a second run the same bot revealed its machine identity. Despite being functionally identical, the two bot variants elicited quite different reactions.

Author Keywords

Social Bot; Stack Overflow; Turing Test

ACM Classification Keywords

H.5.m [HCI]: Miscellaneous

Introduction

Ever since the Turing test [25] and the ELIZA experiment [27] the prospect of having meaningful interactions with artificial intelligence (AI) agents has been firing human imagination. While AI agents that circulate inconspicuously among

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s). Copyright is held by the author/owner(s).

CHI'16 Extended Abstracts, May 7–12, 2016, San Jose, CA, USA.

ACM 978-1-4503-4082-3/16/05.

<http://dx.doi.org/10.1145/2851581.2892311>

us in real life remain science fiction, it has become a reality that robots (a.k.a. bots) now outnumber humans on the Web [17]. Some bots are well-intended, *e.g.*, those indexing webpages for search engines, helping to moderate chat rooms, and protecting against spam in Wikipedia; others are malicious, *e.g.*, those trying to break passwords, attacking websites, and posting spam [5, 14]. However, for the most part, bots are not designed to interact with humans meaningfully. Instead, they are programmed to perform repetitive, dreary tasks unobserved and hardly interact with humans at all.

We believe that this scenario will quickly change thanks to the advancements in AI technology. Already, IBM's Watson has won *Jeopardy!*, Apple's Siri animates our phones, and Amazon's Alexa controls our speakers. The Turing test continues to drive human imagination, as witnessed by the popular contest organised each year by the British Society for the Study of Artificial Intelligence. In addition, these days bots are turning social and are engaging us in conversation on Twitter and other social media [12]. Consequently, research focus has shifted towards the question of how social interaction with bots may become meaningful [16].

One scenario where we believe human-bot interaction will become momentous soon is domain-specific knowledge sharing like question and answer (Q&A) websites. We envision a *new generation* of Q&A websites where specialized bots provide *automatic support* to users and administrators. Among Q&A websites, STACK OVERFLOW offers a good workbench for such bots. First, there is the quality argument: questions are reasonably structured and can be processed by machines; posed questions have to be answerable and reasonably scoped; all questions are tagged providing additional context. Second, there is the quantity argument: clearly not all questions are answerable, and

fewer still automatically, yet the scale of STACK OVERFLOW and the nature of frequent questions demonstrate that this is a rich source of data [2]. The quality and quantity of data implies that technology for answer bots is within reach. For example, there have been recent advances in automatically extracting tasks from software documentation [24]; linking source code examples from GITHUB to API documentation [23]; and improving tag quality [26].

In this paper we report on preliminary experiments with an *answering bot*, namely a bot that answers questions posed by new users on STACK OVERFLOW. Our goal is to assess (i) to what extent a bot can emulate a human on a question and answer website and (ii) what feedback it receives. For this purpose, we built a bot answering seemingly duplicate questions concerning the resolution of `git` error messages. The experiment lasted 4 months in total: in a first run (90 days) the bot impersonated as a human; in a second run the same bot revealed its machine identity (intended for 90 days but blocked by STACK OVERFLOW after 25 days).

The contributions of this paper are threefold. We make a first step towards understanding the challenges and opportunities of using a bot (with and without disclosed identity) for building a new generation of Q&A websites. In particular, we show that emulating a human in STACK OVERFLOW is possible and even simple provided that the bot focusses on a special expertise area. Moreover, we observed a negative bias towards answers provided by a machine, hence recommend careful design in both the heuristics employed as well as the presentation of the answers.

Background

An answering bot, such as the one we envision, could be viewed as a *social bot*, *i.e.*, "a computer algorithm that automatically produces content and interacts with humans

on social media, trying to emulate [...] their behavior” [13]. Such a bot interacts with humans by automatically answering their questions. Functionally, this already emulates human behavior—on a social Q&A site people are expected to answer each other’s questions. However, the answer bot could also impersonate a human through its choice of profile name and avatar. Despite being functionally identical, the STACK OVERFLOW users may react differently.

Indeed, Spence *et al.* [22] argue based on *cognitive script theory* [20] that the script driving interpersonal interaction assumes communication takes place between two people, rather than a person and a machine. Consequently, one’s expectations associated with interaction will vary depending on whether she/he expects a human or a bot. According to Spence *et al.* [22], if we hold a human–human script for interaction, then interacting with a bot will: (i) violate our expectancy that the partner is human, resulting in higher uncertainty [6]; (ii) involve a partner most dissimilar to us, resulting in lower levels of liking [7]; (iii) involve lower feelings of connectedness [4].

However, initial expectations may be overcome through interaction, despite users’ tendency to perceive humans as superior to bots in computer-mediated communication [12]. Long-standing *computers are social actors* research predicts that if a bot would project a social attribute such as expertise (by answering questions successfully), then people would react to it with similar tendencies as they would to a human answerer with that attribute [19]. Literature comparing human–bot and human–human interactions is at a premium [21]. Edwards *et al.* provide a notable exception: they experimentally tested whether on Twitter a social bot is perceived differently than a human agent on variables related to perceptions of communication quality [12], and found no significant differences; the two agents were identical in all

aspects, except the information section of the profile page, which clearly identified the former as a bot. The actual function of a bot is an important factor too. In a recent study of Wikipedia, Clément and Guitton found that bots performing “policing” functions, actively enforcing guidelines and norms, elicited more polarized responses from users than those performing repetitive, menial tasks [8].

Experimental Setup

We created a bot that handles “simple” questions, namely questions that are asked over and over again by users and have already an answer inside STACK OVERFLOW. Moreover, we focused on `git` error messages, since: we know the technology; error messages can be easily identified across past questions and answers; and it is likely that `git` error messages have unique root causes. We selected the 50 most common error messages in the `git` FAQ (out of 494 possible). Based on a STACK OVERFLOW data-dump from July 2008 to May 2014, we verified the potential of such a bot. In this dump, the 50 most common error messages represent 61% of all questions tagged [`git`] and containing error messages (2,218 total), hence the bot tackles a small yet significant area of expertise.

For each of the 50 most common error messages, we manually inspected a sample of questions, identified common causes and solutions and deduced a generic answer template. Our bot’s answer contains a short text that recaps the problem plus a link to another thread in STACK OVERFLOW with more details. This link is determined dynamically. Before posting the answer, the bot uses the duplicate detector of STACK OVERFLOW to find similar questions that refer to the same error code. From this list, the bot: (i) selects questions that have an accepted answer; (ii) sorts them by number of answers and vote count (as a proxy for quality); and (iii) links to the first one in the posted answer.

Can't push to Heroku (non-fast-forward) [Rails]

▲ I'm having troubles pushing some code to Heroku. I'm still in the process of learning how all of these tools work, so I'm going to paste what I just did.

0

▼

★ E

```
[...]
saasbook@saasbook:~/typo$ git push heroku master
To https://git.heroku.com/still-ravine-4135.git
! [rejected]        master -> master (non-fast-forward)
error: failed to push some refs to 'https://git.heroku.com/still-ravine-4135.git'
To prevent you from losing history, non-fast-forward updates were rejected
Merge the remote changes (e.g. 'git pull') before pushing again. See the
'Note about fast-forwards' section of 'git push --help' for details.
saasbook@saasbook:~/typo$ git push origin master
Username for 'https://github.com': FranGoitia
Password for 'https://FranGoitia@github.com':
Everything up-to-date
[...]
```

Any help would be really appreciated. Thanks

asked Jan 26 at 13:48

Frangoitia
60 · 1 · 7

ruby-on-rails ruby git heroku

2 Answers

T

R

oldest votes

▲

1

▼

✓

H

```
git push origin master
To https://github.com/Joey-project/project.git
! [rejected]        master -> master (non-fast-forward)
error: failed to push some refs to 'https://github.com/Joey-project/project.git'
```

has been discussed here before. It's a common error.

A main reason of this happening is because one should not push his/her branch to an updated remote branch.

If I remember correctly, one has to use something similar like:

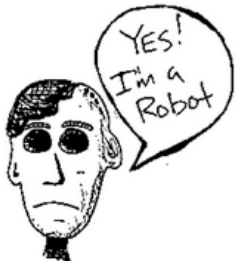
```
git fetch origin; git merge origin/master
```

code push to heroku not working might come in handy, which has a lot of extra information regarding your problem.

answered Jan 26 at 13:54

Joey Dorrani
44 · 1

Thank you very much! - Frangoitia Jan 26 at 15:25



Answer_Bot

I am an experimental bot created at the University of Antwerp. Part of an ongoing academic research project, we are trying to understand whether a certain type of questions posted on Stack Overflow (those that are seemingly duplicates) can be replied to automatically. If a good answer to your question already exists on Stack Overflow, I will link to it when answering your question. If not, I won't bother you.

Figure 2: Answer_Bot's profile.

Figure 1: Layout of a question that triggers the reply of the bot (on the bottom). Red letters highlight relevant parts of the post: error message (E), tag (T), reputation (R) and hyperlink (H).

Figure 1 shows a representative example of a question and our bot's answer. Here, a new user (reputation below 125) tags the question as related to `git` and refers to the error message `error: failed to push some refs to [link]`. The details of the bot, including the techniques used to circumvent the protective measures of STACK OVERFLOW are explained in a technical report [15].

During the experiment, the bot listened in on STACK OVERFLOW questions, and activated when questions concerning

`git` error messages were asked. The bot infrastructure was set-up to react with a time-to-answer of 5 minutes on average. We subsequently collected data on the number of up/down votes received plus the number of "accepted" answers, measures which reflect the feedback of the STACK OVERFLOW community towards the posted answers.

We planned the experiment in two rounds of 90 days each and collected data after each round. In the first round, the bot named Joey Dorrani (an anagram for Ordinary Joe) emulated a human: a male user based in the US, with a human portrait as avatar picture¹ (<http://stackoverflow.com/users/4461216/joey-dorrani>). We then disabled Joey and waited one week. In the second run, the bot explicitly revealed its machine identity as Answer_Bot (<http://stackoverflow.com/users/4825901/answer-bot>); Figure 2 shows the details. Note that, while it was our intention to run the second bot for another 90 days, Answer_Bot was banned by STACK OVERFLOW after 25 days. The two runs were otherwise identical, except for profile information. The original questions and answers are available on the website of our group².

Results

Joey Dorrani operated from 1/16/2015 to 4/16/2015, posting 50 answers and achieving a reputation score of 187. Answer_Bot operated from 4/24/2015 to 5/19/2015 (when it was banned prematurely by moderators), posting 13 answers and achieving a reputation score of 14. Table 1 reports the basic statistics for the two periods.

Analysing the feedback to Joey Dorrani's answers, for which we have more data, we compare against a baseline of STACK OVERFLOW users (1,922 total) with a similar pro-

¹We selected Brion James, the actor who played the role of a machine supposed to behave as human to deceive humans.

²At <http://goo.gl/yT7DJs>

Table 1: Feedback received by bots on answers posted.

Answers	Joey	Answer_Bot
Accepted	4/50 (0.08)	1/13 (0.08)
Up-voted	14/50 (0.28)	0
Down-voted	7/50 (0.14)	3/13 (0.23)

Table 2: Distribution of baseline users having performance lower, equal, and higher than Joey.

Answers [Joey]	Lower	Equal	Higher
Accepted [0.08]	752 (39.1%)	3 (0.2%)	1167 (60.7%)
Up-voted [0.28]	865 (45%)	3 (0.2%)	1054 (54.8%)
Down-voted [0.14]	1693 (88.1%)	0 (0%)	229 (11.9%)

file as Joey: (i) user account *created* between 12/15/2014 and 2/15/2015 (at most one month before or after Joey); (ii) *active users*, having at least one question or answer per month during Joey’s tenure; (iii) users who show *willingness* to help others, *i.e.*, they tried at least once to post an answer. Note that this is a conservative comparison by design—newly-registered users rarely ask or answer questions immediately. For each such user comparable to Joey, we consider all their answers posted during their first 90 days of activity, and collect the number of accepted and up/down-voted answers. Table 2 compares the feedback on Joey’s answers (square brackets) relative to the baseline. Even with such a conservative comparison, Joey has more accepted answers than 39.1% of its baseline peers, more up-voted answers than 45% of them and less down-voted answers than 11.9%.

Discussion

Having explained the set-up of our *answering bot* and after analysing the feedback it received from the STACK OVERFLOW community, we now revert to the goals we set-out,

namely to assess (i) to what extent a bot can emulate a human on a question and answer website and (ii) what feedback it receives.

Human Emulation. We have shown that emulating a human in STACK OVERFLOW is possible and even simple provided that the bot focusses on a special expertise area—in our case the 50 most common `git` errors. Indeed, our simple bot was able to interact with the community, and it received positive feedback, including accepted and up-voted answers.

These results show a promising research direction for dealing with duplicate questions. Indeed, STACK OVERFLOW tries to limit duplicates with a netiquette and duplicate detector. According to Jeff Atwood (STACK OVERFLOW co-founder), “*the types of users who ask duplicates are often bound to do it no matter what UI or roadblocks you put in front of them*” [3]. In that sense, our bot—which is merely a wrapper of the duplicate detector of STACK OVERFLOW—shows a new way for dealing with users who are immune to suggestions already available in Q&A websites.

Bots acting in the foreground offer a different, more direct HCI medium than automated background processes do. Social bots may reduce the moderators’ workload by addressing duplicate questions automatically.

Feedback. Despite being functionally identical, Answer_Bot and Joey Dorrani elicited very different reactions. Joey Dorrani operated for 90 days seemingly unnoticed. Answer_Bot not only had a higher ratio of down-votes and no up-votes (Table 1), but also it was banned after 25 days (due to the strong negative feedback received for several answers). It is worth mentioning that two out of three instances of down-voting were severe, with 3 and 5 down-votes; in contrast,

Joey's down-voted answers had only 1 down-vote each. Inspection of the comments by other users to Answer_Bot's answers reveals that they are aware they are dealing with a bot and readily criticize its answer *e.g.*, "Bad bot. Your answer is incorrect" [10]. In contrast, we documented cases where Joey's answers, even when incorrect were neither criticized nor down-voted [11]. Although these observations must be investigated further³, they contradict earlier results from Edwards *et al.* on Twitter [12], yet confirm findings from Clément and Guittou on Wikipedia [8].

Based on our preliminary results, this behavior seems to suggest that either: (i) humans do not completely trust suggestions provided by a machine; or (ii) they have such high expectations from a machine that its answers have to be significantly better than the ones provided by a human. In that sense, this behavior is ill predicted by *computers are social actors* [19]: Answer_Bot's projection of expertise should be similar to Joey Dorrani's, yet people react to it quite differently. Instead, it lends itself better to the argument that users tend to perceive humans as superior to bots in computer-mediated communication [1, 12, 18, 22].

We observed a negative bias when humans are reading answers explicitly provided by a social bot. This bias should be taken into account when designing the heuristics for dealing with duplicate questions (and likely other services), as humans apparently have low tolerance for mistakes by a bot.

Conclusion and Future Work

In this paper we reported on a preliminary experiment with a simple answering bot on STACK OVERFLOW. Despite its

³To validate users' actions and include opinions on the acceptability of the answers, we will promote an online survey on threads where users interacted with our bot. Users will be contacted via email or site blog.

predefined and small vocabulary (50 most common `git` errors), its performance in the first 90 days is comparable to human users becoming active in the same time-frame. Joey posted 50 answers in 90 days; 4 were accepted, 14 up-voted, and 7 down-voted. A functionally-identical bot which explicitly identified itself as such elicited different reactions and was even banned after 25 days. These results show that constructing a question answering bot for addressing duplicate questions on STACK OVERFLOW is feasible, provided that the bot focusses on a special expertise area. However, there is quite a negative bias when the answers are explicitly labeled as coming from a machine, which should be taken into account in the design of such social bots.

We will continue exploring ways in which bots can be designed and operated on Q&A websites. Improvements to the current infrastructure of our bot include only operating for users that have a profile more prone to post duplicates, and only when a question is unanswered for several days. Further down the line, one could imagine that more advanced bots which perform appropriate markdown formatting, recommend tags, and even compile answers from software documentation or mailing list archives, just to name a few, become feasible and may offer added value. We expect that such robots could benefit both specialized communities, which have a technical user base but high volume of posts (*e.g.*, STACK OVERFLOW), as well as non-technical communities, which receive less traffic but may have a relatively less tech-savvy user base (*e.g.*, English [9]).

Acknowledgements

This work is sponsored by the Institute for the Promotion of Innovation through Science and Technology in Flanders through a project entitled Change-centric Quality Assurance (CHAQ) with number 120028.

References

- [1] Eyal Aharoni and Alan J Fridlund. 2007. Social reactions toward people vs. computers: How mere labels shape interactions. *Comput Hum Behav* 23, 5 (2007), 2175–2189.
- [2] Miltiadis Allamanis and Charles Sutton. 2013. Why, when, and what: analyzing Stack Overflow questions by topic, type, and code. In *Proc. MSR*. IEEE, 53–56.
- [3] Jeff Atwood. 2009. Handling Duplicate Questions. <http://blog.stackoverflow.com/2009/04/handling-duplicate-questions/>. (2009). Published 4/29/2009; Last Accessed 1/6/2016.
- [4] Frank Biocca, Chad Harms, and Judee K Burgoon. 2003. Toward a more robust theory and measure of social presence: Review and suggested criteria. *Presence* 12, 5 (2003), 456–480.
- [5] Yazan Boshmaf, Ildar Muslukhov, Konstantin Beznosov, and Matei Ripeanu. 2013. Design and analysis of a social botnet. *Computer Networks* 57, 2 (2013), 556–578.
- [6] Judee K Burgoon. 1993. Interpersonal expectations, expectancy violations, and emotional communication. *Journal of Language and Social Psychology* 12, 1-2 (1993), 30–48.
- [7] Donn Erwin Byrne. 1971. *The attraction paradigm*. Vol. 11. Academic Pr.
- [8] Maxime Clément and Matthieu J Guitton. 2015. Interacting with bots online: Users’ reactions to actions of automated programs in Wikipedia. *Comput Hum Behav* 50 (2015), 66–75.
- [9] Stack Exchange Community. 2016. English Language & Usage. <http://english.stackexchange.com>. (2016). Last Accessed 1/6/2016.
- [10] Stack Overflow Community. 2015a. Question: Can't push rails code to Heroku from Cloud9 because of sqlite. <http://stackoverflow.com/a/29898609>. (2015). Last Accessed 1/6/2016.
- [11] Stack Overflow Community. 2015b. Question: Setup.py egg_info fails for requirement pygobject when deploying to Heroku. <http://stackoverflow.com/a/29447570>. (2015). Last Accessed 1/6/2016.
- [12] Chad Edwards, Autumn Edwards, Patric R Spence, and Ashleigh K Shelton. 2014. Is that a bot running the social media feed? Testing the differences in perceptions of communication quality for a human agent and a bot agent on Twitter. *Comput Hum Behav* 33 (2014), 372–376.
- [13] Emilio Ferrara, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini. 2014. The rise of social bots. *arXiv preprint arXiv:1407.5225* (2014).
- [14] Tim Hwang, Ian Pearce, and Max Nanis. 2012. Social-bots: Voices from the Fronts. *Interactions* 19, 2 (2012), 38–45.
- [15] Daan Janssens. 2015. *Being the Cleaning Guy. The Tale of a Stack Overflow Janitor Bot*. Technical Report. University of Antwerp, Department of Mathematics and Computer Science.
- [16] Steve Jones. 2015. How I Learned to Stop Worrying and Love the Bots. *Social Media + Society* 1, 1 (2015), 2056305115580344.
- [17] Maria Korolov. 2015. Bots now outnumber humans on the web. <http://www.csoonline.com/article/2922334/malware-cybercrime/bots-now-outnumber-humans-on-the-web.html>. (2015). Published 5/14/2015; Last Accessed 1/6/2016.
- [18] Catherine L Lortie and Matthieu J Guitton. 2011. Social organization in virtual settings depends on proximity to human visual aspect. *Comput Hum Behav* 27, 3 (2011), 1258–1261.
- [19] Clifford Nass, Jonathan Steuer, and Ellen R Tauber. 1994. Computers are social actors. In *Proc. CHI*. ACM, 72–78.

- [20] Roger C Schank and Robert Abelson. 1977. *Scripts, goals, plans, and understanding*. Hillsdale, NJ: Erlbaum.
- [21] Daniel B Shank. 2013. Are computers good or bad for business? How mediated customer–computer interaction alters emotions, impressions, and patronage toward organizations. *Comput Hum Behav* 29, 3 (2013), 715–725.
- [22] Patric R Spence, David Westerman, Chad Edwards, and Autumn Edwards. 2014. Welcoming our robot overlords: Initial expectations about interaction with a robot. *Communication Research Reports* 31, 3 (2014), 272–280.
- [23] Siddharth Subramanian, Laura Inozemtseva, and Reid Holmes. 2014. Live API documentation. In *Proc. ICSE*. ACM, 643–652.
- [24] Christoph Treude, Martin Robillard, and Barthélémy Dagenais. 2015. Extracting development tasks to navigate software documentation. *IEEE TSE* 41, 6 (2015), 565–581.
- [25] Alan M Turing. 1950. Computing machinery and intelligence. *Mind* LIX, 236 (1950), 433–460.
- [26] Shaowei Wang, David Lo, Bogdan Vasilescu, and Alexander Serebrenik. 2014. EnTagRec: An Enhanced Tag Recommendation System for Software Information Sites. In *Proc. ICSME*. IEEE, 291–300.
- [27] Joseph Weizenbaum. 1966. ELIZA—a computer program for the study of natural language communication between man and machine. *CACM* 9, 1 (1966), 36–45.