Factors Influencing Approval of Wikipedia Bots

Ayse Dalgali
Syracuse University School of Information Studies
ayocal@syr.edu

Kevin Crowston
Syracuse University School of Information Studies
crowston@syr.edu

Abstract

Before a Wikipedia bot is allowed to edit, the operator of the bot must get approval. The Bot Approvals Group (BAG), a committee of Wikipedia bot developers, users and editors, discusses each bot request to reach consensus regarding approval or denial. We examine factors related to approval of a bot by analyzing 100 bots' project pages. The results suggest that usefulness, value-based decision making and the bot's status (e.g., automatic or manual) are related to approval. This study may contribute to understanding decision making regarding the human-automation boundary and may lead to developing more efficient bots.

1. Introduction

In the present era, we witness automation in many domains through tools capable of performing tasks much faster than humans. Increasingly though, automated systems are expected to work with and support humans rather than simply replacing them. One of the most widespread examples of such a tool is the bot, a program that perform automated tasks over the Internet. There are different types of bots, such as trading bots (e.g., chatbots in customer service, help bots in commercial company websites), social media bots (e.g., Facebook, Twitter, Reddit bots) and social bots chatting to human users (e.g., Eliza representing a mock Rogerian psychotherapist).

As with any new technology, an important question is user acceptance and factors that predict acceptance. Technology acceptance is one of the most studied concepts in information systems research with a rich literature. However, bots seem likely to have a distinctive set of acceptance factors. For example, ease of use may be less relevant for a tool that works by itself. Accordingly, our goal in this paper is to identify factors in the acceptance of a novel technology.

In this study, we focus on Wikipedia bots, those that support Wikipedia editors by editing articles or managing edits. Bots that edit Wikipedia undertake various routine tasks, such as checking spelling mistakes, moving categories or automatically importing batches of entries from a public/GFDL database. Priedhorsky et al. [1] note that the list of top editors by edit count is filled with bots: in 2014, Wikipedia bots carried out approximately 15% of the edits on all language editions of the encyclopedia [2]. Bots are also used to deal with the more than 155,000 edits made per day,[1] e.g., finding and reverting changes by suspicious new users or protecting pages from vandalism.

In the case of Wikipedia bots, acceptance is a formal process, making the factors predictive of acceptance visible for study. Before a bot can be deployed, the Bot Approvals Group (BAG) must approve the bot's purpose and implementation. The BAG was founded in 2004 and includes Wikipedia bot developers and non-developers. It is tasked with reviewing proposals for new bots for compliance with the community-authored Bots policy [3].

Figure 1 shows the BAG's decision-making process for approval or disapproval of a bot, drawn from the wiki/Help: Creating a bot page[2] and from the Wikipedia: Bots/Requests for approval project pages of the bots. After reviewing proposals, bots may be accepted for a trial implementation. After implementation, BAG members and the operators of the bot discuss the bot's implementation and testing results. Based on those discussions, the bot is finally approved or denied for regular use. Much of this approval process occurs online in Wikipedia-based discussions, such as the Wikipedia: Bots/Requests for approval project page of each bot.

Although the fundamental features that are expected from a bot are presented in the Wikipedia Bots policy, such as being harmless, useful, not consuming resources unnecessarily[3], etc., for bot developers, it may be difficult to understand how the
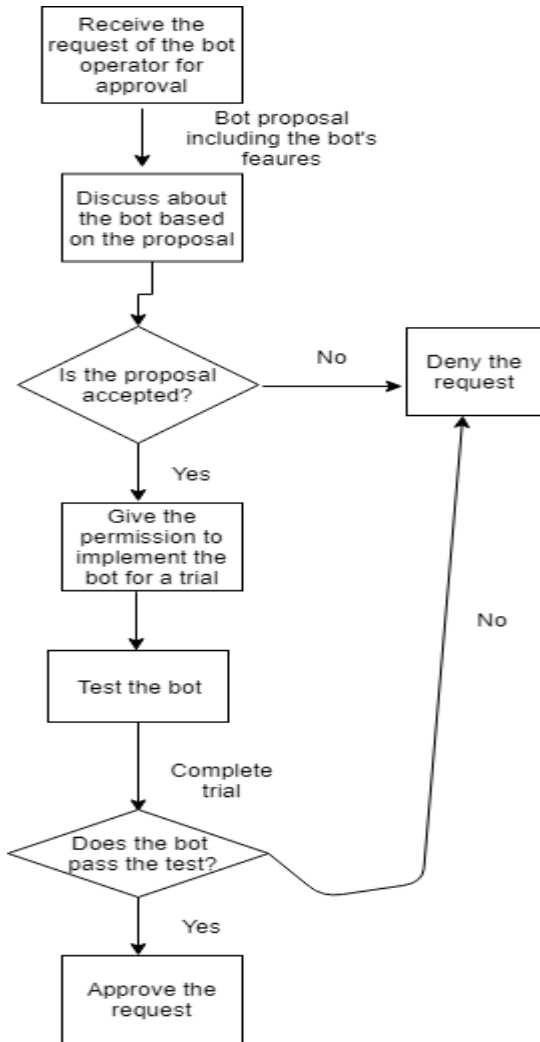

Figure 1. BAG's decision-making process to approve or deny a Wikipedia bot

BAG evaluates whether a proposed bot meets those requirements or whether the criteria expressed in the stated Wikipedia Bots policy are the same as those examined in the discussions. Moreover, the BAG may consider other factors in addition to those fundamental requirements while making decisions. Hence, examining the discussions in which each bot is evaluated will shed light on the actual evaluation factors of the BAG. By analyzing and interpreting 100 discussions in the Wikipedia: Bots/Requests for approval project page of 100 bots, this study investigates how the BAG evaluates the bots, more specially how the BAG makes decisions to approve or deny an operator's request for bot approval. We have two research questions:

R.Q.1. What are the characteristics of discussions in which bot approval is decided?
R.Q.2. What features of a bot are related to approval of the bot?

2. Conceptual Background

In this chapter, we briefly discuss prior work on attitudes towards bots and collective decision making as well as how we developed hypotheses for this study using the previous work and Wikipedia Bots policy. We also developed a model using information obtained from Wikipedia Bots policy and from previous bot studies and theories of collective decision making (see Figure 2).

2.1. Attitudes Towards Bots

As noted, Wikipedia bots are increasingly common and research has started to examine attitudes towards them. Clément and Guitton [4] analyzed a corpus of 6528 interventions of users on talk pages of 50 Wikipedia bots to understand reactions of users depending on the characteristics of the bots' actions. They combined the different characteristics of the bots and classified bots as "servant bots", bots "which mainly do repetitive and laborious work instead of human users", and "policing bots", "which proactively enforc[e] Wikipedia's guidelines and norms" [4, p. 66]. The researchers found that users' attitudes towards the policing bots were either negative or positive rather than neutral. On the other hand, users have positive attitudes towards Wikipedia's servant bots, which help them when the bots are under their control. Users' perceptions are not so different than that Wikipedia's Bots policy aims to allow to produce bots that help humans best, which may articulate the ongoing success of Wikipedia.

Geiger [5] conducted a study of the issues during a bot's uses in Wikipedia with a focus on Wikipedia's Bots policy. He provided examples of specific bots' activities, other users' reactions to these activities and the bot developers' responses to the users. For example, the HagermanBot[4] appends signatures to comments in discussion spaces for those who had 'forgotten' to leave them, was approved. However, several problems occurred regarding the bot's identification algorithms [5], which Hagerman fixed. Then, some users were angry with the bot's normal

---

functioning, since the bot was promptly signing users' comments instead of giving them time to sign themselves, requiring the developer to make further changes.

In other words, even though a bot is approved, problems may still occur to which the operator must respond. Therefore, making careful decisions before approving a bot may help to lessen those problems. Despite the increasing use and popularity of bots, there are not studies focusing on how groups decide to work with bots. Thus, in this paper, the aim is at better understanding how a group decides to work with bots by examining the BAG's decision-making process for approving or denying of Wikipedia bots' deployment.

2.2. Collective Decision Making

To understand decision-making process of the BAG, we employed a collective decision-making approach because the decision about approval of a bot is a group decision. Bose, Reina and Marshall [6, p.30] defined collective decision making as the "subfield of collective behavior concerned with how groups reach decisions." The researchers emphasized the importance of value-based decision making and a speed-value tradeoff in collective decision making. "Value" may vary in different contexts, such as food, prestige or any other reward. A speed-value tradeoff means that a decision-making process may be oriented towards saving time (speed) or maximizing reward (value) [7], i.e., a strategy to choose the best alternative among available options (best value) even if it sometimes takes a lot time (speed tradeoff). Hence, this approach may also be appropriate in making decisions regarding bots' approval or disapproval in terms of considering the amount of a bot's benefits to Wikipedia (value). Namely, in their decisions BAG can approve the bots that can optimize the magnitude of the benefits while minimizing the potential issues that the bot may cause. In addition, the discussions made by the BAG to decide approval of a bot may take a lot of time (speed tradeoff).

2.3. Research Hypotheses and Model

Based on information from wiki/Help: Creating a bot page[5] and collective decision-making approach, explained in the section 2.2, we propose a model for the decision making of the BAG for approval or disapproval of a bot (see Figure 2). Furthermore, we develop hypotheses using the previous work related to attitudes towards bots and collective decision making in addition to Wikipedia's Bots policies to identify the

key factors related to bots' approval. In our research model, collective decision making is the main factor (see Figure 2). Furthermore, we defined two factors that are related to collective decision making clarified in the section 2.2: value-based decision making and speed-value tradeoff, and the bot features included in the data set that may affect approval of a bot: the bot's status (i.e., automatic, supervised, manual, etc.), the number of pages that the bot affects, and how many times the bot is run in a month.
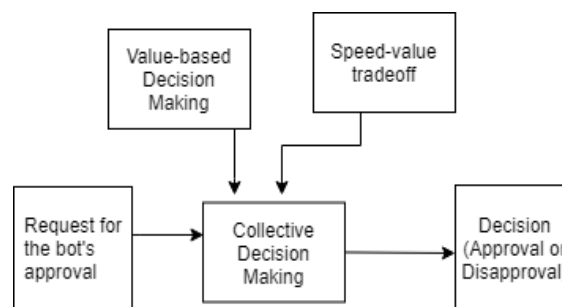

Figure 2. Model for BAG's decision making to approve or deny a Wikipedia bot

To develop our initial hypotheses, we used Wikipedia's Bots policy and previous work related to attitudes toward bots. For example, studies [5,8,9] that focus on problems and concerns regarding bots indicate that harmlessness is an important factor that positively affects attitudes. Other studies [5,10] emphasize the importance of usefulness by pointing out the bots' capability, appropriateness and efficiency for determined tasks. Wikipedia's Bots policy also recognizes harmlessness and usefulness as fundamental requirements for bot approval[6]. Thus, to answer the first research question, we proposed the following hypotheses:

H1. Discussions resulting in approval of a bot mostly include elements indicating that a bot is harmless.

H2. Discussions resulting in approval of a bot mostly include elements indicating that a bot is useful.

Furthermore, [11,12] examine the effects of topic importance in attitudes and agreement. Topic importance was found as a significant factor to reach an agreement; thus, to help answer the first research question, we also proposed the following hypothesis regarding topic importance.

H3. Some topics covered in the discussions are related to approval or disapproval of a bot.

On the other hand, because the decision about approval of a bot is a group decision, theories of collective decision making may also be appropriate in

---

making decisions regarding approval or disapproval of a bot. Moreover, the Wikipedia Bots policy includes the item "bot must perform only tasks for which there is consensus" as a requirement for approval of a bot.[7] Additionally, two factors that are related to collective decision making were clarified in section 2.2: value-based decision making and speed-value tradeoff. We claim that a value-based approach in terms of considering the amount of a bot's benefits to Wikipedia (value) may also be valid in the BAG's decision-making process. Namely, in their decisions the BAG can approve the bots that can optimize the magnitude of the benefits (value) while minimizing the potential problems that the bot may cause. In addition, the discussions made by the BAG to decide on the approval of a bot may take a lot of time (speed tradeoff). Thus, we proposed the following hypotheses concerning collective decision making.

H4. Discussions about the approval or the disapproval of a bot include elements indicating that decisions are made by collective decision making.

H5. Value-based decision making is related to collective decision making about the approval of bots as well.

H6. A speed-value tradeoff is involved in the collective decision making about the approval of bots as well.

Finally, referring to the Wikipedia Bots policy, we developed other hypotheses concerning bots' features. The policy warns that an approval request must include details of the bot's function, the status of the bot (manually assisted or running automatically, when the bot operates continuously, intermittently, or at specified intervals), and its rate.[8] Thus, to answer the second research question, we proposed the following hypotheses regarding bots' features:

H7. The function of a bot is related to approval of the bot.

H8. The bot's status (i.e., automatic, supervised, manual, etc.) is related to approval of the bot.

H9. The number of times a bot is run in a month is related to approval of the bot.

H10. The number of pages that a bot affects is related to approval of the bot.

3. Method

To answer the research questions and to test the hypotheses, we used text data consisting of Wikipedia discussions and bot functions leading to approval or disapproval of a bot, and then bot features described on each bot's project page. Before analyzing the discussions, we preprocessed the text data via several techniques, such as stop word filtering, to clean up the texts and remove the stop words (i.e., commonly used words such as *the, a*, or *an*). We first compared the most common words and two-word phrases (unigrams and bigrams) in discussions resulting in approval or disapproval of a bot. We also used topic modelling to find commonly used topics in these discussions. Finally, we explored correlations between the bots' features and the approval or disapproval of the bots.

3.1. Data Source

Data came from the Wikipedia: Bots/Requests for Approval website. This website includes Wikipedia discussions about the approval or disapproval of bots. It includes instructions for users who want to run a bot on the English Wikipedia website. After the instructions, there are descriptions of bots such as "operator," "time filed," "function overview," "type" (i.e., "automatic", "supervised", or "manual"). After the description each bot, there is a discussion about approving or disapproving it. At the bottom of the page, there are three lists of bot requests: approved, denied, and expired/withdrawn requests.

We extracted data from the project page for each bot linked to the lists and formed a data set that includes a discussion for each bot, a discussion time, each bot's name, each bot's function, and four other features for each bot: the bot's status (whether the bot is automatic, supervised, or manual), the number of runs in a month (how many times the bot is run in a month), and the number of pages edited (how many pages the bot affects). We started to form this data set on 16 March 2019. We finalized the data set on 29 May 2019. It includes 100 bots, their features and the discussions for each of those 100 bots.

3.2. Data Analysis

We used R for the data analysis in this study. We completed text analysis for discussion of each bot and each bot's function. After cleaning the data, we used document-term matrix (dfm) and quanteda package to find the most common words in discussions and functions of the bots. In addition, in the analysis of the discussions we used topic modelling using LDA (Latent Dirichlet Allocation). Furthermore, we conducted chi-square tests and t-tests to examine relationships between bot features (the bot's status, how many times the bot is run, the number of pages that the bot affects) and their approval; and the time of the discussion for a bot and the bot's approval. In the

---

[7] https://en.wikipedia.org/wiki/Wikipedia:Bot_policy

[8] https://en.wikipedia.org/wiki/Wikipedia:Bot_policy

end, we run logistic regression to identify predictors that affect a bot's approval.

4. Results

4.1. Characteristics of Discussions

To test the first and second hypotheses, we interpret the most common words (unigrams) and two-word phrases (bigrams) in the discussions resulting in approval or disapproval of the bots to explore some patterns that may affect approval of a bot.

The unigrams did not yield significant results related to the first two hypotheses. Nevertheless, whereas bigrams indicate important results supporting the second hypothesis, they did not show any clues with the respect to the first hypothesis. For example, in the discussions resulting in approval of the bots, we found that "edits-made," "can-make," "looks-good," "contributions" are some of the most common bigrams, which may be linked with "usefulness" because "usefulness" is defined in [13, p.985] as "using a specific application system will increase his or her job performance." Namely, after a trial is completed, if the results demonstrate the bots' contributions to users, such as making edits, listing categories, placing tags, and fixing errors, that means helping to improve humans' Wikipedia content editing performance by various contributions.

On the other hand, in the discussions resulting in disapproval of the bots, we found that "doesn't make", "can't cope" and "fast-enough" are some of the most common bigrams, that are related to bots' capabilities, and how much they are "useful" for humans. Hence, we can connect them again to "usefulness." Thus, we claim that the second hypothesis regarding the bot's usefulness is supported by the findings. An example from the original discussion for the bot DannyS712 bot 33, which is approved, and a useful bot that made 52 perfect edits, and did not make any errors, also supports that hypothesis:

> *"@TheSandDoctor:* 👁 ***Trial complete.*** *52 edits made -* [1]. *I did the first few manually to perfect the regex, and previewed the rest of the bot edits - didn't see any errors. Thanks, --DannyS712 (talk) 05:10, 29 April 2019 (UTC)"*[9]

On the other hand, while "harmlessness" was a fundamental requirement emphasized in the Wikipedia policy to approve a bot, we did not identify any expressions related to harmlessness among the most common unigrams or bigrams in the discussions.

---

[9] https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/DannyS712_bot_33

In sum, while the results supported the second hypothesis, they did not support the first hypothesis.

H1. Discussions resulting in approval of a bot mostly include elements indicating that a bot is harmless **(not supported).**

H2. Discussions resulting in approval of a bot mostly include elements indicating that a bot is useful **(supported).**

To test the third hypothesis, we applied LDA topic modeling. Topic modelling yields topics based on terms and each term's beta (the probability that a given term appears in a particular topic; the terms have higher beta define the topic best). In this topic modeling, we used all the discussions in our data set (both discussions of approved and disapproved bots). It yielded topics with some terms and from these terms we defined these topic names: "awb" (topic 1), "fixing errors" (topic 2), "bot flag" (topic 3), "approved updates" (topic 4), "commons category" (topic 5), "contributions" (topic 6), and "use request" (topic 7).

Then, using gammas (the probability that a given topic appears in a particular bot's discussion) obtained from LDA topic modelling, a logistic regression was performed to test whether it is possible to predict whether a bot is approved or disapproved based on discussion topics. The logistic regression results showed statistically significant associations of 4 topics (topic 2, topic 3, topic 4, topic 6) with bot approval. If these four topics are included in the discussion about a bot in a positive way, the probability of approving the bot is significantly increased ($p < 0.05$). The odds ratio for topic 4 (approved updates) is 1.3e+07:1 to 1:1, meaning that if a discussion includes that topic "approved updates" in a discussion, the chance that the bot would be approved increased a lot. On the other hand, the first topic "awb" is associated with disapproval of a bot, because the odd for it is 0.65:1 to 1:1, meaning if this topic is increasing one unit in a discussion of a bot, the approving a bot decreasing 0.35 unit.

Thus, the third hypothesis is supported by the findings because topic 2, topic 3, topic 4, topic 6 covered in the discussions are related to approval of a bot.

H3. Some topics covered in the discussions are related to approval or disapproval of a bot **(supported).**

For the sake of helping to test the hypotheses related to collective decision making (H4, H5, H6), running a t-test, we also examined whether the discussion time affects approval of a bot. The t-test showed that a significant relationship between the discussion time in minutes and approval of a bot

(*MA*=30754) and disapproval of a bot (*MD*=17082) and; t (97.918) = 2.009, *p-value* < 0.05.

On the other hand, it is obvious that the BAG is a committee, and it makes the decisions collectively to reach a consensus. Moreover, Geiger [5, p.87] pointed out that rule for Wikipedia bots: "if there was a consensus for performing the task, the bot was approved and began operating; if there was no consensus, the bot was rejected, or suspended if it had already been operating." Our findings also supported this rule because "consensus" was one of the most common words in the discussions both resulting in approval and disapproval of the bots.

Furthermore, in collective decision making, two key factors were emphasized in section 2.2: value-based decision making and speed-value tradeoff. The word clouds, topic analysis and example discussion quotes indicate that efficient bots that make many contributions and fewer errors (for example, as mentioned, DannyS712 bot 33, which was approved, made 52 perfect edits, and did not make any errors), namely useful, got more approval by the BAG. This approach refers to value-based decision making: choosing the optimal options that maximize the rewards. In our situation, the approved bots are maximizing contributions and minimizing the errors (some bots even fix the errors), therefore, provide most benefits and minimize the costs most. In addition, as seen in the analysis results, the discussion time was greater for the approved bots than for the disapproved bots. This can be linked with speed-value tradeoff. The BAG trades off time to make optimal decisions for choosing the most valuable bots to approve. Thus, these findings support our following hypotheses:

H4. Discussions about approval or disapproval of a bot include elements indicating that decisions are made by collective decision making **(supported).**

H5. Value-based decision making is related to collective decision making about bots' approval as well **(supported)**.

H6. A speed-value tradeoff is involved in the collective decision making about the approval of bots as well **(supported)**.

4.2. Features of Bots

In this section, we aimed to explore whether bots' features that we had in the data set as defined in the section 3.1, (bot's function, bots' status, the number of estimated pages that the bot edits, the number of runs

of a bot in a month) are related to approval or disapproval of a bot.

To investigate whether a bot's function, in other words, whether the task that a bot undertakes is related to its approval, we analyzed text data defining the function of each bot. The word clouds did not yield specific indicators regarding the function of the bot affects the bot's approval because most common words in the approved bots' functions or in the disapproved bots' functions did not indicate any specific patterns, both include similar commonly used words. Moreover, looking at the data set, we observed that various bots undertaking different tasks get approval, in other words, there are not specific tasks undertaken only by the approved or disapproved bots. Conversely, some bots undertaking similar tasks get approved, but some others do not. For example, whereas PkbwcgsBot 21 fixing high-priority CW Error #46 (Square brackets without correct beginning) and error 10 (Square brackets without correct end) was disapproved[10], PkbwcgsBot 13 fixing WP:WCW error 101 (Ordinal number found inside <sup> tags) was approved[11], i.e., one of two bots undertaking similar functions, basically fixing errors, got approval and the other did not.

Thus, we conclude that the results did not support the following hypothesis about bots' function:

H7. The function of a bot is related to approval of the bot **(not supported).**

To test other hypotheses related to other mentioned bot features, we used different statistical tests. For example, using chi-square test, we found a statistically significant relationship between the status of a bot and approval of a bot (Pearson's Chi-squared test statistics: *X-squared* (4, *N*=100) =18.4, *p-value* < 0.01). In addition, t-tests were run to assess whether the number of estimated pages that the bot edits and the number of runs of a bot in a month affect approval of a bot; we did not find a significant relationship between them.

Finally, we ran a logistic regression that includes all the predictors (the bot's status, the number of estimated pages that the bot edits, the number of runs of a bot in a month, and the discussion time). The logistic regression results showed that there is only one significant predictor: the status of the bot in prediction of a bot's approval. Among the bots' status conditions, "Status manual" is the only one significant predictor (p < 0.05). The odds ratio for Status manual is 0.085:1 to 1:1, meaning that if a bot is manual, the chance of the bot's approval significantly decreased. If the bot is automatic, the chance to get approval

---

[10] https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/PkbwcgsBot_21

[11] https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/PkbwcgsBot_13

increases. Thus, the following hypothesis is supported by the findings.

H8. The bot's status (i.e., automatic, supervised, manual, etc.) is related to approval of the bot **(supported).**

However, as noted, the findings did not show a significant relationship between the number of estimated pages that a bot edits and approval of the bot. We also did not find a significant relationship between the number of runs of a bot in a month and approval of the bot. Thus, the following hypotheses were not supported.

H9. The number of times a bot is run in a month is related to approval of the bot **(not supported).**

H10. The number of pages that the bot affects is related to approval of a bot **(not supported).**

5. Discussion

This study aimed to investigate how a group decides to work with bots, in particular how a group approves the bots before they are deployed. To this purpose, we examined discussions about Wikipedia bots and the features of bots. Through the lens of previous work related to collective decision making and attitudes towards bots, and Wikipedia Bot polices, we developed hypotheses to understand whether the discussions include some characteristics related to the approval or disapproval of a bot, and whether some features of a bot are related to the approval or disapproval of the bot. As explained in sections 4.1 and 4.2 in detail, the results support hypotheses H2, H3, H4, H5, H6, and H8:

H1. Discussions resulting in approval of a bot mostly include elements indicating that a bot is harmless **(not supported).**

H2. Discussions resulting in approval of a bot mostly include elements indicating that a bot is useful **(supported).**

H3. Some topics covered in the discussions are related to approval or disapproval of a bot **(supported).**

H4. Discussions about approval or disapproval of a bot include elements indicating that decisions are made by collective decision making **(supported).**

H5. Value-based decision making is related to collective decision making about bots' approval as well **(supported).**

H6. A speed-value tradeoff is involved in the collective decision making about the approval of bots as well **(supported).**

H7. The function of a bot is related to approval of the bot **(not supported).**

H8. The bot's status (i.e., automatic, supervised, manual, etc.) is related to approval of the bot **(supported).**

H9. The number of times a bot is run in a month is related to approval of the bot **(not supported).**

H10. The number of pages that the bot affects is related to approval of a bot **(not supported).**

Based on our observations, we suggest some guidelines for Wikipedia bot developers to consider while developing their bots.

1. Proposals and discussions should cover potential harms of the bot before the bot is deployed. As noted, interestingly, although "being harmless" was a fundamental requirement emphasized in the Wikipedia policy for approving a bot, we did not find any indications that potential harm by a bot is covered among the most common words in the discussions.

As recognized in the HagermanBot example, after a bot's implementation, some problems that harmed users (or at least annoyed them) emerged. The discussion on the project page for HagermanBot began at 7:55 am on 1 December 2006, and the bot was approved at 11:22 pm on 2 December 2006. However, in that discussion, the potential harms of the bot were not pointed out. After the bot was deployed, some users mentioned their problems. For example, a Wikipedia user provided his complaint:[12]

"The main problem I see with this bot is that it hides vandalistic or inappropriate comments or spam on Talk pages from people's watchlists…"

Another user expressed his problem as follows:[13]

"I don't really like this bot editing people's messages on other people's talk pages without either of their consent or even knowledge…"

Before using certain technologies, discussions should be conducted to address problems with the morality and norms associated with the use of those technologies, instead of focusing only on the tasks to be done by them. Thus, while making decisions regarding for approval of a bot, the BAG should consider not only the tasks a bot will undertake, but also potential moral issues. In addition, the operators of the bots should list the potential harms that the bots may cause and potential solutions for them in the proposals for the bots; and before the trial, the BAG, the operators, and other users should discuss them to find solutions for the potential problems. If they find the solutions, then they should approve; otherwise they should not. This approach may help the BAG to

---

[12] https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/HagermanBot

[13] https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/HagermanBot

make more careful decisions when approving or disapproving bots.

2. A bot should be useful, and the bot's functions should be clearly expressed in the proposals and checked if the bot does not do exactly what it should after the trial. "Trial" was one of the most common words in the discussions resulting in approval of a bot. At first glance, the word "trial" seems to be an unimportant word. However, the expression "trial was completed" and the positive results seen after the trial (i.e. making contributions, fixing errors, perfectly completing tasks, etc.) are crucial for approving a bot; these are linked to efficiency and usefulness. In addition, the bigrams, "edits-made", "looks-good", "automatic-fixing", and the topics "categories," "contributions of the bot", "tags", and "fixing errors" were seen more in the discussions resulting in the approval of bots. Furthermore, automatic and supervised bots were approved more than manual bots.

These findings offer some insight regarding what kinds of bots are approved. For example, related to "fixing errors," a bot from the data set used in this study, WikiCleanerBot 3, which is automatic and "fix[ing] some simple cases of square brackets without correct beginning,"[14] was approved. An example related to the topic of "categories," Pi bot 4, which is again automatic and "fix[ing] or remov[ing] commons category links that are missing, or are to category redirects or disambiguation categories"[15] was approved. As an example related to topic of "tags," we indicate Ronbot 12. It "tags pages that have broken images, and sends a neutral message to the last editor."[16]Moreover, Ronbot 12 is also automatic. As a final example for an approved bot, we offer PkbwcgsBot 20 which "fixes some broken Wall Street Journal external links."[17] PkbwcgsBot 20 is a supervised bot.

However, as noted in section 4.2, based on our findings, the function of the bot was not related to its approval. Nevertheless, it is important to clearly define a bot's function in its proposal while requesting its approval. For example, on the project page for the DiyarBot,[18] the function of the bot was written as "to make repetitive automated or semi-automated edits that would be extremely tedious to do manually." However, this function is not clear and not specific to that bot, since the purpose for running many bots is to

make repetitive edits; therefore, this bot request was denied by a BAG member:

> *"...I'd note that your request is **far too vague** and tell you to read* WP:BOTPOL*, and also I'd suggest that you might want to spend some time around the English Wikipedia making content edits as a normal editor before coming back with a more specific request."*

Our findings also indicate that the number of pages that a bot edits and how many times a bot is run are not important. The crucial thing is that the bot functions properly and as defined in its proposal. When referring to the HagermanBot example, we explained that point. On the project page of the HagermanBot, its function was described as "inserts the {{unsigned}} template on talk pages when a user forgets to sign a comment." However, the bot was instantly appending signatures to comments in discussion spaces instead of giving users time to sign their own comments. A user left the following message [5]:

> *"HangermanBot keeps adding my signature when I have not signed with the normal four tilde signs. I usually just sign by typing my username and I prefer it that way. However, this Bot keeps appearing and adding another signature. I find that annoying. How do I make it stop?"*

Thus, a bot's function should be clearly described in the proposal, and after the trial, the bot should be checked to see if it functions properly and as defined in the proposal. In addition, the bot developers should pay attention to the usefulness of the bot. Furthermore, the bot developers should also develop automatic or supervised bots as appropriate to the bots' functions, because automatic or supervised bots tend to be preferred by the BAG.

3. A new bot should be proposed if it is needed, and the most appropriate tool or software should be chosen for the proposed bot: There are various Wikipedia bots undertaking many tasks. However, sometimes the operators of the bots propose new bots to undertake the same tasks that some bots are already doing, which often results in disapproval of the new bots. On the other hand, different programming

---

[14] https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/WikiCleanerBot_3
[15] https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/Pi_bot_4
[16] https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/RonBot_12

[17] https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/PkbwcgsBot_20
[18] https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/DiyarBot

languages (e.g., Visual C#NET[19], Java[20], Python[21,22,23] etc. – Python is more common in new bots) and tools (e.g., AWB[24]) are used for operating the bots, but sometimes these tools are not found to be appropriate by the BAG. To strengthen these arguments, we will point out the most common words seen in the discussions that resulted in the disapproval of bots, and in particular, the PkbwcgsBot10 example. The abbreviation *AWB* (AutoWikiBrowser) and topics that include *AWB* were present more in the discussions resulting in the disapproval of bots. We attribute this result to two potential factors. First, because AutoWikiBrowser (AWB) is a semi-automated tool designed to assist with editing on Wikipedia, it can accomplish some tasks instead of running a new bot. Therefore, a newly proposed bot may be disapproved and AWB usage encouraged by the BAG. Second, some bot developers use the AWB tool to run their bots; although it is easy to use, sometimes this tool is not appropriate for a bot or for a determined task. For example, Primefac and Xaosflux, two members of the BAG, were conversing, and for this reason, decided to deny the request for running PkbwcgsBot10,[25] which was proposed to fix double redirects using AWB. Their discussion:

> "@Pkbwcgs: *I think we already have several more robust bots doing this, that also include a hold-down to not 'fix' DR's that are very new and could still be getting worked on. Is there a backlog forming that they can't keep up with?* **I don't think AWB is the best tool for this job** *either as you mentioned.* — xaosflux ^{Talk} *14:16, 23 December 2018 (UTC)*
>
> *I concur.* **The bots mentioned above are fully automatic and do not require AWB to be manually started.** *I see no clear reason for this task.* Primefac(talk) *15:50, 23 December 2018 (UTC)*
>
> ⊗ *Denied. this is* **just the wrong tool for this job and the process is already being well handled by very experienced bots** *(with 1+ million edits).* — xaosflux ^{Talk} *17:49, 23 December 2018 (UTC)"*

This discussion also demonstrates how the BAG makes a collective decision. More specifically, in this discussion, Primefac and Xaosflux use value-based decision-making approach because they emphasize AWB tool is not appropriate for the proposed bot; and they choose other alternatives which are more valuable through expressing "*the process is already being well handled by very experienced bots.*"

As another example, we indicate again PkbwcgsBot 21[26] fixing high-priority WP:WCW error 46 and PkbwcgsBot 13[27] fixing WP:WCW error 101. They both basically fix some errors, belong to the same operator and use AWB. While PkbwcgsBot 21 was disapproved, PkbwcgsBot 13 was approved because AWB was appropriate for the latter one whereas not for the first one. PkbwcgsBot 21 was denied by the following sentences of a BAG member:

> *"... here are just too many CONTEXT issues to blindly attack this with AWB."*

Thus, before proposing a new bot, the operators of the bots should check previous bots to decide if a new bot is really needed. In addition, the operators of the bots should consider the most appropriate tool for running the proposed bots.

Finally, we recognized that some of the guidelines presented in this study are similar to the guidelines presented for designing systems that humans interact with, such as "requirements determination," "evaluation," and "alternative selection" [14]. For example, as noted, new bots should be proposed if there is a need (requirements determination); the proposed bot should be tested by a trial and bot developers should update their bots based on the recommendations of the BAG and other users (evaluation); bot developers should choose the best tool for running their bots among the alternative tools (alternative selection). These points are also critical for the BAG's decision since the group makes decisions considering the best bots among other alternatives as appropriate to Wikipedia's needs after trials by which they test the bots based on some evaluation criteria (usefulness, functions properly, etc.).

---

[19] https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/HagermanBot
[20] https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/WikiCleanerBot_3
[21] https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/JJMC89_bot_17
[22] https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/DeltaQuadBot_7
[23] https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/TheSandBot_3
[24] https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/Muhbot
[25] https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/PkbwcgsBot_10
[26] https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/PkbwcgsBot_21
[27] https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/PkbwcgsBot_13

6. Conclusion

Despite the increasing use and popularity of bots, there are not studies focusing on how groups decide to work with bots; therefore, our study may be novel in terms of contributing to understanding decision making regarding the human-automation boundary and to facilitating to develop more efficient bots. Wikipedia content develops as contributors, editors and users add new content, increasing the content to be edited. For editing this huge data, Wikipedia benefits from bots. Before a Wikipedia bot is run to edit, the developer of the bot must request to get approval for the bot from the BAG. The BAG makes decisions through discussing each bot on the bot's talk page or the bot's project page. In this paper, we investigate how the BAG makes decisions to approve or deny a request of the operator of the bot for approval of the bot. We analyzed 100 discussions for each of 100 bots and interpreted them. The results suggested that usefulness, value-based decision making and bots' status (i.e. automatic, etc.) affect the result of an approval of a bot.

6.1. Limitations and Directions for Future Research

In this study, we focused on Wikipedia bots. However, the usage of various bots in different areas has become widespread, such as trading bots (e.g., chatbots in customer service, help bots in commercial company websites), social media bots (Facebook, Twitter, Reddit), social bots chatting to human users (e.g., Eliza representing a mock Rogerian psychotherapist). The fundamental guidelines presented in this study (e.g., giving importance to the usefulness and harmlessness of a bot, using appropriate tools for running a bot, etc.) may also be of help to different bot developers. However, specific guidelines may vary for different bots used in various areas for various purposes. For example, for a social bot chatting with a human user as a psychotherapist, emotional features that may affect user satisfaction (e.g., trust, intimacy, sympathy, etc.) may also be important. Therefore, effective emojis might be more useful in that bot's conversational flow to perform more human-like interactions with humans, which may increase the trust and sympathy of a user. For other bots, many other features might be more important, depending on the purpose of the bot. Therefore, bots in other domains might be evaluated based on many other criteria. Thus, in the future, more comprehensive studies may be conducted to better understand the decision-making processes for other bots that assist humans in various situations.

References

[1] R. Priedhorsky, J. Chen, S.T. K. Lam, K. Panciera, L. Terveen, and J. Riedl, "Creating, destroying, and restoring value in Wikipedia," In Proceedings of the 2007 international ACM conference on Conference on supporting group work- GROUP '07. Sanibel Island, Florida, USA: ACM Press, 2007, pp. 259-268.

[2] M. Tsvetkova, R. García-Gavilanes, L. Floridi, and T. Yasseri, "Even good bots fight: The case of Wikipedia," *Plos One*, vol. 12, no. 2, 2017.

[3] R. S. Geiger and A. Halfaker, "When the levee breaks: without bots, what happens to Wikipedia's quality control processes?" In Proceedings of the 9th International Symposium on Open Collaboration ACM, 2013, pp.1-6.

[4] M. Clément and M. J. Guitton, "Interacting with bots online: Users' reactions to actions of automated programs in Wikipedia," *Computers in Human Behavior*, vol. 50, pp. 66–75, 2015.

[5] "The Lives of Bots - R. Stuart Geiger." [Online]. Available: https://stuartgeiger.com/posts/2011/03/the-lives-of-bots/. [Accessed: 31-May-2019].

[6] T. Bose, A. Reina, and J.-A. Marshall, "Collective decision-making," *Current opinion in behavioral sciences*, vol. 16, pp. 30-34, 2017.

[7] A. Pirrone, T. Stafford, and J.A. Marshall, "When natural selection should optimize speed-accuracy trade-offs," *Frontiers in neuroscience*, 8, pp.73, 2014.

[8] A. Bessi and E. Ferrara, "Social Bots Distort the 2016 U.S. Presidential election online discussion," *First Monday,* vol. 21, no. 11, 2016.

[9] C. Shao, G. L. Ciampaglia, O. Varol, K.-C. Yang, A. Flammini, and F. Menczer, "The spread of low-credibility content by social bots," *Nature Communications*, vol. 9, no. 1, 2018.

[10] K. Long, J., Vines, S. Sutton, P. Brooker, T. Feltwell, B. Kirman,..and S.Lawson, "Could you define that in bot terms?: Requesting, creating and using bots on reddit," In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems ACM, 2017, pp. 3488-3500. [11] L. C. Gerald, and B. Baldridge, "Interpersonal attraction: The role of agreement and topic interest," *Journal of Personality and Social Psychology* 9.4 (1968): 340.

[12] P.G. Banikiotes, "Interpersonal attraction, topic importance, and proportion of item agreements." *Psychonomic Science* 22.6 (1971): 353-354.

[13] F. D. Davis, R. P. Bagozzi, and P. R. Warshaw, "User Acceptance of Computer Technology: A Comparison of Two Theoretical Models," *Management Science*, vol. 35, no. 8, pp. 982–1003, 1989.

[14] Hoffer, J. A. , Modern Systems Analysis and Design, 6/e. Pearson Education India, 2012.