Exploring Consumers' Discernment Ability of Autogenerated Advertisements

J. Sahlin[*+], H. Sundell[*], G. Mbiydzenyuy[*], H. Alm[*], J. Holgersson[+], C. Suhonen[*], and T. Hjelm[*]

[*]*Department of Information Technology, University of Borås, Sweden*
E-mail: {johannes.sahlin, hakan.sundell, gideon.mbiydzenyuy, hakan.alm}@hb.se, {s182303, s182959}@student.hb.se

[+]*School of Informatics, University of Skövde, Sweden*
E-mail: jesper.holgersson@his.se

Autogenerated Advertisements (AGAs) can be a concern for consumers if they suspect that Artificial Intelligence (AI) was involved. Consumers may have an opposing stance against AI, leading companies to miss profit opportunities and reputation loss. Hence, companies need ways of managing consumers' concerns. As a part of designing such advices we explore consumers' discernment ability (DA) of AGAs. A quantitative survey was used to explore consumers' DA of AGAs. In order to do this, we administered questionnaires to 233 respondents. A statistical analysis including Z-tests, of these responses suggests that consumers can hardly pick out AGAs. This indicates that consumers may be guessing and thus do not possess any significant DA of our AGAs.

Keywords: Autogenerated ads, Discernment ability, Marketing.

1. Introduction

Autogenerated Ads (AGAs) may be a concern to consumers[4] when they discover that Artificial Intelligence (AI) is used in creating SMS and email ads (See Figure 1). Consequently, consumers can develop stances against companies that use AI, and companies risk losing profit opportunities and reputation when AI is suspected through identifiable mistakes or oddities with the AGAs. Thus, companies are hesitant towards using AGAs in their marketing activities. In the literature, AI can give rise to several Consumers' Concerns (CC), for example; i) uniqueness neglection, in other words, AI not being able to account for one's uniqueness[9], ii) AI lacking sympathy compared to human actors[3], iii) data privacy issues[10] like data security[8], misuse of information[8] and prediction of sensitive information, for example, sexual orientation[7], and iv) algorithmic biases favoring companies or disfavoring certain consumer groups[14]. Due to such CC, consumers may balance privacy concerns against the benefits of the personalized recommendations in a privacy-personalization trade-off[1]. AGA pioneers relied on predefined templates to generate readable and attractive sentences to decrease the risk of becoming detected[6]. The approach of using template-based generation methods can vastly reduce human efforts but are rigid, lack diversity, and cannot adapt like modern, sophisticated methods[6]. However, the approach of employing modern, sophisticated methods shows 15% of nonsense generations[6]. Therefore, we suggest combining a template-based approach with sophisticated methods that may capture the best of both directions to meet the high production requirements, avoid reluctant consumers and secure profit opportunities and reputation. Business requirements demand an almost perfect attractiveness and full certainty of AGAs to stay undetected, in order to be accepted. However, combining template-based and sophisticated methods are insufficient to prevent the detection of AGAs. Marketing activities need well adjusted and reliable methods for managing CC and identification of anomalies and oddities from quantitative and qualitative views to improve the underlying generational models. For example, AI can generate anomalies and oddities that methods simply relying on quantitative evaluations may not capture, and thus undesirable ads may pass on to consumers. Thus, this article aims to describe consumers' DA of AGAs. To do so, the study has to first establish an environment for researching consumers' perceptions of AGAs. This is needed to identify CC of AGAs, from which advices can be developed on how to manage the identified CC of AGAs (See Figure 2). We consider *consumer perception* as the consumers' awareness and understanding, which can be further broken down into sensations, discernment, apprehensions, consciousness, and notions of the phenomenon. A *notion* is defined by us as a belief about something. We consider *discernment ability* (DA) as the ability to reveal if an ad is an AGA, and can be viewed as an umbrella term for perceptiveness, insight, awareness and understanding. We argue that in order to understand the consumers' perceptions of AGAs, it is essential to determine how consumers distinguish AGAs. Moreover, studying consumers' notions about AGAs can also help establish knowledge about their beliefs. We consider *concerns* to be the activity of being involved with something, or worried about something - influenced by our perceptions, thoughts and subjective experiences. We consider *managing* to be the activity of succeeding in dealing or maintain control over something. The article describes consumers' DA of AGAs.

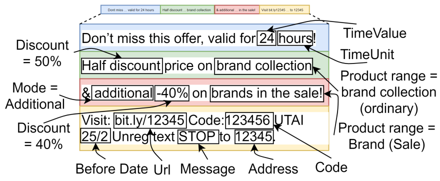


Fig. 1: Example of an autogenerated SMS ad, with variables and values.

2. Method

To capture overall patterns of discernment we adopted a quantitative design when it comes to consumer discernment of AGAs (See Figure 2). The quantitative survey was designed to study consumers' DA of AGAs. To understand consumers' DA an online questionnaire [11] was designed to survey consumers' DA of AGAs. The stop criteria for the data collection were based on time resources and were active for two weeks. Authentic ads were acquired from a Swedish e-commerce retailer. The selected ads upheld the sender's anonymity. The AGAs is a product of Sahlin et al. [12],[13] proposed system for automatically breaking down ads into its components and generating a variety of synonymous messages (See Figure 1) through the use of AI-assisted text generation algorithms. The generated AGAs were chosen randomly. The sampling used for the quantitative research was based on non-probability sampling methods to gain an exploratory sample. Using a combination of self-selection methods and convenience sampling resulted in 102 respondents who self-selected themselves by opting into the questionnaire. The questionnaire was posted on social media feeds (Facebook) on the authors' accounts. These pledges were redistributed three times each week from 2021-04-21 to 2021-05-05. The pledge to participate was later shared further through friends and family. Further attempts were made to increase the amount by sending out a notification through a communication channel used by the University's pedagogical platform. This notification reached a lot of students and increased the sample by 135 respondents (in addition to the 102 above). Lastly, criterion sampling was applied inside the questionnaire to apply criteria as (1) frequent or accustomed to online shopping and (2) having received ads recently as SMS or email. If the participant opted no to these questions, they were not presented with the rest of the questionnaire. Four respondents were excluded, resulting in 233 total respondents (102+135-4 from above). The questionnaire was divided into three parts. The first part consisted of the criterion sampling questions. The second part consisted of demographics collection:

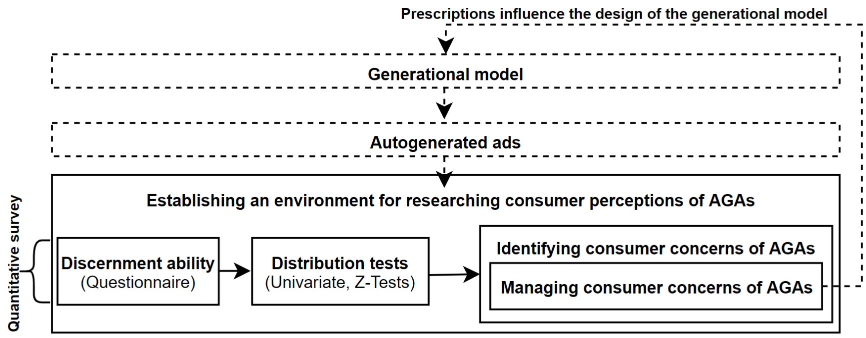


Fig. 2: Illustration of our overall research process.

in other words, age, sex, education, and occupation. The last part consisted of five AGA DA tests. For each test, respondents had a binary choice to discern whether they perceived the first or second option as AGAs or authentic, in other words intentionally disallowing a neutral option. The DA was analysed through univariate and bivariate analysis. To analyse the whole sample's DA, we used univariate analysis. When there were multiple groups, we used bivariate analysis. Here, the respondents were asked to discern which option was the AGA, and answers was coded into binary values where 1 represented proper discernment, and 0 represented faulty discernment. The percentage of correct discernment tests was calculated (total correct discernment divided by total tests). When the data set contained every respondent's percentage of correct DA level, we plotted the DA percentage values for each respondent into a histogram distribution chart. The same procedure was applied to different demographic groups. Z-tests[2] was conducted for inferential statistics regarding the DA of the demographic groups. Z-tests were also used to compare the mean values of two groups and test the acceptance of *the null hypothesis* (H_0). It is defined as that H_0 decides if there is a statistical significance in the mean values according to the groups used and is decided by the p-value relation to the z-value. Rejecting H_0 means the result is not statistically significant and can be achieved by chance, while accepting H_0 implies accepting the alternate hypothesis (H_a). Z-tests are suitable for comparing a numerical variable's averages under the assumption of a normal distribution. Each Z-test consisted of the following comparisons *Gender* (none responded with the option *other*, enabling binary grouping of males and females), *Academics or not*, *Hired or not* (students having a paid job were considered hired). The groups *gender*, *academics or not*, *hired or not* are defined and served to support H_0. The Z-tests are considered with a 95% confidence level.

3. Result and Analysis

The quantitative study yielded the following demographic data. Respondents were 234 in total. Age constituted; 18-29: 35,7% (81), 30-39: 41,9% (95), 40-49: 11,5% (26), 50-59: 8,8% (20), 60-69: 1,8% (4), 70+: 0,4% (1). Gender constituted; Male: 39,6% (90), Female: 60,4% (137), Other: 0% (0). Academics 58.5% (86) and non-academic 41.5% (147). The univariate analysis represents the whole sample, and the z-test results represent a more in-depth view of selected groups in the sample. In Table 1 each z-test is presented with details. Results of the online questionnaire yielded an DA level distribution based on the population, of: 0% (18), 25% (54), 50% (100), 75% (43), 100% (18), total: 233. The accuracy of the responses about DAs suggested that the mean, mode and median were approximately the same (0.5). Thus indicating the responses were indicative of a normal distribution (See Figure 3a). From the mean and standard deviation of the accuracy scores in Table 1, the responses from each group in the sample do not differ significantly.

	Male or Female		**Academic or not**		**Hired or not**	
Size	92	141	86	147	117	116
Average	.505	.476	.509	.588	.504	.467
Std Dev	.228	.270	.237	.260	.249	.254
Skewness	-.043	.185	.039	.087	.175	.012
Skewness shape	Sym	Sym	Sym	Asym	Sym	Sym
Normality	9.257×10^{-7}	1.161×10^{-7}	5.94×10^{-6}	3.847×10^{-7}	4.217×10^{-7}	3.753×10^{-7}
Outliers	5	0	3	0	0	7
P-value	0.402686		0.0515185		0.267159	
Z-score	0.836835		-1.947136		1.109630	
H_0	Accepted		Accepted		Accepted	

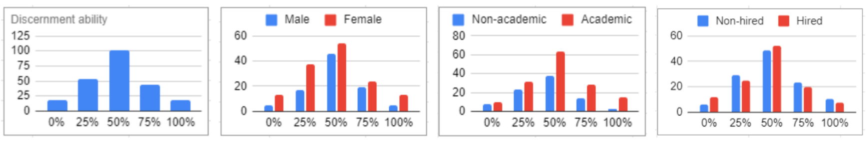
Table 1: Z-test summary. Sym for symmetrical, Asym for asymmetrical.

Z-test - Male and Female. The aim of a Z-test here is to determine the extent to which responses differ as a result of gender. Thus, Z-test allows to test for H_0: There is no difference in respondents' DAs as a result of gender.

To test this, variables representing different samples were introduced; male (92) and female (141). Detail statistics of these variables are summarised in Figure 3b. From the accuracy scores calculated for these groups, the P-value was estimated to be .40, (p(x≤Z)=.79). A large P-value suggest that the chances that there is a difference in DA responses from the population due to gender are low, thus we accept H_0.

Z-test - Academic or not. The aim of a Z-test here is to determine the extent to which responses differ as a result of academic education level. Thus, Z-test allows to test for H_0: There is no difference in respondents' DAs as a result of academic education level. To test this, variables representing different samples were introduced; academic (86) and non-academic (147). Detail statistics of these variables are summarised in Figure 3c. From the accuracy scores calculated for these groups, the P-value was estimated to be .40, (p(x≤Z)=.79). A large P-value suggest that the chances that there is a difference in DA responses from the population due to academic education level are low, thus we accept H_0.

Z-test - Hired or not. The aim of a Z-test here is to decide the extent to which responses differ as a result of employment. Thus, Z-test allows to test for H_0: There is no difference in respondents' DAs as a result of employment. To test this, variables representing different samples were introduced; hired (86) and non-hired (147). Detail statistics of these variables are summarised in Figure 3d. From the accuracy scores calculated for these groups, the P-value was estimated to be .40, (p(x≤Z)=.79). A large P-value suggest that the chances that there is a difference in DA responses from the population due to employment are low, thus we accept H_0.



(a) All respondents (b) Male or female (c) Academic or not (d) Hired or not

Fig. 3: Discernment ability distribution tests.

4. Discussion and Conclusion

The study suggest that consumers cannot discern AGAs from authentic ads, indicating they may be guessing. The results contribute to the existing literature by exploring consumers' perceptions of AGAs. Moreover, adding to the discussion of how to design and implement value-creating AI-enabled marketing services, which is a topic of merit for scholars and practitioners. Through the univariate analysis consumer's DA of AGAs were measured, this study found that the respondents could not discern AGAs from authentic ads, at least in this situation. Also, this is shown by the average, median and mode values that indicate that the DA of the research population is close to 50% which perfectly correlate with a random probability outcome in a normal distribution under independent observations. The result of the z-tests indicated no significant difference between the samples' DA, and using the 95% confidence interval showed that all samples are within statistical boundaries. Also, the presented z-value for all the tests is in accepted boundaries to support H_0, implying no or minor differences between population characteristics. The average scores presented in the z-tests are similar to those presented in the univariate analysis, showing a 50% chance to discern the ad correctly. The law of large numbers (LLN) can be considered for quantitative measurements[5]. LLN states that an observed sample average from a large sample will be close to the actual population average and will get closer the larger the sample is. LLN also states that there is no guarantee that a small sample will reflect the actual population characteristics and that the true population will be balanced in the subsequent sample, and hence, a larger sample will always be preferable for more accuracy. Yet, as seen in the results, even though the sample is relatively small, there is a resemblance of a normal distribution in the DA of the sample population, which might indicate some degree of generalisability. Using a univariate analysis for deciding consumers' DA of AGAs was considered most feasible. Directions for future research include extending the topic of managing CC of AGAs by other views. In a future study, we suggest studying discernment motivations through content analysis with conceptual and co-occurrence analysis. Moreover, a qualitative perspective can be included for giving advice on how to manage consumers' concerns of AGAs.

References

1. Elizabeth Aguirre, Dominik Mahr, Dhruv Grewal, Ko De Ruyter, and Martin Wetzels. Unraveling the personalization paradox: The effect of information collection and trust-building strategies on online advertisement effectiveness. *Journal of retailing*, 91(1):34–49, 2015.
2. George Casella and Roger L Berger. *Statistical inference*. Cengage

Learning, 2021.
3. Noah Castelo. *Blurring the Line Between Human and Machine: Marketing Artificial Intelligence*. PhD thesis, Columbia University, 2019.
4. Thomas Davenport, Abhijit Guha, Dhruv Grewal, and Timna Bressgott. How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science*, 48(1):24–42, 2020.
5. Frederik Michel Dekking, Cornelis Kraaikamp, Hendrik Paul Lopuhaä, and Ludolf Erwin Meester. *A Modern Introduction to Probability and Statistics: Understanding why and how*, volume 488. Springer, 2005.
6. J Weston Hughes, Keng-hao Chang, and Ruofei Zhang. Generating better search engine text advertisements with deep reinforcement learning. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2269–2277, 2019.
7. Michal Kosinski, David Stillwell, and Thore Graepel. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the national academy of sciences*, 110(15):5802–5805, 2013.
8. Nir Kshetri. Big data s impact on privacy, security and consumer welfare. *Telecommunications Policy*, 38(11):1134–1145, 2014.
9. Chiara Longoni, Andrea Bonezzi, and Carey K Morewedge. Resistance to medical artificial intelligence. *Journal of Consumer Research*, 46(4): 629–650, 2019.
10. Kelly D Martin, Abhishek Borah, and Robert W Palmatier. Data privacy: Effects on customer and firm performance. *Journal of Marketing*, 81(1):36–58, 2017.
11. B.J. Oates. *Researching Information Systems and Computing*. ISBN 9781412902243.
12. Johannes Sahlin, Håkan Sundell, Håkan Alm, and Jesper Holgersson. Short message service campaign taxonomy for an intelligent marketing system. In *Developments of Artificial Intelligence Technologies in Computation and Robotics: Proceedings of the 14th International FLINS Conference*, pages 606–613. World Scientific, 2020.
13. Johannes Sahlin, Håkan Sundell, Håkan Alm, and Jesper Holgersson. Evaluating artificial short message service campaigns through rule based multi-instance multi-label classification. In *AAAI 2021 Spring Symposium on Combining Machine Learning and Knowledge Engineering*, volume 2846. Sun SITE Central Europe, 2021.
14. John Villasenor. Artificial intelligence and bias: Four key challenges. *Retrieved February*, 11:2019, 2019.