Artificial Intelligence, Social Computing and Wearable Technologies, Vol. 113, 2023, 205–215
https://doi.org/10.54941/ahfe1004193

The Consistency Between Popular Generative Artificial Intelligence (AI) Robots in Evaluating the User Experience of Mobile Device Operating Systems

Victor K. Y. Chan

Faculty of Business, Macao Polytechnic University, Macao, China

ABSTRACT

This article attempts to study the consistency, among other auxiliary comparisons, between popular generative artificial intelligence (AI) robots in the evaluation of various perceived user experience dimensions of mobile device operating system versions or, more specifically, iOS and Android versions. A handful of robots were experimented with, ending up with Dragonfly and GPT-4 being the only two eligible for in-depth investigation where the duo was individually requested to accord rating scores to the six major dimensions, namely (1) efficiency, (2) effectiveness, (3) learnability, (4) satisfaction, (5) accessibility, and (6) security, of the operating system versions. It is noteworthy that these dimensions are from the perceived user experience's point of view instead of any "physical" technology's standpoint. For each of the two robots, the minimum, the maximum, the range, and the standard deviation of the rating scores for each of the six dimensions were computed across all the versions. The rating score difference for each of the six dimensions between the two robots was calculated for each version. The mean of the absolute value, the minimum, the maximum, the range, and the standard deviation of the differences for each dimension between the two robots were calculated across all versions. A paired sample t-test was then applied to each dimension for the rating score differences between the two robots over all the versions. Finally, a correlation coefficient of the rating scores was computed for each dimension between the two robots across all the versions. These computational outcomes were to confirm whether the two robots awarded discrimination in evaluating each dimension across the versions, whether any of the two robots systematically underrated or overrated any dimension vis-à-vis the other robot, and whether there was consistency between the two robots in evaluating each dimension across the versions. It was found that discrimination was apparent in the evaluation of all dimensions, GPT-4 systematically underrated the dimensions satisfaction ($p = 0.002 < 0.05$) and security ($p = 0.008 < 0.05$) compared with Dragonfly, and the evaluation by the two robots was almost impeccably consistent for the six dimensions with the correlation coefficients ranging from 0.679 to 0.892 (p from 0.000 to 0.003 < 0.05). Consistency implies at least the partial trustworthiness of the evaluation of these mobile device operating system versions by either of these two popular generative AI robots based on the analogous concept of convergent validity.

Keywords: Artificial intelligence, Robots, Perceived user experience, Mobile device operating system versions

INTRODUCTION

Mobile devices (e.g., phones) have become as prevalent as probably any necessity in life. The widespread adoption of mobile devices led to a few prominent mobile device operating systems, notably Android and iOS (in alphabetical order for neutrality) and to a lesser extent HarmonyOS, Windows Phone, BlackBerry, etc. (Ali, 2023) Despite their basic objectives to communicate, take photographs, run application programs (i.e., apps), they may differ in their features, capabilities, etc. Evaluating such features and capabilities of mobile device operating systems is indispensable for selection among them for any particular potential users. Evaluation of mobile device operating systems has been a hot topic in both academia and industry for over a decade (Ahmad Sheikh *et al.*, 2013; Haris *et al.*, 2017; Shahdi Ahmad *et al.*, 2013; Shimada *et al.*, 2006; Wukkadada, Nambiar and Nair, 2015). Such traditional means of evaluation as surveys, user feedback, weighted scoring, etc. are time-consuming and somewhat subjective and can usually represent a limited sample of respondents' or evaluators' opinions. Generative artificial intelligence (AI) robots emerge as a potentially alternative means of evaluation through which operating systems can be alternatively evaluated and compared by robots in an automated manner. Having said that, there exists no absolutely impartial "yardstick" to assess the trustworthiness of any particular evaluation modality. At best, one can derive inkling of which evaluation to trust by gauging the consistency between multiple evaluations. If all evaluations happen to be somewhat consistent, chances are that all of them are trustworthy to an extent, although in theory all being coherently spurious can never be ruled out. This way to assess the trustworthiness of evaluation modalities is analogous to the concept of convergent validity for an operationalized instrument to measure an abstract construct. This article aims to scrutinize such consistency, among other auxiliary comparisons, between as many popular generative AI robots as possible in the evaluation of the perceived user experience with various mobile device operating system versions. It is noteworthy that perceived user experience as opposed to "physical" technology is the focus of the study.

GENERATIVE AI ROBOTS FOR EVALUATING AND COMPARING MOBILE DEVICE OPERATING SYSTEMS

The fundamentals of generative AI is summarized in two short paragraphs by Aydın and Karaarslan (2023):

> "Generative AI is an artificial intelligence field that concentrates on generating new and original information by machine learning on massive databases of experiences. There are several potential applications for generative AI, such as the generation of new pictures, text, and music, as well as computer vision, natural language processing, and speech recognition. As an example, generative models may be used to create realistic images for video games, simulations, and virtual reality, as well as novel chemical compounds for medicinal research.

> This is performed by using a model that has been trained on a large dataset of examples and constructing new instances that are comparable to the training dataset. Natural language processing generative models may be used to produce new material that is equivalent to the style and substance of a specific author or subject. Among the most popular generative AI models are GANs [Generative Adversarial Networks] (Wang et al., 2017), VAEs [Variational Autoencoders] (Sønderby et al., 2016), and Transformer-based models (Bouschery, Blazevic and Piller, 2023)."

For the evaluation of the perceived user experience with various mobile device operating systems, generative AI robots can be programmed or otherwise instructed to analyze the operating systems' such perspectives or dimensions as (1) efficiency, (2) effectiveness, (3) learnability, (4) satisfaction, (5) accessibility, and (6) security, among some others of less relevance. Thereby, each robot generates an objective and standardized rating score for each dimension of each mobile device operating system for the purpose of evaluation.

In recent years, researches have studied the application of generative AI robots to phenomena concerning mobile device operating systems. For example, Amin *et al.* (2022) proposed technique to cater for malware detection, which was by design a deep learning model making use of generative adversarial networks. It was responsible for detecting Android malware by means of famous two-player game theory for a rock-paper-scissor problem. The researchers used three state-of-the-art datasets and a large-scale dataset of opcodes extracted from the Android Package Kit bytecodes. The technique achieved an F1 score of 99% with a receiver operating characteristic of 99% on the bytecode dataset.

Another example is that Huang *et al.* (2022) proposed Android-SEM, which was an Android source code semantic enhancement model based on transfer learning. The proposed model was built upon the Transformer architecture to achieve a pre-training framework for generating code comments from malware source codes. The performance of the pre-training framework was optimized using a generative adversarial network. The proposed model relied on a novel regression model-based filter to retain high-quality comments and source codes for feature fusion pertinent to semantic enhancement. Creatively, and contrary to conventional methods, a quantum support vector machine (QSVM) was incorporated for classifying malicious Android codes by combining quantum machine learning and classical deep learning models. The results proved that Android-SEM achieved accuracy levels of 99.55% and 99.01% for malware detection and malware categorization, respectively.

Having said that, the author is not aware of any extant literature specific to the evaluation of mobile device operating systems by means of generative AI robots. This is exactly the gap that this article is to fill. In particular, this article deals with the two most popular mobile device operating systems, namely, Android and iOS. The versions concerned include Android 7.0, 7.1, 8.0, 8.1, 9, 10, 11, 12, 12.1, and 13 and iOS 9.3.5, 9.3.6, 10.3.3, 10.3.4, 11, 12, 13, 14, 15, 16.4.1, and 16.5 RC. These 21 versions were selected simply by referencing some commonplace commercial websites on Android and iOS version history, for example, Raphael (2023) and Casserly (2023).

METHODOLOGY

Data and Materials

The present study started off experimenting with six very popular generative AI robots, namely, Claude+, Dragonfly (de Souza, de Andrade Neto, and Roazzi, 2023), GPT-4 (Zhang *et al*., 2023), Sage (de Souza, de Andrade Neto, and Roazzi, 2023), Claude-instant (de Souza, de Andrade Neto, and Roazzi, 2023), and NeevaAI (de Souza, de Andrade Neto, and Roazzi, 2023) as candidates for the evaluation and comparison, all of them having been incorporated into the AI portal poe.com. Six major dimensions to evaluate the perceived user experience with any mobile device operating system were identified as (1) efficiency (Raptis *et al*., 2013), (2) effectiveness (Raptis *et al*., 2013), (3) learnability, (4) satisfaction, (5) accessibility, and (6) security, which were to be rated by the robots in this study. Efficiency refers to the speed and accuracy with which users can complete their tasks using a mobile device operating system. It is influenced by factors such as the speed of the system, responsiveness to user inputs, and the design of the user interface. The efficiency of a mobile device operating system can be measured by metrics such as the average time for users to perform a certain number of specified tasks on the operating system (Nielsen, 1993). Users complete tasks fast on mobile device operating systems with high efficiency. Effectiveness refers to the chance of users completing tasks successfully and correctly on a mobile device operating system. Effectiveness of a mobile device operating system can be measured by metrics such as users' task success rate and the number of errors made by users on the operating system. Users' task success rate is high and users' number of errors is low for mobile device operating systems with high effectiveness. Learnability refers to the ease with which users can learn how to use a mobile device operating system. It is influenced by factors such as the availability of documentation, the simplicity of the user interface, and the consistency of the user interface. The learnability of a mobile device operating system can be measured by metrics such as the time for novice users to reach a specified level of proficiency in using the operating system or for them to be able to complete a certain task successfully on the operating system (Nielsen, 1993). Novice users' time to be able to complete a certain task successfully on mobile device operating systems with high learnability is short. Satisfaction refers to how pleasant it is to use a mobile device operating system. It is influenced by factors such as the visual appeal of the user interface, the ease of use, and the availability of features. The satisfaction of users can be measured by metrics such as subjective ratings in a questionnaire given to users as part of debriefing after user tests or after field deployment (Nielsen, 1993). Users' satisfaction is high when using mobile device operating systems is conducive to pleasure. Accessibility refers to the ease with which users with disabilities can use a mobile device operating system (Lazar, Goldstein and Taylor, 2015). It is influenced by factors such as the availability of assistive technologies, the design of the user interface, and the compatibility with screen-reading software. The accessibility of a mobile device operating system can be measured by metrics such as the time to complete specific tasks using assistive technologies and the number of errors made by users with disabilities. Mobile device operating systems enabling short times to complete tasks using assistive technologies with minimal errors have high accessibility. Security refers to the ability of a mobile device operating system to protect user data and prevent unauthorized access. It is influenced by factors such as the availability of security features, the strength of encryption, and the compatibility with security software. The security of a mobile device operating system can be measured using metrics such as the number of security vulnerabilities, the ability to detect and prevent malware, and the compatibility with security software. Security of mobile device operating systems is high if they have small numbers of security vulnerabilities, are capable of detecting and preventing malware, and are highly compatible with security software.

Then, the following request for rating scores was submitted to each of the six robots:

> "For each of the six dimensions (1) efficiency, (2) effectiveness, (3) learnability, (4) satisfaction, (5) accessibility, and (6) security, please give a rating score to each of the mobile phone operating system versions iOS 9.3.5, 9.3.6, 10.3.3, 10.3.4, 11, 12, 13, 14, 15, 16.4.1, and 16.5 RC and Android 7.0, 7.1, 8.0, 8.1, 9, 10, 11, 12, 12.1, and 13 based on a scale of 1 to 10 (1 being the worst and 10 the best). Please derive your scores from global users' textual comments on these six dimensions of these versions as appear all around the web." Sage's reply can be summarized by its statement "Newer versions of operating systems tend to have better scores in terms of efficiency, effectiveness, learnability, satisfaction, accessibility, and security," which does not contain any rating scores and precludes further study. Likewise, Claude+, Claude-instant, and NeevaAI outright declined to provide any rating scores.

Dragonfly replied with complete rating scores for all the six dimensions and all the 21 mobile device operating system versions enumerated in the above request whereas GPT-4 yielded the corresponding rating scores for only 17 operating system versions, skipping those of iOS 16.4.1, iOS 16.5 RC, Android 12.1, and Android 13. In other words, only Dragonfly's and GPT-4's rating scores, respectively, for 21 and 17 operating systems were amenable to further analysis. Please note that both the request above explicitly underscores "...derive your scores from global users' textual comments on these six dimensions of these versions as appear all around the web." Stated differently, each robot presumably derived its rating scores from global users' textual comments appearing on the worldwide web instead of simply echoing any analogous scores already existing somewhere.

Analysis

For each of the two robots Dragonfly and GPT-4, the minimum, the maximum, the range, and the standard deviation of the rating scores for each of the six dimensions were calculated across all the 21 (for Dragonfly) and 17 (for GPT-4) mobile device operating system versions. If there is a considerable range and standard deviation for a particular dimension, it is ascertained that the corresponding robot accords discrimination in rating the dimension across the operating system versions.

Subsequently, the rating score difference for each of the six dimensions between the two robots was computed for each of the 17 operating system versions (common to both Dragonfly and GPT-4). The mean of the absolute values, the minimum, the maximum, the range, and the standard deviation of the differences for each dimension between the two robots were computed across all the 17 operating systems versions. If the mean of the absolute values, the range, and the standard deviation for a particular dimension are sufficiently small, it is revealed that the two robots neither overrate nor underrate erratically with respect to each other the dimension across the operating system versions. A paired sample *t*-test was then applied to each dimension for the rating score differences between the two robots over all the 17 operating system versions. If the *t*-test is significant for a particular dimension and the corresponding mean difference is positive (negative), it is attested that the first robot systematically overrates (underrates) the dimension with respect to the second robot.

Finally, a correlation coefficient of the rating scores was calculated for each of the six dimensions between the two robots across the 17 operating system versions. If the correlation coefficient is positively high enough (for instance, over 0.6) for a particular dimension, it is verified that there is consistency between the two robots in rating the dimension across the operating system versions.

RESULTS

Table 1 lists the minimum, the maximum, the range, and the standard deviation of the rating scores for each of the six dimensions across the 21 and 17 mobile device operating system versions as rated by each of the two robots Dragonfly and GPT-4, respectively. Both Dragonfly and GPT-4 rated all the six dimensions with appreciable discrimination as indicated by the substantial ranges and standard deviations. Have said that, Dragonfly rated the two dimensions effectiveness and learnability with comparatively more discrimination than the remaining four dimensions as evidenced by the difference in the ranges and the standard deviations between these two groups of dimensions. In contrast, GPT-4 rated the dimension learnability with comparatively less discrimination than all the other five dimensions as substantiated by the former's smaller range and standard deviation.

Table 2 enumerates the mean of the absolute values, the minimum, the maximum, the range, and the standard deviation of the rating score differences for each of the six dimensions across the 17 mobile device operating system versions (common to both Dragonfly and GPT-4) between the two robots. The disparity between Dragonfly and GPT-4 in rating each dimension is rather even across all the six dimensions. Notwithstanding, one may

Table 1. The minimum, the maximum, the range, and the standard deviation of the rating scores for each of the six dimensions across the 21 and 17 mobile device operating system versions as rated by each of the two robots Dragonfly and GPT-4, respectively.

Robot (sample size n)	Minimum/ maximum/ range/ standard deviation	Efficiency	Effectiveness	Learnability	Satisfaction	Accessibility	Security
Dragonfly ($n = 21$)	Minimum	7	6	6	7	7	7
	Maximum	9	9	9	9	9	9
	Range	2	3	3	2	2	2
	Standard derivation	0.6690	0.9437	0.8646	0.6690	0.5732	0.6761
GPT-4 ($n = 17$)	Minimum	6	6	7	6	6	6
	Maximum	9	9	9	9	9	9
	Range	3	3	2	3	3	3
	Standard derivation	1.1789	1.2632	0.8828	1.1757	1.1311	1.2632

be in the opinion that such disparities between Dragonfly and GPT-4 in rating the dimensions effectiveness and learnability are marginally smaller than in rating the remaining four dimensions given that the means of the absolute values of the corresponding rating score differences for the former two dimensions are somewhat less than those for the latter four dimensions though the ranges and the standard deviations of the corresponding rating score differences for the former two dimensions are not far less than those for the latter four dimensions. In other words, the two robots overrate or underrate erratically with respect to each other the former two dimensions across the operating system versions only slightly less than the latter four dimensions.

Table 2. The mean of the absolute values, the minimum, the maximum, the range, and the standard deviation of the rating score differences for each dimension between the two robots.

Differences (sample size n)	Mean of the absolute values/ minimum/ maximum/ range/ standard deviation of the differences	Efficiency	Effectiveness	Learnability	Satisfaction	Accessibility	Security
Dragonfly – GPT-4 ($n = 17$)	Mean of the absolute values	0.6471	0.3529	0.4706	0.6471	0.7059	0.7059
	Minimum	-1	-1	-1	0	-1	-1
	Maximum	2	1	1	2	1	2
	Range	3	2	2	2	2	3
	Standard derivation	0.8703	0.6002	0.7071	0.7019	0.8314	0.7952

Table 3 depicts the paired sample t-tests of the rating score differences for each of the six dimensions between the two robots over the 17 mobile device operating system versions (common to both Dragonfly and GPT-4). With respect to Dragonfly, GPT-4 appears to underrate the two dimensions satisfaction and security at the 1% significance level ($p < 0.01$).

Table 3. The paired sample *t*-test of the rating score differences for each of the six dimensions between the two robots.

Differences (sample size *n*)	Dimension	Mean difference / [95% confidence interval]	*t* (*p*-value) / degrees of freedom
Dragonfly – GPT-4 ($n = 17$)	Efficiency	.412 / [-0.036, 0.859]	1.951 (.069) / 16
	Effectiveness	-.118 / [-0.426, 0.191]	-.808 (.431) / 16
	Learnability	.000 / [-0.364, 0.364]	.000 (1.000) / 16
	Satisfaction	.647 / [0.286, 1.008]	3.801 (.002**) / 16
	Accessibility	.235 / [-0.192, 0.663]	1.167 (.260) / 16
	Security	.588 / [0.179, 0.997]	3.050 (.008**) / 16

** $p < 0.01$

Table 4 illustrates the correlation coefficient of the rating scores for each of the six dimensions between the two robots over the 17 platforms (common to both Dragonfly and GPT-4), the 95% conference interval for the correlation coefficient, and the *p*-value to test whether the coefficient differs from zero. The two robots are highly, positively correlated and thus consistent in rating all the six dimensions as affirmed by their corresponding positive correlation coefficients and by the corresponding *p*-values being significant at the 1% significance level ($p < 0.01$). It is worth noting that a highly positive correlation coefficient and thus consistency for a particular dimension imply a high rating score of an operating system version for the dimension by one robot being generally associated with a high rating score of that operation system version for that dimension by another robot and vice versa albeit these two rating scores may not necessarily be the same or not even close.

Table 4. The correlation coefficient of the rating scores for each of the six dimensions between the two robots.

Dimension	Correlation coefficient / [95% confidence interval]	*p*-value
Efficiency	.687 / [0.3081, 0.8778]	.002**
Effectiveness	.892 / [0.7201, 0.9607]	.000**
Learnability	.679 / [0.2945, 0.8743]	.003**
Satisfaction	.852 / [0.6289, 0.9455]	.000**
Accessibility	.714 / [0.3553, 0.8894]	.001**
Security	.827 / [0.5749, 0.9357]	.000**

** $p < 0.01$

In summary, it may be rather safe to trust and rely on Dragonfly's and GPT-4's ratings of mobile device operating system versions for all the six dimensions efficiency, effectiveness, learnability, satisfaction, accessibility, and security in view of the consistency between these two robots in all the six dimensions and thus by analogy with the concept of convergent validity.

CONCLUSION AND DISCUSSION

There are conceivably a multitude of factors contributing to any inconsistency, however small, between generative AI robots' rating scores in the evaluation of mobile device operating systems (and virtually anything). These factors can be epitomized by the additional caveat in Sage's reply to the aforementioned request for rating scores:

> "As an AI language model, I can analyze textual comments on the six dimensions of the mobile phone operating system versions you mentioned, but I cannot provide a rating score for each version. The reason is that the analysis of textual comments is subjective and can be influenced by various factors such as the source of the comments, the language used, and the context in which the comments were made. Additionally, I cannot guarantee the accuracy and reliability of the comments found all around the web."

In summary, the factors are:

1. Different robots may derive their rating scores from different textual comments on the web, yielding disparate scores.
2. The textual comments on the web may themselves be subjective, so any robots' rating scores based on such comments may concomitantly be subjective and specific to the robots concerned as well.
3. The textual comments on the web may be inaccurate and unreliable and subject to influences like their sources, the (human) languages presenting them, and the contexts, intensifying the deviation of a robot's rating scores from those of another robot as all such scores are dependent on the different textual comments accessible to different robots.

After all, the consistency between Dragonfly and GPT-4 renders their ratings of mobile device operating system versions trustworthy to a certain extent in the evaluation of such operating system versions though the plausibility of both being coherently erroneous can never be presumed non-existent.

REFERENCES

Ahmad Sheikh, Aijaz, Tehseen Ganai, Prince, Ahmad Malik, Nisar and Ahmad Dar, Khursheed. (2013) Smartphone: Android vs iOS, SIJ TRANSACTIONS ON COMPUTER SCIENCE ENGINEERING & ITS APPLICATIONS (CSEA) Volume 1 No. 4. Website: 10.9756/SIJCSEA/V1I4/0104600401

Ali, Ray. (2023) What are the different mobile operating systems? U SWITCH. Website: https://www.uswitch.com/mobiles/guides/mobile-operating-systems/

Amin, Muhammad, Shah, Babar, Sharif, Aizaz, Ali, Tamleek, Kim, Ki-Il and Anwar, Sajid (2022). Android malware detection through generative adversarial networks, TRANSACTIONS ON EMERGING TELECOMMUNICATIONS TECHNOLOGIES Volume 33 No. 2. Website: https://doi.org/10.1002/ett.3675

Aydın, Ömer and Karaarslan, Enis. (2023) Is ChatGPT Leading Generative AI? What is Beyond Expectations? Website: https://ssrn.com/abstract=4341500 or https://dx.doi.org/10.2139/ssrn.4341500

Bouschery, Sebastian G., Blazevic, Vera and Piller, Frank T. (2023) Augmenting human innovation teams with artificial intelligence: Exploring transformer-based language models, JOURNAL OF PRODUCT INNOVATION MANAGEMENT Volume 40 No. 2. Website: https://doi.org/10.1111/jpim.12656

Casserly, Martyn. (2023) iOS versions: Every version of iOS from the oldest to the newest, MACWORLD. Website: https://www.macworld.com/article/1659017/ios-versions-list.html

de Souza, C., de Andrade Neto, A. S. and Roazzi, A. (2023). Are the New AIs Smart Enough to Steal Your Job? IQ Scores for ChatGPT, Microsoft Bing, Google Bard and Quora Poe. Website: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4412505

Haris, Muhammad, Jadoon, Basit, Yousaf, Madiha and Khan, Farhan Hassan. (2017) "Evolution of Android Operating System: A Review", proceedings of the 2nd International Conference on Advanced Research, Melbourne, Australia. Website: https://apiar.org.au/wp-content/uploads/2017/12/16_ICAR_Nov_BRR708_ICT-125-135.pdf

Huang, Yizhao, Li, Xingwei, Qiao, Meng, Tang, Ke, Zhang, Chunyan, Gui, Hairen, Wang, Panjie and Liu, Fudong. (2022) Android-SEM: Generative Adversarial Network for Android Malware Semantic Enhancement Model Based on Transfer Learning, ELECTRONICS Volume 11 No. 5. Website: https://doi.org/10.3390/electronics11050672

Lazar, Jonathan, Goldstein, Daniel F. and Taylor, Anne (2015) Ensuring Digital Accessibility through Process and Policy. Morgan Kaufmann.

Nielsen, Jakob. (1993) Usability Engineering. San Diego: Academic Press.

Raphael, J. R. (2023) Android versions: A living history from 1.0 to 14, COMPUTERWORLD. Website: https://www.computerworld.com/article/3235946/android-versions-a-living-history-from-1-0-to-today.html

Raptis, Dimitrios, Tselios, Nikolaos, Kjeldskov, Jesper and Skov, Mikael. B. (2013) "Does size matter? Investigating the impact of mobile phone screen size on users' perceived usability, effectiveness and efficiency", proceedings of the 15th International Conference on Human-computer Interaction with Mobile Devices and Services, Munich, Germany. Website: https://doi.org/10.1145/2493190.2493204

Shahdi Ahmad, Mohd, Emyra Musa, Nur, Nadarajah, Rathidevi, Hassan, Rosilah and Effendy Othman, Nor. (2013) "Comparison between Android and iOS Operating System in terms of security", proceedings of the 2013 8th International Conference on Information Technology in Asia (CITA), Kota Samarahan, Malaysia. Website: https://ieeexplore.ieee.org/abstract/document/6637558/authors#authors

Shimada, Michihiro, Minato, Takashi, Itakura, Shoji and Ishiguro, Hiroshi. (2006) "Evaluation of Android Using Unconscious Recognition", proceedings of the 2006 6th IEEE-RAS International Conference on Humanoid Robots, Genova, Italy. Website: https://ieeexplore-ieee-org.mpu.idm.oclc.org/search/searchresult.jsp?searchWithin=%22Publication%20Number%22:4115555&searchWithin=%22Document%20Title%22:Evaluation%20of%20Android%20Using%20Unconscious%20Recognition

Sønderby, Casper Kaae, Raiko, Tapani, Maaløe, Lars, Sønderby, Søren Kaae and Winther, Ole. (2016) Ladder Variational Autoencoders, ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS Volume 29 (NIPS 2016). Website: https://proceedings.neurips.cc/paper/2016/hash/6ae07dcb33ec3b7c814df797cbda0f87-Abstract.html

Wang, K., Gou, C., Duan, Y., Lin, Y., Zheng, X. and Wang, F.-Y. (2017) Generative Adversarial Networks: Introduction and Outlook, IEEE/CAA JOURNAL OF AUTOMATICA SINICA Volume 4 No. 4. Website: 10.1109/JAS.2017.7510583.

Wukkadada, Bharati, Nambiar, Ramith and Nair, Amala. (2015) Mobile Operating System: Analysis and Comparison of Android and iOS, INTERNATIONAL JOURNAL OF COMPUTING AND TECHNOLOGY Volume 2 No. 7. Website: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=101f723d06082a6c06d327077607aded1e5c7a56

Zhang, C., Zhang, C., Zheng, S., Qiao, Y., Li, C., Zhang, M., Dam, S. K., Thwal, C. M., Tun, Y. L., Huy, L. L., Kim, D., Bae, S.-H., Lee, L.-H., Yang, Y., Shen, H. T., Kweon, I. S. and Hong, C. S. (2023) A Complete Survey on Generative AI (AIGC): Is ChatGPT from GPT-4 to GPT-5 All You Need? arXiv:2303.11717v1 [cs. AI] Website: https://arxiv.org/abs/2303.11717