An Innovative Approach to Develop Persona from Application Reviews

Dylan Clements, Elysia Giannis, Fi Crowe, Mike Balapitiya, Jason Marshall, Paul Papadopoulos and Tanjila Kanij[a]

Department of Software Systems and Cybersecurity, Monash University, Clayton, Melbourne, Australia

Keywords: Persona, E-commerce, Application Reviews.

Abstract: Software end users are diverse by nature and their different facets influence the way they use software. An understanding of the users and their needs are achieved by engaging with the users during requirement engineering. However, sometimes recruiting users during requirement engineering phase can be very challenging. An accessible way to understand a user's perspective and traits is through user application reviews. This research paper proposes an innovative approach to develop user personas from a data set of e-commerce application user reviews by using GPT-3 and PATHY. This enables the development teams to see different demographic data, as well as overall frustrations and expectations that users of their platform possess, so developers know how to enhance their software solutions. This is also helpful to developers of new e-commerce applications.

1 INTRODUCTION

Most software solutions provide an inadequate user experience due to the developers lacking the understanding of enduser needs, as Mathews et al. states "end user diversity is not sufficiently contemplated" (Mathews et al.,). This ultimately limits the succession of these platforms. As such understanding the endusers of a software is essential. Generally this is done through rigorous user research during the requirement engineering phase. However, the enduser characteristics, as well as their preferences, keep changing and software needs to take these into account throughout the development as well as after deployment.

Understanding end users through rigorous research, is most of the times time, cost and effort-consuming. Another major challenge is to establish communication with the end users and get valuable insight from them. To address these challenges, we propose an innovative approach to understanding the end user of software from an alternative source of information - application reviews provided by the end users. By collecting a large volume of application reviews and evaluating a range of characteristic traits of users, a chance to visualise and assimilate what a particular user of software would resemble, is presented. This is what a user persona entails. A persona is a description of a fictional character who will use the software (Cooper, 1999). By collating particular user facets and outlining their frustrations, we are able to develop insightful personas which describe the key user information and their feelings towards the software. This then helps software developers understand their users better.

To develop the proposed framework we selected e-commerce as our application domain. It has been shown that age bias is found in various e-commerce software solutions, meaning it is harder for people of certain ages to proficiently use these applications (McIntosh et al., 2021; El Shamy and Hassanein, 2018). Yet, it is unclear which other facets beyond age affect the way people use e-commerce platforms. E-commerce is a domain where people of different demographics purchase goods and services and these users can also share their experiences of the application being used (Obie et al., 2021). This information can be leveraged, and act as a valuable source of problem statements, ideas and requests from users, which could ultimately help development teams understand potential issues with their software, as well as find ways in which they can be mitigated. Additionally, e-commerce application usage has risen vastly due to the evolution of online marketing and changes in economic and environmental factors (Grundy et al., 2018). For instance, the COVID-19 pandemic forced many people to use these platforms out of obligation. With the population's increasing interaction with e-

[a] https://orcid.org/0000-0002-5293-1718

Clements, D., Giannis, E., Crowe, F., Balapitiya, M., Marshall, J., Papadopoulos, P. and Kanij, T.
An Innovative Approach to Develop Persona from Application Reviews.
DOI: 10.5220/0011996000003464
In *Proceedings of the 18th International Conference on Evaluation of Novel Approaches to Software Engineering (ENASE 2023)*, pages 701-708
ISBN: 978-989-758-647-7; ISSN: 2184-4895

commerce applications, it is imperative that the user experience is critically considered when developing and improving these applications to reduce any negative impacts of insufficiently understanding users.

Given the volume of reviews sent to app stores, many developers do not have the capacity to extract and analyse meaningful insights from all of them (Mathews et al.,). Therefore, an automated machine learning tool that can be used to accurately capture the user's experience to get a clearer understanding of the platform's end users would be advantageous to development teams. This would enable software developers to focus more on implementing improvements for the application, rather than manually analysing their platforms' reviews (Mao et al., 2005). To assist e-commerce developers in understanding their end users more extensively, we can extract certain facets from the user application reviews of the platforms being assessed using GPT-3, and segment the responses based on their characteristics to investigate possible trends between the sentiment and user facets. This will help developers discern what types of users are avidly or apathetically using their application, which can lead to further investigation on how the sentiment for users that show lower interest to the platform can be restored. This also increases the potential for accommodating more diverse users.

The remainder of this paper is structured as follows: Section 2 details the review of related research and background of this research. Section 3 describes our approach in detail. Section 4 presents the results of analysing e-commerce app reviews and developing persona. Section 5 illustrates the results of the evaluation of our proposed method, Section 6 demonstrates the possible threats to the validity of the research. Finally, section 8 concludes our findings and outlines any future research.

2 REVIEW OF LITERATURE

2.1 Automation

Gao et al. attempted to analyse application reviews to identify existing issues (Gao et al., 2018) and found that "noise words" (e.g., misspelled words) made it challenging to identify new app issues through automation. They derived an approach information extraction removing the noisy words. Similarly, Malgaonkar et al. reduced the burden of human involvement by using natural language processing, feature engineering, and word sense disambiguation, to automatically generate the taxonomy(Malgaonkar et al., 2022). Mcilroy et al. analysed large app reviews and found that up to 30% of the reviews raise various issues in a single review, including *"feature requests"* and *"bug reports"* (Mcilroy et al.,).

2.2 Facet Research

Grundy outlines that age and gender are some of the characteristics that need to be better incorporated into requirements engineering and design (Grundy, 2021). Tekin and Sebnem support Grundy's analysis of the characteristics that influence one's use of e-commerce and amount of online expenditure, with research displaying the effects of gender and age on users (Kose and Arslan, 2020). An initial study has looked at the facets of risk aversion, technical proficiency, visual impairment, and attention among people of different ages while interacting with e-commerce (McIntosh et al., 2021). Middle-aged individuals were found to be more likely to engage in e-commerce activities, compared to relatively younger and older individuals by Tekin and Sebnem (Kose and Arslan, 2020).

2.2.1 GPT-3 Overview

Generative Pre-Trained Model 3 (GPT-3) by OpenAI is the largest language model constructed to date (Dale, 2021). The model is trained on 499 billion tokens of web content, including all of Wikipedia, a variety of books, and a large portion of web pages (Singh et al., 2021). GPT-3 consists of 12 layers of transformer decoder blocks with 175 billion trainable parameters (Acheampong et al., 2021). Text weight embeddings and positional embeddings are passed as inputs into 96 attention layers, then into a feedforward layer, which then outputs a probability distribution (Acheampong et al., 2021).

Advantages. Although GPT-3's main application is in automated text generation (Acheampong et al., 2021; Dale, 2022), GPT-3 can be used in a range of scenarios, including text classification (or sentiment analysis) (Mathews et al.,). The extensive training data set for GPT-3 (Mathews et al.,) and its computational power (Alexandridis et al., 2021) alone makes it a prime candidate for our application of NLP on bulk user review data sets.

Limitations. Despite GPT-3's computational power and access to a large training set, the model does have some limitations. GPT-3's large training set means that text classification is very costly to run and requires significant processing power (Dale, 2021). This is why GPT-3 is currently only available via API (Alexandridis et al., 2021; Ashraf and Perez, 2020), as the typical machine could not handle the large & complex processes of GPT-3 (Acheampong et al., 2021). In addition to this, given that the training data set is comprised of a wide collection of books, all of Wikipedia, and an abundance of web pages, the model will be slightly biased, and reflect predictions similar to that of the training data (Mathews et al., ; Dale, 2021; Ashraf and Perez, 2020).

2.3 Persona Development

The fundamental idea behind persona development is gathering information about the users and grouping those into personas. Guo and Ma use a technique that focuses mainly on three pillars of a persona: biographic background; archetype and personality (Guo and Ma, 2018). Ferreira et al. and Marr summarise that the key steps to developing personas for e-commerce platforms are to gather demographic data, general attitudes toward the application, main goals and frustrations they bear, and finally commonalities between the respective facets so they can be grouped into personas (Ferreira et al., 2017)(Marr, 2020). All these techniques require rigorous user research.

PATHY is another technique of persona development, that is based on empathy mapping (Ferreira et al., 2016). The PATHY technique bridges the gap between empathy maps and personas themselves and elicits theoretical user requirements for an application (Ferreira et al., 2016). To improve the support provided to software developers, a second component was added in the PATHY technique to deal with issues related to identifying application features and characteristics (Ferreira et al., 2016). An advantage of the PATHY technique is that it found more potential application requirements than Acuña et al.'s technique (Acuña et al., 2012). Given its clear strengths, we have chosen to use the PATHY technique for this research.

3 METHODOLOGY

Application reviews were extracted with an automated script written in Python, from 25 different e-commerce applications. The applications domains included: departmental stores, supermarket, pharmacy, retailer, food, fashion, and so on. A variety of types of domains ensures a variation of reviews and subsequently a variation of personas.

3.1 Development Environment

A development environment was configured to allow for collaborative execution and near-full automation of the execution of the research methodology. This environment included a private GitHub repository, Python installations (including dependencies) and JSON files. The OpenAI python API (OpenAI,) was used to communicate with GPT-3 to perform the analysis.

3.2 Facet Extraction

In order to extract user facets and identify their age, gender, expectations and frustrations from the extracted reviews, we used Open AI's deep learning platform, Generative Pre-trained Transformer 3, or GPT-3.

For setting up the environment we turned the temperature to zero that ensures most deterministic output is produced by GPT -3. We also selected "*text-davinci-2*" algorithm since this was the most advanced algorithm implemented by GPT-3 at that time. The following code snippet shows the queries we executed against each review.

```
prompts = [
"Rate the author's sentiment on a scale of -1 to 1: ",
"Predict the author's age: ",
"Predict the author's gender: ",
"What are the author's expectations?: ",
"What are the author's frustrations?: "
]
```

For example review "*I'm incredibly unhappy with how slow this app runs. I can barely use it.*", if we run the following code "*Rate the author's sentiment on a scale of -1 to 1:* ", we get output as "-1", which indicated negative sentiment.

Three of the facets we intended to capture were to be expressed quantitatively in our review characteristics data set. We found that age, gender and sentiment were most of the times expressed quantitatively however sometimes those were expressed qualitatively. e.g.: for age, instead of "30-40", the output would be *"The author's age is likely to be between 30 and 40"*. This was highly unusual despite setting GPT-3's temperature to 0, which was said to produce entirely deterministic results.

To work around this challenge, we developed intermediate analysis scripts using natural language processing tools such as NLTK. The intermediate scripts were used to extract the key points from each of the qualitative responses. Where any qualitative responses lacked keywords such as age, gender or sentiment scores, it was assumed that there were none found for that review.

Another challenge was that, some facets review characteristics were empty. This is most likely due to two main reasons. The first being that requests to the OpenAI API could have failed, despite the back-off

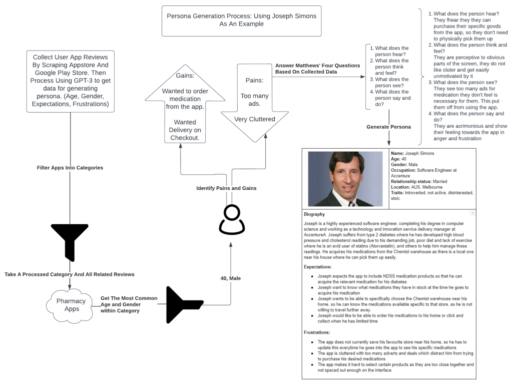

Figure 1: Persona development Diagram.

mechanism working successfully. The second, being that GPT-3 outputted a result that could not be captured as a valid review characteristic. For example, GPT-3 could output "There is not enough information to determine the author's gender". As a result, the author's gender would be marked as null and discounted from further analysis. 3.28% of review facets analysed reported null values.

3.3 Persona Development

Given that various persona development techniques are complex to implement and do not explicitly guide designers in identifying development relevant information, we decided to utilise a more methodical technique - PATHY. PATHY utilises empathy map checklist questions to create customer segment profiles, and a template to simplify its implementation. An empathy map reveals the rationale underlying users' actions, decisions and choices; therefore it helps in designing for users' real needs (Ferreira et al., 2016).

Matthews (Gray et al., 2010) proposed four different areas that should be covered when creating an empathy map: What does the person hear? What does the person think and feel? What does the person see? What does the person say and do? Bratsberg, H.M. (Bratsberg, 2012) mentioned Pain and Gain as important areas to look for. Based on these we adopted the following approach for persona development:

1. Found the most common age and gender within each e-commerce category and added this information to a persona.
2. Identified the pains and gains of said persona from the most common expectation and frustrations in user reviews.
3. Given these facets, the team prepared answers to Matthews' questions which contributed to their biography and character understanding.

4 RESULTS

We extracted 4999 app reviews from open source app store and play store. Among those, 4931 reviews were analysed. 68 reviews were discarded due to extraneous errors. Figures 2 outline the number of reviews analysed per category.

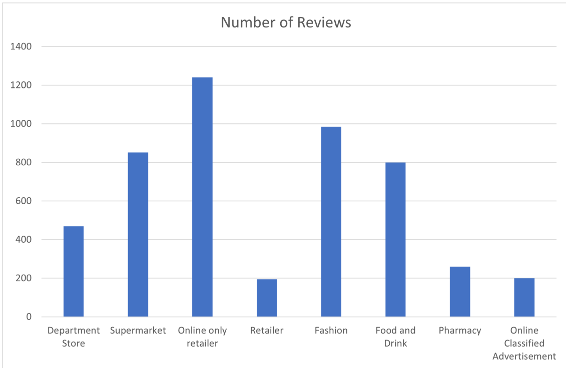

Figure 2: Number of reviews per category.

4.1 Facet Extraction

Three of the facets we intended to capture were to be expressed quantitatively in our review characteristics data set. We found that age, gender and sentiment were most of the times expressed quantitatively however sometimes those were expressed qualitatively. e.g.: for age, instead of "30-40", the output would be *"The author's age is likely to be between 30 and 40"*. This was highly unusual despite setting GPT-3's temperature to 0, which was said to produce entirely deterministic results.

To work around this challenge, we developed intermediate analysis scripts using natural language processing tools such as NLTK. The intermediate scripts were used to extract the key points from each of the qualitative responses. Where any qualitative responses lacked key words such as age, gender or sentiment scores, it was assumed that there were none found for that review.

Another challenge was that, some facets review characteristics were empty. This is most likely due to two main reasons. The first being that requests to the OpenAI API could have failed, despite the back-off mechanism working successfully. The second, being that GPT-3 outputted a result that could not be captured as a valid review characteristic. For example, GPT-3 could output "There is not enough information to determine the author's gender". As a result, the author's gender would be marked as null and discounted from further analysis. 3.28% of review facets analysed reported null values.

4.2 Persona Development

We developed eight personas for the eight categories of e-commerce applications. In order to develop the Pharmacy persona in Figure 3, we began by identifying the most common age and gender of reviewers of the pharmacy mobile applications that we analysed. We found that this was 40-year-old men. Hence, our persona was given this gender and age.

To create this persona's pains and gains, we selected common expectations and frustrations which were generated from relevant apps. This included that the pharmacy should stock NDSS products and that they should be able to add a favourite store to get the product.

Given this information, we were able to determine what this person hears, thinks, feels, sees, says and does. This included giving him a medical condition (diabetes) that would require NDSS products. Additionally, we outlined his occupation which might lead to him expecting certain features such as the ability to add a favourite store. Also, we believe that in order to maintain a job and have a condition like this, one most likely would be quite stoic, think logically, and be inactive and introverted. While these were assumptions made by us, and may not be precise, they are guided by the information we extracted and add to the character of the developed persona. Consequently, the developed persona has increased empathy from developers and allows them to gain a deeper understanding of the end user.

Due to space limitations, all the personas are not presented in this article. Another example persona is the departmental store user persona - Makaylah 4.

5 EVALUATION

Assessing the accuracy of our results, and hence the accuracy of our subsequently developed personas requires consideration of two aspects. Firstly, we must assess the accuracy of GPT-3. To do so we must determine the precision and accuracy of GPT-3's Natural Language Processing model. Secondly, we need to ensure that our extracted reviews have enough data in them to extract the information we need, and the quality of discussion within these reviews needs to be high enough to add value to our persona.

GPT-3 has extracted four facets for us, therefore we must assess the accuracy of each of these aspects. The four aspects were the reviewers age, gender, sentiment, frustrations, and expectations. Thankfully these tests have already been conducted which verify the accuracy of transformer-based models. Since

Name: Joseph Simons
Age: 40
Gender: Male
Occupation: Software Engineer at Accenture
Relationship status: Married
Location: AUS, Melbourne
Traits: Introverted, not active, disinterested, stoic

Biography

Joseph is a highly experienced software engineer, completing his degree in computer science and working as a technology and innovation service delivery manager at AccentureA. Joseph suffers from type 2 diabetes where he has developed high blood pressure and cholesterol reading due to his demanding job, poor diet and lack of exercise where he is an avid user of statins (Atorvastatin), and others to help him manage these readings. He acquires his medications from the Chemist warehouse as there is a local one near his house where he can pick them up easily.

Expectations:

- Joseph expects the app to include NDSS medication products so that he can acquire the relevant medication for his diabetes
- Joseph want to know what medications they have in stock at the time he goes to acquire his medication
- Joseph wants to be able to specifically choose the Chemist warehouse near his home, so he can know the medications available specific to that store, as he is not willing to travel further away.
- Joseph would like to be able to order his medications to his home or click and collect when he has limited time

Frustrations:

- The app does not currently save his favourite store near his home, so he has to update this everytime he goes into the app to see his specific medications
- The app is cluttered with too many adverts and deals which distract him from trying to purchase his desired medications
- The app makes it hard to select certain products as they are too close together and not spaced out enough on the interface

Figure 3: A user persona of Joseph.

Name: Makaylah Martin
Age: 35
Gender: Female
Occupation: Unemployed
Relationship Status: Married
Location: Casuarina, Darwin, Australia
Traits: Suzy Homemaker, Organised, People Pleaser, Creative

Biography
Makaylah considers herself as a Suzy Homemaker and while she is unemployed, she considers herself to be an expert interior designer. When she is not hosting dinner parties, she is constantly looking for ways to improve the feng shui of her home. She is always on the lookout for a deal and considers it a waste of money if she doesn't capitalise on one.

Expectations:

- Instagram posts to show me cool new items or ideas
- Deals and promotions clearly shown and promoted
- Promote items based on previously shopped
- Able to filter products to find the best fit for a broad item that was searched

Frustrations:

- No quick buy option, thus preventing impulse buying
- Customer service being too slow to assist
- Long delivery times with no way of tracking

Figure 4: A user persona of Makaylah.

GPT-3 is the largest and most advanced transformer-based model at the time of writing we can assume the findings of previous studies would carry over. In particular we can look at the PAN Author Profiling task. As discussed earlier this competition proved, through 2016-2019 (Rangel et al., 2016; Rangel et al., 2017; Rangel et al., 2018; Rangel and Rosso, 2019), that age, gender and sentiment could be determined from twitter tweets. Therefore, we know that we are able to determine implicit information such as age and gender from short term text. Hence, GPT-3 can be utilized to estimate a reviewer's demographics and sentiment. Next is determining the explicit information of the reviewer's expectations and frustrations. Since this information is explicit, to determine GPT-3's accuracy, we would simply have to compare the expectation/frustration against the review and determine whether that aspect was present. Therefore, our verification for this process involved sampling 100 reviews and manually identifying their expectations and frustrations. We then analysed the expectations and frustrations derived by GPT-3. We found that in 87% of cases, the expectations and frustrations manually identified aligned with those identified by GPT-3.

Apart from the accuracy of GPT-3, we also had to ensure that there was enough detail in our reviews to afford GPT-3 the highest probability of extracting accurate information. Therefore, to assess the quality of our reviews, we used information power. In doing so we would assess our extracted reviews, before being passed into GPT-3, against five criteria items: *"study aim, sample specificity, use of established theory, quality of dialogue, and analysis strategy"* (Malterud et al., 2016). Firstly, the study aim in this case was to develop user personas from app reviews. As a persona consists a user's demographics and expectation/frustration and the nature of a review is for a user to express their expectation/frustration, and we have determined that we can accurately find their demographics, hence the information power is higher. Secondly, our sample specificity was sparse as situations occur where insufficient data was provided in the review, hence this lowers our information power. Thirdly, for established theory, we considered the quality of dialog or in other words how detailed the reviews were. In many cases we found that reviews were one sentence at maximum, hence we had a very low quality of dialog. This severely impacted the information power. Lastly for analysis strategy, considering all of the above, we opted to conduct cross-case studies. This involved collating all of the data in each app into one or two identifiable personas. Again, this reduced our information power. Therefore, with these considerations, identifying the most prevalent age, gender, sentiment, frustrations, and expectations, and creating larger, overarching categories for the apps was the strategy we followed. In doing so, our developed personas will more accurately represent the most prominent aspects we could extract.

To further access the accuracy of our developed personas, we can analyse how well it resembles existing personas that developers have already developed. We collected some e-commerce personas and compared those with the ones we developed. For example, the retail customer persona of *"Suggestible Sally"* (biz, 2015) shared many commonalities with our developed Makaylah persona, seen in Figure 4, who represented a department store shopper. Some key items were the positive response to marketing and engagement with support staff. Therefore, we can have higher confidence in the information we have extracted from our reviews.

6 THREATS TO VALIDITY

Although our methods produced reasonable results, there are a few threats to the validity of our findings.

6.1 Internal

Due to the training data set used for GPT-3, the results will contain some bias. Given that GPT-3 is trained on almost all of the public internet, it will contain the biases found in public web content (Mathews et al., ; Dale, 2021; Ashraf and Perez, 2020). Furthermore, we were only able to analyse a sample set of 4931 reviews from various categories of mobile apps. This implies that our findings may not apply to all e-commerce applications. Moreover, some of the GPT-3 analysis has required further manual analysis of the data set. This has caused potential for further internal errors on the data set due to incorrect classification of information. As we are using the PATHY persona development technique (Ferreira et al., 2016), there is some subjective analysis of information. However, we have taken steps to ensure these risks are minimized by automating as much analysis as possible. Moreover, we have also ensured that all personas were reviewed and agreed upon by all researchers.

6.2 External

In terms of the external validity of our research, this method is not directly transferable to be used on reviews written in languages other than English. However, the tooling we used to extract reviews can be used to obtain reviews from app stores across the world, and GPT-3 is capable of identifying the language in which a text is written (Chiriatti, 2020). Therefore the tooling we used can be extended to work with reviews in different languages, however, it would require changes to the methodology.

7 DISCUSSION

Implication for Software Development. To improve e-commerce applications and ensure that end users are satisfied with their experience, it is beneficial to understand the characteristics and pain points of users that might effect their use of the platform. Grundy states that characteristic factors of users should be incorporated in engineering and design requirements at an early stage to ensure the success of the product after its development life cycle (Grundy, 2021). It is also supported by Kose and Arslan that when these characteristics are taken into account, it can directly impact the use of e-commerce and online expenditure, showing how particular facets had a blunt correlation to them (Kose and Arslan, 2020). To assist e-commerce developers in understanding their end users more extensively, we could extract certain facets from the user application reviews of the platforms being assessed using GPT-3, and segment the responses based on their characteristics to investigate possible trends between the sentiment and user facets. We developed eight personas based on the findings. We believe this will help developers discern what types of users are avidly or apathetically using their application, which can lead to further investigation on how the sentiment for users that show lower interest in the platform can be restored. This will also be beneficial for developers of new e-commerce platforms to understand the diverse user facets and their needs.

Implication for Research. The research presents an important perspective of collecting enduser data and effectively using those for software development. We conducted a small-scale research with one selected domain such as - e-commerce. The findings of this approach indicate a number of things - firstly, the use of existing machine learning tools (eg. GPT-3) make it easy, efficient and less time consuming for analysing small-scale app reviews in order to understand end users' facets, frustrations and expectations. The initial promising results indicate that a customised machine learning algorithm can be also developed for this purpose if needed. Secondly, it was apparent that the app reviews contain a lot of information about end users, especially their expectations and frustrations about the software. This can be an excellent source of information about the end users. This framework can be further investigated and developed as a fully automated tool that can give developers information about their end users.

8 CONCLUSIONS

This research has shown that it is possible to develop user personas from e-commerce mobile application reviews. By using the PATHY persona development technique and the GPT-3 machine learning model, we were able to extract user facets, expectations and frustrations from a data set of 4931 Australian e-commerce app reviews. These findings can be used by software developers to better understand the users of their platforms. However, there are some threats to the validity of these findings, which should be considered in future work.

There are a number of ways in which this research can be extended. Firstly, by amending the methodology, we could extend this model to work with reviews written in different languages. This ability would allow us to work with a more diverse data set, resulting in more widely-application findings. Secondly, with a larger sample size, the validity of the research can be improved. If more reviews from different mobile applications are used to develop personas, it would lead to stronger results. Thirdly, our research takes a holistic approach to developing personas from mobile application reviews. More specific research can be conducted by sorting results by each application. This would allow us to investigate how personas relate to specific applications, and potentially see trends in personas from similar application categories. Finally, while we focused on e-commerce mobile application reviews, this model could be extended to process and analyse reviews from different platforms such as web applications.

ACKNOWLEDGEMENTS.

Kanij is supported by ARC Laureate Fellowship FL190100035.

REFERENCES

(2015). The eight personas of retail customer bases.

Acheampong, F. A., Nunoo-Mensah, H., and Chen, W. (2021). Transformer models for text-based emotion detection: a review of bert-based approaches. *The Artificial intelligence review*, 54(8):5789–5829.

Acuña, S. T., Castro, J. W., and Juzgado, N. J. (2012). A hci technique for improving requirements elicitation. *Inf. Softw. Technol.*, 54:1357–1375.

Alexandridis, G., Varlamis, I., Korovesis, K., Caridakis, G., and Tsantilas, P. (2021). A survey on sentiment analysis and opinion mining in greek social media. *Information (Basel)*, 12(8):331.

Ashraf, Muhammad Adnan, R. M. A. N. F. N. D. P. V. S. and Perez, F. (2020). A study of deep learning methods for same-genre and cross-genre author profiling. *Journal of Intelligent & Fuzzy Systems*, 39(15):2353–2363.

Bratsberg, H. M. (2012). Empathy maps of the foursight preferences.

Chiriatti, L. F. . M. (2020). Gpt-3: Its nature, scope, limits, and consequences. *Minds & Machines*, 30(1):681–694.

Cooper, A. (1999). The inmates are running the asylum: Why high-tech products drive us crazy and how to restore the sanity pearson education.

Dale, R. (2021). Gpt-3: What's it good for? *Natural language engineering*, 27(1):113–118.

Dale, R. (2022). Nlp: How to spend a billion dollars. *Natural Language Engineering*, 28(1):125–136.

El Shamy, N. and Hassanein, K. (2018). The Impact of Age and Cognitive Style on E-Commerce Decisions: The Role of Cognitive Bias Susceptibility. In Davis, F. D., Riedl, R., vom Brocke, J., Léger, P.-M., and Randolph, A. B., editors, *Information Systems and Neuroscience*, Lecture Notes in Information Systems and Organisation, pages 73–83, Cham. Springer International Publishing.

Ferreira, B., Santos, G., and Conte, T. (2017). Identifying Possible Requirements using Personas - A Qualitative Study:. In *Proceedings of the 19th International Conference on Enterprise Information Systems*, pages 64–75, Porto, Portugal. SCITEPRESS - Science and Technology Publications.

Ferreira, B. M., Barbosa, S. D., and Conte, T. (2016). Pathy: Using empathy with personas to design applications that meet the users' needs. In *International Conference on Human-Computer Interaction*, pages 153–165. Springer.

Gao, C., Zeng, J., Lyu, M. R., and King, I. (2018). Online app review analysis for identifying emerging issues. In *Proceedings of the 40th International Conference on Software Engineering*, ICSE '18, pages 48–58, New York, NY, USA. Association for Computing Machinery.

Gray, D., Brown, S., and Macanufo, J. (2010). *Gamestorming*. O'Reilly Media, Sebastopol, CA.

Grundy, J. (2021). Impact of End User Human Aspects on Software Engineering:. In *Proceedings of the 16th International Conference on Evaluation of Novel Approaches to Software Engineering*, pages 9–20, Online Streaming, — Select a Country —. SCITEPRESS - Science and Technology Publications.

Grundy, J., Mouzakis, K., Vasa, R., Cain, A., Curumsing, M., Abdelrazek, M., and Fernando, N. (2018). Supporting Diverse Challenges of Ageing with Digital Enhanced Living Solutions. *Telehealth for our Ageing Society*, pages 75–90. Publisher: IOS Press.

Guo, A. and Ma, J. (2018). Archetype-Based Modeling of Persona for Comprehensive Personality Computing from Personal Big Data. *Sensors*, 18(3):684.

Kose, T. and Arslan, S. (2020). Turkish consumer participation in e-commerce. *Journal of electronic commerce in organizations*, 18(4):30–50.

Malgaonkar, S., Licorish, S. A., and Savarimuthu, B. T. R. (2022). Automatically generating taxonomy for grouping app reviews — a study of three apps. *Software Quality Journal*, 30(2):483–512.

Malterud, K., Siersma, V. D., and Guassora, A. D. (2016). Sample size in qualitative interview studies: Guided by information power. *Qualitative Health Research*, 26(13):1753–1760. PMID: 26613970.

Mao, J.-Y., Vredenburg, K., Smith, P. W., and Carey, T. (2005). The state of user-centered design practice. *Communications of the ACM*, 48(3):105–109.

Marr, K. (2020). How to Create Buyer Personas for Your Online Store.

Mathews, C., Ye, K., Grozdanovski, J., Marinelli, M., Zhong, K., Khalajzadeh, H., Obie, H., and Grundy, J. AH-CID: A Tool to Automatically Detect Human-Centric Issues in App Reviews. page 12.

Mcilroy, S., Ali, N., Khalid, H., and Hassan, A. E. Analyzing and automatically labelling the types of user issues that are raised in mobile app reviews | SpringerLink.

McIntosh, J., Du, X., Wu, Z., Truong, G., Ly, Q., How, R., Viswanathan, S., and Kanij, T. (2021). Evaluating Age Bias In E-commerce. In *2021 IEEE/ACM 13th International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE)*, pages 31–40, Madrid, Spain. IEEE.

Obie, H. O., Hussain, W., Xia, X., Grundy, J., Li, L., Turhan, B., Whittle, J., and Shahin, M. (2021). A First Look at Human Values-Violation in App Reviews. In *2021 IEEE/ACM 43rd International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS)*, pages 29–38, Madrid, ES. IEEE.

OpenAI. openai: Python client library for the OpenAI API.

Rangel, F., Montes-y-Gómez, M., Potthast, M., and Stein, B. (2018). Overview of the 6th Author Profiling Task at PAN 2018: Cross-domain Authorship Attribution and Style Change Detection. In Cappellato, L., Ferro, N., Nie, J.-Y., and Soulier, L., editors, *CLEF 2018 Evaluation Labs and Workshop – Working Notes Papers, 10-14 September, Avignon, France*. CEUR-WS.org.

Rangel, F. and Rosso, P. (2019). Overview of the 7th Author Profiling Task at PAN 2019: Bots and Gender Profiling. In Cappellato, L., Ferro, N., Losada, D., and Müller, H., editors, *CLEF 2019 Labs and Workshops, Notebook Papers*. CEUR-WS.org.

Rangel, F., Rosso, P., Potthast, M., and Stein, B. (2017). Overview of the 5th Author Profiling Task at PAN 2017: Gender and Language Variety Identification in Twitter. In Cappellato, L., Ferro, N., Goeuriot, L., and Mandl, T., editors, *Working Notes Papers of the CLEF 2017 Evaluation Labs*, volume 1866 of *CEUR Workshop Proceedings*.

Rangel, F., Rosso, P., Verhoeven, B., Daelemans, W., Potthast, M., and Stein, B. (2016). Overview of the 4th Author Profiling Task at PAN 2016: Cross-Genre Evaluations. In Balog, K., Cappellato, L., Ferro, N., and Macdonald, C., editors, *Working Notes Papers of the CLEF 2016 Evaluation Labs*, volume 1609 of *Lecture Notes in Computer Science*.

Singh, R., Garg, V., and GPT-3 (2021). Human Factors in NDE 4.0 Development Decisions. *Journal of Nondestructive Evaluation*, 40(3):71.