

## MSIN0231 Candidate #: XRKQ3

### Machine Learning Individual Assignment

Q1. State the service and generative language model that you used.

The service and generative language model used is OpenAI's language model GPT-3 (Generative Pretrained Transformer 3). ChatGPT has a vast range of capabilities regarding natural language tasks, including but not limited to text generation, language translation, and question answering. According to OpenAI, ChatGPT is a fine-tuned version of the GPT – 3.5 series using Reinforcement Learning from Human Feedback. (Biyang Guo et al., 2023)

Q2. Provide 5 prompts that you used to develop a business model for an AI start-up. Rank them from high to low in terms of how useful the generated text was for you.

The following prompts were fed to ChatGPT to develop a business model for a social media retail application operating with a recommendation engine.

1. Develop a business plan proposal for an AI start-up specializing as a social media application with an e-commerce recommendation engine
2. Create a revenue model for an AI start-up specializing as a social media application with an e-commerce recommendation engine
3. Outline the marketing strategy necessary to launch an AI start-up specializing as a social media application with an e-commerce recommendation engine
4. Provide an overview of the logistics for an AI start-up specializing as a social media application with an e-commerce recommendation engine
5. Market analysis for an AI start-up specializing as a social media application with an e-commerce recommendation engine

Q3. Provide the most useful generated text (i.e., from the prompt ranked highest)

The following prompt was the most useful generated text by ChatGPT. However, while it was the most useful, I found that I was unable to develop the specific details of the product. Thus, the details were further elaborated in the "Who, What, How" section.

"Our AI-powered retail social media application start-up aims to revolutionize the online shopping experience for consumers. Our cutting-edge AI technology is designed to learn about a customer's preferences, shopping history, and budget to provide customized recommendations for their every need. Our personal shopping assistant will also be able to communicate with customers through voice and text interfaces, making it easier to use than ever before.

In terms of revenue, we plan to monetize our AI shopping assistant through pay-per click targeted advertisements and commission from sales made through our platform. We believe this model will allow us to scale our user base and revenue quickly. In addition, we plan to partner with leading e-commerce companies to offer our AI shopping assistant as a value-added service to their customers.

Our marketing strategy will focus on building brand awareness through targeted digital advertising and influencer partnerships. We will also leverage social media and word-of-mouth to grow our user base organically. Our team is made up of experienced professionals in AI, e-commerce, and marketing, and we will continue to invest in these areas as we scale.

In conclusion, our AI-powered personal shopping assistant has the potential to disrupt the online shopping market and provide significant value to consumers. With the growing trend of AI technology, the timing is right for our start-up to capture this opportunity and deliver a game-changing solution to the market.”

Q4. Provide a name for your start-up.

The name of this start up is “ShopTok”

Q5. For your AI start-up, describe the business model using the who-what-how dimensions. You may use the generated text as basis. (Approx. 150-300 words.)

Who?

ShopTok targets busy customers looking to maximize the efficiency of their shopping, otherwise known as the “Utilitarian shopper” (Ashraf et al., 2016). The utilitarian shopper seeks functional and instrumental benefits, these benefits are obtained through systematic and time-offering convenience during the online shopping experience (Brusch et al., 2019). Shopping convenience has become a major influencer in motivating consumers to online shop (Clemes et al., 2014). Elements such as reduced shopping time and minimized effort have offered shoppers savings of both time and money (ibid). The Utilitarian consumer shops online based on rational necessity with a specific target in mind, thus seeking a systematic, sensible and planned online shopping experience (Wolfenbarger and Gilly, 2001).

Moreover, as ShopTok is a social media platform offering a personalized shopping experience, the application aims to target Gen Z and millennials who are tech-savvy. ShopTok aims to target this segment of consumers as studies show that in the United States alone, 22% of twitter users purchase a product after being influenced by their peers through a tweet or a retweet, and 33% of Facebook users purchase an item after sharing, liking or making a comment. (Bennett, 2013). According to McQuail’s classification of motivations (McQuail, 1983), remuneration motivation is when individuals utilize social media to earn economic benefits, such as purchasing an item with a discount or incentive (Wang et al., 2003). Shopping behavior on social media displays a

more hedonistic approach to retail consumption, in which consumers are attracted to the socialization, community, and recreational aspects of shopping through social media (Palmer et al., 2009). Thus, ShopAi aims to appeal to the consumer looking for both an engaging and efficient consumer experience.

## What?

It is evident that online shopping is sweeping the retail industry, with companies such as Amazon and Alibaba, consumers are heading towards abandoning brick-mortar altogether. Online retailing has experienced a growth rate of 22% with a sales figure of 200 billion euros in Europe (Chetioui et al., 2021). In addition to this, the rise of social media has become a major driver in influencing sales numbers. Studies show that consumers shopping through social networking sites have found the experience to be both enjoyable and convenient. (Dennis et al., 2010). Therefore, inspiring ShopTok purpose to combine social media and shopping into one convenient mobile application.

The application will follow a similar model and algorithm to tiktok in which, the application offers an infinite scroll of personalized fashion and apparel recommendations. The algorithm would be generated through information gathered through cross-application tracking, internet searches, shopping history, as well the consumer's interaction with the application. Further, consumers will be able to follow their favorite brands, stores, influencers, and friends to create a community. Consumers will be able to like, comment, save, share, and repost items they've found on the application. Hence, allows consumers to interact with the item by gauging its popularity, reading reviews left in the comments, understanding sizing preferences, and opening up communication lines with the brand.

The application will be split up into different types of users; shoppers and sellers. In which, shoppers are unable to post (only repost) and sellers are able to post items for sale. The revenue model will follow targeted advertisements with pay-per-click as well as a 3% commission on sales made through the application.

## How?

In order to move forward with the development of the algorithm for this application, consumer socialization must be defined as the first step. Consumer socialization is the process by which "people acquire skills, knowledge, and attitudes relevant to their functioning as consumers in the marketplace" (Bandura, 1969). According to Bandura's social learning theory, through socialization agents, modeling, reinforcement, and social interaction can be influenced. (Rohm et al., 2009) Thus to create the algorithm the modeling will occur when the user imitates the behavior of the socialization agent (such as liking, reposting, and interacting with the content) and the reinforcement mechanism will occur when there is positive or negative reinforcement (Lueg et al., 2007). In understanding the psychology behind the model an endless scroll recommendation engine algorithm can be created.

*The following steps will be taken to create the recommendation engine model:*

1. Data Collection – gathering data on user preferences based on their shopping history, purchase behavior, and platform interaction.
2. Pre-processing - the data will then be cleaned in order to remove any irrelevant information and format it to a usable form.
3. Feature engineering – relevant features and information from the data will be extracted to make predictions such as purchase frequency, product category, and budget.
4. Model selection and training – there are numerous models which can be created in this step. One potential model would be a deep learning model such as Convolutional Neural Networks and Recurrent Neural Networks which can be used to analyze images and text data to understand the style and context of fashion items. Another model that can be employed is a Reinforcement Learning model, which can be used to optimize the recommendation process over time, by learning customer interactions and feedback. Lastly, a Collaborative Filtering model can be selected and can be used to recommend items based on customer preferences and purchase history.
5. Model Evaluation – evaluate the performance of the model using metrics such as accuracy, precision, and recall.
6. Model Deployment – Deploy the trained model in the personal shopping assistant application, where it can make recommendations for users in real time.

Q6. For your AI start-up, who are the rivals and what are the substitute products / services?

Established e-commerce companies utilizing AI algorithms for predictive technology, targeting the same consumer segment; such as ASOS, Amazon, and StitchFix may serve as ShopTok's primary competition. With over 24 million customers and 85,000 products, ASOS is a leader in driving AI technology in retail forward. To gain a deep understanding of their customers for personalization features, ASOS uses deep neural networks to classify product images and descriptions based on specific attributes such as item type, color, print, length, and occasion (Mixon, 2021). Furthermore, ASOS utilizes other predictive technology to create recommendation engines such as collaborative filtering and multilayer neural networks. Thus, deploying similar technology and machine learning algorithms to that of ShopTok (Mixon, 2021).

*The following diagram depicts how ASOS identifies features extracted from customer data for analysis and development of their personalized shopping experience (Mixon, 2021):*



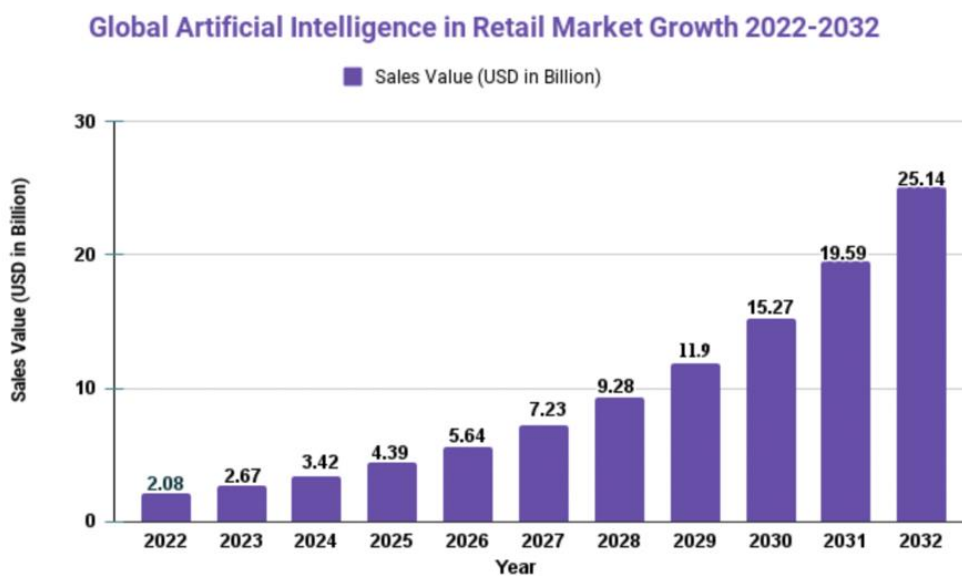
Similarly, Amazon is the largest e-commerce consumer platform worldwide, with 200 billion monthly visitors and 200 million prime members, gaining its rapid success through innovative AI technology. Amazon utilizes personalized recommendation engine algorithms analyzing variables such as location, recent purchases, saved items/lists, user reviews, and viewing activity. Amazon's product recommendation engines lead 35% of their total sales. (Manole, 2022). Conversely, StitchFix, a competitor to ShopTok in terms of personalized shopping experience, does not utilize complex algorithms to determine their consumer needs. Instead, opting for the human touch by assigning personal stylists to their users.

ShopTok differentiates itself from its competitors by incorporating the consumer socialization and community aspect to online shopping. By developing a social media platform of an endless scroll recommendation engine, consumers are able to shop and interact with brands and other users. Emphasis on the social media aspect will set this product apart from its rivals and substitutes.

Q7. As an investor, would you invest in this AI start-up? Why? (Approx. 150-300 words.)

As an investor there are multiple aspects that I would consider prior to investing. Firstly, I would seek to understand what the business does and how it will create shareholder value (Peterson, 2018). Understanding, that ShopTok employs innovative-complex AI technologies to combine e-commerce with social media, I would consider this a unique value proposition and a potentially disruptive product in the retail industry. Moreover, as ShopTok offers a clear revenue model of targeted pay-per click advertisements and 3% commission on sales, I can be confident that ShopTok is adapting strategic decisions that maximize expected value for long-term earnings (Rappaport, 2006).

The next aspect I would consider as an investor is the sector in which the business operates in as well as what unmet market need or problem this product solves within this sector (Peterson, 2018). Acknowledging the industry trends and growth potential as well as the total market value will help me understand ShopTok's capacity to succeed (ibid). Industry trends and potential growth for ShopTok can be seen in the increase in demand for personalized shopping experiences, the rise of e-commerce, and rise of social media consumption. As consumers continue to seek personalized shopping experiences, ShopTok provides a solution by using machine learning algorithms to analyze customer data and offer an endless scroll of tailored product recommendations. The growing popularity of fashion on social media with the rise of e-commerce has created a need for a platform that can offer both entertainment and community to influence shopping decisions. Furthermore, the global artificial intelligence in retail market size is expected to reach approximately \$25.14 billion by 2032 from \$2.08 billion in 2022, with a compound annual growth rate (CAGR) of 28.3% between 2022 and 2032 (John, 2023). Thus, proving ShopTok's potential for growth within the industry and high total market value.



*Image Reference ((John, 2023))*

As ShopTok is still in its initial ideation phase and there are no financial projection reports, I would take a Berkus Approach to understand the valuation of this startup, evaluating several key success factors such as basic value, technology, execution, strategic relationships, market scope, production, and revenue model (Peterson, 2018). With all factors considered, it is clear that there is a large market opportunity, potential for growth, a unique value proposition offering a competitive advantage, and a clear return on investment revenue model, I would invest in this start up.

Q8. Based on your experience, provide one advantage and one disadvantage of using a generative language model for developing a business model.

One advantage of using a generative language model for developing a business model is that it can help generate a large volume of ideas quickly and efficiently. Compared to manual chats, ChatGPT can produce more accurate responses. This is due to the fact that it has been trained on a sizable sample of conversational data, enabling it to comprehend the conversational context and produce pertinent responses (Deng et al., 2022). ChatGPT uses a recurrent neural network (RNN) architecture, allowing it learn from previous conversations, thus making it possible to make more accurate predictions than rule-based methods used by chatbots such as ELIZA and AIML (George et al. 2023). Moreover, the ability of CGA to learn from its mistakes, enables it to adapt to new settings and deliver more accurate results and generative capacities (Deng et al., 2022). This was particularly helpful when brainstorming new business models, as it allowed me to explore a variety of potential options in a short amount of time. This helped save time and resources compared to traditional methods of business model development, such as manual research and analysis.

However, one disadvantage of using a generative language model for developing a business model is that the output can be limited by the quality and quantity of the input data. Since ChatGPT does not have access to external information (Deng et al., 2022) and is trained exclusively on human-generated data, the results can often be outdated or biased (Azaria, 2022). Biased results from ChatGPT can be evaluated as such, the most frequent digit generated by ChatGPT, matches humans' most favorable number, 7 and the least frequent digit generated by ChatGPT, matches humans' least favorable number, 1 (ibid). Therefore, the model may not be able to generate responses on complex or unconventional questions (Deng et al., 2022). Biased or incomplete input data, can lead to inaccurate or flawed business models. Additionally, because ChatGPT does not utilize the internet, it fails to fully capture the nuances and complexities of the business environment, such as evaluating the economic conditions, competitive pressures, and customer preferences. As a result, the generated output may not fully reflect the realities of the business environment, which could lead to a suboptimal business model.

Although adopting a generative language model can be an effective technique for creating a business model, it is crucial to be aware of its restrictions and potential drawbacks. Businesses should take care to ensure that the input data is of high quality and that the result is thoroughly assessed and modified through additional analysis. By doing so, businesses can leverage the benefits of generative language models while minimizing the potential risks.

Q9. Provide 5 prompts for a job posting for your start-up. Rank them from high to low in terms of how useful the generated text was for you.

1. Generate a job listing for a Data Scientist at an AI powered shopping assistant start up. Candidate should have a bachelor's degree specializing in a technical field and 2+ years



machine learning experience. Candidate is expected to analyze large, complex datasets to identify trends, patterns, and insights. Candidate is also expected to conduct feature engineering and create machine learning models such as recommendation engines and convolutional neural networks.

2. Generate a job listing for a Data Scientist at an AI powered shopping assistant start up. Candidate should have a bachelor's degree specializing in a technical field and 2+ years machine learning experience. Candidate is expected to analyze large, complex datasets to identify trends, patterns, and insights.
3. Generate a job listing for a Data Scientist at an AI powered shopping assistant start up. Candidate should have a bachelor's degree specializing in a technical field and 2+ years machine learning experience
4. Generate a job listing for a Data Scientist at an AI powered shopping assistant start up. Candidate should have 2+ years of programming experience.
5. Generate a job listing for a Data Scientist at an AI powered shopping assistant start up

Q10. Provide the most useful generated text (i.e., from the prompt ranked highest).

**Job Title:**

Data Scientist at AI-powered Shopping Assistant Start-up

**Job Description:**

We are seeking a highly motivated Data Scientist with a bachelor's degree in a technical field and 2+ years of experience in machine learning. The ideal candidate will be responsible for analyzing large and complex datasets to identify trends, patterns, and insights. In addition, the candidate will work closely with the development team to design and implement machine learning models, such as recommendation engines and convolutional neural networks.

**Responsibilities:**

- Analyze large and complex datasets to identify trends, patterns, and insights.
- Develop and implement machine learning models such as recommendation engines and convolutional neural networks.
- Conduct feature engineering to improve model performance.
- Work closely with the development team to integrate machine learning models into the shopping assistant product.
- Collaborate with other data scientists and engineers to improve data quality and infrastructure.

**Qualifications:**

- Bachelor's degree in a technical field such as computer science, statistics, mathematics, or a related field.
- 2+ years of experience in machine learning and data analysis.
- Strong understanding of machine learning algorithms and data structures.
- Proficiency in programming languages such as Python, R, and SQL.



- Experience with deep learning frameworks such as TensorFlow, Keras, or PyTorch.
- Strong problem-solving and analytical skills.
- Excellent written and verbal communication skills.

We offer a competitive salary package, flexible work arrangements, and a collaborative work environment. If you are passionate about data science and want to work on cutting-edge technology, please apply with your resume and cover letter.

Q11. Based on your experience, does the generative language model provide better text for the business model or job posting? Why?

As, ChatGPT is a platform for digital transformation that can provide support to improve business processes, models, domains, organizational and cultural structures, it can be very effective in developing a business model and generating job descriptions (George et al., 2023). For business models, a generative language can be helpful in generating ideas such as product offerings, marketing strategies and other key aspect of the business. However, it is important to note that the generated text may not take into account important external factors such as competition, regulation, or consumer behavior, which can have a significant impact on the success of the business. Therefore, the results may not always be coherent or relevant to the specific business.

In my experience, I struggled to refine my specific business idea with the generative language model. Many of the details I was intending to extract from the model were missing, such as the concept for a shopping social media platform application. The model offered me a very general idea for an AI personal shopping assistant, however, the idea I had in mind was more refined. It was difficult to create the correct prompts to develop the idea of an endless scroll recommendation engine that doubles as a social media platform. Since, ChatGPT relies on data from previous conversations, it is clear that the platform lacks originality with generating business ideas.

Job postings tend to be more broad to appeal to a larger number of potential applicants. Therefore, the ambiguity in the generative language model's response may serve as beneficial to expanding the scope for candidates. Since, prompting ChatGPT to create a business model, produces general generated text, the prompt can be further refined for the requirements of the role. For example, including a bachelor's degree qualification or experience with a specific program. I found it easier to prompt ChatGPT with the specific requirements for a job posting than a business model, as I can specify the qualifications and return the generated text.

Q12. Some language models can also generate code. Provide 3 prompts for a function in Python that would be useful for your start-up. Rank them from high to low in terms of how useful the generated text was for you.

1. Generate a Python function for Principal Component Analysis that can reduce the dimensionality of the data and recognize patterns within the data
2. Create a Python function for natural language processing that can identify key phrases in a text document.
3. Build a Python function for recommendation systems that predicts the most likely item a user will purchase based on their previous purchases.

Q13. Provide the most useful generated code (i.e., from the prompt ranked highest).

The most useful generated code was prompt 1 “Generate a Python function for Principal Component Analysis that can reduce the dimensionality of the data and recognize patterns within the data.” This function runs a Principal Component Analysis to identify the most important features within the data.

```
from sklearn.decomposition import PCA
import pandas as pd

# Load the data
df = pd.read_csv("fashion_data.csv")

# Split the data into features and target
X = df.drop("id", axis=1)
y = df["id"]

# Normalize the data
X = (X - X.mean()) / X.std()

# Create a PCA object with 2 components
pca = PCA(n_components=2)

# Fit the PCA model to the data
X_pca = pca.fit_transform(X)

# Print the explained variance ratio
print(pca.explained_variance_ratio_)

# Plot the data
import matplotlib.pyplot as plt
```

```
plt.scatter(X_pca[:, 0], X_pca[:, 1], c=y)
plt.xlabel("PC1")
plt.ylabel("PC2")
plt.show()
```

Q14. Does the code work as intended? Show how you verified the code.

To verify the code generated by ChatGPT, I ran the code on an e-commerce styles dataset. Prior to running the principal component analysis on this dataset, I ran some data preprocessing steps. I first cleaned the data by dropping unnecessary columns. I then ran a one-hot encoding on the dataset to convert the categorical data into numerical data for analysis.

*The Principal Component Analysis is useful for the development of ShopTok in a number of ways:*

1. Dimensionality reduction: PCA can reduce the dimensionality of the data by identifying the most important features (i.e., principal components) that explain the majority of the variance in the data. This can help to simplify the data and make it more manageable for machine learning algorithms.
2. Pattern recognition: PCA can reveal underlying patterns and relationships in the data that may not be immediately apparent. By identifying these patterns, ShopTok can make more accurate recommendations to customers based on their preferences and past behaviors.
3. Feature engineering: PCA can be used to create new features (i.e., principal components) that capture the most important information in the data. These new features can then be used as input to machine learning algorithms, which can improve their performance.

Overall, running a PCA analysis on fashion data can help ShopTok to make more accurate and personalized recommendations to customers, which can ultimately lead to higher customer satisfaction and retention.

```
# Split the data into features and target
X = df_encoded.drop("id", axis=1)
y = df_encoded["id"]
```

```
# Normalize the data
X = (X - X.mean()) / X.std()
```

```
# Create a PCA object with 2 components
pca = PCA(n_components=2)
```

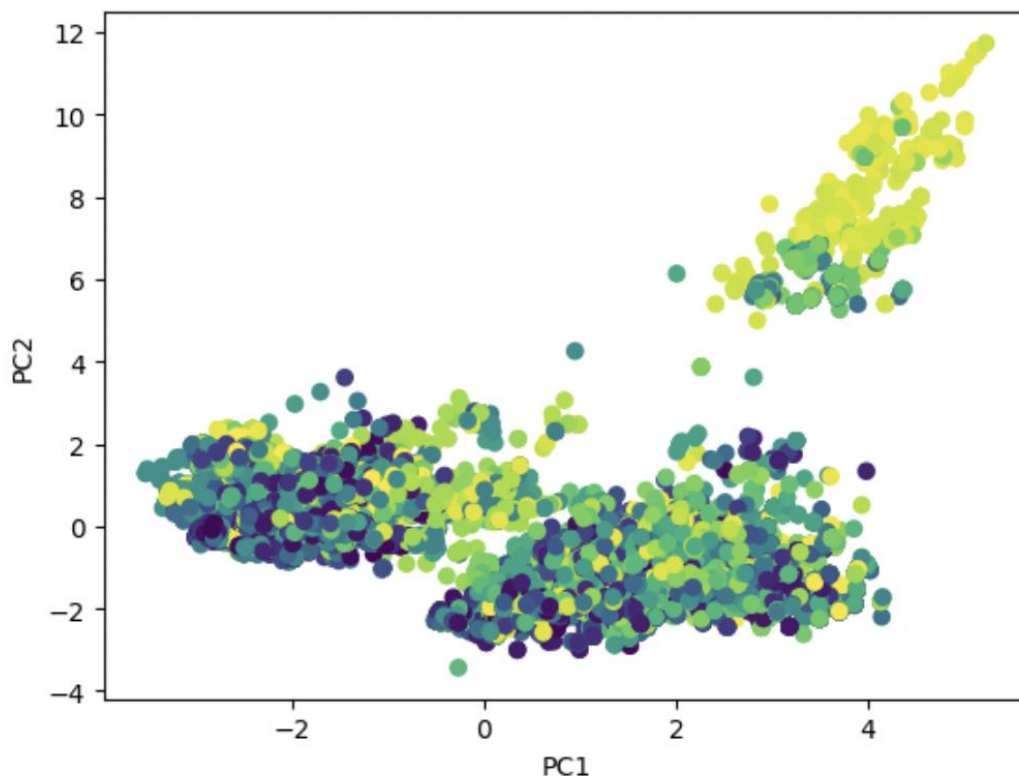
```
# Fit the PCA model to the data
X_pca = pca.fit_transform(X)
```

```
# Print the explained variance ratio
print(pca.explained_variance_ratio_)
```

```
# Plot the data
import matplotlib.pyplot as plt
```

```
plt.scatter(X_pca[:, 0], X_pca[:, 1], c=y)
plt.xlabel("PC1")
plt.ylabel("PC2")
plt.show()
```

```
[0.0186547  0.01662361]
```



The above figure represents the two most important features from the styles dataset. For future reference, ShopTok can take a clustering approach to identify the most important features (as represented by the figure) and implement this unsupervised learning technique to refine the machine learning algorithms as well as develop strategic business decisions.

Q15. Reflect on your experience. Next time that you are developing a business model for a start-up would you use a generative language model or not? Why?

Utilizing a generative language model to develop a business model served many benefits. The biggest proponent for using ChatGPT to develop a business model was the speed at which it developed ideas and produced research results. Moreover, I've found that using ChatGPT to develop a business model helps the business improve its products by generating information about current trends in user behavior and in-depth analysis (George et al., 2023). Thus, helping me to understand what kinds of features are needed by the target group and tailoring my product as such (ibid).

One issue that I encountered when using ChatGPT to develop my business model is its ambiguity and lack of originality. Since, ChatGPT does not have access to external resources, it is unable to provide the full business scope for the model. When developing my business model, I felt as though many key aspects were missing and needed human touch to further expand on the idea. Bearing in mind the generative language model's limitations, I would use ChatGPT along with other resources to develop a business model in the future.

## Bibliography:

Guo, B. *et al.* (2023) *How close is CHATGPT to human experts? comparison corpus, evaluation, and detection*, *arXiv.org*. Available at: <https://arxiv.org/abs/2301.07597>.

Billewar, S.R. *et al.* (2021) *The rise of 3D e-commerce: The online shopping gets real with virtual reality and augmented reality during COVID-19*, *World Journal of Engineering*. Emerald Publishing Limited. Available at: <https://www.emerald.com/insight/content/doi/10.1108/WJE-06-2021-0338/full/html>.

Srivastava, A. and Thaichon, P. (2022) *What motivates consumers to be in line with online shopping?: A systematic literature review and discussion of Future Research Perspectives*, *Asia Pacific Journal of Marketing and Logistics*. Emerald Publishing Limited. Available at: <https://www.emerald.com/insight/content/doi/10.1108/APJML-10-2021-0777/full/html?skipTracking=true>.

Ward, S. (1974) "Consumer socialization," *Journal of Consumer Research*, 1(2), p. 1. Available at: <https://doi.org/10.1086/208584>.

Berger, M. (1971) "Principles of behaviour modification. by Albert Bandura. London: Holt, Rinehart and Winston. 1969. pp. 677.," *British Journal of Psychiatry*, 119(548), pp. 102–102. Available at: <https://doi.org/10.1192/bjp.119.548.102>.

Köhler, C.F. *et al.* (2011) "Return on interactivity: The impact of online agents on newcomer adjustment," *Journal of Marketing*, 75(2), pp. 93–108. Available at: <https://doi.org/10.1509/jmkg.75.2.93>.

Moschis, G.P. and Churchill, G.A. (1978) "Consumer socialization: A theoretical and empirical analysis," *Journal of Marketing Research*, 15(4), pp. 599–609. Available at: <https://doi.org/10.1177/002224377801500409>.

Dennis, C. *et al.* (2010) "The influences of social e-shopping in enhancing young women's online shopping behaviour," *Journal of Customer Behaviour*, 9(2), pp. 151–174. Available at: <https://doi.org/10.1362/147539210x511353>.

Pookulangara, S. and Koesler, K. (2011) "Cultural influence on consumers' usage of social networks and its' impact on online purchase intentions," *Journal of Retailing and Consumer Services*, 18(4), pp. 348–354. Available at: <https://doi.org/10.1016/j.jretconser.2011.03.003>.

Bennett S. (2013) "How social media is changing the way your customers shop online," *AdWeek*. [www.adweek.com/socialtimes/social-media-shopping/491924](http://www.adweek.com/socialtimes/social-media-shopping/491924)

Muntinga, D., Moorman, M. and Smit, E. (1970) *Introducing cobras: Semantic scholar*, *International Journal of Advertising*. Available at:

<https://www.semanticscholar.org/paper/Introducing-COBRAs-Muntinga-Moorman/a80192a082c74181438f517f5223da8af8aee8ab>

Muntinga DG, Moorman M, Smit EG. "Introducing COBRAs: exploring motivations for brand-related social media use," *International Journal of Advertising* 2011; 30:13–46.

YOUCHENG, W.A.N.G. and FESENMAIER, D.R. (2003) "Assessing motivation of contribution in online communities: An empirical investigation of an online travel community," *Electronic Markets*, 13(1), pp. 33–45. Available at: <https://doi.org/10.1080/1019678032000052934>.

Palmer, A. and Koenig-Lewis, N. (2009) "An entended, community focused, experiential framework for relationship marketing," *Journal of Customer Behaviour*, 8(1), pp. 85–96. Available at: <https://doi.org/10.1362/147539209x414407>.

Park, N., Kee, K.F. and Valenzuela, S. (2009) "Being immersed in social networking environment: Facebook groups, uses and gratifications, and social outcomes," *CyberPsychology & Behavior*, 12(6), pp. 729–733. Available at: <https://doi.org/10.1089/cpb.2009.0003>.

Muralidharan, S. and Men, L.R. (2015) "How peer communication and engagement motivations influence social media shopping behavior: Evidence from China and the United States," *Cyberpsychology, Behavior, and Social Networking*, 18(10), pp. 595–601. Available at: <https://doi.org/10.1089/cyber.2015.0190>.

Mixson, E. (2022) *A look at Asos's fashionable approach to data science, AI, Data & Analytics Network*. Available at: <https://www.aidataanalytics.network/data-science-ai/articles/a-look-at-asoss-fashionable-approach-to-data-science>.

*How Amazon uses Artificial Intelligence: Zentail Blog* (no date) RSS. Available at: <https://www.zentail.com/blog/amazons-flywheel-approach-how-amazon-uses-ai>

Limited, P.P. (2023) *Artificial Intelligence in retail market [USD 25.14 bn by 2032]*, *Enterprise Apps Today*. Available at: <https://www.enterpriseappstoday.com/news/artificial-intelligence-in-retail-market-to-cross-usd-25-14-bn-in-2032-28-3-cagr.html#:~:text=12%20Contact%20us-,Artificial%20Intelligence%20in%20Retail%20Market%20Overview,28.3%25%20between%202022%20and%202032>.

Peterson, C. (2023) *Questions to ask when investing in a startup*, GCV. Growth Capital Ventures. Available at: <https://www.growthcapitalventures.co.uk/insights/blog/26-questions-to-ask-when-investing-in-a-startup-business#:~:text=Building%20a%20strong%20foundation%20of,the%20startup's%20capacity%20to%20succeed>.

Rappaport, A. (2006) *Ten ways to create shareholder value*, *Harvard Business Review*. Available at: <https://hbr.org/2006/09/ten-ways-to-create-shareholder-value>.



Azaria, A. (2022) “CHATGPT usage and limitations.” Available at: <https://doi.org/10.31219/osf.io/5ue7n>.

Deng, J. and Lin, Y. (2023) “The benefits and challenges of chatgpt: An overview,” *Frontiers in Computing and Intelligent Systems*, 2(2), pp. 81–83. Available at: <https://doi.org/10.54097/fcis.v2i2.4465>.

basel99 (2022) *Fashion products recommendation system*, Kaggle. Kaggle. Available at: <https://www.kaggle.com/code/basel99/fashion-products-recommendation-system/notebook>.

George, S., George, H., Martin, G. (2023) “A Review of ChatGPT AI's Impact on Several Business Sectors,” *Partners Universal International Innovation Journal(PUIIJ)*. Available at: <https://puuij.com/index.php/research/article/view/11/5>

## Appendix:

### Data Preprocessing Steps:

```
In [41]: df = pd.read_csv('/Users/nadinealkaraghali/Downloads/archive/styles.csv',error_bad_lines=False)

/var/folders/fw/kv94vtc97fv4nsc6rlxl_2080000gn/T/ipykernel_16664/573934008.py:1: FutureWarning: The error_bad_lines argument has been deprecated and will be removed in a future version. Use on_bad_lines in the future.
```

```
df = pd.read_csv('/Users/nadinealkaraghali/Downloads/archive/styles.csv',error_bad_lines=False)
b'Skipping line 6044: expected 10 fields, saw 11\nSkipping line 6569: expected 10 fields, saw 11\nSkipping line 7399: expected 10 fields, saw 11\nSkipping line 7939: expected 10 fields, saw 11\nSkipping line 9026: expected 10 fields, saw 11\nSkipping line 10264: expected 10 fields, saw 11\nSkipping line 10427: expected 10 fields, saw 11\nSkipping line 10905: expected 10 fields, saw 11\nSkipping line 11373: expected 10 fields, saw 11\nSkipping line 11945: expected 10 fields, saw 11\nSkipping line 14112: expected 10 fields, saw 11\nSkipping line 14532: expected 10 fields, saw 11\nSkipping line 15076: expected 10 fields, saw 12\nSkipping line 29906: expected 10 fields, saw 11\nSkipping line 31625: expected 10 fields, saw 11\nSkipping line 33020: expected 10 fields, saw 11\nSkipping line 35748: expected 10 fields, saw 11\nSkipping line 35962: expected 10 fields, saw 11\nSkipping line 37770: expected 10 fields, saw 11\nSkipping line 38105: expected 10 fields, saw 11\nSkipping line 38275: expected 10 fields, saw 11\nSkipping line 38404: expected 10 fields, saw 12\n'
```

```
In [42]: df.head()
```

```
Out[42]:
```

	id	gender	masterCategory	subCategory	articleType	baseColour	season	year	usage	productDisplayName
0	15970	Men	Apparel	Topwear	Shirts	Navy Blue	Fall	2011.0	Casual	Turtle Check Men Navy Blue Shirt
1	39386	Men	Apparel	Bottomwear	Jeans	Blue	Summer	2012.0	Casual	Peter England Men Party Blue Jeans
2	59263	Women	Accessories	Watches	Watches	Silver	Winter	2016.0	Casual	Titan Women Silver Watch
3	21379	Men	Apparel	Bottomwear	Track Pants	Black	Fall	2011.0	Casual	Manchester United Men Solid Black Track Pants
4	53759	Men	Apparel	Topwear	Tshirts	Grey	Summer	2012.0	Casual	Puma Men Grey T-shirt

```
In [43]: #Step 1: Drop columns that you do not need
#Keep articleType, baseColour, season, usage
df = df.drop(['productDisplayName', 'year'], axis=1)
# Display the updated dataframe
df.head()
```

```
Out[43]:
```

	id	gender	masterCategory	subCategory	articleType	baseColour	season	usage
0	15970	Men	Apparel	Topwear	Shirts	Navy Blue	Fall	Casual
1	39386	Men	Apparel	Bottomwear	Jeans	Blue	Summer	Casual
2	59263	Women	Accessories	Watches	Watches	Silver	Winter	Casual
3	21379	Men	Apparel	Bottomwear	Track Pants	Black	Fall	Casual
4	53759	Men	Apparel	Topwear	Tshirts	Grey	Summer	Casual

```
In [49]: # Select the columns to be encoded
cols_to_encode = ['gender', 'masterCategory', 'subCategory', 'articleType', 'baseColour', 'season', 'usage']

# Create an instance of OneHotEncoder
encoder = OneHotEncoder()

# Fit and transform the data
encoded_data = encoder.fit_transform(df[cols_to_encode])

# Convert the encoded data to a dataframe
encoded_df = pd.DataFrame(encoded_data.toarray())

# Rename the columns of the encoded dataframe
encoded_df.columns = encoder.get_feature_names(cols_to_encode)

# Concatenate the encoded dataframe with the original dataframe
df_encoded = pd.concat([df.drop(cols_to_encode, axis=1), encoded_df], axis=1)

# Print the encoded dataframe
df_encoded

/Users/nadinealkaragholi/opt/anaconda3/lib/python3.9/site-packages/sklearn/utils/deprecation.py:87: FutureWarning: Function get_feature_names is deprecated; get_feature_names_out is deprecated in 1.0 and will be removed in 1.2. Please use get_feature_names_out instead.
  warnings.warn(msg, category=FutureWarning)
```

Out[49]:

	id	gender_Boys	gender_Girls	gender_Men	gender_Unisex	gender_Women	masterCategory_Accessories	masterCategory_Apparel	masterCategory_I
0	15970	0.0	0.0	1.0	0.0	0.0	0.0	1.0	
1	39386	0.0	0.0	1.0	0.0	0.0	0.0	1.0	
2	59263	0.0	0.0	0.0	0.0	1.0	1.0	0.0	
3	21379	0.0	0.0	1.0	0.0	0.0	0.0	1.0	
4	53759	0.0	0.0	1.0	0.0	0.0	0.0	1.0	
...	...	...	...	...	...	...	...	...	...
44419	17036	0.0	0.0	1.0	0.0	0.0	0.0	0.0	
44420	6461	0.0	0.0	1.0	0.0	0.0	0.0	0.0	
44421	18842	0.0	0.0	1.0	0.0	0.0	0.0	1.0	
44422	46694	0.0	0.0	0.0	0.0	1.0	0.0	0.0	
44423	51623	0.0	0.0	0.0	0.0	1.0	1.0	0.0	

44424 rows × 262 columns