FIT5212 Data analysis for semi-structured data - S1 2023
Assignment 2
Name: Tsz Yan CHUNG
Student ID: 32973381

Task 1: Recommender System Challenge

1. Introduction

This report aims to present the finding of a recommender system built around books. The objective is to build an effective model to make accurate prediction on users' rating on some unseen records of books, based on their previous rating habit. This report will mainly cover this recommender system and the rating prediction model under the collaborative filtering setting.

2. Methodology

In the first step of this report, comparisons between algorithms will be discussed, and the best performing models will be picked for further analysis. For clarification, files applied in this report are:
- trainset: 56,199 x 4, columns: [ID, user_id, item_id, book_name], no null value.
- testset: 286,136 x 4, columns: [user_id, item_id, rating, book_name], no null value.
- metadata: 1,636,235 x 9, columns: [Name, pagesNumber, Publisher, CountsOfReview, PublishYear, PublishYear, Language, Authors, Rating, item_id], with null values, where data is missing at random.

Different algorithms were applied in this report. The memory-based collaborative filtering algorithm adopted different detailed algorithms and tested with various epoch and learning rate combinations. As for model-based collaborative filtering algorithm, it adopted to a basic matrix factorization neural network model, and an advanced matrix factorization neural network model that takes on extra features, separately.

3. Algorithms

3.1 Memory-based collaborative filtering

Step 1: Deciding algorithm (Hug, n.d.; Li 2018).

In the tutorial activity, the SVD algorithm was introduced. However, there are other algorithms not covered in the tutorial while offered by the surprise library. In this step, we will go through some of them and decide on the two best algorithms to further investigate their result.

Algorithms adopted include:

- Baseline model: a basic model that computes for the relationship between user and items
- Singular value decomposition (SVD) method covered in the tutorial
- KNN Basic: basic nearest-neighbors algorithm covered in the tutorial

- KNNWithMeans: In addition to KNNBasic, the mean rating of each user was also taken into consideration.
- KNNWithZScore: Apply z-score normalization to each user’s KNN prediction results
- KNNBaseline: KNN prediction, while taking the baseline linear regression’s prediction result into account
- Non-negative Matrix Factorization (NMF): A similar matrix factorization algorithm to the SVD, with the constraints that the latent features must not be negative
- NormalPredictor: Basic predictor that assumes that data is normally distributed with no learning process.

By iterating through the above algorithms, the following table of evaluation result was returned:

Algorithm	test_rmse	test_mae	fit_time	test_time
BaselineOnly	0.885706	0.706468	1.035147	0.373926
SVD	0.893029	0.708788	5.417379	0.606191
KNNBaseline	0.922535	0.725123	2.072465	3.3104
KNNWithZScore	0.950551	0.748652	1.328606	3.305496
KNNWithMeans	0.950937	0.751212	1.000784	3.531999
NMF	0.996664	0.792071	13.23311	0.328419
KNNBasic	0.999318	0.789699	0.868566	3.218344
NormalPredictor	1.344041	1.074687	0.37043	0.512161

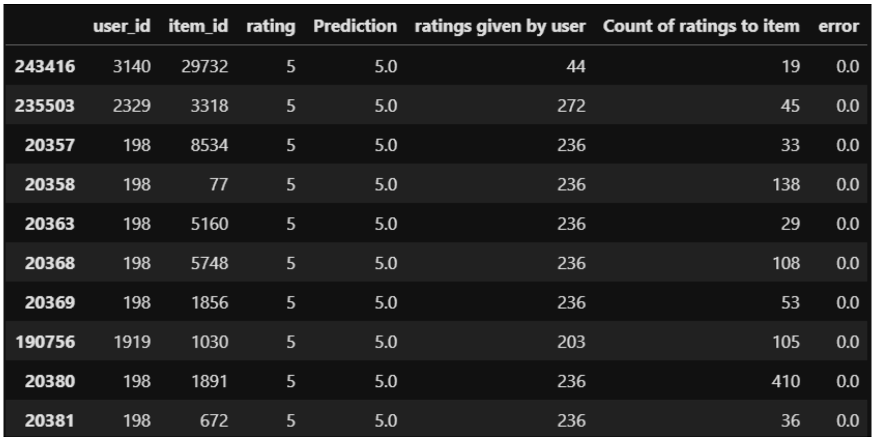
Which shows the BaselineOnly model and the SVD are the best performing algorithms for the train dataset, with the lowest value of RMSE and MAE, possibly due to its relatively simple patterns and relationship between users, items, and ratings.

As a result, in the following memory-based Collaborative Filtering analysis, the BaselineOnly model and the SVD algorithms will be applied.

Step 2: Apply algorithms.

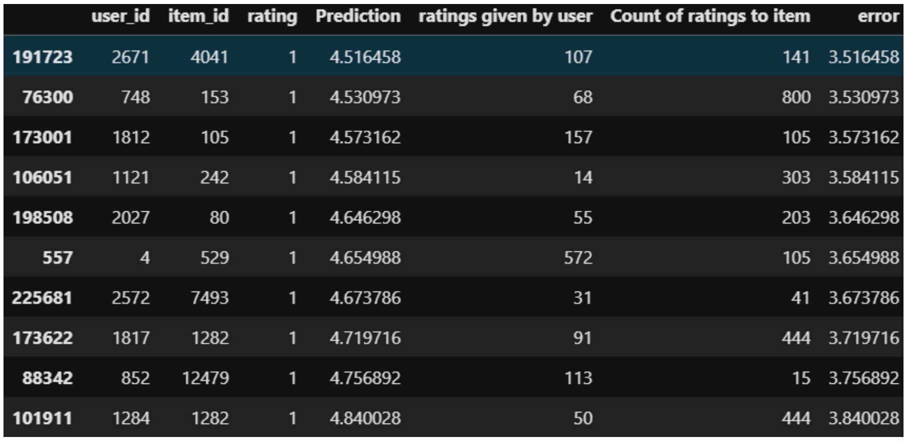
<u>BaselineOnly:</u>

Best predictions from BaselineOnly model:

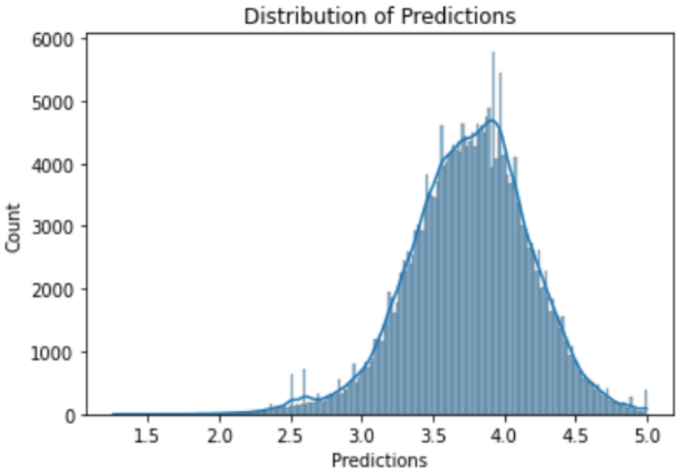


	user_id	item_id	rating	Prediction	ratings given by user	Count of ratings to item	error
243416	3140	29732	5	5.0	44	19	0.0
235503	2329	3318	5	5.0	272	45	0.0
20357	198	8534	5	5.0	236	33	0.0
20358	198	77	5	5.0	236	138	0.0
20363	198	5160	5	5.0	236	29	0.0
20368	198	5748	5	5.0	236	108	0.0
20369	198	1856	5	5.0	236	53	0.0
190756	1919	1030	5	5.0	203	105	0.0
20380	198	1891	5	5.0	236	410	0.0
20381	198	672	5	5.0	236	36	0.0

Worst predictions from BaselineOnly model:



	user_id	item_id	rating	Prediction	ratings given by user	Count of ratings to item	error
191723	2671	4041	1	4.516458	107	141	3.516458
76300	748	153	1	4.530973	68	800	3.530973
173001	1812	105	1	4.573162	157	105	3.573162
106051	1121	242	1	4.584115	14	303	3.584115
198508	2027	80	1	4.646298	55	203	3.646298
557	4	529	1	4.654988	572	105	3.654988
225681	2572	7493	1	4.673786	31	41	3.673786
173622	1817	1282	1	4.719716	91	444	3.719716
88342	852	12479	1	4.756892	113	15	3.756892
101911	1284	1282	1	4.840028	50	444	3.840028



From the prediction analysis, results from the BaselineOnly model are left-skewed and unable to give accurate prediction on low rating results.

SVD

For this algorithm, the singular value decomposition (SVD) method was applied. The SVD does not require extra features to be input, and evaluate imply based on users' and items' feedback, so, only the test set and train set were utilized. In other words, this is relatively easy to implement and computationally effective, while maintaining decent prediction result.

Alterable parameters include:

- Learning rate of model
- Number of training iteration
- Magnitude of regularization term to relieve the issue of overfitting.

The train set was iterated with different combinations of model parameters (with regularization term remained constant), and resulted in the following two graphs regarding its relationships with Root Mean Square Error (RMSE) and Mean Absolute Error (MAE):

Fig 1. Average MAE vs learning rate for different number of iterations

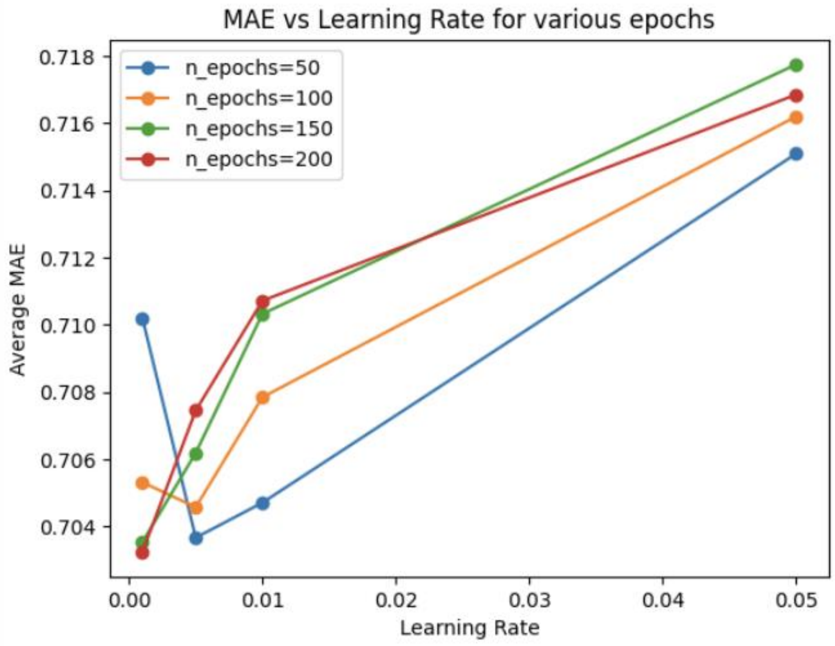
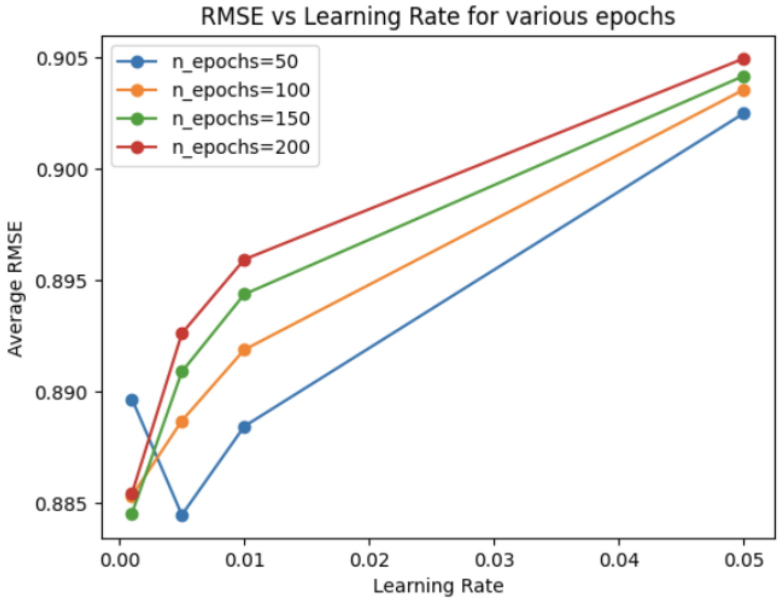


Fig 2. Average RMSE vs learning rate for different number of iterations



From the above graphs, it can be concluded that the following combination of parameters outperformed the rest of the other combination, with significant lower RMSE and lower MAE values:

1. Learning rate = 0.001, number of iterations = 200 (best performance mode)
2. Learning rate = 0.001, number of iterations = 150
3. Learning rate = 0.005, number of iterations = 50

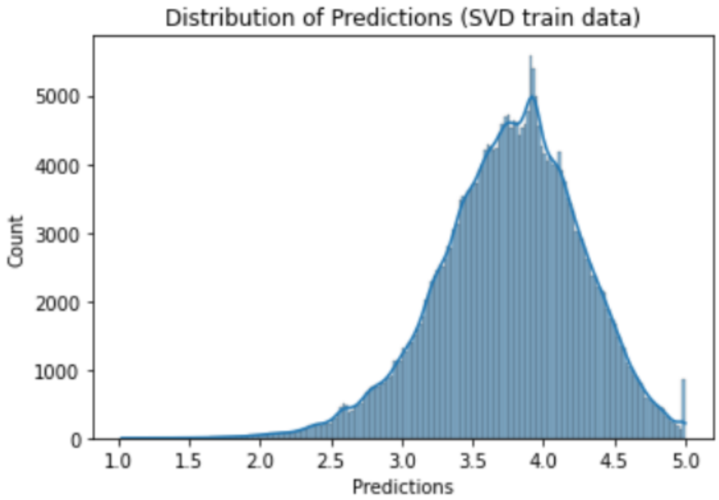
Step 3: Summary of memory-based Collaborative Filtering results

Algorithms	Kaggle result (MAE)
BaselineOnly	0.70462
SVD (lr = 0.001, n_epochs = 200)	0.69876
SVD (lr = 0.001, n_epochs = 150)	0.7033
SVD (lr = 0.005, n_epochs = 50)	0.69905

Generally, the memory-based algorithms have very similar result on the test set. Further post-processing methods were applied to the prediction result, for instance, rounding predictions results to integer values or 2 decimal places. However, it was not able to see significant improvement in the MAE results, the results are even worse for most of the time. Therefore, no further post-processing will be applied.

Prediction from the best SVD model

Predictions made from the best performing SVD model experienced similar bias to the BaselineOnly model, prediction results are left skewed and low rating predictions were not made accurately. Indicating that memory-based models are potentially influenced easily to make accurate predictions on biased data.



3.2 Model-based collaborative filtering

Algorithm 1: Basic Matrix Factorization neural network (Basic MF)

For the Basic MF model, only the users, items, and rating columns in train set were considered. The Basic MF with bias neural network class from tutorial activity were the model adopted in this task.

Learning rate was kept at 0.01, embedding size kept at 100, with regularization term as 5 x 10^{-5}, in order to compensate the potential overfitting issue caused by the high learning rate.

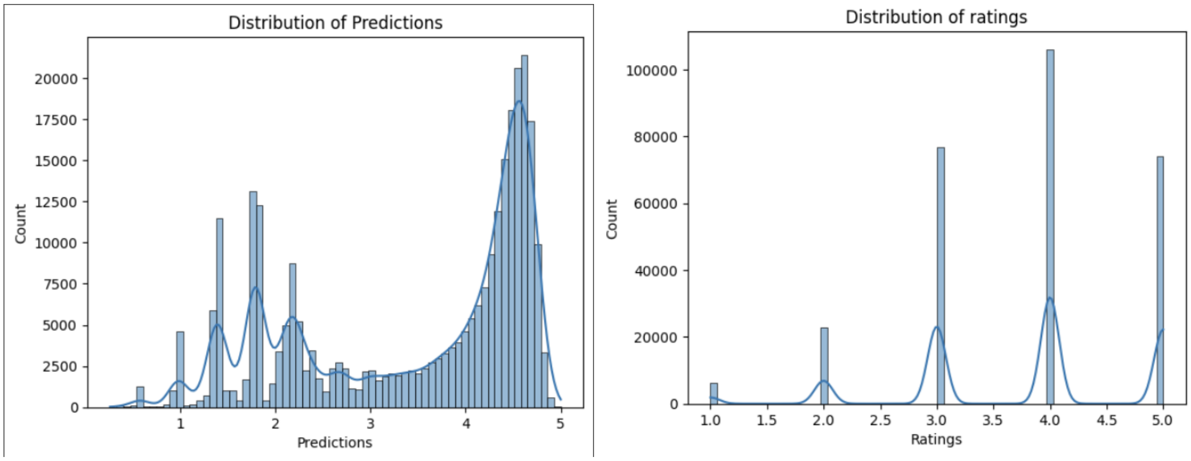
- Normalization

  As the prediction results given by the neural network model ranged outside of the 1 – 5 zone, a basic rescaling method was applied, where:

$$\hat{y}_{\text{rescaled},i} = \hat{y}_i * 5/argmax(\hat{y})$$

  And no further processing was conducted on the prediction results, and the prediction results on the test set will be concluded together with the Advanced MF model.

Fig 3. Comparison of prediction results to their true label in the train set



Algorithm 2: Advance Matrix Factorization neural network with extra features (Advance MF)

In this model, on top of the users, items, and rating columns in the train set, information in the metadata file will also be considered.

Step 1: Metadata preprocessing

Data in the metadata file is not organized and contain various data type, the following preprocessing steps were applied:

Item_id: dropping all records where item_id is null.

Language:

1. Grouping all variation of English (en-US, en-GB, en-CA) to “eng”.
2. Impute for null value from author column, assuming an author only write in one language.
3. Remove rare language (count less than 50) or null value as “other”
4. Assign categorical value to the remaining language value

Authors:

1. Get count of books written by each author in the dataset.
2. Divide the range from 0 to max count in 100 bins, assign counts to bins accordingly.
3. Further convert them as categorical value, where null values were represented by “0”.

CountOfReview:

1. Divide the range from 0 to max count in 100 bins, assign counts to bins accordingly.
2. Convert them into categorical value, where null values were represented by “0”.

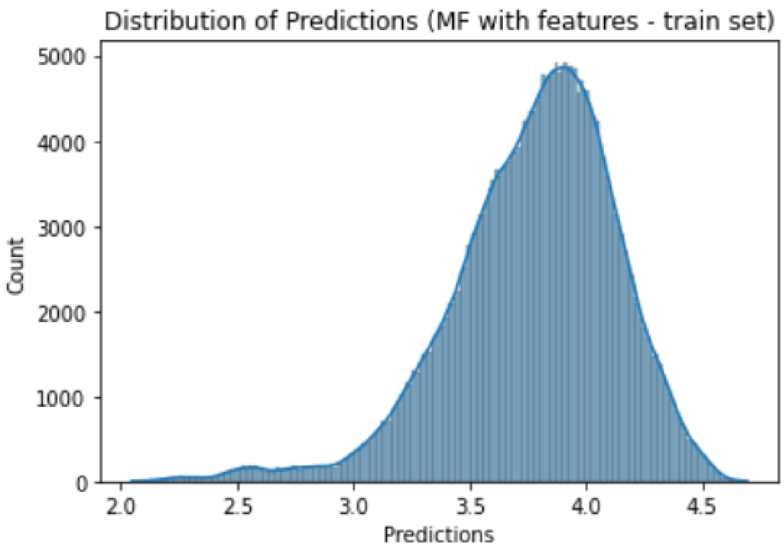
Dropped columns: pagesNumber, Publisher, PublishYear, Name

Step 2: Define MFWithFeatures class

The code of this class referenced the model built in the tutorial activity. However, part of it in the __init__ and forward method were altered so that it matches with the parameters in this book rating prediction work case.

The model was run with 50 training epochs with a learning rate of 0.01, and a regularization term of 5 x 10^{-5}. The same normalization method applied in Basic MF model was also applied to the Advance MF results.

Distribution of prediction results on the trainset from Advance MF model:



The results from the Advance MF model present

Kaggle results:

Algorithms	Kaggle results (MAE)
Basic MF	1.0528
Advance MF	0.7789

Comparing Basic MF and Advance MF

For Basic MF, the distribution of prediction results demonstrated a similar pattern to the true labels of the trainset, potentially indicating an overfitting issue. Thus, leading to relatively poor performance on the test set.

As for Advance MF, it exhibits similar characteristics to the memory-based models, with making majority of the predictions on the far end on the scale. One possible reason is the categorical values in the merged dataset are in range far larger than the range of rating (1 to 5), hence resulting in bias results. If all categorical values are limited to 5, it is likely that the biasness in the prediction results could be relieved.

4. Conclusion

To conclude, this report has covered various models: BaselineOnly, SVD, BasicMF, and AdvancedMF with extra features.

Surprisingly, the simplest model, BaselineOnly, outperformed the more complex ones slightly in our experiments. While the performances of all models were generally similar, BasicMF demonstrated some overfitting issues, suggesting it may not generalize well to unseen data.

One potential explanation for the BaselineOnly model's superior performance is that the patterns in the dataset might be simpler than expected, reducing the advantage of more complex models. It's also noteworthy that the AdvancedMF did not perform as expectedly better over the other models.

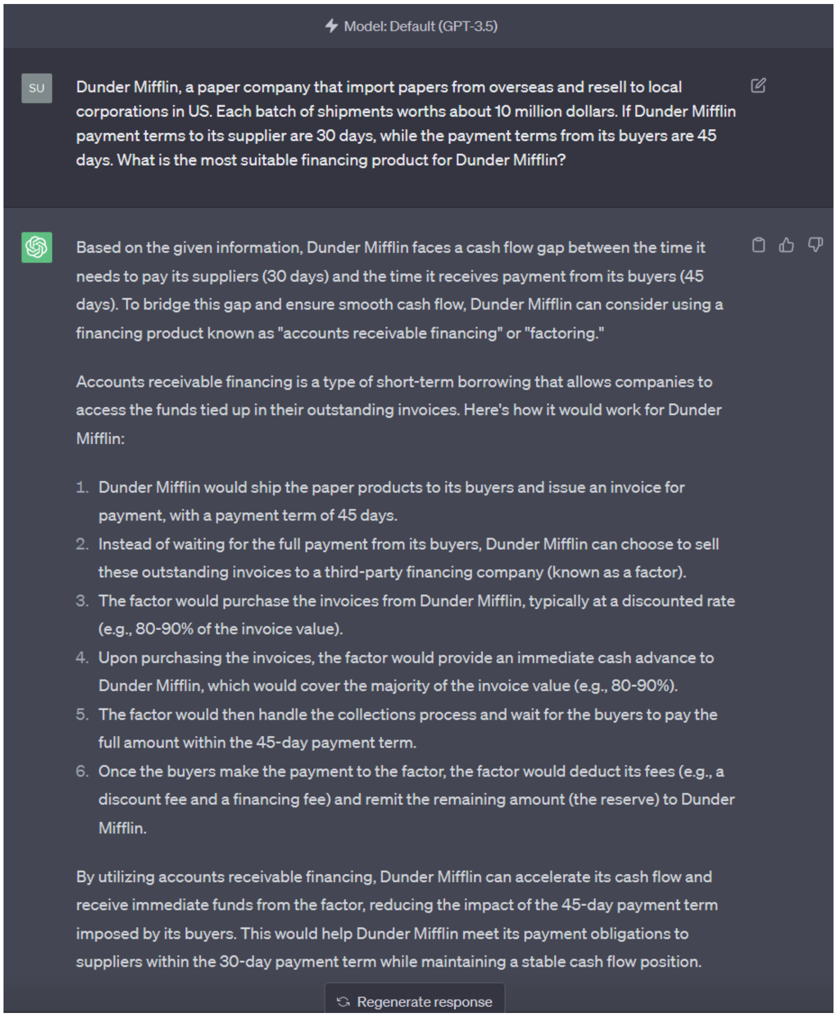
This could be indicative of improper implementation or insufficient tuning of the model's hyperparameters or may suggest that the extra features included did not significantly contribute to the prediction accuracy.

Future work should focus on reviewing the implementation of AdvancedMF and exploring other potential features implementation method or more complex modeling approaches that could better capture the patterns in the data.

Task 2: ChatGPT

Domain 1: Banking and Finance industry

1. Wednesday 24 May 2023 18:25



Dunder Mifflin, a paper company that import papers from overseas and resell to local corporations in US. Each batch of shipments worths about 10 million dollars. If Dunder Mifflin payment terms to its supplier are 30 days, while the payment terms from its buyers are 45 days. What is the most suitable financing product for Dunder Mifflin?

Based on the given information, Dunder Mifflin faces a cash flow gap between the time it needs to pay its suppliers (30 days) and the time it receives payment from its buyers (45 days). To bridge this gap and ensure smooth cash flow, Dunder Mifflin can consider using a financing product known as "accounts receivable financing" or "factoring."

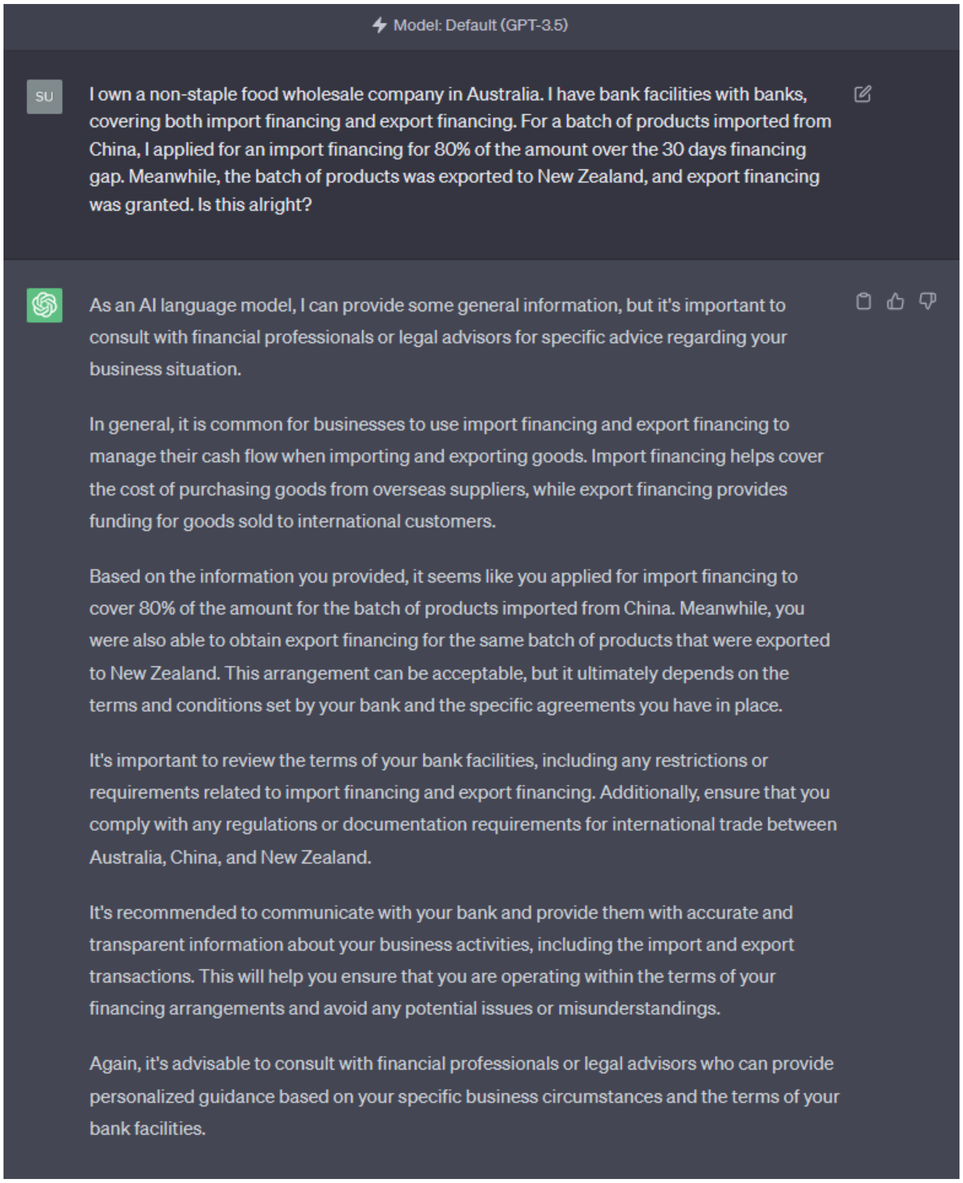
Accounts receivable financing is a type of short-term borrowing that allows companies to access the funds tied up in their outstanding invoices. Here's how it would work for Dunder Mifflin:

1. Dunder Mifflin would ship the paper products to its buyers and issue an invoice for payment, with a payment term of 45 days.
2. Instead of waiting for the full payment from its buyers, Dunder Mifflin can choose to sell these outstanding invoices to a third-party financing company (known as a factor).
3. The factor would purchase the invoices from Dunder Mifflin, typically at a discounted rate (e.g., 80-90% of the invoice value).
4. Upon purchasing the invoices, the factor would provide an immediate cash advance to Dunder Mifflin, which would cover the majority of the invoice value (e.g., 80-90%).
5. The factor would then handle the collections process and wait for the buyers to pay the full amount within the 45-day payment term.
6. Once the buyers make the payment to the factor, the factor would deduct its fees (e.g., a discount fee and a financing fee) and remit the remaining amount (the reserve) to Dunder Mifflin.

By utilizing accounts receivable financing, Dunder Mifflin can accelerate its cash flow and receive immediate funds from the factor, reducing the impact of the 45-day payment term imposed by its buyers. This would help Dunder Mifflin meet its payment obligations to suppliers within the 30-day payment term while maintaining a stable cash flow position.

In the banking industry, factoring is considered as a financing method that requires excessive commitments from the banks and the client during the audit and execution process. While it is definitely a possible option, it is not the common practice to adopt factoring. Due to the nature of factoring requires the company to pledge their account-receivable during the financing period, which may significantly stagnant cash flow and potentially cause issues.

2. Wednesday 24 May 2023 18:38



I own a non-staple food wholesale company in Australia. I have bank facilities with banks, covering both import financing and export financing. For a batch of products imported from China, I applied for an import financing for 80% of the amount over the 30 days financing gap. Meanwhile, the batch of products was exported to New Zealand, and export financing was granted. Is this alright?

As an AI language model, I can provide some general information, but it's important to consult with financial professionals or legal advisors for specific advice regarding your business situation.

In general, it is common for businesses to use import financing and export financing to manage their cash flow when importing and exporting goods. Import financing helps cover the cost of purchasing goods from overseas suppliers, while export financing provides funding for goods sold to international customers.

Based on the information you provided, it seems like you applied for import financing to cover 80% of the amount for the batch of products imported from China. Meanwhile, you were also able to obtain export financing for the same batch of products that were exported to New Zealand. This arrangement can be acceptable, but it ultimately depends on the terms and conditions set by your bank and the specific agreements you have in place.

It's important to review the terms of your bank facilities, including any restrictions or requirements related to import financing and export financing. Additionally, ensure that you comply with any regulations or documentation requirements for international trade between Australia, China, and New Zealand.

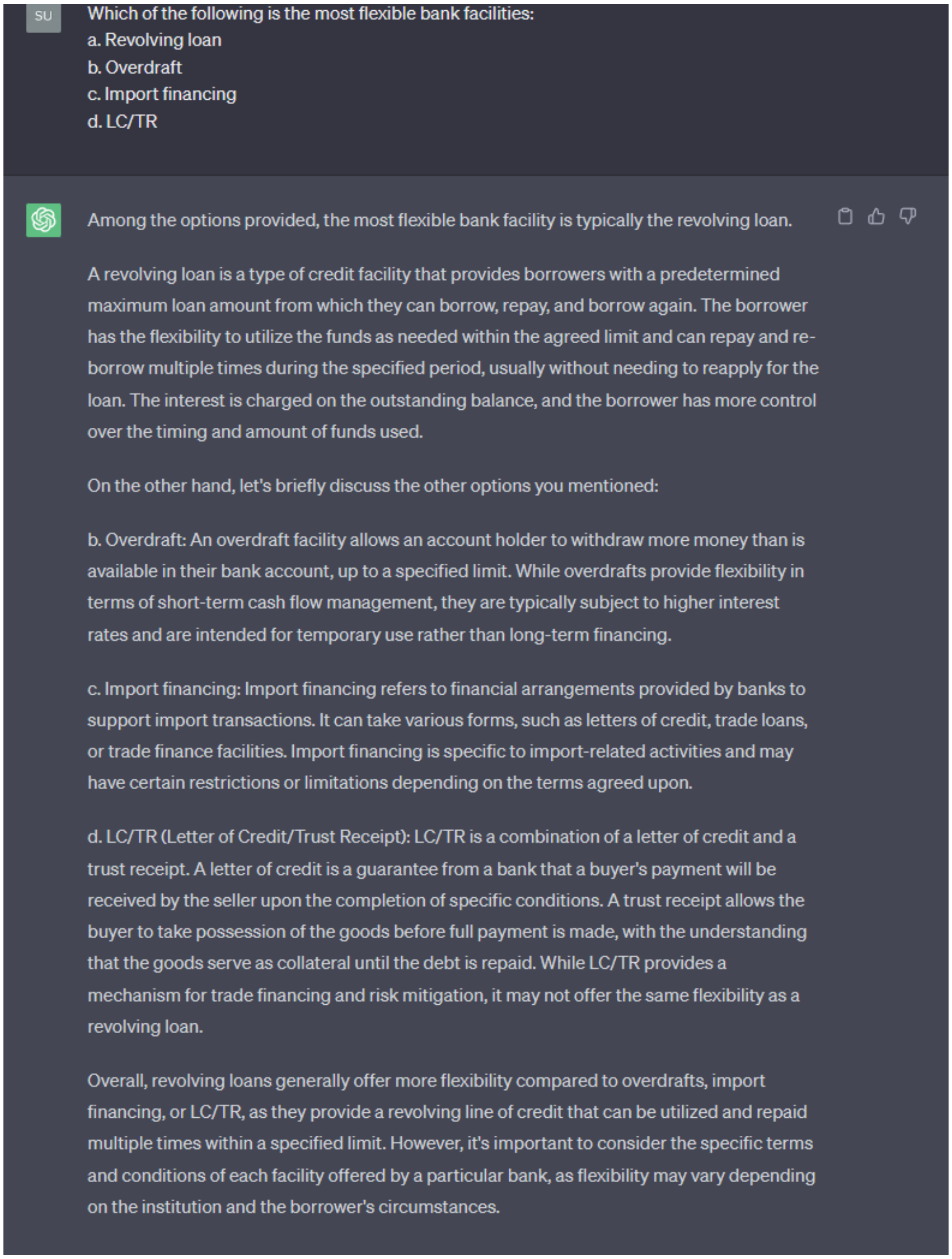
It's recommended to communicate with your bank and provide them with accurate and transparent information about your business activities, including the import and export transactions. This will help you ensure that you are operating within the terms of your financing arrangements and avoid any potential issues or misunderstandings.

Again, it's advisable to consult with financial professionals or legal advisors who can provide personalized guidance based on your specific business circumstances and the terms of your bank facilities.

ChatGPT has failed to notice the potential double financing issue in the example, where a company is receiving financing from a same batch of goods, that may violate facility contract with the financial institute and may cause severe issue.

This confusion could also be caused by the attention mechanism embedded in the model, leading the model to focus on other sections of the prompt and failed to notice other elements in the text.

3. Wednesday 24 May 2023 18:51



Which of the following is the most flexible bank facilities:
a. Revolving loan
b. Overdraft
c. Import financing
d. LC/TR

Among the options provided, the most flexible bank facility is typically the revolving loan.

A revolving loan is a type of credit facility that provides borrowers with a predetermined maximum loan amount from which they can borrow, repay, and borrow again. The borrower has the flexibility to utilize the funds as needed within the agreed limit and can repay and re-borrow multiple times during the specified period, usually without needing to reapply for the loan. The interest is charged on the outstanding balance, and the borrower has more control over the timing and amount of funds used.

On the other hand, let's briefly discuss the other options you mentioned:

b. Overdraft: An overdraft facility allows an account holder to withdraw more money than is available in their bank account, up to a specified limit. While overdrafts provide flexibility in terms of short-term cash flow management, they are typically subject to higher interest rates and are intended for temporary use rather than long-term financing.

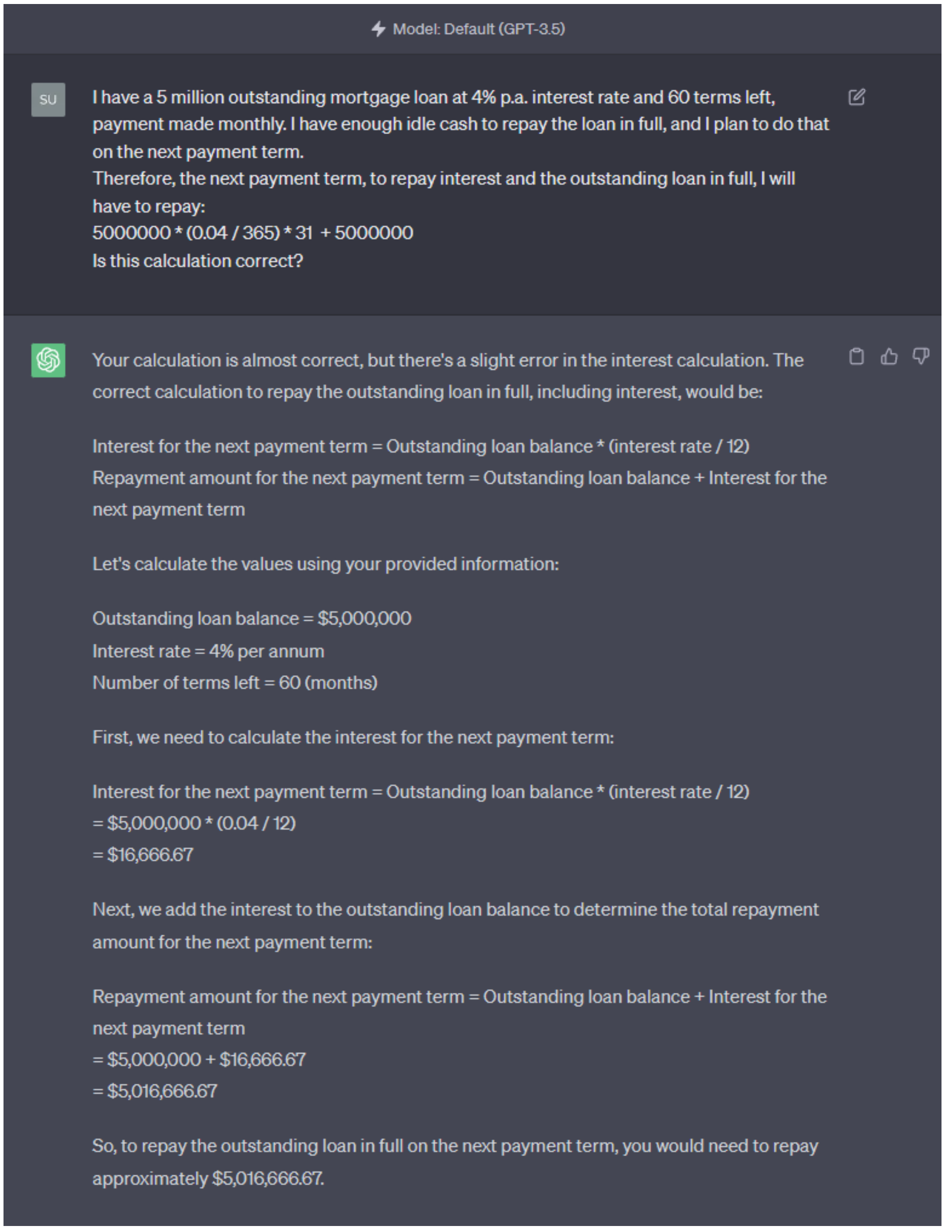
c. Import financing: Import financing refers to financial arrangements provided by banks to support import transactions. It can take various forms, such as letters of credit, trade loans, or trade finance facilities. Import financing is specific to import-related activities and may have certain restrictions or limitations depending on the terms agreed upon.

d. LC/TR (Letter of Credit/Trust Receipt): LC/TR is a combination of a letter of credit and a trust receipt. A letter of credit is a guarantee from a bank that a buyer's payment will be received by the seller upon the completion of specific conditions. A trust receipt allows the buyer to take possession of the goods before full payment is made, with the understanding that the goods serve as collateral until the debt is repaid. While LC/TR provides a mechanism for trade financing and risk mitigation, it may not offer the same flexibility as a revolving loan.

Overall, revolving loans generally offer more flexibility compared to overdrafts, import financing, or LC/TR, as they provide a revolving line of credit that can be utilized and repaid multiple times within a specified limit. However, it's important to consider the specific terms and conditions of each facility offered by a particular bank, as flexibility may vary depending on the institution and the borrower's circumstances.

While revolving loan is also flexible, in the list of four, overdraft has to be the most flexible facility of them all. The model gives hallucinating responses which also take into consideration of interest rate, repayment terms, neglecting the main element in the question is about the flexibility.

4. Wednesday 24 May 2023 19:12



I have a 5 million outstanding mortgage loan at 4% p.a. interest rate and 60 terms left, payment made monthly. I have enough idle cash to repay the loan in full, and I plan to do that on the next payment term.
Therefore, the next payment term, to repay interest and the outstanding loan in full, I will have to repay:
5000000 * (0.04 / 365) * 31 + 5000000
Is this calculation correct?

Your calculation is almost correct, but there's a slight error in the interest calculation. The correct calculation to repay the outstanding loan in full, including interest, would be:

Interest for the next payment term = Outstanding loan balance * (interest rate / 12)
Repayment amount for the next payment term = Outstanding loan balance + Interest for the next payment term

Let's calculate the values using your provided information:

Outstanding loan balance = $5,000,000
Interest rate = 4% per annum
Number of terms left = 60 (months)

First, we need to calculate the interest for the next payment term:

Interest for the next payment term = Outstanding loan balance * (interest rate / 12)
= $5,000,000 * (0.04 / 12)
= $16,666.67

Next, we add the interest to the outstanding loan balance to determine the total repayment amount for the next payment term:

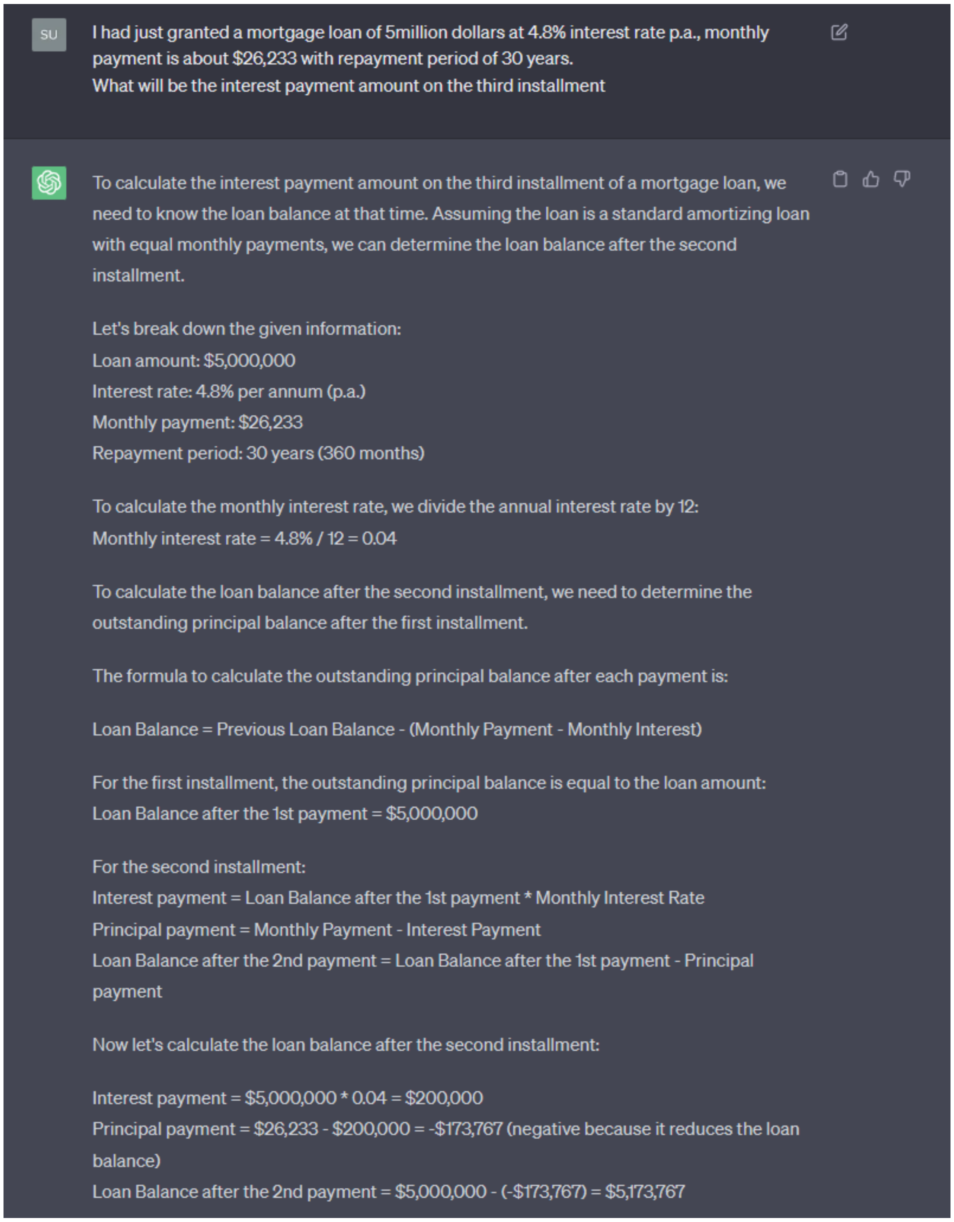
Repayment amount for the next payment term = Outstanding loan balance + Interest for the next payment term
= $5,000,000 + $16,666.67
= $5,016,666.67

So, to repay the outstanding loan in full on the next payment term, you would need to repay approximately $5,016,666.67.

ChatGPT did its calculation correctly, while this is not primarily a mathematical calculation question.

In the banking industry, an early prepayment penalty is very common. My input prompt may have misled it to focus only on the calculation part, while the main issue lies within the concepts behind the action of repaying a loan before maturity.

5. Wednesday 24 May 2023 19:23



I had just granted a mortgage loan of 5million dollars at 4.8% interest rate p.a., monthly payment is about $26,233 with repayment period of 30 years.
What will be the interest payment amount on the third installment

To calculate the interest payment amount on the third installment of a mortgage loan, we need to know the loan balance at that time. Assuming the loan is a standard amortizing loan with equal monthly payments, we can determine the loan balance after the second installment.

Let's break down the given information:
Loan amount: $5,000,000
Interest rate: 4.8% per annum (p.a.)
Monthly payment: $26,233
Repayment period: 30 years (360 months)

To calculate the monthly interest rate, we divide the annual interest rate by 12:
Monthly interest rate = 4.8% / 12 = 0.04

To calculate the loan balance after the second installment, we need to determine the outstanding principal balance after the first installment.

The formula to calculate the outstanding principal balance after each payment is:

Loan Balance = Previous Loan Balance - (Monthly Payment - Monthly Interest)

For the first installment, the outstanding principal balance is equal to the loan amount:
Loan Balance after the 1st payment = $5,000,000

For the second installment:
Interest payment = Loan Balance after the 1st payment * Monthly Interest Rate
Principal payment = Monthly Payment - Interest Payment
Loan Balance after the 2nd payment = Loan Balance after the 1st payment - Principal payment

Now let's calculate the loan balance after the second installment:

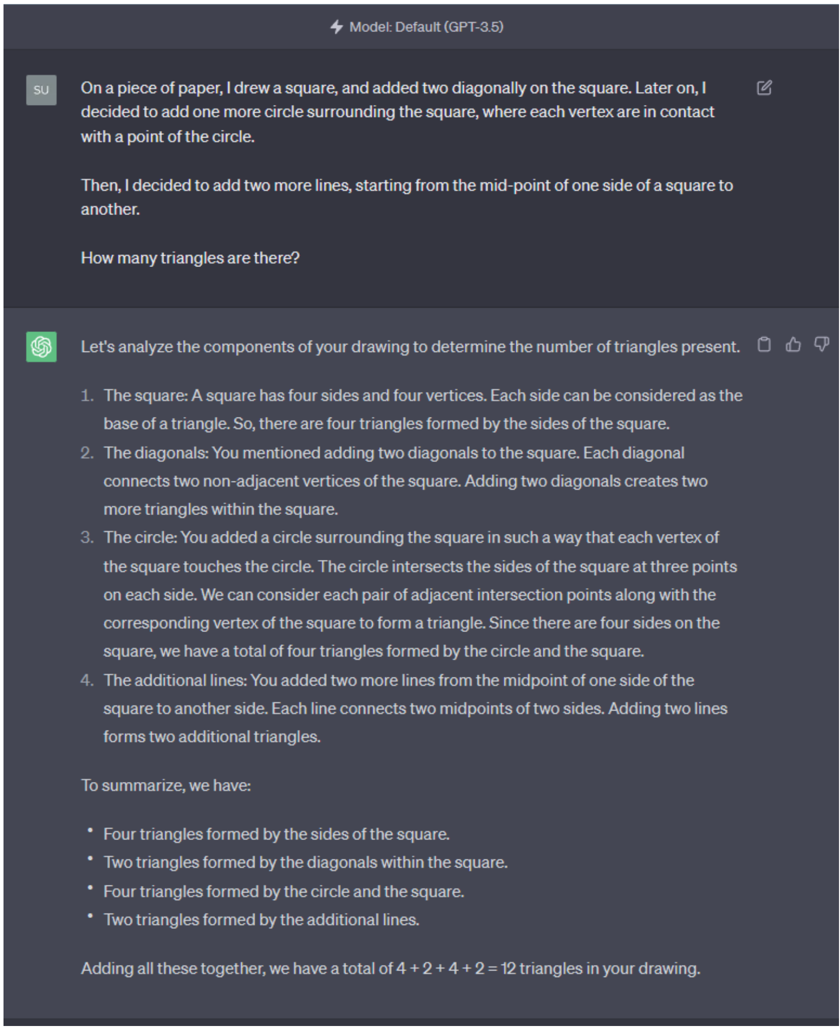
Interest payment = $5,000,000 * 0.04 = $200,000
Principal payment = $26,233 - $200,000 = -$173,767 (negative because it reduces the loan balance)
Loan Balance after the 2nd payment = $5,000,000 - (-$173,767) = $5,173,767

Finally, mathematical calculations question to challenge ChatGPT while it is known that it is relatively weak at. The error was made at very early stage where monthly interest 4.8%/12 should be 0.004 instead of 0.04.

As ChatGPT model was primarily built as a large language model with the objective to understand natural language and to generate human-like responses. To achieve that, Bubeck et al. suggested the issue of reverse reasoning and verification, implying the model’s tendency to generate a full output even if the initialization phase is already incorrect.

Domain 2: General (Commonsense, Logical, Temporal

1. Wednesday, 24 May 2023 17:22



On a piece of paper, I drew a square, and added two diagonally on the square. Later on, I decided to add one more circle surrounding the square, where each vertex are in contact with a point of the circle.

Then, I decided to add two more lines, starting from the mid-point of one side of a square to another.

How many triangles are there?

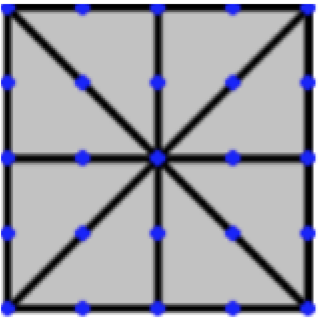
Let's analyze the components of your drawing to determine the number of triangles present.

1. The square: A square has four sides and four vertices. Each side can be considered as the base of a triangle. So, there are four triangles formed by the sides of the square.
2. The diagonals: You mentioned adding two diagonals to the square. Each diagonal connects two non-adjacent vertices of the square. Adding two diagonals creates two more triangles within the square.
3. The circle: You added a circle surrounding the square in such a way that each vertex of the square touches the circle. The circle intersects the sides of the square at three points on each side. We can consider each pair of adjacent intersection points along with the corresponding vertex of the square to form a triangle. Since there are four sides on the square, we have a total of four triangles formed by the circle and the square.
4. The additional lines: You added two more lines from the midpoint of one side of the square to another side. Each line connects two midpoints of two sides. Adding two lines forms two additional triangles.

To summarize, we have:

- Four triangles formed by the sides of the square.
- Two triangles formed by the diagonals within the square.
- Four triangles formed by the circle and the square.
- Two triangles formed by the additional lines.

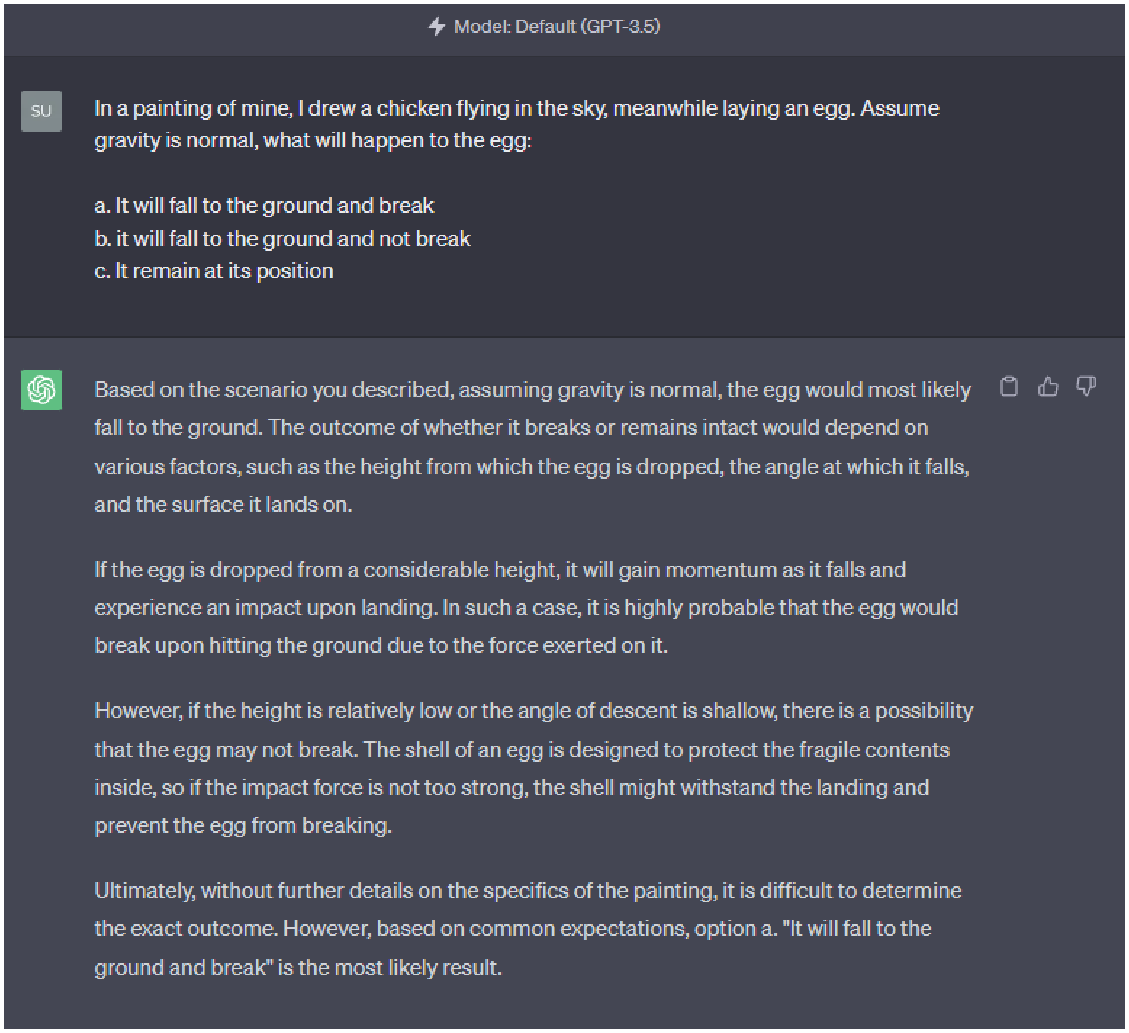
Adding all these together, we have a total of 4 + 2 + 4 + 2 = 12 triangles in your drawing.



Bubeck et al. (2023) suggested that Text-based games pose a significant challenge for language models as it is challenging for them to comprehend natural language, reason about current game state, while generating valid comments all at the same time.

They may also encounter similar traits to a RNN model that may lose track of elements during the computation process (Kiddon, 2016), and hence, lead to the reasoning failure occurred in this example.

2. Wednesday, 24 May 2023 17:26



In a painting of mine, I drew a chicken flying in the sky, meanwhile laying an egg. Assume gravity is normal, what will happen to the egg:

a. It will fall to the ground and break
b. it will fall to the ground and not break
c. It remain at its position

Based on the scenario you described, assuming gravity is normal, the egg would most likely fall to the ground. The outcome of whether it breaks or remains intact would depend on various factors, such as the height from which the egg is dropped, the angle at which it falls, and the surface it lands on.

If the egg is dropped from a considerable height, it will gain momentum as it falls and experience an impact upon landing. In such a case, it is highly probable that the egg would break upon hitting the ground due to the force exerted on it.

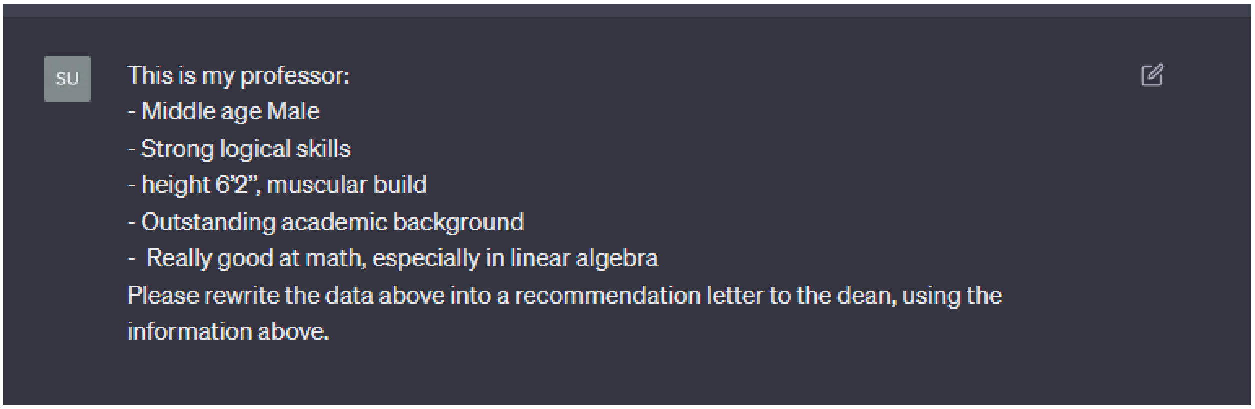
However, if the height is relatively low or the angle of descent is shallow, there is a possibility that the egg may not break. The shell of an egg is designed to protect the fragile contents inside, so if the impact force is not too strong, the shell might withstand the landing and prevent the egg from breaking.

Ultimately, without further details on the specifics of the painting, it is difficult to determine the exact outcome. However, based on common expectations, option a. "It will fall to the ground and break" is the most likely result.

In this hypothetical drawing scenario, GPT, as a large language model, is not able to recognize the influence of physics in painting, indicate its limitation on processing spatial ideas, in a structure that could be rare in its training data.

According to Davis (2023), commonsense knowledge and reasoning have unclear scope and involved highly intelligent reasoning process which is easily affect by noise in data. Due to the massive dataset that ChatGPT was built upon, it may have suffered from some subpar quality of data in this response that led to this reasoning error.

3. Wednesday, 24 May 2023 17:37 (Bubeck et.al., 2023)



This is my professor:
- Middle age Male
- Strong logical skills
- height 6'2", muscular build
- Outstanding academic background
- Really good at math, especially in linear algebra

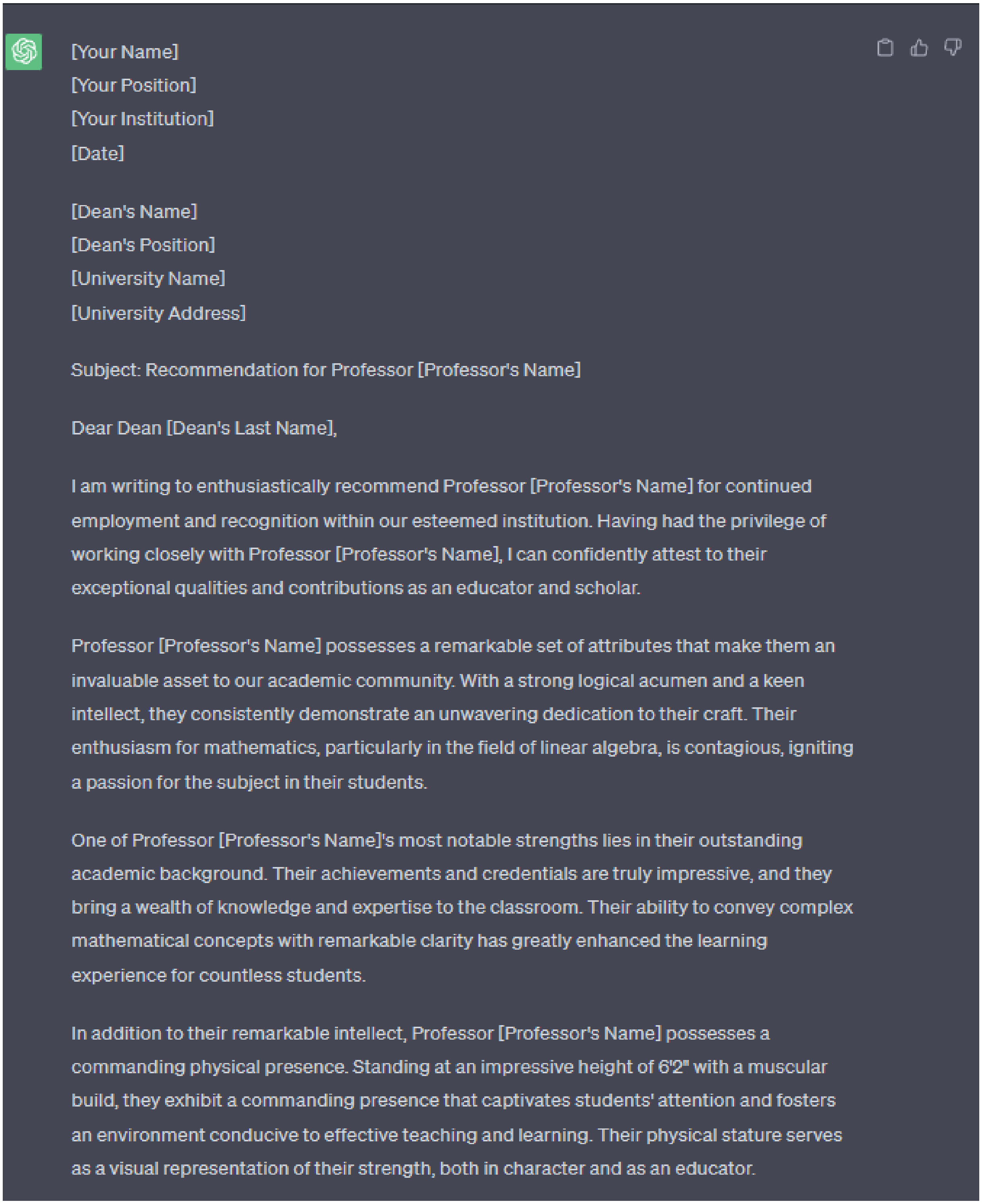
Please rewrite the data above into a recommendation letter to the dean, using the information above.

(As GPT's response is long, I will analysis before giving the screenshot)

This question has referenced Bubeck et.al. (2023) example in their paper about employing GPT 4 to produce medical record of a female patient.

Their paper refered such situation as hallucination, imply the language model creating new information on his own, in this case, creating information of the professor's leadership ability and other personal traits, that wasn't include in the prompt.

While hallucination could be useful in some situations, especially when it is asked to create some creative contents, it could cause potential issue when the industry requires only precise and accurate information, like the healthcare sector.



[Your Name]
[Your Position]
[Your Institution]
[Date]

[Dean's Name]
[Dean's Position]
[University Name]
[University Address]

Subject: Recommendation for Professor [Professor's Name]

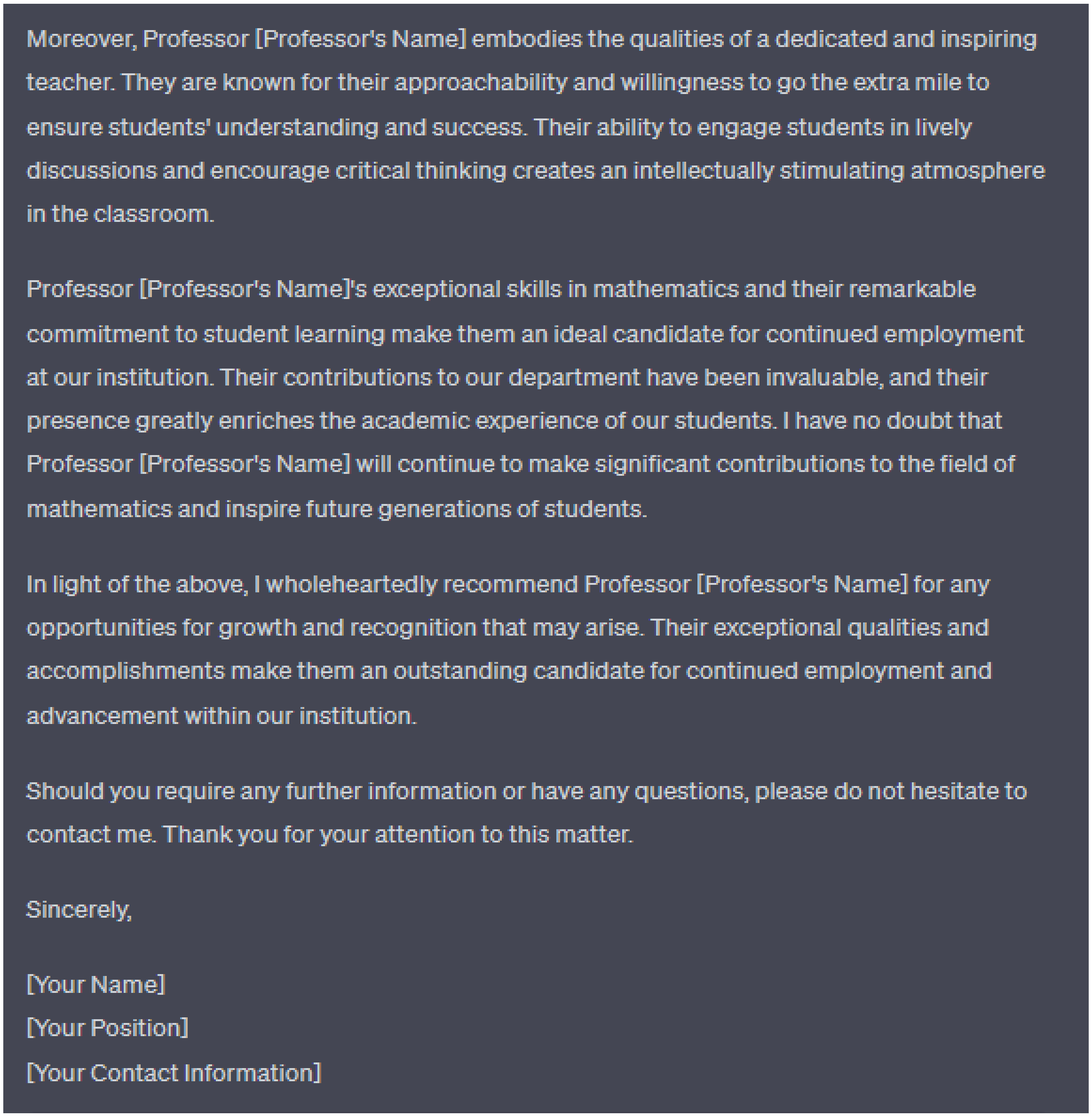
Dear Dean [Dean's Last Name],

I am writing to enthusiastically recommend Professor [Professor's Name] for continued employment and recognition within our esteemed institution. Having had the privilege of working closely with Professor [Professor's Name], I can confidently attest to their exceptional qualities and contributions as an educator and scholar.

Professor [Professor's Name] possesses a remarkable set of attributes that make them an invaluable asset to our academic community. With a strong logical acumen and a keen intellect, they consistently demonstrate an unwavering dedication to their craft. Their enthusiasm for mathematics, particularly in the field of linear algebra, is contagious, igniting a passion for the subject in their students.

One of Professor [Professor's Name]'s most notable strengths lies in their outstanding academic background. Their achievements and credentials are truly impressive, and they bring a wealth of knowledge and expertise to the classroom. Their ability to convey complex mathematical concepts with remarkable clarity has greatly enhanced the learning experience for countless students.

In addition to their remarkable intellect, Professor [Professor's Name] possesses a commanding physical presence. Standing at an impressive height of 6'2" with a muscular build, they exhibit a commanding presence that captivates students' attention and fosters an environment conducive to effective teaching and learning. Their physical stature serves as a visual representation of their strength, both in character and as an educator.



Moreover, Professor [Professor's Name] embodies the qualities of a dedicated and inspiring teacher. They are known for their approachability and willingness to go the extra mile to ensure students' understanding and success. Their ability to engage students in lively discussions and encourage critical thinking creates an intellectually stimulating atmosphere in the classroom.

Professor [Professor's Name]'s exceptional skills in mathematics and their remarkable commitment to student learning make them an ideal candidate for continued employment at our institution. Their contributions to our department have been invaluable, and their presence greatly enriches the academic experience of our students. I have no doubt that Professor [Professor's Name] will continue to make significant contributions to the field of mathematics and inspire future generations of students.

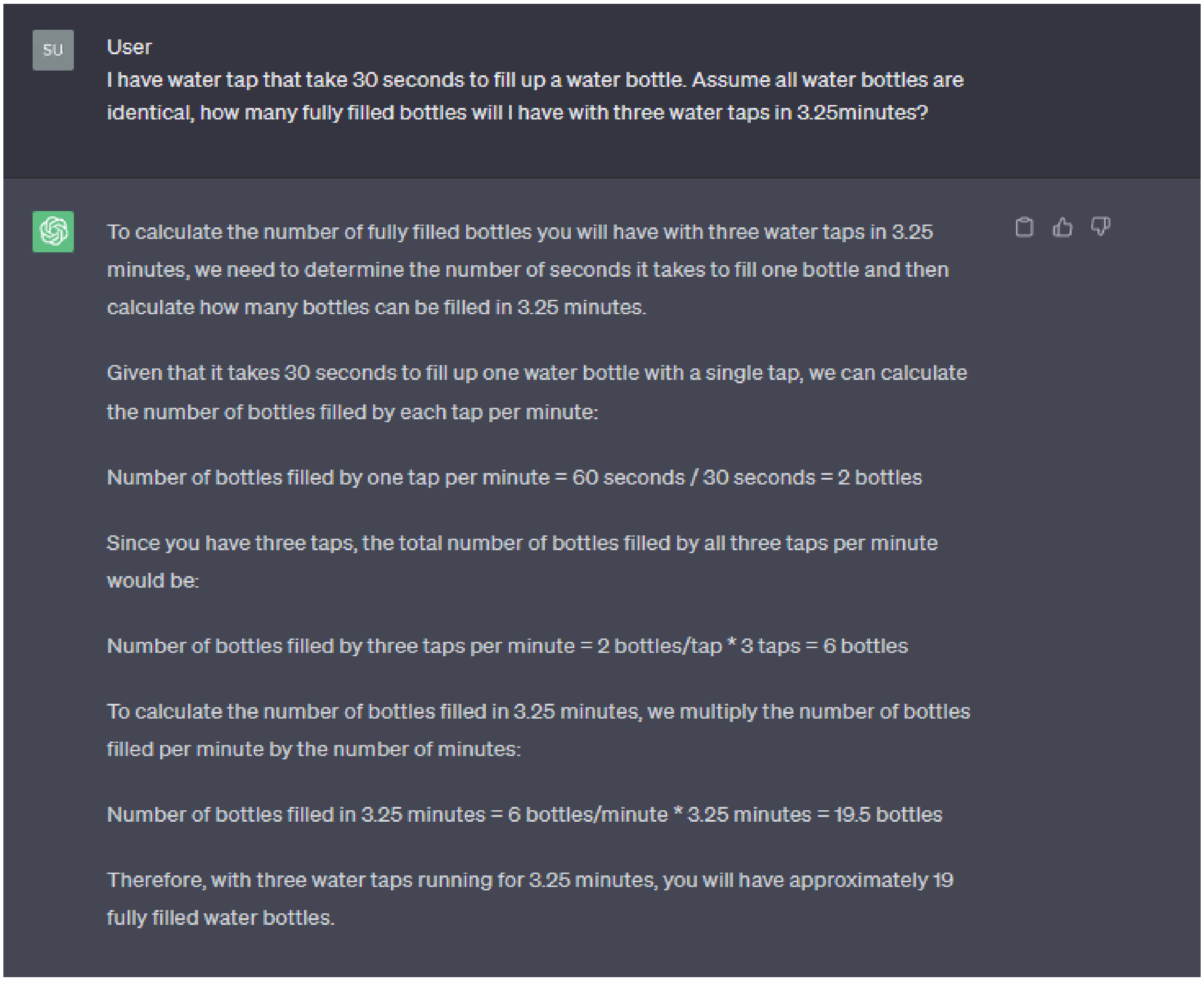
In light of the above, I wholeheartedly recommend Professor [Professor's Name] for any opportunities for growth and recognition that may arise. Their exceptional qualities and accomplishments make them an outstanding candidate for continued employment and advancement within our institution.

Should you require any further information or have any questions, please do not hesitate to contact me. Thank you for your attention to this matter.

Sincerely,

[Your Name]
[Your Position]
[Your Contact Information]

4. Wednesday, 24 May 2023 17:50



User

I have water tap that take 30 seconds to fill up a water bottle. Assume all water bottles are identical, how many fully filled bottles will I have with three water taps in 3.25minutes?

To calculate the number of fully filled bottles you will have with three water taps in 3.25 minutes, we need to determine the number of seconds it takes to fill one bottle and then calculate how many bottles can be filled in 3.25 minutes.

Given that it takes 30 seconds to fill up one water bottle with a single tap, we can calculate the number of bottles filled by each tap per minute:

Number of bottles filled by one tap per minute = 60 seconds / 30 seconds = 2 bottles

Since you have three taps, the total number of bottles filled by all three taps per minute would be:

Number of bottles filled by three taps per minute = 2 bottles/tap * 3 taps = 6 bottles

To calculate the number of bottles filled in 3.25 minutes, we multiply the number of bottles filled per minute by the number of minutes:

Number of bottles filled in 3.25 minutes = 6 bottles/minute * 3.25 minutes = 19.5 bottles

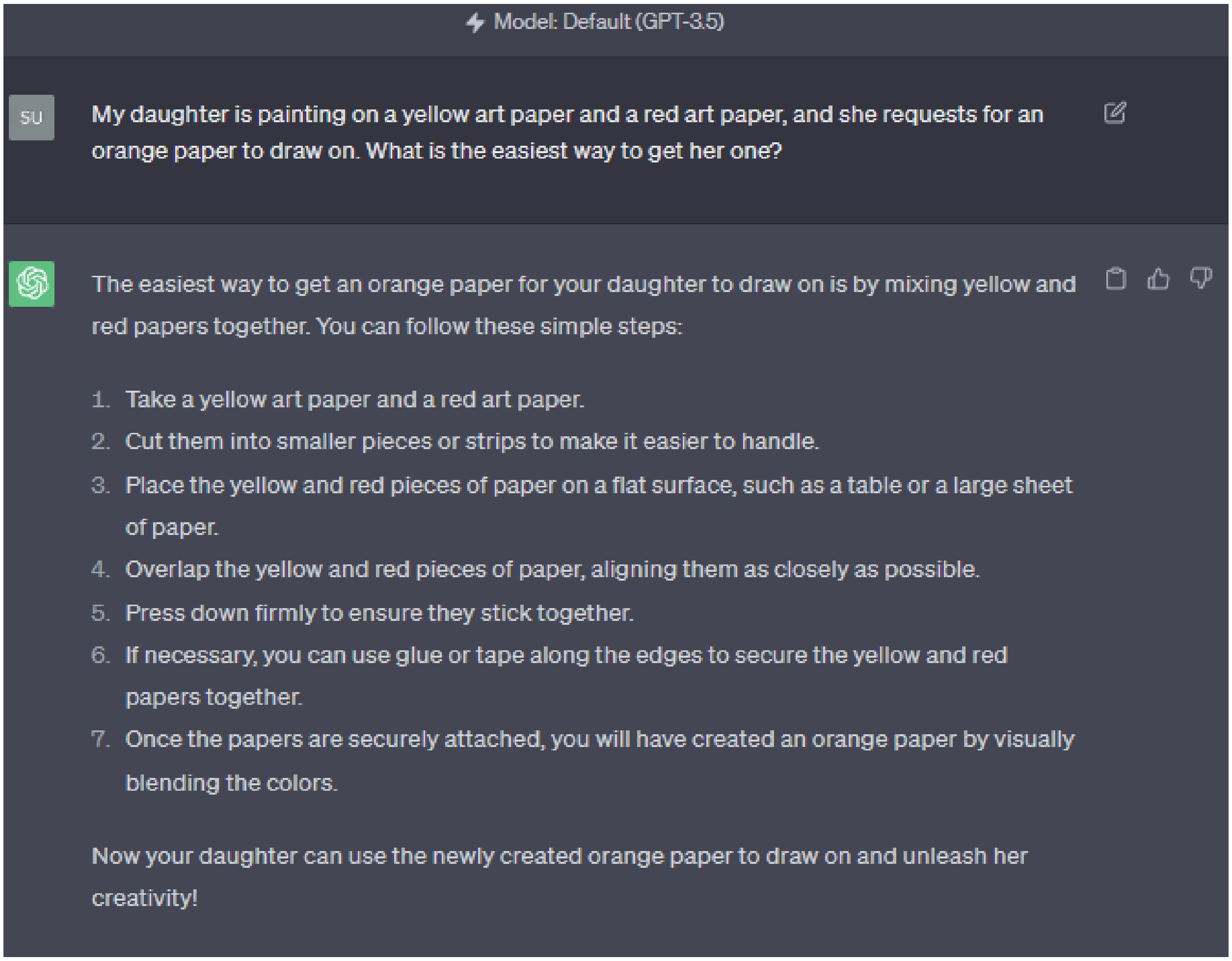
Therefore, with three water taps running for 3.25 minutes, you will have approximately 19 fully filled water bottles.

Liu et.al. (2023) suggest that ChatGPT can solve well-known logical comprehension reason task accurately, but may struggle when the question is “out-of-distribution”.

In this example, I specifically asked for the amount of fully filled bottles, while ChatGPT is not able to distinguish between commonsense reasoning problem and numerical calculation question, and hence lead to such inaccurate output.

It directly jumped into calculation to handle my question, which failed to recongize the nature of the question.

5. Wednesday, 24 May 2023 18:02



My daughter is painting on a yellow art paper and a red art paper, and she requests for an orange paper to draw on. What is the easiest way to get her one?

The easiest way to get an orange paper for your daughter to draw on is by mixing yellow and red papers together. You can follow these simple steps:

1. Take a yellow art paper and a red art paper.
2. Cut them into smaller pieces or strips to make it easier to handle.
3. Place the yellow and red pieces of paper on a flat surface, such as a table or a large sheet of paper.
4. Overlap the yellow and red pieces of paper, aligning them as closely as possible.
5. Press down firmly to ensure they stick together.
6. If necessary, you can use glue or tape along the edges to secure the yellow and red papers together.
7. Once the papers are securely attached, you will have created an orange paper by visually blending the colors.

Now your daughter can use the newly created orange paper to draw on and unleash her creativity!

This example is another demonstration of ChatGPT's commonsense reasoning failure.

From its training data, it is able to recognize the combination of colours, as formula. Therefore, when it comes to this question, it directly apply its knowledge that the mixture of yellow and red would become orange, possibly tricked by the attention mechanism that correlated yellow red and orange too quickly while disregarding the physical constraints on the objects itselves.

References:

Bubeck, Sebastien, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, et al. 2023. “Sparks of Artificial General Intelligence: Early Experiments with GPT-4.” *ArXiv Preprint ArXiv:2303.12712*. https://arxiv.org/abs/2303.12712.

Davis, Ernest. 2023. “Benchmarks for Automated Commonsense Reasoning: A Survey.” *ArXiv Preprint ArXiv:2302.04752*, February. https://arxiv.org/pdf/2302.04752.pdf.

Hug, Nicolas. n.d. “Home.” Surprise Home Page. https://surpriselib.com/.

Li, Susan. 2018. “Building and Testing Recommender Systems with Surprise, Step-By-Step.” Medium. Towards Data Science. December 26, 2018. https://towardsdatascience.com/building-and-testing-recommender-systems-with-surprise-step-by-step-d4ba702ef80b.

Liu, Hanmeng, Ruoxi Ning, Zhiyang Teng, Jian Liu, Qiji Zhou, and Yue Zhang. 2023. “Evaluating the Logical Reasoning Ability of ChatGPT and GPT-4.” *ArXiv Preprint ArXiv:2304.03439,* April. https://arxiv.org/pdf/2304.03439.pdf.