

# OPTIMIZING PRICING FRAMEWORK USING REASONING LLMs AND AGENTS FOR ENHANCED DEAL NEGOTIATIONS

Chirag Soni <sup>1</sup> and Swati Shah <sup>2</sup>

<sup>1</sup> Senior Product Manager, Data, AI and Insights, PayPal, Bangalore India

<sup>2</sup> Head of Margin Optimization Product, PayPal, San Jose USA

## ABSTRACT

*Generative AI and the potential it carries has enabled businesses to incorporate intelligent automation and optimization across domains. With this paper we present a methodology to enhance b2b pricing and deal negotiation process where langchain framework and agents could be used to summarize critical information regarding businesses (B2B customers) in real time to indirectly calculate their willingness to pay in the form of a score generated by reasoning LLMs. This score will act as an indicator to suggest the customer's ability to accept the price point being negotiated in the given period of time based on their financial health, market sentiments and internal to company performance.*

## KEYWORDS

*Generative AI, LLM, LangChain, AI Agents, Reasoning models, Risk, Sales*

## 1. INTRODUCTION

Deal negotiation is a very complex process in the sales domain specially in B2B segment. It indicates how much a Business is willing to pay for the services your company offers, and directly correlates to how you price your products. It is a complex metric because there is no set standardized framework to assess end to end Customer health overview including their internal and external performance and the inputs to assess health parameters varies individual to individual. Many times the Sales teams would not be going deeper into the customer health in that given point in time and stand the risk of losing the given deal as the pricing does not work out for both the parties or the deal may be called off due to higher price perception by the target Business (customer). This would also come under the domain of assessing Willingness to Pay, and this metric is very complicated to arrive at. There is also a very critical component in assessing overall health of the customer, which is Time. The Business's health will be majorly depending on the time during which deal negotiations are happening, and there are multiple inputs that need to be looked at before concluding. It may be possible that the Business was not doing well financially but during the deal negotiation the latest financial metrics might be very positive and the negotiation could take a totally different direction. Also, even if the Sales representatives have all the required information for data based deal negotiation, they still would need to gather information from multiple sources which may take several days resulting in delayed sales and onboarding cycles impacting revenue generation. This paper discusses an approach that could be used for B2B sales to assess at a given point in time what would be the WTP for a Business and when is the best time to have the sales process kicked off for these customers. The motivation behind this research was to enable faster Business onboarding and kick start revenue stream sooner since there is considerable time being spent in manual research

which is often inconsistent. Also, most pricing decisions consist of data driven strategies to study internal performance and offer products with customized pricing to Businesses based on how well they are already engaged, and usually does not include estimates of deal negotiation parameters coming from external data and insights. Customer health is a metric that's very dynamic in nature and external factors such as Financial health and ongoing market sentiments add a critical element to this framework. The objective was to automate the entire process to a certain extent by removing dependency on manual research and consolidate all the major factors affecting deal negotiation at one single platform.

## **2. DEAL NEGOTIATION ENHANCEMENT FRAMEWORK**

### **2.1.Uncovering Willingness to Pay**

Let's look at the most important inputs that are needed to define and assess WTP. In the next section we will discuss how to consolidate everything and score the Businesses on these parameters–

- Internal performance rank for the Business: for already onboarded Businesses which have existing product penetration, how has their engagement been with your company historically. For e.g. High product penetration, long term relationship, higher profit margins coming from these Businesses to your company etc. are all metrics that indicate higher rank for these Businesses for within your company performance.
- Financial Health from P&L statements: this indicates how a given Business is performing on Financial metrics like Sales, Opex, Capex, EBITDA, Revenue, Margin etc. These are a very strong indicator of how healthy a business is in given point in time compared to previous financial statements released
- Market and Investor Sentiments: Are the end customers for these Businesses happy? Are there any news articles indicating customer dissent and risk factors due to the Business's long-term decisions, or bankruptcy indicators etc.? All these factors are also important to consider when assessing the WTP of a business at given point in time

### **2.2. Comsolidating the Required Data**

- Each of the scoring pillars mentioned above will be an input to the final Deal Negotiation score we will be generating. Below is the framework to prepare for scoring:
- Internal Ranking: Performance Score generated over historical transactional data internal to your company that rates the businesses on all parameters which affect engagement and incoming revenue.
- Financial Health Ranking: this data needs to be fetched from external sources. The most reliable source (for listed companies) is their publicly released Financial statements. The financial metrics discussed in previous paragraphs would usually be available on Google/ Yahoo Finance and a trained sales agent would need to fetch this manually at the current time while deal is being framed. They would then need to compare with historical data and form an opinion about given Business's financial health
- Market Sentiment Ranking: this data should also come from external sources like news articles, press releases, finance outlook articles from data providers like Google

Finance etc. A trained sales agent would need to read through this data and understand the performance based on their individual analysis

### 2.3. Using LLMs to Score the Internal and External Performance

For the internal to company performance scoring, it would be a direct approach to rank businesses on their performance. However, for pulling external data and analysing the Financial and Sentiment performance is manually intensive activity which can be automated using AI Agents as below

1. Design prompts integrated with Agents that will call Google Finance API by passing the Business name as input
2. Receive a large Jsonfile with a host of financial metrics for the given Business
3. Pass this into another pre-created prompt and feed the information to an LLM. This prompt is designed to summarize financial metrics and consolidate information in a textual format
4. This summary will again be passed into another prompt that will call a reasoning LLM to generate a score based on given information and rank the Business on a predetermined scale (provide specific examples on how the scoring framework should look like)
5. Similar call is made to News APIs, and the articles are summarized to build Market Sentiment score by repeating steps 1 through 5

### 2.4. Combining the Scores Together

For the final step, consolidate all scores generated from above steps. This is not a direct summation as there will be different weightage assigned to each scoring pillar. Based on our research, a Business's current Financial Health compared with historical metrics is one of the most important parameters to assess deal score, as good financial health is direct indicator of ability to pay higher, hence this gets the highest weightage. Next will be the internal to company performance score since even though Business is in good financial health, it may be multihoming onto different platforms, meaning product penetration may not be with a single company. Therefore, it is important to compare Financial Score with Internal Score and a large gap between both may indicate operational challenges with the Business, bringing the deal negotiation score down.

Now what remains is the Market Sentiment Score. This also is important because this is an indicator of future risks as financially healthy businesses with negative investor sentiments directly contribute to challenges with relationship in future.

Based on above context, we have come up with an estimate of the equation that would work best for calculating Deal Negotiation Score of a business at any given point in time:

$$\text{Deal Score} = (0.65 * \text{Financial Health Score (1 to 10)}) + (0.25 * \text{Internal to Company Performance Score (1 to 10)}) + (0.1 * \text{Market Sentiment Score (1 to 10)})$$

This equation works as a best case estimate that incorporates 360-degree overview of the Business's health, both internally and externally, if the Deal Score is high it indicates a very good Merchant Health and negotiation could favor the Business as we definitely want to onboard a very healthy customer, whereas low deal score would indicate the Sales teams need to be cautious while negotiating with this customer as there could be some hidden challenges with the business.

### **3. ESTABLISHING THE PROCESS**

The above equation will be used to rank order your customers on a 30-point scale. It does not indicate any amount or percentage but is just an indicative score. But how do we use this? And for what use cases? Below points provide a detailed context of how and why we should use this equation in the pricing process:

#### **3.1. Reading the Score**

Deal score acts as an AI based assistant to prioritize pricing negotiation and deal conversations with Businesses. Without this score, there will be inconsistent and incomplete frameworks where mostly the internal to company data will be used to drive sales whereas extremely critical external metrics will be completely missed.

### **4. BUSINESS USE CASES**

This Score will have multiple use cases within the companies as follows:

1. Critical metric for the Risk teams to keep a check on customers and identify leading indicators of Margin challenges – any significant score drops in Financial and Sentiment health of businesses may indicate upcoming challenges in Business' performance, thereby helping them to take actions and mitigate risks in advance
2. Sales teams can use this to prioritize customer conversations – from the list of Companies that you do business with, which are the ones that have a good Deal Score as of today, so that conversations and sales processes could be prioritized and targeted for them
3. Pricing teams can determine which price points for their products work best, for e.g. If many customers are scored low then it could indicate challenges in Sales process or even higher price perception compared to competitors in the market
4. Leadership teams to assess which Markets are healthy by plotting the scores in regional graph and identify new opportunities to grow the business

These are a few direct use cases, however there may be more based on usage by teams within the company.

### **5. SCORE GENERATION**

#### **5.1. Internal Performance Score**

Ranking businesses based on internal to company performance will majorly involve three types of data sources:

1. Historical transactions data
2. Contact center data
3. Product purchase history

Above datasets need to be combined to generate a performance score using below framework:

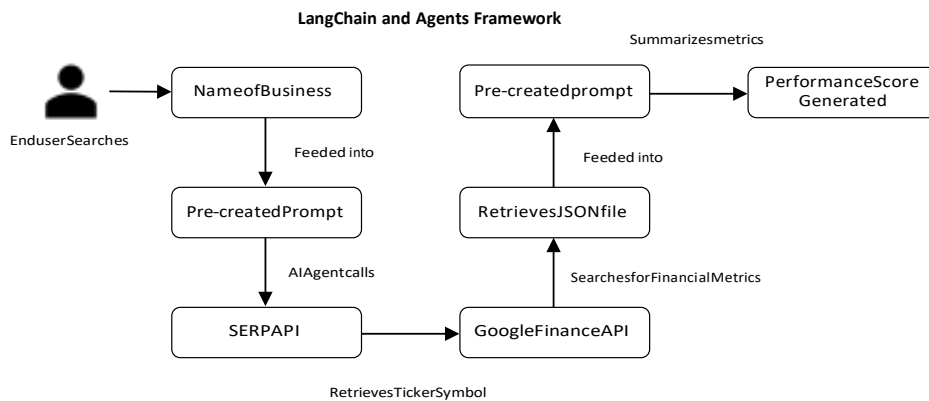
- Transactions Data can be used to generate RFM score (Recency Frequency Monetary). This includes looking at how recent has the businesses transacted with your company,

what is the frequency of these transactions, and what was the applicable monetary value for these transactions

- Customers reaching our to Contact Center for any kind of service or product related queries indicates their active engagement.
- Product penetration from the products that your company offers, or the increase in penetration indicates growth in engagement and good internal performance

## 5.2. Financial Health Score

This involves multiple steps which could be automated via LLMs, Agents and LangChain framework as below



The pre-created prompts form the LangChain where one output from a prompt is used as an input to another prompt. Scoring the Business will be done using Reasoning LLMs like GPT o1 which are capable of using Chain of Thoughts on the financial metric summary provided via the previous prompt and then generate a final score.

## 5.3. Market Sentiment Score

The framework is same as above, where instead of calling SERP API for company Ticker symbol and then Google Finance API for financial metrics, the agents will call API from news service providers. These news articles will then be summarized to pass onto the reasoning LLM which will then generate the final Sentiment score.

## 5.4. Consolidation and Data Backend Creation

Final step will be to feed all scores and summaries into backend tables from where users will be able to access the Scores and use them to track performance. Internal teams can be provided access to this data for their use cases.

## 6. COST IMPLICATIONS

- Usage of LLMs for designing the prompts and analyzing the output generated, using the data feed from finance vendors to generate health score, analyzing the sentiments from news articles would consume LLM infrastructure and add up to the cost
- Charges from API providers (Google Finance, News Vendors)
- Infrastructure costs for storing and maintaining the consolidated database will need to be added.

## 7. LIMITATIONS OF LLM BASED APPROACH

This approach works only for publicly listed businesses that are listed through any stock exchange and share their financial data with the market. Nonregistered business' financial data availability is a challenge and there are no reliable sources for trustworthy information.

Market Sentiment Scoring is a highly complex process where a lot of training and maintenance of models will be required. There are possibilities of bias in inferring sentiments from news articles, but we will need to train the models to look for signals that can cause financial distress to the Businesses, for e.g. High refund rates from customers, extremely poor product reviews, filing for insolvency proceedings, high tax implications etc.

Generating scores depends on API calls made to specific vendors. Because of this dependency, the scores would not be generated if API calls fail or there are data outages from the data providers

LLMs are subject to hallucinations. There may be cases where a LLM could incorrectly infer the financial growth of a Business by focusing more on certain parameters and ignoring others, or it may infer the growth incorrectly. Similar issues could happen from Market Sentiment scoring too. There are techniques to train the model against this bias, but these are not 100% efficient and hence a human in the loop is needed to ensure false positives are removed and algorithms are enhanced accordingly

## 8. OVERVIEW WITH EXAMPLE

Let's look at an example output from the model for latest quarter's data. There are 4 columns, first 3 for the three pillars of scoring we have discussed about and the last one for the Overall Score based on equation we discussed. The last two columns are an output from the LLM based scoring algorithm based on how these companies have performed (scores are directional and have been generated based on their financial and news data from Google Finance using LLM prompts). However, the first column is dummy data and depends on how well engaged these two businesses are with your company (let's suppose this is generated by the Internal Performance scoring methodology we discussed), let's say if Microsoft has a very good product penetration and engagement with your company and scores good on other parameters for internal scoring, its Internal Performance Score will be high (row 2):

	Internal Performance Score (depends on Company to Company)	Financial Performance Score Q4 2024	Market Sentiment Score Q4 2024	Overall Score
Microsoft	5	8	7	7.15
Microsoft	9	8	7	8.15
Intel	5	4	5	4.35
Intel	9	4	5	5.35

Now suppose Microsoft has not had a good engagement with your company that quarter (row 1), then the overall score will reflect that and score Microsoft lower, meaning even though Microsoft has a good health as a company, they would have less willingness to pay for your products given they had a decrease in engagement with your company in that Quarter.

Similarly, we see Intel has a low score on both Financial and Market Sentiment health given the challenges company is facing, and even if they have a good engagement with your company based on Internal Performance scoring, it will always be ranked lower than Microsoft since it is penalized for inconsistent market performance.

The same score for Businesses could be generated over continuous time (four quarters for e.g.) and we can assess the overall performance by comparing couple of data points together. This score exposes other aspects of the business that are critical to be looked at while negotiating a deal with them, rather than just relying only on the Internal Performance.

These scores could also be plotted across regional graphs to detect the hotspots for Sales and get leading insights into which regions could pose a challenge in near future if many Merchants seem unhealthy on external parameters.

## **9. COST EFFECTIVENESS OF ENHANCED DEAL NEGOTIATION**

There are several inputs to Manual Research for getting a directional deal score. Manually browsing through financial statements and recent news articles about the business, extracting important metrics and consolidating them into a meaningful analysis, inferring the details and getting approvals for the same etc. involves a lot of time and efforts. Deal negotiation is a long process and delays severely impact Business onboarding. There is also opportunity cost of Sales Representative Bandwidth that goes into research which could have been used for ongoing deal conversation with the Business otherwise. Deploying additional sales resources is always costly as it involves a lot of training and education.

The methodology discussed above takes away challenges in Time to Onboard as it reduces the cycle time for research and enables faster Revenue Generation from the Business for your company. This in turn reduces the dependency on Sales bandwidth and covers the costs for additional resources deployed for same research activity across Businesses. LLMs and the Generative AI domain is getting advanced day by day and new technologies further reduce the cost to implement language models into research activities, reducing manual work dependency.

Based on our extensive experimentation running the methodology discussed, it costs ~ \$ 0.02 per query (LLM used – GPT 3.5 Turbo), which includes the prompt, incoming Json file from Google finance, tokens generated for summary and score generation prompt. This would mean ~ \$20 for 1,000 Businesses searched and scores generated. Repeating this exercise for the same businesses for 4 quarters will require \$80 as costs from the LLM. This does not include costs from API providers or infrastructural costs to store and maintain the data, which depends on a lot of factors.

## **10. CONCLUSION**

WTP is a complex but critical metric which is often ignored due to the manually intensive processes and frameworks. This methodology enables Sales teams to quickly assess the WTP metric and make conclusive and efficient decisions. The costs are justified given the enhancement to overall process this would bring in and provide ability to look at multiple angles since we can track the changes in WTP at both micro and macro level.

The methodology proposed above automates accurate detection of WTP for any Business at a given point in time by looking at three major pillars of information – Internal to company performance, Financial Health performance and Market Sentiment performance. These three

parameters are consolidated together to generate a single score that reflects true assessment of WTP for that business. The process involves automation of external insights (finance and sentiment) integration to internal performance and does this through a weighted average equation discussed through the paper. This score could be used across multiple process to enhance existing logic of Business assessment for customized offerings and also help prioritize Sales teams in driving conversations with the Businesses.

## **AUTHORS**

**Chirag Soni** is an AI product Manager in Data and Insights team at PayPal and has worked as hands on Data Scientist in previous roles

**Swati Shah** is the head of Margin Optimization product team at PayPal and has varied experience across digital payments and products, and leading AI teams