BrokerChooser Homework Discussion

How to Pick and Prioritize Languages for Cost Optimization

I'd take a data-driven approach. We should consider the size of the market for each language, within that the percentage of people who don't speak English, and weigh that by analysis on possible adoption size and speed. We could also use other metrics like median income within the population that doesn't speak English – or even summed income of such a population controlled by how likely they are to use BrokerChooser (this itself could be modeled with Machine Learning).

At the end of the day, we need to consider which translations would bring the highest amount of profit, and I'd use a data- and model-driven approach to measure this using a combination of features and ML models.

Improving Translation Quality in the Future

I'm currently using OpenAI's models, which are among the best in the market nowadays. However, these are not specialized for machine translation.

Using a specialized model for multilingual machine translation might give us better scores, or at the very least, provide cheaper translations. Using specific models for non-Latin script languages like Japanese and Hindi might provide significantly better translations for those languages.

A custom embedder and language model trained on our own texts might also improve translation quality, as it will better capture the specialized concepts used in finance/brokerages.

I'd also implement language-specific embedders, because using an embedder pretrained on English text may not capture the nuances and meaning of other languages.

I'd also crowd-source corrections, so that users could provide custom translations or notify us if a translation doesn't make sense.

Regular retraining of these models and embedders with the new texts we produce plus the crowd-sourced corrections will probably produce the best possible translation quality.

Evaluation Report Insights

I evaluated the solution's translation against the translations provided in the CSV file. Translation scores usually measure machine translations against a human reference. I treated the supplied translations as the human reference. I also added a Google Translate column to the CSV, so that I can compare my solution's score against the Google Translate translation quality, too. Both my solutions and GT translations were measured against the human reference. The scores are as follows:

Metric	Model Translations	Google Translate
BLEU	58.65	34.6
CHRF	82.24	65.2
ROUGE-1 F1	71.59	56.67
ROUGE-2 F1	47.85	29.47
ROUGE-L F1	69.61	55.17
BERTScore F1	96.23	93.22

You can find a short description of the metrics in the README.md document.

What can be said is that my solution outperforms Google Translate, as it gives better scores for the translations altogether than GT, when compared to the human reference.

I'd highlight the BERTScore, as that's based on a vector representation of the text that better describes the closeness in *meaning* of the translations, as the other metrics are based around character/word/n-gram similarity but don't directly consider the meaning of each sentence. It shows the GPT-4o translations are very close in meaning to the human reference.

AI Product Ideas for BrokerChooser

Risk mitigation and mental health support

It's well known that some brokers specialize – based on their pricing structure and available functionality – in providing tools for speculative trading. This can easily turn into a gambling addiction.

While the user spending more money on the broker's platform creates value for the broker, it doesn't create value for BrokerChooser, and an unfortunate match might even make BrokerChooser's public perception worse ("BC recommended me this trader, now I lost everything!").

I propose to develop a model that looks for certain markers and categorizes potential users into risk categories based on their probable gambling addiction and does not recommend them brokers that specialize in speculative trading.

This will allow BrokerChooser to have a more socially responsible business while also protecting the end user's interests.

Here the risk is if the model's not good enough, it would impact our recommendation quality for users and they might not be matched with the best broker for their needs.

Regulatory Compliance Advisor

People either spend a lot of time or some money on making a valid tax report. While many government solutions may exist, those usually don't work for income unrelated to full-time employment.

I propose an AI-based assistant (built with RAG and agentic capabilities) that can advise users on how to stay tax compliant and how to fill out a valid tax report.

This can be a cheaper way for users to make investments and might be a paid service if we can price it right (not free but cheaper than a tax advisor).

The risk could be model inaccuracy, as even with RAG hallucinations won't become non-existent. We could mitigate this by also providing human oversight, e.g. a tax report that's pre-filled by AI based on knowledge about laws stored in a RAG system, and using actual tax advisors for oversight. Since the bulk of the work would be done, the tax advisors' time would need to be less than without the AI assistant, and would hence be more affordable.

Personalized Learning Pathways for Investors

We could create an AI-based learning assistant that specializes in understanding the user's knowledge, and responsively and interactively introduce them to more financial instruments.

Understanding these instruments is complicated, especially for a newbie, and could prevent them from using many instruments. The tool would also be able to help calculate the possible wins and losses, recommend safe strategies and explain potential downsides.

My hypothesis is that users of such a tool would be incentivized to use more instruments (and thus spend more money), and if this is true, we could "sell" these users as high spending potential users to brokerages.

The risk would be the quality of the recommendations, as we know LLMs can hallucinate. We absolutely must avoid making a tool that gives bad advice.