Baymax T&C

User Manual

Authors: Lasya Ednuri, Samantha Michael, Hiranya Vsishnavi Bandi, Uma Chavali

Github: https://github.com/samanthajmichael/baymax_tc/tree/main

About Baymax T&C

What is Baymax T&C?

Baymax T&C is an LLM application designed to simplify complex Terms of Service agreements using the power of Large Language Models (LLMs). It aims to help users understand the legalese in these often daunting documents by providing clear and concise explanations.

How does it work?

Baymax T&C works by combining a large repository of Terms and Conditions agreements with a powerful LLM. When you ask a question about a specific T&C, the application searches its document store for relevant information (RAG) . This information is then processed by the LLM, which translates the legal language into plain English and generates a concise and easy-to-understand explanation.

Who can use Baymax T&C?

Baymax T&C is designed for anyone who wants to better understand the Terms and Conditions of online services they use. This includes:

- ★ Anyone! Consumers who want to be informed about their online rights and responsibilities.
- ★ Do you know your employer? Employees who need to understand company policies or review contracts with other businesses.
- ★ Are you still learning? Students studying law or business who are learning about legal contracts and agreements.

Is Baymax T&C free to use?

Currently, Baymax T&C is free to use for all users. However, we may introduce different subscription tiers in the future to offer additional features and benefits.

How do I use Baymax T&C?

Using Baymax T&C is simple:

- ★ Enter the URL of the Terms of Service agreement you want to understand.
- ★ Ask your question in plain language. For example, "What are the rules about sharing personal information on this website?"
- ★ Baymax T&C will analyze the document and provide you with a clear and concise explanation.

What are the Features of Baymax T&C?

Baymax T&C offers several key features:

- ★ User-Friendly Interface: A simple and intuitive interface allows users to easily input queries and receive clear explanations.
- ★ Comprehensive Coverage: The application covers a wide range of Terms of Service agreements, from popular social media platforms to online marketplaces.
- ★ Accurate and Reliable: The LLM is trained on a vast dataset of legal documents, ensuring accurate and reliable explanations.
- ★ Customizable: Users can customize the level of detail in the explanations to suit their needs and understanding.

What Problems can Baymax T&C Address?

Baymax T&C addresses several problems associated with understanding Terms and Conditions agreements:

- ★ Legal jargon: Many T&Cs are written in complex and technical legal language that can be difficult for non-lawyers to understand.
- ★ Length and complexity: T&Cs can be very long and complex, making them time-consuming and difficult to read through entirely.
- ★ Lack of clarity: Often, T&Cs lack clear explanations of important terms and concepts.

What types of Terms of Service agreements does it cover?

Baymax T&C covers a wide range of Terms of Service agreements, from popular social media platforms like Facebook and Twitter to online marketplaces like Amazon and eBay. We are constantly expanding our collection to include more services and platforms. Currently in our database: (HireVue, Llama2, Mettler Toledo, Open AI, Pentair, Qdoba Mexican Eats, Bank of America, Ulta Beauty, Verizon, Truist)

Baymax T&C vs Our Competitors?

Baymax T&C differentiates itself by focusing on providing clear and concise explanations of legal jargon in a user-friendly interface. It leverages the power of LLMs for accurate and reliable information retrieval and processing, offering a more comprehensive and personalized experience compared to traditional methods of understanding T&Cs.

- Our competitor - TOSDR (Terms of Service; Didn't Read) offers a simplified version of T&Cs by highlighting key sections and providing brief summaries. While TOSDR is helpful for quickly scanning through lengthy documents, it lacks the depth of explanation that Baymax T&C provides.
- Key Differences:
    - TOSDR focuses on highlighting important sections and providing brief summaries, while Baymax T&C provides detailed explanations of legal jargon in plain language.
    - TOSDR relies on human-generated summaries, while Baymax T&C utilizes an LLM for more accurate and personalized explanations.
    - TOSDR is a web-based tool, while Baymax T&C can potentially offer a more integrated user experience through a dedicated app.

Is Baymax T&C accurate and reliable?

The LLM powering Baymax T&C is trained on a massive dataset of legal documents and is designed to understand and interpret complex legal language. This ensures that the

explanations provided are accurate and reliable. However, it's important to remember that Baymax T&C is not a substitute for legal advice.

Can I customize the level of detail in the explanations?

Yes, Baymax T&C allows you to customize the level of detail in the explanations provided. You can choose to receive a brief overview or a more detailed breakdown of the relevant information.

What if I can't find the specific Terms of Service I'm looking for?

If you can't find the specific Terms of Service you're looking for, please contact us through the website and we'll do our best to help you.

How do I provide feedback or report an issue?

You can provide feedback or report any issues through the "Contact Us" page on the Baymax T&C Github website. Your feedback is invaluable to help us improve the application.

Can I use Baymax T&C for my business?

While Baymax T&C is designed for individual users, it can be a valuable tool for businesses as well. For example, you can use it to help employees understand company policies or to review contracts with other businesses. However, it's essential to remember that Baymax T&C is not a substitute for legal advice and you should consult with a lawyer for any legal matters.

What are the limitations of Baymax T&C?

It's important to remember that Baymax T&C is a tool for understanding legal language, not a replacement for legal advice.

- It might not capture every nuance of legal language.
- The LLM is constantly learning and improving, but it may not always be perfect.
- The information provided by Baymax T&C should not be used as a substitute for professional legal advice.
- We are constantly working to improve Baymax T&C and address these limitations, but it's important to be aware of them when using the tool.

What are the future plans for Baymax T&C?

We are constantly working to improve Baymax T&C and add new features. Our future plans include:

★ Expanding our document store to cover even more services and platforms.
★ Adding support for different languages.
★ Integrating with other applications and services.
★ Developing new features to enhance the user experience.

FAQs For the Business

How will you measure the success of your product, Baymax T&C?

We will measure the success of Baymax T&C through various metrics:

- ★ User engagement: Number of active users, frequency of use, and time spent on the platform.
- ★ User satisfaction: Feedback from users regarding ease of use, accuracy of explanations, and overall experience.
- ★ Growth and adoption: Rate of new user acquisition and adoption across different target demographics.
- ★ Business impact: Number of users who report understanding T&Cs better after using Baymax T&C, leading to improved online safety and informed decision-making.

How will you measure model performance?

We will measure model performance using metrics from deepeval, which automatically assess the generated answers across three key dimensions:

- ★ **Correctness**: How factually accurate are the answers?
- ★ **Faithfulness:** How well do answers align with the information in the source document? This includes identifying potential "hallucinations" (information not present in the source).
- ★ **Contextual Relevancy:** How well do the answers address the specific question and its context within the source document?

These metrics provide us with scores, thresholds, strictness levels, and reasoning for each evaluation, allowing us to gauge the model's overall performance and identify areas for improvement.

What Technical Challenges can Baymax T&C face?

- ★ Maintaining accuracy: Ensuring the LLM remains up-to-date with changes in legal language and constantly evolving T&Cs.
- ★ Handling large amounts of data: Managing and processing the massive amount of data from different T&C documents.
- ★ Scalability: Ensuring the platform can handle a growing number of users and requests without compromising performance.
- ★ Security: Protecting user data and ensuring the platform is secure from malicious actors.

What Non-Technical Challenges can Baymax T&C face?

- ★ User trust and adoption: Gaining user trust and encouraging widespread adoption of the platform.
- ★ Competition: Competing with other platforms and services that provide similar functionalities.
- ★ Legal and ethical considerations: Ensuring compliance with legal and ethical standards in using and interpreting legal documents.
- ★ Maintaining neutrality: Avoiding biases and ensuring explanations are objective and unbiased.

In-scope features: What functionalities will you definitely implement?

- ★ Core summarization functionality: Providing concise and clear explanations of T&Cs based on user queries.
- ★ Document repository: Maintaining a comprehensive and up-to-date database of T&Cs from various platforms.
- ★ User-friendly interface: Designing a simple and intuitive interface for easy navigation and interaction.

★ Customizable explanations: Allowing users to choose the level of detail they require in explanations.
★ Feedback mechanisms: Implementing user feedback mechanisms to gather insights and continuously improve the platform.

Out-of-scope features: What ideas or functionalities have you decided not to pursue, and why?

★ Legal advice: Baymax T&C will not provide legal advice as it is not a legal professional.
★ Automated contract negotiation: This feature is complex and requires a deeper understanding of legal contracts and negotiation strategies.
★ Integration with other platforms: While integration is possible, it will be explored in future iterations after establishing the core functionalities.

Can Baymax T&C Ethically Go Wrong? How can we address them?

Potential ethical concerns with Baymax T&C include:

- Misinterpretation: The LLM could misinterpret legal language, leading to inaccurate explanations.
- Bias: The LLM might exhibit biases in its processing of legal documents, potentially affecting the objectivity of explanations.
- Privacy concerns: Handling and storing user data, including T&C documents, requires robust security measures to protect privacy.

We address these concerns by:

★ Continuously improving the LLM: Constantly training and refining the LLM to improve accuracy and reduce bias.

- ★ Transparency: Providing users with clear information about how the platform works and the limitations of the technology.
- ★ User feedback mechanisms: Implementing feedback mechanisms to identify and address potential issues related to accuracy and bias.
- ★ Data privacy: Implementing robust security measures to protect user data and comply with privacy regulations.

We believe that addressing these ethical considerations is crucial for the long-term success and responsible development of Baymax T&C.

User Testing Plan

- **Target User Group:**
    - Demographics: We will recruit a diverse group of testers reflecting our target audience, encompassing various age groups, backgrounds, and levels of technical proficiency.
    - Expertise Level: Our target group includes:
        - General users: Individuals with varying degrees of familiarity with online services and T&Cs, from casual users to tech-savvy individuals.
        - Students: Students studying law, business, or related fields who are learning about legal contracts.
        - Business professionals: Employees who regularly encounter and need to understand T&Cs in their work.
- **Recruitment Strategy:**
    - Social Media: We will leverage targeted advertising on social media platforms like Facebook, Twitter, and LinkedIn to reach potential testers within our desired demographics.
    - Online Communities: We will engage with relevant online communities, forums, and groups related to technology, law, and online services to identify and recruit testers.

    - Email Campaigns: We will create targeted email campaigns to reach individuals in our database and solicit their participation in user testing.
    - Personal Networks: We will utilize our personal networks to reach out to individuals who might be interested in testing and providing feedback on Baymax T&C.
- **Testing Methodology:**
    - Guided Tasks: We will provide testers with specific tasks, such as searching for specific information in a T&C document or asking questions about particular clauses, to evaluate the usability and effectiveness of the application.
    - Open Exploration: We will encourage testers to freely explore Baymax T&C, using it in their own way and providing feedback based on their experience.
- **Feedback Collection:**
    - Google Forms: We will use online surveys and questionnaires via Google Forms to gather structured feedback from testers, covering aspects like ease of use, clarity of explanations, and overall satisfaction.
    - In-Person Interviews: We will conduct individual interviews with a select group of testers to gather more detailed and qualitative feedback on their experience using Baymax T&C.
    - Screen Recordings: We will request testers to record their screens while interacting with Baymax T&C to identify potential usability issues and areas for improvement.
- **Metrics:**
    - User Experience:
        - System Usability Scale (SUS): A standardized survey to measure overall usability and satisfaction with the application.
        - Task Completion Rate: The percentage of testers successfully completing assigned tasks.

        - Time on Task: The time it takes testers to complete specific tasks, indicating efficiency.
- Application Effectiveness:
    - BERT, ROUGE, BLEU: We will use these metrics to evaluate the quality and accuracy of the LLM-generated explanations, comparing them to human-written summaries.
    - Likert Scale: We will use Likert scale questions to assess user satisfaction with the clarity and accuracy of explanations.
- Timeline:
    - Week 1: Recruitment and initial testing setup
    - Week 2-3: Conducting user tests with a diverse group of participants.
    - Week 4: Analyzing gathered data, identifying areas for improvement, and refining the application.
    - Week 5: Implementing feedback and re-testing with a smaller group to validate changes.
- Iteration:
    - This user testing process will be iterative, meaning we will conduct user testing throughout the development cycle to gather feedback and continuously improve Baymax T&C.

Github Workflow Enhancement

The process for the PR:

- We decided to implement development tools like Black, isort, and bandit, along with pre-commit, to ensure our commits were rigorously checked before being pushed to GitHub.
- We started by creating a new issue on our GitHub repository and branching off from the main branch.

- Then, we used pip to install Black, bandit, isort, and pre-commit, followed by running pre-commit install to activate the hooks.
- Next, we added the necessary configuration files to our local repository – .bandit, pyproject.toml, and .pre-commit-config.yaml – and updated our README file with instructions on how to use these tools.
- Once everything was in place, we staged the changes, committed them to our remote repository, and created a pull request to close out the issue. This workflow ensures that our code is consistently formatted, secure, and adheres to quality standards.
- GitHub Issue Link: https://github.com/samanthajmichael/baymax_tc/issues/20

Code Quality Tools

This project prioritizes maintaining high code quality through a suite of automated tools for formatting, style checking, and security analysis. We leverage Python 3.10.12 as our runtime environment and rely on the following tools for ensuring clean and secure code:

Code Formatting and Style:

- Black: This uncompromising code formatter enforces consistent style across the project, ensuring readability and maintainability. It enforces a line length limit of 88 characters.
- isort: This tool automatically organizes and formats import statements, ensuring consistent import ordering and making code easier to understand.

Security Analysis:

- Bandit: This security linter helps identify common security vulnerabilities in Python code, improving the overall security posture of the project.

These automated checks contribute to a robust and reliable development process, ensuring that our code meets the highest standards of quality and security.

Reflections

Challenges

Using Modal and successfully hosting our local LLaMA 3 model has remained a persistent obstacle. Despite our efforts, we haven't been able to overcome the technical hurdles associated with this integration. This has led us to temporarily utilize OpenAI's API to ensure we could move forward with testing and development.

RAG Process and its Impact on Application Performance:

Our journey with RAG has been a learning experience, filled with both successes and challenges. While we've made significant progress, there are areas we're actively working on to improve the performance and user experience of our application.

The Shift to OpenAI API: Initially, we encountered difficulties getting our local LLaMA 3 model running seamlessly on Modal. To ensure we could start testing and evaluating our RAG pipeline quickly, we decided to utilize OpenAI's API for this milestone. This strategic shift allowed us to focus on building a usable and functional application, enabling us to gather crucial feedback and refine our approach.

RAG Pipeline Progress: Our RAG pipeline has demonstrated promising results, particularly with the 10 cleaned files we've incorporated so far. However, we recognize that scalability and robustness are key, and we're continuously working on enhancing the testing process to ensure its effectiveness.

Optimizing Answer Length: A major hurdle we've encountered is finding the optimal balance in answer length. We've found that overly long answers can sometimes hinder performance, especially when evaluated against metrics like correctness, faithfulness, and contextual relevancy using deepeval. Conversely, overly short answers can also lead

to failing these metrics. We're actively refining our approach to identify the "sweet spot" that maximizes accuracy and relevance while remaining concise and impactful.

Looking Ahead

- Our next milestone is to implement the RAG pipeline using Modal and LLaMA 3, which will require further exploration and optimization. We're confident that by leveraging the learnings from our current iteration, we can create a more efficient, robust, and user-friendly system.
- To ensure the accuracy and robustness of our RAG pipeline, we plan to significantly improve our unit tests by incorporating a wider range of metrics, including ROUGE, BLEU, and hallucination scores from deepeval. This will enable us to more comprehensively assess the quality of generated answers, identify potential biases, and refine our model's performance in generating accurate and relevant responses.