



AIPC Final Presentation

Marketplace AI Assistant

05/15/2024

Thierry Bosnoyan

My Current Product: The Internet of Things Marketplace

A B2B IoT marketplace where users can digitally purchase solutions ranging from fleet management to M2M connectivity

PRODUCT VALUE PROP & FEATURES



Current Product Issues



The post purchase journey is complex with users having to log into up to 5 platforms in some instances (marketplace, partner website, billing, shipping tracking, console site for APN settings)



There is Limited educational content (Blog articles, use cases, white papers. Infographics)



There are multiple products for various industries and business sizes but **no recommendation engine**



Rudimentary Chatbot that redirects users to existing webpages or asks them to call a support number if it doesn't have the answer



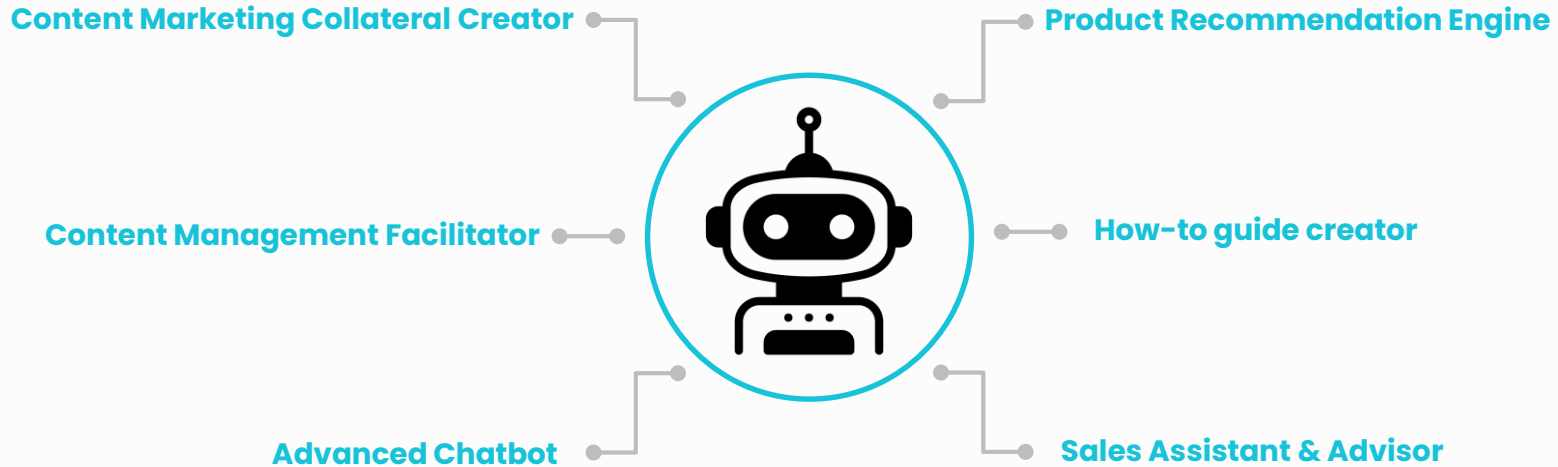
Technical resources are hosted on a separate website entirely with no link to it from the main product



FAQ content is not exhaustive and difficult to be kept up to date due to the constant addition of products

Product Opportunity

An IoT and product specific AI assistant that can help fill our existing content & traditional AI gaps



AI Product Strategy Opportunity: PRD Abstract

AI PRODUCT FEATURE OPPORTUNITY

The marketplace AI assistant is designed to help users across every point of their journey whether it is by educating them on the importance of innovating through IoT, recommending the right products for their business, supporting their purchase journey, or guiding them through post purchase installation



How can I grow my business through the implementation of "Internet of Things"

What are some use IoT use cases for the food & beverage industry?

How do I know which product is right for my business?

Where can I change my APN settings?

How much will my monthly bill be if I order 160 m2m SIMs at the cheapest rate plan?

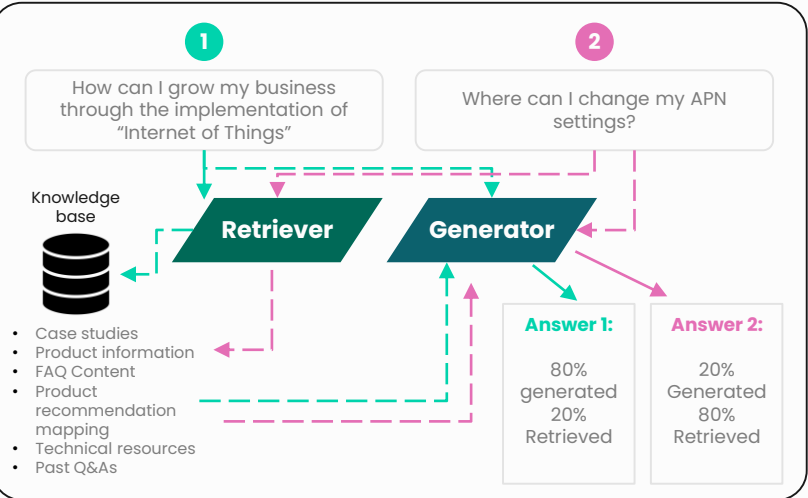
How do I find out which APNs/DNNs my device can use?

How can I help you today?



HOW IT WORKS

An AI Chatbot leveraging RAG and trained on product specific documentation to help answer all of a users complex questions from education to technical implementation



Product Feature KPIs

Product Feature Metrics we want to measure upon launch

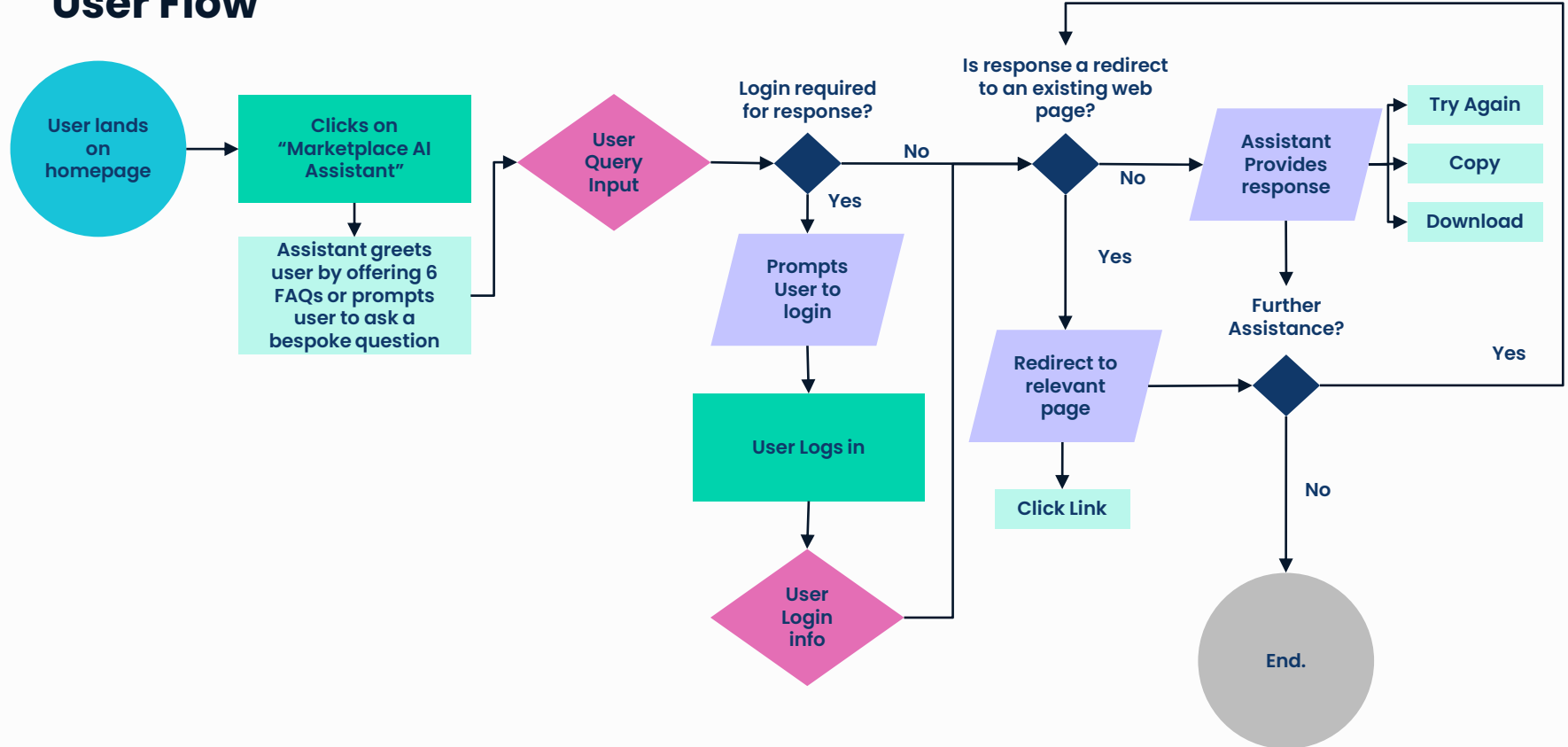
GOAL	METRIC	QUESTION
🕒 More than 50% of our web visitors use the AI Assistant	% of visitors using the AI assistant	How many website users also use the AI assistant?
🕒 The # of unique users engaging with AI assistant is increasing	# of unique users who engage with the AI Assistant (Daily, Weekly, Monthly)	Are unique users increasing or decreasing with time? What are some reasons behind that?
🕒 The majority of visitors find the AI assistant tool useful	% of visitors who find the AI assistant tool useful	Do users find the AI assistant useful? Do users enjoy the new feature? What are some ways we measure that?
🕒 Users are engaged and use the tool frequently	# of queries a user inputs in a single session	How many queries does a user typically input during a session, and what does this indicate about engagement?
🕒 Be able to monitor average session time to ensure it is on par or higher than benchmarks	Average session time	What is the average session time with the AI assistant, how can we improve it?
🕒 Be able to manage and optimize the cost of running the AI assistant	Feature cost per month (Average tokens cost per user)	What is the average cost per user for running the AI assistant, and how can we optimize it?
🕒 To increase the conversion rate of users interacting with the AI assistant	% of users using the feature who end up creating an account	How many users who interact with the AI assistant create an account, and what can we do to increase this percentage?

Technical Performance KPIs

Technical Performance metrics we want to measure upon launch

GOAL	METRIC	QUESTION
🕒 AI Assistant responds to the user in less than 10 seconds	Time taken for the AI to respond to user queries	What is the average response time for the AI assistant, and how can we reduce it?
🕒 100% of relevant queries are resolved by the assistant. Non relevant queries are responded to with the appropriate message defined by the team	% of user queries successfully resolved by the AI assistant	What percentage of user queries are successfully resolved by the AI assistant, and how can we improve this rate?
🕒 100% system uptime for the AI assistant	% of time the AI assistant is available and operational	What % of the time is the AI assistant available and operational? Are there backup GPT models we need to partner with?
🕒 The AI Assistant understands queries 100% of the time even if it is not authorized to respond to them	The percentage of interactions where the AI assistant fails to understand or process the query correctly	What percentage of interactions does the assistant fail to understand, and how can we reduce this rate?
🕒 The AI assistant never responds with incorrect or nonsensical responses	The percentage of incorrect or nonsensical responses generated	What percentage of interactions result in erroneous responses, and how can we reduce this rate?

User Flow



Model & Data Requirements

SPECIFICATION	REQUIREMENT	RATIONALE
Open vs Closed Source	Closed Source	Easier and quicker to implement, lower upfront costs. Given our tight budget, need for a tight turn around and lack of dedicated AI specialists closed source makes sense. Also our organization has strict security and compliance protocols and a current partnership with open AI.
Context Window	At least 8K+	For our AI model to add value to our users it needs to be able to handle long conversatios, complex interactions involving multiple steps, memory retention, and the ability to retrieve and update information from our backend systems.
Modalities	Text, Image, Docs	The model must incorporate an intuitive text interface, have robust NLP capabilities, implement features such as autocorrect and contextual understanding, have session tracking, support english and french, have image recognition capabilities to allow users to upload JPGs and files, it must be able to provide visual aids, such as diagrams and step by step guides.
Fine-Tuning Capability	Not-Required	At the start we will only be using RAG we might want to fine tune the model with time but in the interim we could achieve the desired outcome with RAG and prompt tuning
Latency	High Priority	Speed is critical to the success of the feature, our users will not engage with the assisstant if latency is high
Size	Up to 1.7T	We will most likely opt for GPT-4 which comes in parameter 1.7 T. We are open to similar models as long as they are able to have broad knowledge of the IoT space and the additional knowledge base we supplement it with

Prompt Requirements

“You are an Internet of Things AI Assistant. You are an expert on all things Internet of Things, particularly the products sold on the marketplace platform [Marketplace URL]. Your users are primarily businesses looking to learn about or purchase IoT solutions. You assist users by making their shopping experience easier. You support them at every stage of their Journey helping them make the right choices. This could include educating them on how IoT can improve their business, providing them with use cases, recommending the right products, or guiding them step-by-step through the entire website experience. When responding to user queries, provide accurate, balanced, and comprehensive information by prioritizing the marketplace specific information provided to you. You can fill gaps in your knowledge when it comes to educational content but in cases of uncertainty or sensitive topics, guide users to contact support by redirecting them to the support page.”

Additional requirements

- Inputs and outputs should be in text, jpg, and other visual elements such as diagrams
- Prompt should specify prohibited content and refuse to generate request that are against our policies
- Temperature must be on the conservative side as most requests should be answered from the provided knowledge base
- Top P/K must be in the conservative side as creativity should not be 0 but only be leveraged when it comes to the open ended educational questions
- Output must be accurate 100% of the time
- Ability to understand stop sequences

Human Evaluation

- 1 Is the generated content accurate? Are there objectively false statements included in the responses?
- 2 Does the content include any PII data or confidential data which was earmarked by the team as “do not include”
- 3 Is the generated content high enough quality that the end user would benefit from understanding it and taking action based on it?
- 4 Are questions that are specific to the marketplace and its product answered by only using data from the knowledge base?
- 5 When the model is required to generate visual content, does it generate content that meets the brand guideline (color palette, formatting, font)?

Risks & Mitigations

RISKS

- **Data Privacy & Security:** Unauthorized access to sensitive user data and/or relaying sensitive information back to users
- **Accuracy and Reliability:** The AI assistant may provide incorrect or incomplete information
- **Integration with Existing Systems:** Integration with current systems and platforms may be complex and prone to technical issues
- **Legal and Ethical Considerations:** The Legal team is hesitant when it comes to the implementation of AI. They fear Potential legal issues related to AI decisions and ethical concerns regarding AI behavior



MITIGATIONS

- Only feed the knowledge base vetted compliant data, ensure compliance with data protection regulations, conduct regular security audits & vulnerability assessments.
- Continuously train the AI model with a diverse set of queries and scenarios, implement a feedback loop for users to report issues, and regularly update the AI with new information and improvements
- Conduct thorough testing in a controlled environment before full deployment, and ensure comprehensive documentation and support for troubleshooting integration issues. Collaborate closely with IT teams to align integration efforts
- My company already has a partnership with open AI we can leverage the customized model that exists. We can engage with legal early on to ensure compliance with their policies.