Travel Copilot: A Generative AI Solution for Gen Z Itinerary Planning

1. Product Requirements Document (PRD)

1.1 The Opportunity: Market and Problem Validation

Area	Current Market Context (Based on 2024-2025 Data)	Problem Statement
Target User	Gen Z (18-27): Takes 3-5 leisure trips annually, prioritizing international travel (up to 79%), authenticity, and sustainability. They are digital natives who use social media (90%) for inspiration.	Gen Z travelers are overwhelmed by the sheer volume of fragmented information (blogs, TikTok, booking sites) and struggle to synthesize it into a personalized, logistically sound, cost-effective, and unique 3-day itinerary.
AI Trend	Adoption is Growing: Nearly half of Gen Z either use or plan to use AI for travel planning. They use it for inspiration but are often disappointed by the generic nature or factual errors of current LLM outputs.	Existing AI tools (generic LLMs) fail to meet the high standards of low latency (≤1.5s) and high accuracy (≤2% Hallucination Rate) required for a reliable travel companion.
Competition	The Travel Planner App Market is projected to grow at a CAGR of over 11%. Major competitors (Google Trips, Skyscanner) are integrating AI, but often focus on booking or general search, lacking the hyper-focus on Gen Z's need	The market lacks a highly specialized, fine-tuned AI solution that can deliver unique, non-mainstream itineraries that are inherently cost-effective and socially relevant.

	for authentic, budget-aware, 3-day micro-cations.	

Proposed Solution: Build a generative AI-powered "Travel Copilot" mobile application that leverages a fine-tuned foundation model to instantly create personalized, 3-day itineraries within a specified budget and interest profile (Adventure, Local Culture, Sustainability) with low latency.

2. Goals, Metrics, and Guardrails (Define Success)

This section directly outlines how the project's Goal is measured and maintained.

2.1 Success Metrics

Metric	Target	Rationale
Side-by-Side (SxS) Win Rate	≥70%	The primary measure of product quality and user delight. This proves the AI-generated itinerary is superior to a baseline (e.g., a manually planned itinerary or one from a generic LLM), justifying the fine-tuning investment.
Hallucination Rate	≤2%	The critical trust metric. This low threshold minimizes the risk of factual errors (wrong addresses, closed sites), which destroy the user experience and violate the safety guardrail.
Latency (Time to First Token)	≤1.5 seconds	The core UX metric. Gen Z users demand speed. This low-latency target ensures the planning process feels instant and effortless, addressing the "overwhelm" problem.

Retention	Track User Engagement (Goal: 20% returning within 60 days)	Measures product stickiness. A high retention rate confirms the itineraries are valuable enough for users to rely on for future trips.

2.2 Guardrails (Responsible AI & Safety)

1. Safety: The model will undergo pre- and post-generation filtering to prevent outputting information related to:
    - Dangerous or closed sites (e.g., restricted military zones, abandoned buildings).
    - Illegal activities (e.g., drug tourism, prohibited activities in the target country).
    - Unsafe logistical advice (e.g., directions through high-crime areas late at night).
2. Ethical Data Use: The curated dataset must be strictly audited for bias to ensure recommendations are not skewed towards specific demographics, income levels, or popular tourist sites. The goal is to promote cultural diversity, local businesses, and inclusive travel options.

3. Technical Decisions and Improvement

3.1 Data Strategy & Tokenization

- Data Sources:
    - POI/APIs: Real-time, factual data for accuracy and logistics (addresses, hours, transit links).
    - Niche Travel Blogs/Forums: Qualitative data for authentic experiences, local culture, and unique perspectives that resonate with Gen Z's preference for off-the-beaten-path travel.
    - User Feedback Loop: Data generated from corrected 'Thumbs-Down' itineraries (crucial for model retraining).
- Data Cleaning: Rigorous removal of spam, outdated information, and known biased content. Data must be geo-temporal tagged to ensure the model knows if a suggestion is a winter activity or a year-round landmark.
- Tokenization: Select a tokenizer (e.g., specialized SentencePiece or a custom travel-vocabulary extension) that is highly efficient at handling long-tail travel entities and budget terms ("50K," "hostel," "bleisure"). This improves both accuracy and latency.

3.2 Technical Choice

Technical Choice: LoRA (Low-Rank Adaptation)

Justification:

LoRA is the ideal Parameter-Efficient Fine-Tuning (PEFT) technique for this project due to its focus and efficiency:

1. Domain Focus: A general LLM is a great language model, but it is a poor travel agent. LoRA allows us to efficiently inject a specialized travel "skillset" into the model, training it to master the nuances of itinerary generation, budget calculation, and logical sequencing that a user needs. This is what drives the SxS Win Rate.
2. Agility and Iteration: Given the fast-changing nature of Gen Z trends and travel logistics, the model must be easily updated. LoRA reduces the training time and required GPU resources by up to 70%, allowing the team to quickly retrain the model with fresh data from the User Feedback Loop every few weeks, ensuring continuous relevance and improvement of the Hallucination Rate.
3. Low Latency Requirement: By only updating and serving small LoRA adapter weights, the inference process (generating the itinerary) is significantly faster than using a fully fine-tuned model, directly supporting the non-negotiable ≤1.5s Latency target.

3.3 User Feedback Loop

- Implementation: A simple, persistent Thumbs-Up/Thumbs-Down widget on every generated itinerary.
- Purpose: The feedback, along with the user's prompt, is captured. Any 'Thumbs-Down' events are prioritized for human review to identify the root cause (factual error/hallucination, poor personalization, or logistical flaw). Corrected, high-quality examples are then added to the custom training data for subsequent LoRA fine-tuning rounds. This is the mechanism for continuous model improvement and sustained high performance.

4. Prototyping & UX

The user experience prioritizes minimal design, speed, and shareability, aligning with Gen Z expectations.

4.1 Workflow: User Journey Map (Initial Query to Final Itinerary)

We will use Excalidraw to map this fast, four-step flow:

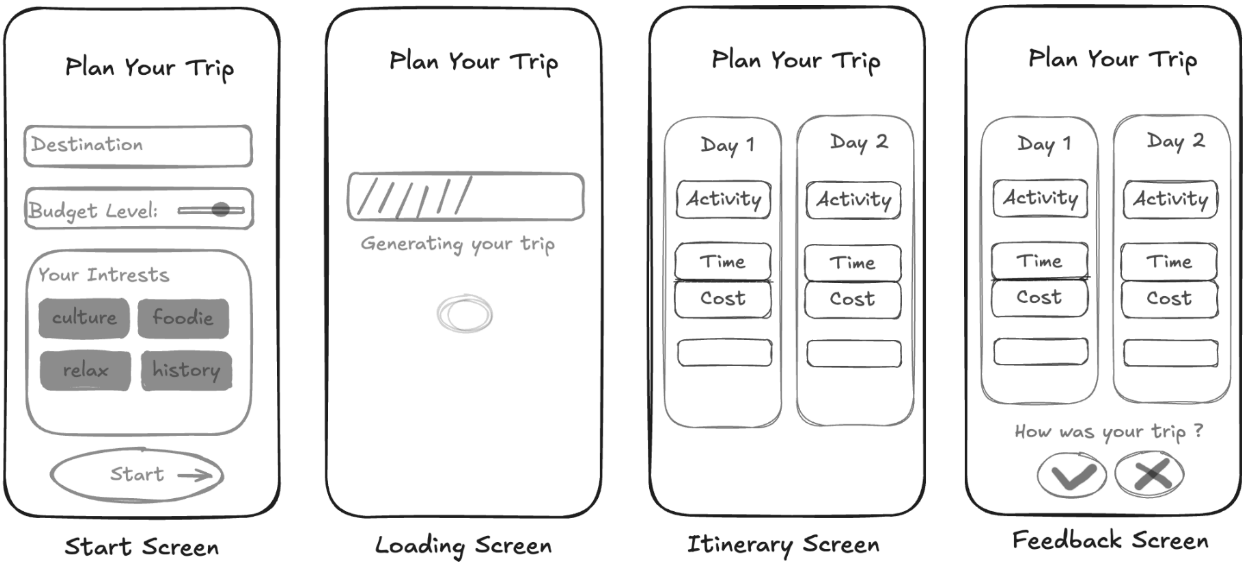
1. Input: User opens app → Quick Chat/Form (Destination + 3 Interests: e.g., "Culture," "Budget," "Hiking" + Budget Level: $/$$/$$).
2. Generation: User taps 'Generate' → AI runs fine-tuned model → Low Latency (Screen shows a fun travel fact/progress bar for ≤1.5 seconds).
3. Review (Core Itinerary Screen): The 3-day itinerary is displayed (Day 1, Day 2, Day 3).
4. Action/Feedback: User interacts with Map/Budget views or submits Thumbs-Up/Thumbs-Down.

4.2 Mockups (High-Fidelity Clickable Prototype)

Using a tool like Llizard or Figma, we will create three key screens that meet the UX requirements:

1. Itinerary Generation Screen (The Core Value):
   - Design: Clean, mobile-first, and highly scannable, using clear iconography for activity type.
   - Content Detail: Each activity block must include: POI Name, a Social Media-Ready Description, Estimated Cost ($), Estimated Time, and Travel Time to Next Point (critical for logistical trust).
   - Action: Persistent Thumbs-Up/Thumbs-Down widget for instant feedback.
2. Map View Screen (The Logistics Check):
   - Function: Visualizes all 3 days on a single map. Uses color-coding per day (e.g., Blue pins for Day 1, Green for Day 2).
   - AI Integration: Demonstrates how the AI has optimized the route to be geographically efficient, minimizing transit time (a key Gen Z need).
3. Budget Breakdown Screen (The Constraint Check):
   - Function: Presents a visual breakdown (bar or pie chart) of the specified budget vs. the itinerary's estimated cost.
   - AI Feature: Includes a visible "Cost Cutter" button powered by the AI that, when tapped, instantly suggests cheaper, but equally relevant, alternatives for high-cost items (e.g., switching a mid-range restaurant for a top-rated local food stall).

Low fi Wireframe



High fi Wireframe