SkillSprint Design Document

Project Title: SkillSprint

Team Members:

- Arshia Bakhshayesh
- Maher Athar Ilyas

Mentors:

- Mughees Asif
- Julia Gurel

Date: 18.08.2024

1. Introduction

1.1. Problem Statement:

Current online learning platforms are largely human-curated, leading to limitations in content variety and personalization. This often results in a one-size-fits-all approach that doesn't cater to individual learning goals or preferences. Users struggle to find relevant, high-quality resources that align with their specific needs, making it difficult to effectively acquire new skills.

1.2. Project Objective:

SkillSprint overcomes these limitations by using AI to create personalized learning roadmaps, continuously integrating the latest and most relevant educational resources. By leveraging AI models and web scraping techniques, SkillSprint ensures that users have access to an infinite variety of learning possibilities tailored precisely to their goals and skill levels. This approach not only provides high-quality resources but also guarantees that the content remains current and comprehensive.

1.3. Description:

SkillSprint is a personalized learning platform designed to help users acquire new skills efficiently. Users specify the skills they want to learn, and the platform generates a customized learning roadmap. This roadmap includes daily quizzes and small knowledge segments to maintain engagement. The platform uses web scraping to gather high-quality resources, such as videos and courses from reputable websites, which are then integrated into quizzes. This process supports an active learning methodology.

SkillSprint incorporates gamification, allowing users to compete with friends to enhance motivation. The platform also provides daily notifications and follows a systematic approach to help users develop consistent learning habits. The aim is to offer a user-friendly and effective tool for continuous skill development.

1.4. Target Audience:

- **Students and Lifelong Learners**: Individuals enhancing their education or pursuing lifelong learning.
- **Professionals**: Those advancing their careers through skill acquisition in new technologies and practices.
- **Hobbyists and Creative Enthusiasts**: People exploring or improving skills in creative fields.
- **Entrepreneurs and Business Owners**: Entrepreneurs learning skills to manage and grow their businesses.
- **Corporations and Organizations**: Enterprises training employees to adapt to market changes and improve productivity.
- **Academic Institutions**: Schools and universities enhancing curriculum with online resources.

2. Key Features

2.1 Completed Features

2.1.1 Personalized Roadmap Generation

- **Use Case:**
    1. The user inputs their learning goals, including the skill they want to acquire, their current skill level, desired skill level, estimated learning duration, and daily time commitment.
    2. The user requests the generation of a personalized learning roadmap.
- **Workflow:**
    1. The user fills out a form with the required roadmap information.
    2. The system processes the input and generates a personalized roadmap, displaying it to the user.
    3. The user can view the overall roadmap structure, including phases and topics.
- **Process:**
    1. The system utilizes a large language model (LLM) to analyze the user's input and generate a personalized learning roadmap.

2. A roadmap skeleton is created, outlining the phases and topics based on the skillName and other user inputs.
3. For each topic, the LLM generates specific content, including infobits (small chunks of information) and quizzes.
4. The system also uses web scraping to gather relevant resources, such as articles and videos, from reputable educational websites based on search terms derived from the LLM.

2.1.2 Interactive Learning Experience

Use Case:

- The user interacts with the roadmap by exploring phases and topics, reading infobits, taking quizzes, and tracking their learning progress.
- The user checks quiz results to see if their answers are correct or incorrect, and can redo similar quizzes if needed.
- The user accesses additional learning resources, such as websites and videos, for each lesson or topic.
- User can see their progress for each roadmap.

Workflow:

1. The user selects a phase or topic from the roadmap to start learning.
2. The user reads infobits associated with the selected topic.
3. The user takes quizzes to reinforce their understanding.
4. The system provides immediate feedback on quiz results.
5. If the user answers too many questions incorrectly, they must retake the quiz before they can progress to the next lesson.
6. Once the user passes the quiz, their progress is updated, and they can advance to the next topic.
7. The user can access web resources related to each lesson or topic.
8. The system continuously tracks the user's progress, showing completed lessons and remaining tasks on the roadmap interface.

Process:

- Quizzes are automatically generated based on the infobits for each topic, ensuring they are directly relevant to the content.
- The system monitors quiz performance and enforces a passing threshold. If the user fails to meet this threshold, they must retake the quiz to demonstrate sufficient understanding before moving on.
- As the user completes lessons, the system updates their progress in real-time, visually representing it on the roadmap.
- The system retrieves and displays web resources, including websites and videos, by performing targeted web scraping using search topics generated by the LLM.
- Quiz feedback is provided instantly, allowing the user to assess their understanding, take corrective actions if necessary, and track their progress throughout their learning journey.

2.1.3 Resume and Background Information Upload

Use Case:

- The user uploads their resume and background information, such as a list of current skills, in PDF format.
- This information is used to tailor the personalized learning roadmap more accurately to their existing knowledge and experience.

Workflow:

1. The user uploads their resume and background information as a PDF file.
2. The system processes the uploaded document to extract relevant information, such as current skills, work experience, and education.
3. The extracted data is used to refine the personalized roadmap, ensuring it aligns with the user's existing skillset and learning goals.
4. The user then proceeds with the roadmap generation, incorporating the uploaded information.

Process:

- The system parses the uploaded PDF to extract key details about the user's skillset and background.
- The LLM utilizes this extracted information to adjust the roadmap, modifying the phases and topics based on the user's existing skills and experience level. This ensures that the new learning roadmap is not redundant and builds on the user's previous knowledge.
- The infobits within each topic are mostly personalized, allowing users to better relate to the content and contextualize new information based on their prior experience.
- The system then proceeds with the standard roadmap generation process, incorporating customized content, infobits, quizzes, and relevant web resources.

2.1.4 Base Roadmap Selection

Use Case:

The user selects a pre-made roadmap from a collection of base roadmaps, either provided by the system or shared by other users, and personalizes it to their background.

Workflow:

1. The user browses available base roadmaps based on different skills, learning paths, or recommendations.
2. The user selects a base roadmap they would like to follow.
3. The user has uploaded their background information (e.g., resume, skills) to the profile section for further personalization.
4. The system personalizes the selected roadmap by adjusting phases, topics, infobits, and quizzes based on the user's background.

Process:

- Base roadmaps are predefined by the system or created by users who share their personalized roadmaps with the community.

- Once the user selects a roadmap, the system analyzes the user's background and adjusts the content to avoid redundancies and to align with their skill level.
- The LLM updates the roadmap to ensure personalized learning experiences while maintaining the structure of the chosen base roadmap.

2.1 Features nearing completion

Use Case: The system would scrape books and video transcripts, create embedding summarizations based on the scraped content for the skills users provide, and categorize and save them for future use. These embedding summarizations would then be utilized in roadmap generation to provide a more comprehensive and well-informed learning experience tailored to the user's skills.

Process:

- Scraping books and video transcripts relevant to the skills that users input.
- Creating embedding summarizations (turning the content into vectorized representations that capture the meaning and context) based on specific chapters or sections of books.
- Categorizing and saving these embeddings into a knowledge base for future use.
- When generating a roadmap for a skill, the system would pull relevant embeddings from the knowledge base, aligning the roadmap with expert content from books and videos.

Incompleteness Reason: This feature wasn't fully implemented due to the following roadblocks:

- Book formats were inconsistent and not reliable for content extraction.
- Scraping blocking mechanisms from various websites limited the ability to gather content.
- Attempts to circumvent the scraping blocks using multiple Lambdas with different IDs were only partially successful and did not work reliably in this case.

3. Low-Level Design

3.1 AWS Architecture Diagram

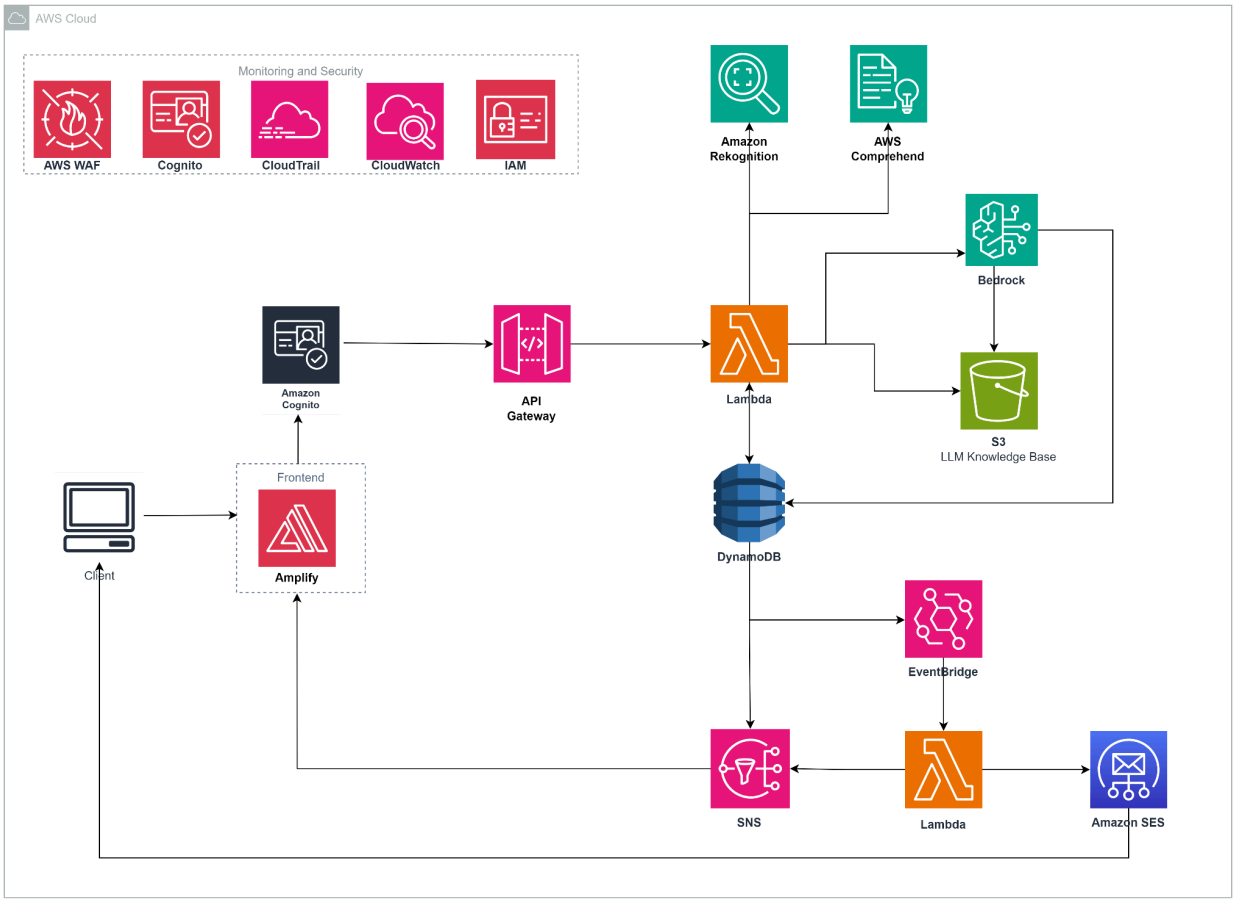

Figure 1. Low level AWS Architecture Diagram

3.2 Learning Path Generation Activity Diagram

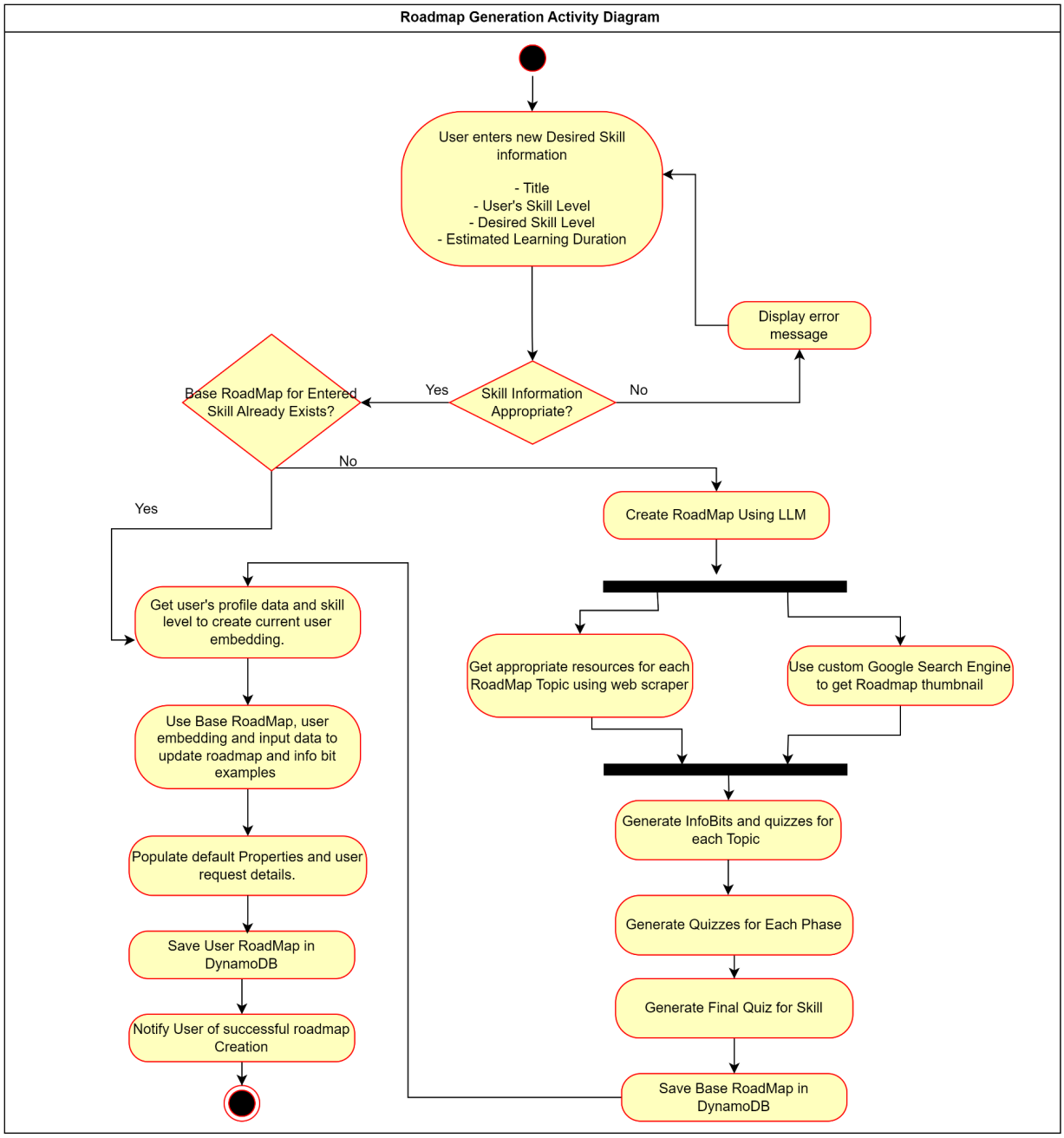

Figure 2. Learning Path Generation Activity Diagram

3.3 WebScrapping for Resources Activity Diagram

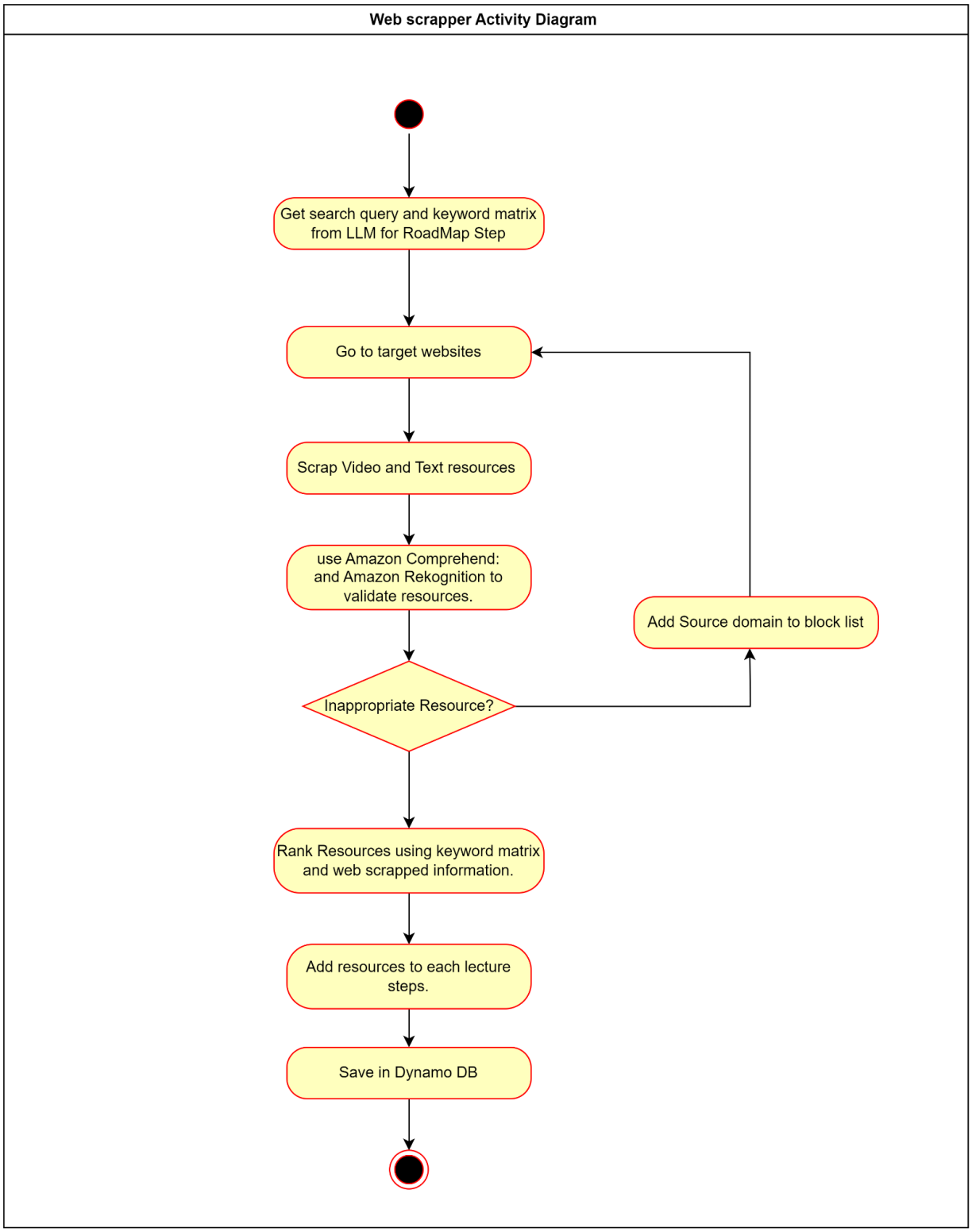

Figure 3.WebScrapping for Resources Activity Diagram

4. Technologies

4.1. Programming Languages

- **Python:** We chose Python for our backend because of its familiarity, flexibility, and simplicity, allowing faster development and debugging. Python is also well-supported in the Amplify backend environment, making it an ideal fit for deployment. Additionally, it offers a rich ecosystem of libraries for tasks such as web scraping and GenAI, making it a comprehensive language for our project.

- **TypeScript:** Enhances JavaScript by adding type safety, leading to more robust code. TypeScript's integration with modern development tools, especially in frameworks like Next.js, improves debugging and maintenance capabilities, making it ideal for building scalable applications. We use it together with Eslint to further enhance code quality.

- **Next.ts:** A React-based framework that provides functionalities like server-side rendering and generating static websites, which are crucial for SEO and optimal performance. Next.ts offers a streamlined development experience with automatic routing and optimized prefetching.

4.2. Frameworks and Libraries

- **Libgen-Scraper:** For scraping books and gathering educational resources from Libgen which later would be used for generating a knowledge base.
- **PyPDF2, pytesseract, pdf2image:** For processing and extracting text from PDFs, including resumes for the personalized roadmap and books for generating a knowledge base.
- **Boto3:** Used for interacting with AWS Bedrock and DynamoDB, enabling user data and roadmap data management and LLM integration in the application.
- **Parsel:** for parsing HTML/XML during web scraping from Google books and good reads.
- **Requests**, **JSON**, **RE:** for handling web requests, data manipulation, and regular expressions during scraping for knowledge base and receiving data from the frontend.

- **youtubesearchpython:** To search for YouTube videos and access their transcripts which later would be used for the knowledge base.
- **youtube_transcript_api:** For retrieving and formatting YouTube video transcripts, even managing issues like disabled transcripts.

- **React and Material-UI**: Used for most frontend components; Material-UI also manages dark and light modes, ensuring a uniform, themed, and responsive UI.
- **Microlink**: Used for previewing resources and taking HTML component screenshots for image sharing on social media.
- **dom-to-image**: Converts DOM elements to images.
- **React-apexcharts**: Displays charts and data.

4.3. Tools and Services

- **AWS Amplify**
  Manages backend resources, supporting deployment and scaling.
- **AWS Lambda**
  Executes backend functions dynamically in a serverless environment. All of our main flow for database operations, roadmap generation, scraping, and knowledge base generation take place in lambdas.
- **API Gateway**
  Routes API calls between frontend and backend, handling interactions like roadmap generation and quizzes.
- **AWS Cognito and Cognito User Pools**: Manages all authentication processes.
- **AWS Amplify Auth UI**: Manages UI components integrated with Cognito.
- **ESLint**: Maintains coding standards.
- **Prettier**: Ensures code reliability and formatting consistency.

3. Market Research:

3.1. Competitors in the Market:

- **Udemy**: Offers a wide range of topics with an extensive library and self-paced courses.

- **Coursera**: Collaborates with global universities and organizations, offering courses with optional certificates and degrees.
- **Skillshare**: Provides project-based courses in creative fields, fostering hands-on learning in a community setting.
- **Brilliant**: Focuses on interactive STEM education, emphasizing problem-solving and active learning.
- **Khan Academy**: Delivers free courses in math, science, and programming and is known for its comprehensive and accessible content.

3.2. Competitive Advantage:

- **Personalized Learning Paths**
  SkillSprint generates customized learning paths tailored to individual goals and skill levels, using skill lists and resume information to enable adaptive learning that surpasses the static offerings of competitors.
- **Microlearning for All Subjects**
  We use microlearning to break down complex concepts into bite-sized lessons, providing flexibility that enhances user retention, unlike the rigid traditional models offered by the majority of online learning platforms.
- **Unlimited, Up-to-Date Resources**
  Continuous integration of the latest resources, including up-to-date videos and textual content, ensures our material remains current and comprehensive, setting us apart from platforms limited by outdated materials.
- **Innovative Gamification and Habit Formation**
  Our unique approach to gamification, featuring progress tracking and structured timelines, promotes habit formation through daily engagement, providing a motivational learning experience that competitors lack.
- **Social Motivation**
  By incorporating social challenges and competitions, SkillSprint enhances user motivation and engagement, offering a community aspect rarely matched by others. Users are encouraged to share their milestones on social media, further amplifying the learning experience.