Comic book AI application first thoughts/dump

1. **Problem statement**
    a. **Vision** - Multimodal AI application to generate unique comic books that we know of but even more so with auditory stimulus.
    3 main modules:
        1. image(udit-cv) (good opportunity for getting up to speed with the domain.
        2. music(who tf doesn't like music): Now we get to know we can create royalty free original music with one click of a button.
        3. Narration: We can have Amitabh Bacchan saying “gaand me namak waali hari mirch”, we can have anything essentially.

    b. **Goal** (create unique, your own stories from scratch. Endless possibilities, very interesting).
        - Explore and push boundaries of existingGANs and LLMs.
        - We are not trying to reinvent the wheel but rather ‘put the pieces of a scattered puzzle together to create something new’.
        - To have a compound increase in overall effectiveness/throughput of individual generators when put together. Current LLMs are basically useless on their own.

    c. Who is this for?
        - kids‘ entertainment and learning too
        - Fanfiction
        - Tool for graphic designers and/or comic book creators who might be suffering from creative block.

    d. Why are we doing this?
        - We (Udit and I) feel like we are lagging behind the current innovations in AI technology.
        - People have saturated the generative AI market with thin LLM wrappers that are posed as ‘GenAI’ and something that is

valuable. We would wanna redefine that trend by showcasing that if correctly done, you can leverage existing tech pretty well.

- Generative AI is the new boom and we want to leverage it ourselves.

- **WHY COMIC BOOKS? WHY NOT SOMETHING ELSE:**
    1. Resonates with us cause we are nerds and products of the superhero content that we have consumed in life. Eg - udit listening to highway to hell to get the tony stark dawg in him.
    2. People have stopped going out and purchasing comic books. This rediscovers the beauty of comics and their power of storytelling.
    3. This is more or less the one of the few segments left which is not saturated by every other AI enthusiast. Opportunity to stand out.
    4. Will help Ipad kids not be Ipad kids. Ipad is a metaphor here for inhibition of actual creativity which kids have the most of.
    5. The idea of a multimodal application to essentially create entertainment out of thin air that forces one to go inside a rabbit hole of creativity.
    6. Fans of different superhero universes possess deep knowledge of the domain but might lack the creativity to execute it. More than this, it takes either a long time for one person or a team of people specializing in different things such as writing ability, graphic designing ability, narration and even music generation. This solves this problem by allowing the user to do these tasks using simple prompts.

- **Value Proposition:**

1. Accessibility for Non-Artists
2. Customization
3. Fast Production
4. Multimodal Integration
5. Cost-Effective
6. Adaptability to Trends

- **What are we trying to avoid?**
    1. We do not want this just be another LLM wrapper
    2. We want to leverage LLM wrapper to create something nuanced, engaging and entertaining.
    3. We don't want this to be a one-time visit for people. This obviously doesn't mean that it caters to every human being but only those we care about.

2. **Tech/Workflow**:

Modules:

a) LLM: Generate the story and split them into scenes and generate the required prompts to feed to the further modules. The existing stories, its plots, characters, relations and traits will be used as a context to generate a new story that does not feel out of place for the traits and relations of existing characters (eliminate plot holes) and maintain a certain degree of continuity.

Input: User enters what kind of story, what characters (additional future expansion of mapping your own face into the comic and creating a new character), overall tone and length of the story.

Output: Prompts for the following modules split according to scenes and format required for each module

b) Stable Diffusion: Leveraging existing character models and appearances as context variables to keep the appearances consistent across the comic, generates the graphics for the story. (Stan Lee ki mkc, we do it with a click of a button)

Inputs: Generated prompts per scene from LLM + Context Variables

Output: Described scene as an image in a comic book appearance/style.

c) Music Generation: Generates an appropriate background score according to each scene keeping in mind the plot and the overall tone of the story as well as the tone of the particular scene. If need for generation of background score for pre-existing characters, the generated melody will sound like it(example - If i want to generate a background score for spiderman with the scene being inspirational, it should sound something similar to Sam Raimi's spider man 1's scene's background score where Uncle Ben says 'With great power, come great responsibility'.

Inputs: Generated prompts per scene from LLM + Context Variables

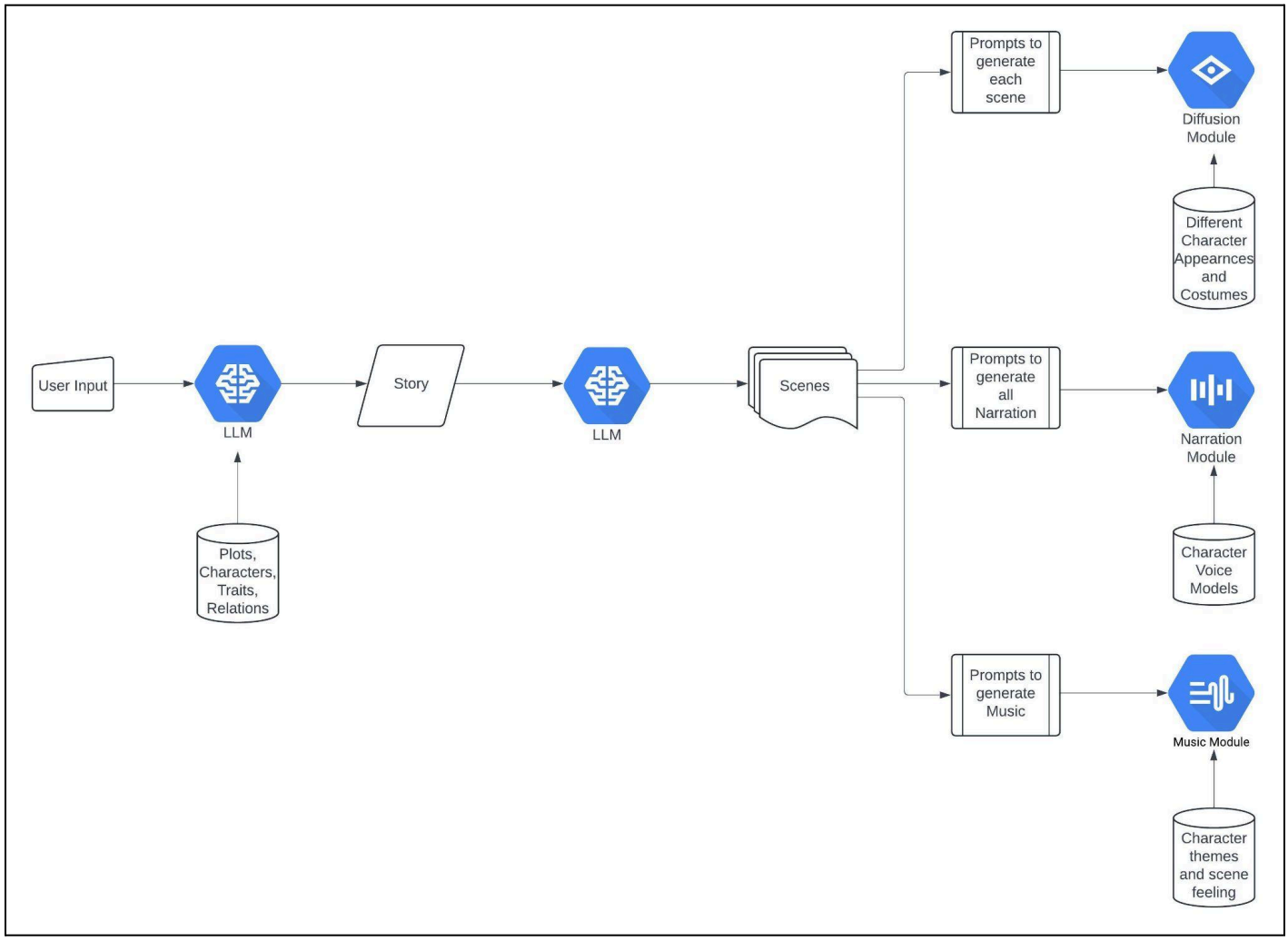
Output: Hans Zimmer ki mkc we do it with one click

d) Text to Speech: Keeping the characters in mind, we would require a per character based mapped voice model. Eg: Batman : deep raspy voice, Bruce Wayne: cocky yet relaxed voice. The dialogues would be fed as an input generated from the previous LLM.

Inputs: Generated prompts per scene from LLM + Context Variables

Output: Per character lines per scene + Narrator (if required)

High level workflow below



Feasibility:

LLM and Diffusion modules are pretty much sorted. Preliminary checks come out promising.

Music and Narration may be janky currently but the rate at which DL is expanding, confident we can see near perfect modules in the near future (6 months) that can be easily leveraged.

A with a cloud-based deployment in mind, we want to have this modules be serverless (no clue kaise karenge)

—----------

Market analysis -
https://docs.google.com/document/d/1MBBZ5j0gW33apg6wK6inUmuzfNxDNmoGMZnofTwA1FY/edit

—----------

—----------

Competition -
https://docs.google.com/document/d/1MBBZ5j0gW33apg6wK6inUmuzfNxDNmoGMZnofTwA1FY/edit

—----------