**Echoes of Imagination**

**Define the Problem (25–30s)**  
Stories have been one of humanity’s oldest ways of sharing knowledge, emotions, and imagination. But today, most stories are still limited to text on a page. Readers, especially younger audiences, often crave more immersive experiences that combine visuals and sound. While we have text-to-image or text-to-music systems, there is no seamless way to transform a narrative into a fully multimodal experience that engages sight and hearing together.

**Describe Your Solution (25–30s)**  
My project addresses this gap by creating a multimodal AI pipeline that generates both images and background music directly from a story. As the story is read in short paragraphs, the system first produces an illustration that reflects the scene. Then, using both the image and the text, it generates music that matches the mood and context. This approach makes storytelling far more engaging, turning static words into vivid multimedia experiences.

**Explain the Technology (25–30s)**  
The system is built in two stages. First, a text encoder extracts rich semantic meaning from the story and passes it into a variational autoencoder’s decoder to generate an image. In the second stage, both the text and the image are encoded together using multimodal techniques—similar to CLIP—and the resulting joint representation is fed into a music generation model. This layered design ensures that visuals and sound align closely with the narrative flow.

**Highlight the Impact (25–30s)**  
The impact of this project goes beyond entertainment. Imagine classrooms where children’s stories come alive with matching pictures and soundtracks, or therapy sessions where personalized narratives are enriched with calming imagery and music. By bridging text, visuals, and audio, this system has the potential to redefine how we consume and interact with stories. My goal is to push storytelling into the next era of immersive AI experiences.