Why AI Video Generation Struggles with Text Rendering

The Technical Problem

AI video generation models, including Veo 3, face significant challenges with accurate text rendering due to several technical limitations:

1. Pixel-Level Generation vs. Symbolic Understanding

- Video generation models work at the pixel level, essentially "painting" each frame
- They don't have a separate text rendering engine like traditional video editing software
- Text becomes just another visual pattern to learn, not a symbolic system with rules

2. Training Data Limitations

- Models are trained on millions of videos, but most don't contain clear, readable text overlays
- When text appears in training data, it's often:
- Blurry or low resolution
- Partially obscured
- In various fonts, sizes, and orientations
- Sometimes misspelled in the original source

3. Compression and Resolution Issues

- Video compression during training can degrade text quality
- Models learn to approximate text appearance rather than reproduce exact letterforms
- Fine details like text are often lost in the generation process

4. Lack of Spell-Check Integration

- Unlike traditional video editing tools, AI models don't have built-in spell-checkers
- They generate text based on visual patterns, not linguistic rules
- Common words might be rendered more accurately than proper nouns or brand names

Specific Challenges with Brand Names

Brand names like "SUNAIVA" are particularly problematic because:
- They're not common words in the training data
- The model has no reference for the correct spelling
- It tries to approximate based on phonetic similarity
- Unique letter combinations are harder to render accurately

Practical Solutions

Option 1: Generate Video Without Text, Add Text in Post-Production

Best Approach for Professional Results:
1. Generate the video focusing on visuals, characters, and scenes
2. Remove text overlay requests from the prompt
3. Use video editing software to add professional text overlays
4. This ensures perfect spelling and typography

Option 2: Multiple Generation Attempts

- Generate several versions with text
- Select the one with the most accurate text rendering
- This is hit-or-miss and not guaranteed to work

Option 3: Hybrid Approach

- Generate video with placeholder text areas
- Replace with professional graphics in post-production
- Maintains the AI-generated visual flow while ensuring text accuracy

Option 4: Simplified Text Requests

- Use very common, short words in prompts
- Avoid brand names or complex terminology
- Request simple, large text that's easier for the model to render

Recommended Workflow for Your SUNAIVA Video

Given the importance of accurate branding, I recommend:

1. **Generate a clean version without text overlays**

2. **Focus the prompt on the visual storytelling and character journey**
3. **Add professional text overlays using video editing software**

This approach will give you:
- Perfect spelling of "SUNAIVA"
- Professional typography
- Consistent branding
- Full control over text timing and positioning

Would you like me to generate a revised version of your video without the text overlays, so you can add professional text in post-production?