A. Failure Modes in Multi-Shot Video Generation (e.g., Track S: Semantic Leap)

B. The DIAL Pipeline & Diagonal Semantic Alignment (DSA)

Failure 1: Static Trap / Failure 2: Identity Amnesia

Input Narrative Prompts

Shot 1:
"A cybernetic samurai in a traditional Japanese garden..."

Shot 2:
"A cybernetic samurai in neon Cyberpunk city..."

Similarity Matrix M

Column-wise Softmax (Contrastive Isolation)

High Continuity S... But ignores p...

DSA Heatmap

$$
\begin{bmatrix} 0.85 & 0.40 \\ 0.30 & 0.75 \end{bmatrix}
$$

$$P_{i,j} = \frac{\exp(\tau \cdot M_{i,j})}{\sum_{k=1}^{K} \exp(\tau \cdot M_{k,j})}$$

0.95 0.05

0.05 0.95

Absolute Scores
(Vulnerable to Bleeding)

C. Legacy vs. DIAL Metrics: Unveiling the "Double-Kill" Dilemma

Legacy Metrics (Holistic)

FVD
(Static Model)

CLIP
(Amnesia M...

*Masks failures by
rewarding uniform continuity*

DIAL Pareto Frontier (Track S)

Unoccupied
Ideal Zone

Static
Trap

Identity
Amnesia

Subject Consistency

Background Diversity