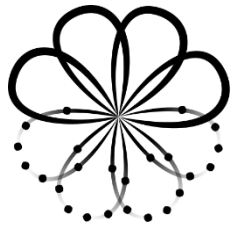
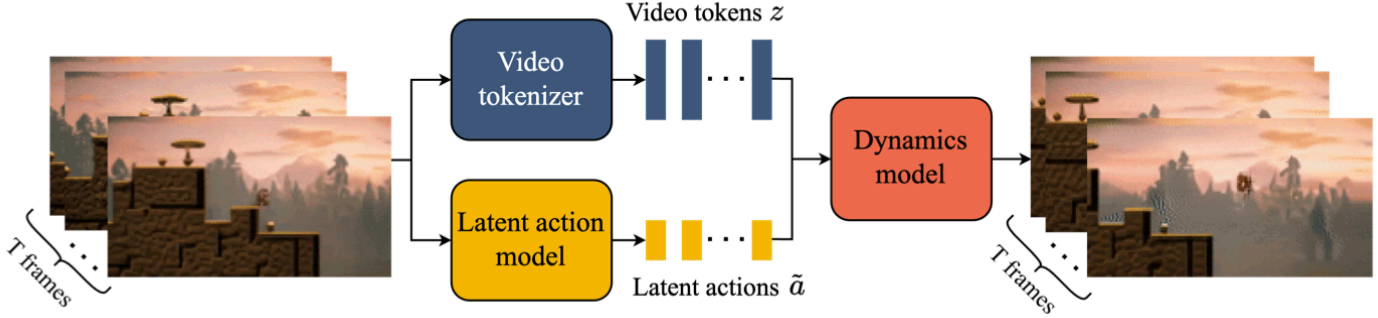


Institute of
Artificial Intelligence

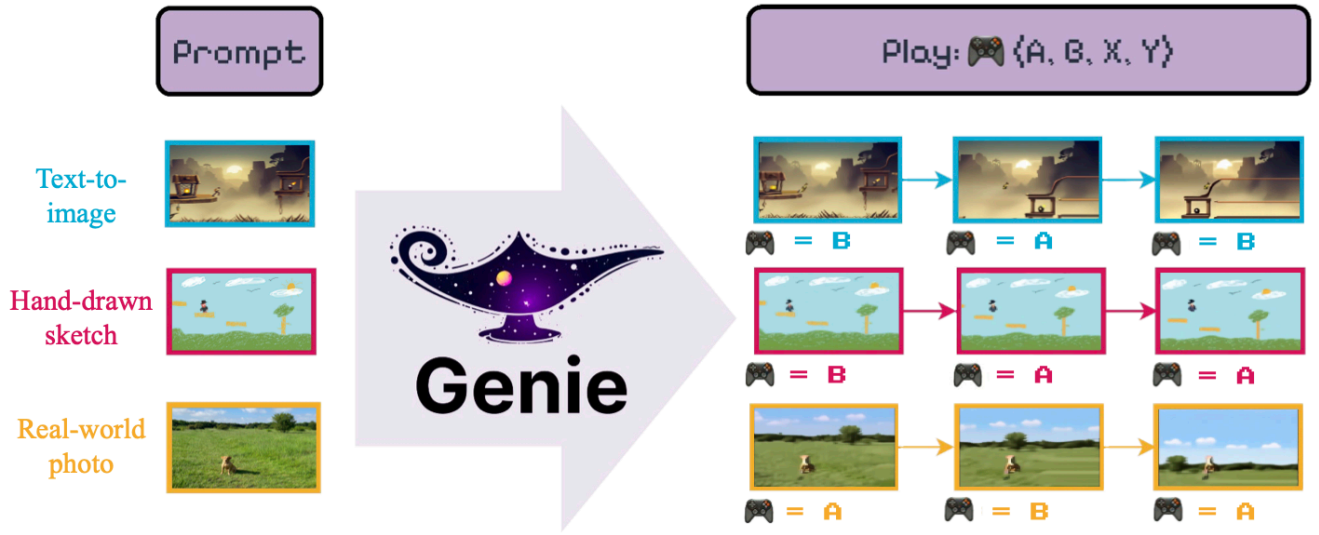
Leibniz
Universität
Hannover

Introduction to Genie [BDE+24]

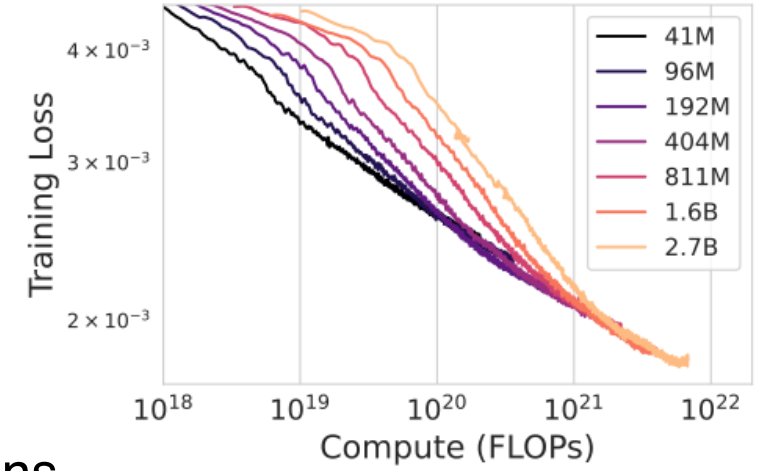
Introduction to Genie



- Genie is a generative interactive environment trained from unlabelled internet videos
- Capable of generating interactive, controllable virtual worlds from various prompts (e.g., text, synthetic images, photos, sketches)
- **Model Composition:**
  - Spatiotemporal video tokenizer
  - Autoregressive dynamics model
  - Scalable latent action model
- **Features:**
  - 11B parameters
  - Allows frame-by-frame interaction in generated environments
- **Examples:** [Genie]



Methodology and Applications



- **Architecture:**
  - Video Tokenizer: Converts raw video frames intro discrete tokens
  - Latent Action Model (LAM = unlabelled actions): Infers latent actions between frames
  - Dynamics Model: Predicts the next frame using tokenized video and latent actions
- **Applications:**
  - Interactive Environment Generation:
    - Users can create and explore diverse virtual worlds using simple prompts
    - Potential for training generalist agents by imitating behaviors from unseen videos
- **Training Agents:**
  - Genie can be used to train agents in generated environments, creating an generalist agent