Generative AI Tools for Digital Communication

Combined Teaching Notes

Summarized by Dr. Sobel Chan Ka Lok

Based on the teaching notes provided by RTHK AI Labs
from the project
"Enhancing Digital Communication through Generative AI: A Pilot Scheme for Innovative Pedagogies"

About These Notes

These notes compile and translate the key learning points from RTHK AI Labs teaching kits into one English handout. The aim is to help students understand practical generative AI workflows for image generation, video generation, voice creation, lip-syncing, and avatar-based communication production.

Teaching purpose
The notes are designed for classroom learning, workshop demonstration, self-study, and guided practice in digital communication, media production, animation, and creative content courses.

Source Materials Summarized

- Adobe Firefly Generative AI Image and Video
- ComfyUI Fundamentals for Image Generation
- HeyGen- focusing on AI voice cloning, text-to-speech, lip sync, and avatar creation.

Important Internal Use Statement

This document is an internal teaching aid. It is not a replacement for the original workshop demonstrations, platform terms of use, copyright policies, or institutional guidelines. Students should always check the latest tool interfaces, credit systems, and ethical requirements before publishing AI-assisted work.

Contents and Learning Outcomes

Part	Topic	Main Learning Goal
Part I	Generative AI Production Mindset	Understand how AI tools support creative communication workflows.
Part II	Adobe Firefly	Use accessible, commercially safer AI tools for images and videos.
Part III	ComfyUI	Understand node-based workflows and advanced control.
Part IV	HeyGen	Understand AI voice, TTS, lip-sync, avatars, and dubbing.
Part V	Integrated Workflow and Ethics	Plan an AI-assisted media production project responsibly.
Part VI	Motion Capture	For body motion capturing and mimic of human facial expression
Part VII	Topaz AI	Image and Video Restoration by AI

By the End of These Notes, Students Should Be Able To:

- Explain the basic differences between Firefly, ComfyUI, and HeyGen.
- Write clearer prompts using subject, action, style, camera, lighting, and detail.
- Describe the difference between text-to-image, image-to-image, text-to-video, and image-to-video.
- Identify the role of model checkpoints, prompts, latent space, KSampler, VAE, and ControlNet in ComfyUI.
- Prepare clean audio and video material for AI voice cloning, TTS, lip sync, and avatar generation.
- Evaluate AI outputs for quality, ethics, copyright, authenticity, and suitability for communication goals.

Part I - Generative AI Production Mindset

Generative AI is not simply a shortcut for making content. In professional communication, it is best understood as a production assistant that helps creators ideate, prototype, test, revise, and localize media. The strongest results still require human intention, aesthetic judgment, ethical awareness, and editorial control.

Student-friendly idea
Think of generative AI as a creative studio assistant. It can generate options quickly, but you are still the creative director, designer, editor, and final decision maker.

Three Levels of AI Use in Media Production

Level	Example	Student Skill Required
Basic generation	Create an image or voice from a prompt.	Clear prompting and basic quality checking.
Guided generation	Use image references, pose references, or style references.	Visual analysis, composition control, iteration.
Integrated production	Combine image, video, voice, lip sync, and editing tools.	Workflow planning, consistency management, ethical review.

A Useful Production Loop

1. Define the communication goal: What should the audience feel, understand, or do?
2. Design the concept: subject, story, mood, platform, and target audience.
3. Generate drafts: use text, image, video, or audio tools to create options.
4. Evaluate outputs: check accuracy, style, originality, ethics, and technical quality.
5. Refine and integrate: edit, upscale, dub, subtitle, animate, or redesign as needed.
6. Disclose and archive: keep records of prompts, sources, rights, and AI involvement.

Part II - Adobe Firefly: Overview

Adobe Firefly is a generative AI platform designed for creative work such as digital imaging, graphic design, video production, marketing, and 3D-related design support. In the workshop material, Firefly is presented as a user-friendly tool for turning ideas into images, video clips, soundtracks, speech, and editable creative assets.

Why Firefly Matters for Students

- It is easier to learn than node-based systems because the interface is more guided.
- It connects with the Adobe Creative Cloud ecosystem, which many design and media students already use.
- It supports multilingual prompting, including Traditional Chinese input, although some prompt suggestion features are English-focused.
- It emphasizes commercial safety and content transparency through Adobe training data choices and Content Credentials.

Firefly Function	What It Does	Typical Student Use
Generate Image	Creates still images from text prompts.	Poster concepts, social media visuals, thumbnails.
Edit Image	Changes image elements such as color, object, or background.	Retouching, object replacement, visual variation.
Generate Video	Creates short video clips from ideas or prompts.	Mood shots, campaign teasers, prototype scenes.
Generate Soundtrack	Creates music for video use.	Background music for student video drafts.
Generate Speech	Creates voice-over or narration.	Guide narration, learning videos, explainer content.
Firefly Boards	An infinite canvas for ideation with AI models.	Mood boards and campaign planning.

Adobe Firefly: Interface, Models, and Credits

The Firefly home interface organizes tools for image, video, audio, vector, and Adobe app-based creation. Students should learn the interface as a creative dashboard: choose a task, enter a prompt, adjust settings, generate variations, then save or continue editing.

Model Selection

The workshop compares different Firefly image models. Students should not assume that the newest model is always the best for every task. A model should be selected according to quality needs, speed, realism, and the type of image required.

Model	Core Strength	Suggested Use
Firefly Image 3	Fast generation of four variations; useful for early ideation.	Brainstorming and rapid concept exploration.
Firefly Image 4 / 4 Ultra	Higher detail and realism; stronger handling of complex subjects.	Portraits, products, small architecture, refined visuals.
Firefly Image 5 Preview	Higher resolution, stronger editing control, improved lighting and character rendering.	High-quality drafts, campaigns, and advanced creative testing.

Generative Credits

Adobe generative functions consume credits. Higher-resolution video generation uses more credits. Students should plan experiments carefully, test with lower-cost settings first, and reserve higher-quality generations for final versions.

Practical habit
Before a workshop, check your account, credit balance, file storage, and export settings. This prevents workflow interruption during class.

Adobe Firefly: Text-to-Image Workflow

Text-to-image is the most accessible Firefly workflow. Students describe a visual idea in words and Firefly produces image options. The most important skill is not using many words, but using the right categories of information.

Basic Prompt Formula

Core formula

Subject + action or mood + style or medium + details such as lighting, color, camera, composition, and texture.

Prompt Element	Question to Ask	Example
Subject	What is the main thing in the image?	A whale, a knight, a fantasy castle, a student presenter.
Action or Mood	What is happening? What is the feeling?	Floating, charging, peaceful, mysterious, energetic.
Style or Medium	What should it look like?	Photorealistic, cinematic, cartoon, origami, steampunk.
Details	What visual details improve specificity?	Pink sunset, 3D lighting, wide-angle lens, neon colors, high detail.

Example Prompt

```
A whale gracefully floating through clouds under a pink sunset sky, physically accurate 3D lighting and atmospheric realism.
```

Common Mistakes

- Using vague prompts such as "a sunset" without subject, style, or composition.
- Asking the AI to "generate" or "create" rather than simply describing the image.
- Ignoring aspect ratio, content type, and style presets.
- Accepting the first result without reviewing details, hands, text, lighting, or brand suitability.

Adobe Firefly: Image-to-Image, Text-to-Video, and Image-to-Video

Firefly supports workflows that begin not only with text but also with visual references. These reference-based workflows are useful because they help students control composition, style, and continuity.

Image-to-Image

- Use an outline image, sketch, photo, or previous AI output as a reference.
- A composition reference controls layout and structure.
- A style reference controls visual look, texture, color, and artistic feeling.
- Strength settings determine how closely the new image follows the reference.

Text-to-Video

Text-to-video turns a written idea into a short video clip. Students can select resolution and aspect ratio, and may set camera motion such as pan, zoom, tilt, or directional movement. This makes Firefly useful for creating mood shots, short campaign visuals, or prototype scenes.

Image-to-Video

Image-to-video begins with a still image and animates it into a video clip. The workshop example uses first and last frames, then gives a prompt describing how the motion should move from one image to another. This is especially useful for turning AI concept images into short cinematic transitions.

Workflow	Starts With	Best For
Text-to-Image	Prompt only	Visual ideation.
Image-to-Image	Reference image + prompt	Style transfer and composition control.
Text-to-Video	Prompt only	Short moving shots.
Image-to-Video	Image or frames + prompt	Animating stills or matching existing footage.

Adobe Firefly: Prompting, Iteration, and Responsible Use

Firefly encourages experimentation. Good prompting is a cycle: write a clear prompt, generate options, inspect the results, then refine the prompt or use editing features such as Generate Similar and Generative Fill.

Refinement Tools

- Generate Similar keeps one result and produces similar new versions.
- Generative Fill allows users to insert, remove, or replace content inside a selected area.
- Prompt Enhancement can automatically enrich English prompts, making them more detailed and vivid.
- Firefly Gallery can be used for inspiration and prompt learning.

Responsible Use Guidelines for Students

- Do not use generated output to train AI or machine learning systems unless permitted.
- Do not create, upload, or share illegal, harmful, or rights-violating content.
- Respect third-party rights, including copyright, likeness, brand identity, and privacy.
- Review and verify all outputs before submission or publication.
- Use Content Credentials and clear disclosure when appropriate.
- Check whether beta outputs are limited to personal use and not commercial use.

Classroom reminder

AI output is not automatically correct, ethical, or publishable. Students must act as editors and responsible communicators.

Part III - ComfyUI: Overview

ComfyUI is an open-source, node-based visual AI engine for building custom workflows for text-to-image, image-to-image, video, audio, and other generative tasks. Unlike one-click AI tools, ComfyUI shows the workflow as connected nodes. This makes it more technical, but also much more flexible and repeatable.

Why ComfyUI Is Important

- It runs locally and can work offline if the correct models and dependencies are installed.
- It is highly modular and supports many models and custom nodes.
- It gives advanced users precise control over generation workflows.
- It supports repeatable production pipelines, style consistency, and integration with traditional media production.

Advantage	What It Means for Students
Free and open-source	Students can learn the logic of AI workflows without being limited to one commercial interface.
Local workflow	Useful for experimentation, privacy-conscious workflows, and offline practice.
Node graph	Students can see how generation is constructed step by step.
Customizable	Advanced users can add ControlNet, upscaling, image editing, and specialized models.

Main challenge
ComfyUI has a steeper learning curve and may require stronger hardware, especially an NVIDIA GPU for fast results.

ComfyUI: Installation and Workspace Basics

The RTHK material introduces Windows installation through the ComfyUI Windows Portable build. This is useful for students because the portable build includes its own Python environment, so it does not depend on system Python.

Windows Installation Steps

1. Prepare 7-Zip or use the built-in Windows unzip tool.
2. Use an NVIDIA GPU if possible. CPU mode is supported but much slower.
3. Download the latest ComfyUI Windows Portable build from the official documentation or GitHub repository.
4. Unzip the downloaded file.
5. Run run_nvidia_gpu.bat.
6. Open a browser and go to http://127.0.0.1:8188.
7. Keep the command window open while using ComfyUI; closing it stops the local server.

Workspace Concept

A ComfyUI workspace is a canvas where nodes are connected into a workflow. The user does not need traditional coding for basic workflows, but must understand what each node does and how data moves between nodes.

Analogy
A ComfyUI workflow is like a production line: model loading, prompt reading, latent preparation, sampling, decoding, and saving all happen in order.

ComfyUI: Node-Based Workflow

A node is a functional block. Each node performs one task, such as loading a model, reading a prompt, sampling an image, decoding a latent image, or saving the result. Nodes usually have inputs on the left, outputs on the right, and adjustable settings inside the node.

Why Node-Based Learning Helps

- It makes the hidden AI generation process visible.
- It allows students to modify one part of the workflow without rebuilding everything.
- It improves understanding of cause and effect: changing a seed, step count, sampler, or denoise value changes the output.
- ComfyUI saves time by re-executing only the graph parts that changed.

Node Concept	Simple Explanation
Input	Information entering a node, such as a model, prompt, or latent image.
Output	Information leaving a node and moving to the next node.
Widget	A setting inside the node, such as seed, steps, CFG, or denoise.
Graph	The full network of connected nodes.
Workflow	A saved graph that can be reused and shared.

Classroom tip
Ask students to describe each node aloud. If they can explain what data enters and leaves a node, they understand the workflow better.

ComfyUI: Custom Nodes and ComfyUI-Manager

Custom nodes are community-made Python extensions that add new nodes, models, and utilities. They are essential for advanced workflows such as ControlNet variants, upscalers, special samplers, and workflow management tools.

ComfyUI-Manager

ComfyUI-Manager is a practical tool for installing, updating, and managing custom nodes. It reduces the difficulty of adding community features and helps students maintain a more usable ComfyUI environment.

Task	Why It Matters
Install missing custom nodes	Many shared workflows require additional nodes.
Update nodes	Keeps workflows compatible and reduces errors.
Manage models	Helps organize models and dependencies.
Install by URL	Allows users to add specific node repositories.

Suggested Good Practice

- Record which custom nodes are used in a project.
- Avoid installing unnecessary nodes before a live workshop.
- Restart ComfyUI after installing major custom nodes.
- When a workflow fails, check missing nodes first.

ComfyUI: Basic Text-to-Image Workflow

A text-to-image workflow generates images from natural language prompts. The model interprets the prompt and synthesizes visual information in latent space before producing the final image.

Five Main Components

Component	Function	Student Explanation
Model Checkpoint	Loads the trained AI model.	The model brain that contains visual knowledge and style ability.
Positive Prompt	Describes what the image should include.	What you want the AI to emphasize.
Negative Prompt	Describes what to avoid.	What you want the AI to reduce or remove.
Empty Latent Image	Defines the starting latent canvas.	The hidden compressed image space before pixels appear.
KSampler	Controls denoising and image construction.	The engine that turns noise into an image.

Typical Image Sizes

- Stable Diffusion 1.x workflows commonly begin around 512 x 512.
- SDXL workflows commonly begin around 1024 x 1024.
- Final size can be changed through generation settings and upscaling.

Key idea
The AI usually does not draw pixels directly from the prompt. It works in latent space first, then the VAE decoder turns the latent result into a visible image.

ComfyUI: Understanding KSampler

KSampler is one of the most important nodes in ComfyUI. It controls the denoising process that turns a noisy latent image into a coherent final image, guided by the model and prompts.

Parameter	Meaning	Practical Starting Point
Seed	Controls reproducibility.	Same seed + same graph = same output.
Steps	Number of denoising iterations.	Start around 20 to 30.
CFG / Guidance	How strongly the image follows the prompt.	Start around 6 to 9.
Sampler Name	The algorithm used for denoising.	Try Euler, Euler a, or DPM++ 2M SDE.
Scheduler	Controls denoising schedule.	Try Karras or simple, depending on model guidance.
Denoise	Strength of change compared to the input latent.	1.0 means full generation from noise.

Mini Experiment

1. Keep the same prompt and seed, then change the number of steps.
2. Keep the same prompt and steps, then change CFG from low to high.
3. Keep everything the same, then change the seed.
4. Compare the differences in composition, detail, color, and prompt accuracy.

Learning outcome
Students learn that AI images are not magical accidents. They are the result of interacting settings inside a controlled workflow.

ComfyUI: Image-to-Image Workflow

Image-to-image transforms an existing picture into a new one. The input image provides structure or content guidance, while the prompt changes style, detail, or concept. This is useful when students want to preserve a layout while changing the appearance.

Text-to-Image vs Image-to-Image

Text-to-Image	Image-to-Image
Starts from a prompt and empty latent image.	Starts from an existing image.
More creative freedom.	More structure and layout control.
Useful for new concepts.	Useful for redesigning or restyling existing visuals.
Uses empty latent image.	Requires VAE Encode to convert the input image to latent space.

Denoise rule

Lower denoise values preserve more of the original structure. Higher denoise values change more of the input image. A denoise value of 1.0 behaves like almost full generation from noise.

Classroom Exercise

Students can use a simple sketch, a phone photo, or a storyboard frame and test different denoise values. They should record how the image changes at low, medium, and high denoise settings.

ComfyUI: Advanced Guidance - ControlNet and Image Editing Models

Advanced image generation focuses on reliability and control. In professional production, creators often need consistent composition, pose, camera angle, and shot-to-shot continuity. ComfyUI can guide generation through structured conditions and image-editing models.

Method	What It Controls	Best Use
ControlNet Canny	Edges and contours.	Preserving silhouettes, object shapes, and composition.
ControlNet OpenPose	Human skeleton and body pose.	Controlling full-body posture and gesture.
Image Editing Models	Natural-language edits to existing images.	Changing camera angle, hair, clothing, pose, or selected details.
JSON-Style Prompting	Structured prompt categories.	Improving reliability, predictability, and consistency.
Model-Based Upscaling	Resolution and reconstructed detail.	Preparing images for posters, slides, production stills, or social media.

ControlNet in Simple Words

ControlNet is like giving the AI a blueprint. Canny says: follow these outlines. OpenPose says: follow this body pose. The prompt still controls the look, but the control map helps the AI respect structure.

Important limitation
Guided generation improves control but does not guarantee perfection. Students must still review anatomy, continuity, and unwanted artifacts.

ComfyUI: JSON-Style Prompting and Upscaling

JSON-style prompting organizes image instructions into clear key-value sections. It is especially helpful when students need consistent visual outputs across multiple shots or characters.

Example JSON-Style Prompt Template

```
{
  "image_purpose": "cinematic character portrait",
  "subject": {"type": "young detective", "pose": "standing", "expression": "serious"},
  "environment": {"setting": "rainy city street", "atmosphere": "mysterious"},
  "style": "photorealistic cinematic film still",
  "camera": {"shot_type": "medium close-up", "lens": "50mm"},
  "lighting": {"type": "neon side light", "quality": "soft and moody"}
}
```

Upscaling

Model-based upscaling increases resolution by intelligently adding pixels and reconstructing details, rather than merely stretching the image. This is useful for improving draft images before using them in presentations, posters, storyboards, or video assets.

- Use upscaling after choosing a strong image, not before selecting the final concept.
- Check whether upscaling invents unwanted details or changes faces, logos, or text.
- Keep original files and record which upscaler was used.

Part IV - HeyGen: AI Voice, Lip Sync, and Avatars

HeyGen is introduced in the teaching kit as a tool for AI voice cloning, text-to-speech, lip-sync technology, avatar generation, and video translation. It combines voice generation with face and mouth movement, allowing users to produce avatar or dubbed video content.

Core Concepts

Concept	Explanation	Use Case
AI Voice Cloning	Replicates a voice tone, pitch, and speaking style.	Personalized narration or presenter voice.
Voice Mirroring	Uses recorded audio to mirror speech style.	More natural voice performance.
TTS	Turns text into spoken audio.	Multilingual narration and explainer videos.
Lip Sync	Matches mouth movement and expression to generated or recorded voice.	Dubbing and avatar video production.
Avatar Training	Uses video or photo input to create a speaking avatar.	Virtual presenters, training videos, internal communications.

Production formula
Avatar or video dubbing = AI voice + lip sync + suitable source video or photo.

HeyGen: AI Voice Cloning and Text-to-Speech

AI voice cloning requires clean source audio. The kit recommends around two minutes of clear speech recorded with a high-quality microphone in a quiet environment. The speaker should choose an easy topic, speak clearly, and use natural pauses between sentences.

Voice Cloning Checklist

- Record around two minutes or more of uninterrupted speech.
- Use a high-quality microphone when possible.
- Record in a quiet environment without echo or background noise.
- Speak about something easy and natural.
- Use clear articulation; slightly exaggerated enunciation can help.
- Pause between sentences so the voice model learns natural rhythm.

Voice Director Prompting

The teaching kit gives an example of directing voice performance through prompt categories: voice, tone, punctuation, delivery, and phrasing. Students should learn that vocal quality is not only about words, but also rhythm, emotion, pace, and emphasis.

Voice Direction Element	Example Instruction
Voice	Warm, lively, confident, and suitable for storytelling.
Tone	Playful, clever, inviting, and upbeat.
Punctuation	Use natural pauses before important points.
Delivery	Relaxed and clear, with light emphasis.
Phrasing	Simple, expressive, and easy to follow.

HeyGen: Lip Sync and Avatar Creation

Lip-sync technology connects the voice to the mouth movement and facial expression of a person or avatar. This allows a video or photo-based avatar to speak in a new voice or language. The kit presents a basic production pipeline: source input, avatar training, and video synthesis.

Step	Action	Student Attention Point
1. Source Input	Provide video or photo material.	Use clear face visibility and stable framing.
2. Consent	Confirm the subject agrees to avatar use.	Consent is essential for ethical and legal use.
3. Avatar Training	Upload footage and train the avatar.	Better input produces a more reliable avatar.
4. Video Synthesis	Generate speech and synced video.	Review mouth movement, facial expression, and timing.

Consent reminder
Never clone or animate a person without permission. AI voice and avatar tools can affect identity, trust, and personal rights.

Video or Photo Avatar?

A video-based avatar usually provides more movement information, while a photo avatar can be faster for simple speaking-head content. Students should choose according to project goals, available materials, time, and ethical approval.

HeyGen: Filming and Recording Tips for Better Avatars

Avatar quality depends heavily on source quality. The kit emphasizes stable footage, direct eye contact, controlled hand movement, natural pauses, and emotionally engaging speech.

Filming Tip	Reason
Record at least two minutes of uninterrupted speech.	The system needs enough examples of the person speaking.
Be natural and stable.	The avatar will copy performance habits.
Maintain eye contact with the camera.	Creates a more direct and professional presenter style.
Keep hand movements below chest level.	Avoids confusing motion and visual artifacts near the face.
Pause with lips closed between sentences.	Helps produce cleaner sentence transitions.
Add appropriate emotion.	A slightly expressive delivery can make the avatar more engaging.

Common Problems to Watch For

- Background noise affects audio quality and cloning accuracy.
- Large gestures may create unstable body or face movement.
- Poor lighting or low-resolution footage can reduce avatar realism.
- Incorrect pronunciation may require manual pronunciation adjustments.
- Lip sync may look unnatural if speech speed, emotion, or language rhythm is unsuitable.

Part V - Integrated AI Production Workflow

The strongest classroom application is not learning each tool separately, but understanding how tools can connect in a production pipeline. Firefly can generate accessible visuals, ComfyUI can refine controlled image workflows, and HeyGen can add voice, lip sync, and avatar communication.

Production Stage	Possible Tool	Output
Concept Development	Firefly Gallery, Firefly Boards, prompts	Mood board, visual direction, reference images.
Image Creation	Firefly Text-to-Image or ComfyUI T2I	Character, setting, poster, storyboard frame.
Image Control	ComfyUI ControlNet, I2I, JSON prompts	Consistent composition, pose, and style.
Video Motion	Firefly Text-to-Video or Image-to-Video	Short clips, transitions, campaign teaser shots.
Voice and Narration	HeyGen TTS or voice cloning	Narration, presenter voice, multilingual speech.
Avatar / Dubbing	HeyGen lip sync and avatar tools	Talking-head video, localized communication.
Final Review	Human editing and ethical checking	Publishable learning or campaign asset.

Workflow principle

Use the easiest tool for the task, not the most complicated tool. Professional AI production is about choosing the right tool at the right stage.

Choosing the Right Tool: Firefly vs ComfyUI vs HeyGen

Students often ask which tool is best. The better question is: what is the task? Each tool has a different role in the production process.

Tool	Best For	Strength	Main Limitation
Adobe Firefly	Fast image and video ideation; Adobe-friendly workflow.	Accessible interface, guided controls, commercial-safety emphasis.	Credit limits and less low-level control than node systems.
ComfyUI	Advanced and repeatable image workflows.	Deep control through nodes, custom workflows, local execution.	Steeper learning curve and stronger hardware needs.
HeyGen	Voice, avatar, dubbing, and lip-sync videos.	Fast presenter-style communication and multilingual output.	Requires high-quality source material and careful consent handling.

Decision Guide

- Choose Firefly when you need quick creative prototypes, design variations, or accessible image and video creation.
- Choose ComfyUI when you need repeatable workflows, custom nodes, ControlNet, or technical control.
- Choose HeyGen when your project needs speech, presenter avatars, translation, or lip-synced communication.
- Combine tools when you need a complete content pipeline from concept to final delivery.

Part VI. From Traditional Filming to Motion Capture

The motion capture kit presents a historical movement from early cinema to advanced real-time character animation. In early cinema, filmmakers relied on static camera setups, puppetry, miniatures, and practical effects to create visual illusions. These techniques were creative but limited in the complexity and realism of character movement.

With the growth of computer graphics in the late twentieth century, films began combining live-action footage with computer-generated imagery. CGI made impossible worlds more believable, but character animation still required artists to manually animate movement frame by frame or key pose by key pose.

Motion capture changed this relationship between body and image. Instead of inventing every movement manually, filmmakers could record the actor's body performance and translate it into digital motion data.

Key idea

Motion capture does not replace performance. It records performance and transfers it into a digital character, object, or environment.

HISTORICAL DEVELOPMENT

Six Stages in the Development of Motion Capture

Stage	What changed?	Why it matters
1. Early filming techniques	Practical effects, puppetry, static setups, and optical tricks were used to create illusion.	Cinema learned to tell impossible stories before digital tools existed.
2. Computer graphics	CGI entered mainstream cinema and could be combined with live action.	Digital images expanded what could appear on screen.
3. Emergence of mocap	Actors wore suits with markers or sensors so their movement could be digitized.	Human movement became usable data for animation.
4. Refinement and integration	Better accuracy and software allowed more lifelike digital characters.	Digital characters could interact more naturally with live-action footage.
5. Real-time applications	Direct visualization made it possible to see animated characters during capture.	Production became faster and more iterative.
6. Broader media use	Mocap expanded to games, VR, virtual events, and live broadcast.	Motion capture became a cross-media communication technology.

Class discussion

Ask students: Which stage changed storytelling the most - CGI, mocap, or real-time visualization? Why?

CORE DEFINITION

What Is Motion Capture?

Motion capture, often shortened to mocap, is a technology that records the movement of people, objects, or facial expressions and converts that movement into a digital format. The captured movement can then drive a 3D character, a virtual presenter, a game avatar, or other animated forms.

Basic process

1. A performer moves, acts, gestures, or speaks in front of a capture system.
2. Sensors, markers, cameras, or algorithms track the movement.
3. The system converts physical movement into digital data.
4. The data is cleaned, edited, and mapped to a digital character or model.
5. The final character appears to move with the performer's timing, rhythm, and physical intention.

Student analogy

Mocap is like recording a dance, but instead of saving only the video image, it saves the movement information so that another digital body can perform the same movement.

TECHNOLOGY AND DATA

How Motion Becomes Digital Data

Motion capture is useful because it converts physical performance into data. The performer's movement is no longer only a visual recording. It becomes a set of coordinates, rotations, timing values, and body-position relationships that software can process.

Production element	What it captures	Example output
Body capture	Whole-body movement: walking, jumping, fighting, dancing, gestures.	A 3D character walks or performs action with realistic timing.
Facial capture	Eyebrows, eyes, cheeks, mouth shapes, and emotional expression.	A digital face smiles, speaks, or reacts believably.
Hand capture	Finger movement, grip, pointing, object interaction.	A virtual presenter holds or touches objects naturally.
Camera / scene sync	Camera perspective and relation between performer and environment.	A virtual scene can match live-action camera movement.

Important

Captured data usually requires cleaning. Raw capture can include noise, missing frames, sliding feet, unstable facial points, or unnatural body positions.

TYPES OF MOCAP

Marker-Based and Markerless Motion Capture

The RTHK teaching kit identifies two major types of motion capture: marker-based systems and markerless systems. Both aim to track movement, but they differ in how movement is detected.

Type	How it works	Strengths	Possible limitations
Marker-based	Performer wears physical markers or sensors tracked by cameras or inertial systems.	High precision; strong for professional animation and demanding production.	Needs equipment, setup time, calibration, and sometimes a controlled space.
Markerless	Software uses computer vision and algorithms to estimate body movement from video.	More accessible; does not require a full suit; useful for fast experimentation.	May be less accurate in complex motion, occlusion, or unusual camera angles.
Facial markerless	Camera-based facial capture tracks expression and mouth movement.	Good for virtual avatars and real-time facial performance.	Needs good lighting, clear face visibility, and careful calibration.

Practical choice

For a classroom project, markerless mocap is often easier to begin with. For professional broadcast or high-precision character work, a controlled marker-based or inertial system may be more reliable.

APPLICATIONS

Where Motion Capture Is Used

Motion capture is now a cross-media production technology. It is not limited to cinema. The teaching kit highlights film, animation, video games, virtual reality, and live broadcast as key fields.

Field	Common use	Student example
Film and animation	Creating realistic creature, human, or fantasy character movement.	A student directs an actor and applies the movement to a digital detective character.
Video games	Animating playable characters, fight actions, sport motion, and interactive gestures.	A character's run, jump, and attack movements are captured as an animation set.
Virtual reality	Creating embodied interaction between user and avatar.	A VR guide waves, walks, and responds to the user.
Live broadcast	Driving real-time virtual hosts, dynamic graphics, or animated presenters.	A live event adds a virtual co-host or digital mascot.
Education and training	Demonstrating movement, body language, sport technique, or performing arts.	Students analyze gesture, posture, and expression in a communication course.

Media communication angle

Motion capture helps communication students understand not only images, but also embodiment, performance, presence, and audience engagement.

WORKFLOW

A Simple Motion Capture Production Pipeline

A motion capture project should be planned like a production workflow, not just a technical experiment. Students should begin with the communication purpose, character design, and performance style before touching the equipment.

1. Define the story purpose: What does the character need to express?
2. Prepare the performer: rehearse action, blocking, rhythm, and emotion.
3. Set up the capture system: body capture, facial capture, or both.
4. Calibrate the performer and software to reduce tracking errors.
5. Record the performance in short, manageable takes.
6. Review the capture data immediately and repeat if necessary.
7. Clean and retarget the data to the digital character.
8. Add camera, lighting, environment, sound, and editing decisions.
9. Review whether the final result communicates the intended emotion and action.

Production advice

Do not record one long performance first. Capture short takes. Short takes are easier to review, clean, and repeat.

LIVE DEMO LOGIC

Understanding the Xsens and Faceware Demonstration

The RTHK motion capture kit includes a live demo with Xsens and Faceware. For teaching purposes, students should understand the logic behind this demonstration even if they do not operate the exact same system.

Component	Role in the pipeline	Teaching focus
Xsens	Captures body movement using a professional motion capture system.	How body performance becomes digital movement data.
Faceware	Captures facial movement and expression for digital characters.	How facial acting and speech expression shape character believability.
Character rig	Receives captured motion and applies it to a digital body.	How body structure and rig quality affect final movement.
Real-time preview	Shows the character response during or soon after capture.	Why immediate visual feedback improves production decisions.

Key learning point

A live demo is not only about watching equipment. It helps students see the relationship between performance, data, software interpretation, and screen output.

RTHK APPLICATIONS

Current Applications and Visual Communication

The motion capture slides show RTHK examples involving virtual characters, presenter-like figures, robotic movement, and stage or broadcast integration. These examples suggest that mocap can support new forms of visual communication in public media, event production, and digital storytelling.

Teaching interpretation

- A virtual presenter can extend the visual identity of a programme or event.
- A digital character can appear in environments where physical filming is difficult or expensive.
- Robotic or stylized characters can communicate abstract technological themes.
- Live stage integration can create mixed reality experiences between human performers and digital characters.
- Motion capture makes animation more responsive to real human timing and body language.

Class prompt

Ask students to design a virtual RTHK-style public service character. What should the character look like? How should it move? What emotion and values should the movement communicate?

FUTURE DIRECTION

AI-Driven Real-Time Character Animation

One future direction highlighted in the motion capture kit is AI-driven real-time character animation. This means using AI to help interpret performance data, generate character movement, improve animation continuity, or drive an animated character quickly during production.

In a media education context, this is important because it moves animation closer to live performance. A character can potentially react, speak, gesture, and appear on screen with very short delay. This can change virtual hosting, live-streaming, education videos, news explainers, and immersive communication.

Opportunity	Example	Risk to consider
Faster production	A virtual host can be animated quickly for a short video.	Speed may reduce careful performance direction.
Improved accessibility	Smaller teams can create digital characters.	Students may over-rely on default movement styles.
Real-time interaction	A character can respond during a live session.	Unexpected output needs moderation and backup planning.
Expressive experimentation	Different acting styles can be tested quickly.	Uncanny or unnatural motion can distract audiences.

FUTURE DIRECTION

Motion Databases and Motion Control

Another future direction is the use of motion databases. A motion database is a collection of reusable movement clips such as walking, turning, waving, jumping, dancing, fighting, or gesturing. Instead of capturing every movement from scratch, creators can select, combine, and modify motion clips.

Why motion databases matter

- They reduce production time because common actions can be reused.
- They allow non-specialists to test animation ideas before arranging a full capture session.
- They support continuity because a character can maintain a consistent movement language.
- They can be combined with AI tools to generate transitions between actions.
- They help students understand movement as an editable media asset.

Critical thinking

A motion database is convenient, but it can also make characters look generic. Students should adapt movement to character personality, cultural context, age, body type, and story situation.

AI-ASSISTED POST-PRODUCTION

From Capture to Enhancement: Why Post-production AI Matters

The second teaching kit focuses on AI technology assisted post-production, especially the way AI tools can improve editing efficiency. The examples include AI photo restoration, AI video restoration, AI slow-motion enhancement, and video upscaling to higher resolution.

For students, the central lesson is that AI can help editors complete work that once required long processing time, expensive resources, or specialist manual repair. However, the editor still needs to judge whether the output is accurate, ethical, and visually appropriate.

Post-production principle

AI enhancement should serve editorial intention. Sharper, smoother, or cleaner does not always mean better. The correct result depends on context, archive value, visual style, and audience expectations.

Main AI-assisted post-production areas

- Photo restoration
- Video restoration
- Resolution upscaling
- Motion smoothing and slow motion
- Workflow efficiency and quality control

PHOTO RESTORATION

Part VII: TopazAI: Image and Video Restoration

The Topaz AI teaching slides show before-and-after examples of photo restoration. Old or low-quality photographs may contain blur, noise, low resolution, faded detail, or damaged facial features. AI restoration tools can reconstruct visual detail, sharpen faces, reduce noise, and improve overall image quality.

Common problem	AI-assisted correction	Editor's judgement
Blurred face	Sharpen and reconstruct facial detail.	Does the restored face still look like the original person?
Low resolution	Upscale and add texture detail.	Is the new detail believable or artificially invented?
Noise and grain	Reduce digital noise or film grain.	Does removing grain destroy the historical texture?
Faded or unclear photo	Improve contrast, edges, and local detail.	Does the image become too modern or over-processed?

Important ethical note

Restoration can change how people remember a person, event, or historical moment. Students should record what was changed and avoid presenting AI reconstruction as untouched evidence.

VIDEO RESTORATION

AI Video Restoration for Archival Footage

The slides use old Hong Kong video footage to demonstrate AI video restoration. Archive footage may be blurry, noisy, low contrast, scratched, unstable, or limited by old recording formats. AI tools can help improve clarity and make historical footage more usable for modern screens.

Typical AI restoration tasks

- Reduce noise and grain while keeping meaningful texture.
- Improve sharpness and readability of buildings, signs, faces, and objects.
- Stabilize or reduce jitter when footage is shaky.
- Improve contrast and recover visual separation between foreground and background.
- Upscale old footage for HD, 4K, or larger presentation formats.

Archive thinking

When working with historical materials, the editor should not erase the time period completely. The goal is often to improve visibility, not to make the past look fake or overly polished.

BEFORE / AFTER ANALYSIS

How to Evaluate an AI Restoration Result

Students should learn to evaluate AI restoration through visual comparison rather than trusting the software automatically. A useful method is to compare before and after versions side by side and check specific areas of the image.

Area to inspect	What to look for	Warning sign
Face	Natural skin texture, recognizable identity, eyes and mouth not distorted.	Plastic skin, strange eyes, false age, or changed identity.
Text and signage	Improved readability without invented letters.	AI creates fake words or symbols.
Edges	Buildings, clothing, and objects become clearer.	Over-sharpened halos or jagged outlines.
Motion	Frame-to-frame consistency.	Flicker, unstable details, or changing faces.
Historical texture	Preserved visual character of original source.	Image looks too modern or loses archive feeling.

Student task

Choose one restored image or video frame. Mark three areas that improved and two areas that may need human correction.

VIDEO UPSCALING

AI Video Enhancement and Upscaling

Video upscaling increases resolution, for example from HD 1080p toward 4K or 8K. Traditional scaling stretches pixels, while AI-based upscaling attempts to reconstruct sharper details. The Topaz AI teaching slides compare different upscaling outcomes and show that AI can produce clearer results for modern delivery.

Concept	Simple explanation	Production use
Upscaling	Increasing video resolution while trying to preserve or reconstruct detail.	Preparing older footage for modern screens.
Enhancement	Improving sharpness, denoising, color stability, or texture.	Making footage cleaner and more presentable.
Model selection	Choosing an AI model based on content type and problem.	Different models may work better for faces, landscapes, or noisy footage.
Preview and compare	Testing short segments before processing the full video.	Prevents wasted time and over-processing.

Best practice

Always test on a short clip first. Processing a full video before checking settings can waste time and create inconsistent output.

SLOW MOTION

AI Slow Motion and Frame Interpolation

The teaching slides show an example of AI slow-motion enhancement for action movement. AI slow motion usually works through frame interpolation: the software analyzes motion between two frames and generates additional in-between frames. This can make fast action appear smoother when slowed down.

Use case	How AI helps	What to check
Sports or action shots	Creates extra frames to make movement smoother.	Look for warped limbs, unstable objects, or broken backgrounds.
Dance or performance	Emphasizes gesture and rhythm.	Check that body motion remains graceful and natural.
Dramatic moment	Extends a key emotional or visual beat.	Make sure slow motion supports story, not just spectacle.
Educational demonstration	Allows viewers to study movement step by step.	Avoid generating misleading movement details.

Editing reminder

Slow motion is a storytelling choice. It should highlight meaning, emotion, skill, or visual detail - not simply show that a tool can create extra frames.

EDITORIAL WORKFLOW

AI-Assisted Editing Workflow

AI tools can significantly improve efficiency, but they should be inserted into a controlled post-production workflow. A responsible workflow helps avoid quality problems and keeps creative decisions visible.

1. Ingest source material and identify the problem: blur, noise, low resolution, unstable motion, or damaged details.
2. Choose a tool and model suitable for the problem.
3. Make a short test export using a small section of the material.
4. Compare original and AI output side by side.
5. Check faces, text, edges, movement, and historical texture.
6. Adjust settings and repeat testing if necessary.
7. Process the full file only after test results are acceptable.
8. Keep a version history and note what AI processing was applied.
9. Export in a format suitable for the final platform or screening context.

Professional habit

Do not overwrite the original file. Keep the source, the AI-processed version, and the final edited version separate.

ETHICS AND QUALITY CONTROL

Ethical and Production Considerations

AI-assisted restoration and enhancement can be powerful, but students need to understand the ethical and production risks. AI may invent visual details that were not present in the original. It may change identity, alter historical evidence, or create a misleading impression of authenticity.

Issue	Why it matters	Recommended action
Authenticity	AI may reconstruct details that did not exist in the source.	Label AI-restored or AI-enhanced material when appropriate.
Consent	Faces and personal images may be modified or clarified.	Use authorized material and respect privacy.
Copyright	Footage and photos may have rights restrictions.	Confirm internal permission before processing or sharing.
Historical integrity	Archive footage can lose its original texture.	Balance clarity with respect for source material.
Over-processing	Too much enhancement can look artificial.	Use visual comparison and peer review.

Internal-use reminder

For HSUHK internal teaching, students should use approved materials, document processing steps, and avoid publishing test outputs without permission.

REVISION AND TAKEAWAYS

Final Revision Questions and Key Takeaways

Revision questions

- What is the difference between practical effects, CGI, and motion capture?
- Why is motion capture useful for creating believable digital characters?
- What are the advantages and disadvantages of markerless mocap?
- How can AI video restoration support archive-based storytelling?
- What visual problems should editors check after AI upscaling?
- Why should students keep the original source files when using AI tools?
- How can AI production tools improve efficiency without weakening creative judgement?

Key takeaways

- AI and motion technologies are most powerful when linked to clear storytelling and communication goals.
- Motion capture preserves human timing, rhythm, and intention inside digital characters.
- AI post-production tools can save time, recover image quality, and support archive reuse.
- Human review remains essential because AI can invent details, distort identity, or over-process visual material.
- A professional workflow includes testing, comparison, documentation, ethics, and final creative judgement.

End note

These notes are for internal educational use at HSUHK and are based on the RTHK AI Labs teaching kits connected to the project "Enhancing Digital Communication through Generative AI: A Pilot Scheme for Innovative Pedagogies."