

Putting AI to Work

7

Generating Video and Audio

Learning Objectives

- Enhance still images by using AI tools to restore color and animate visual elements for dynamic presentations
- Generate video content by translating written prompts into short, animated clips using text-to-video AI models
- Create lifelike avatars and insert custom objects into video environments through AI-generated imagery
- Produce accurate subtitles and synthesize multilingual voice narration using AI-powered video tools
- Convert written content into natural-sounding speech and explore ethical and technical aspects of voice cloning
- Compose original music, generate sound effects, and layer audio tracks to build professional-quality audio using AI tools

Module 7.1: Colorizing and Adding Motion to Images

- AI colorization adds realistic or artistic color to black-and-white or faded photos.
- Tools analyze content like people, objects, and environments to infer appropriate colors.
- Applications:
 - Restoring historical images
 - Enhancing archives
 - Reinvigorating marketing materials
- Image-to-motion simulates visual movement from a single static image.
- Motion types:
 - Camera pan/zoom (parallax)
 - Element motion
 - Depth simulation
 - Looping animations
- AI detects layers, depth, and composition to create movement without actual video.

Module 7.1: Ethics in Action

- Colorizing historical images may distort meaning through incorrect colors on uniforms, skin tones, or surroundings.
- Animating realistic people without their consent can alter expressions or suggest emotions that were never present.
- Always disclose when images have been altered and consider potential misrepresentation.

Module 7.1: Techie Dive

- AI colorization uses convolutional neural networks (CNNs) trained on large image datasets.
- Motion simulation uses depth estimation, optical flow, and layer segmentation.
- Optical flow tracks the motion of individual pixels between frames to create realistic movement.
- Pre-trained models detect and animate separate image parts like the foreground vs. the background.

Module 7.1: Business Lens

- Motion and colorization boost engagement and emotional impact in marketing.
- Animated images perform better on social media and make historical events feel more real.
- Free tools often include branding or lack precision; premium options ensure quality.
- Media restoration attracts new audiences to archival footage.

Module 7.2: Text-to-Video Transformations

- Using text-to-video for creating clips from descriptions is a rapidly evolving area in generative AI.
- Applications:
 - Marketing
 - Entertainment
 - Education
 - Prototyping without camera crews or actors
- Technical approaches involve diffusion models, frame-by-frame generation, and motion modeling.
- Diffusion models start with random noise and refine through multiple steps guided by prompt.
- Motion modeling maintains smooth transitions and realistic movement between frames.
- Free tiers have limitations, such as watermarks, credit limits, and low-resolution outputs.

Module 7.2: Ethics in Action

- Text-to-video transformation may unintentionally produce disturbing or biased content from vague prompts.
- Developers must implement moderation filters to prevent inappropriate results.
- Users are responsible for reviewing and validating AI-generated video before distribution.

Module 7.2: Techie Dive

- Advanced systems use text-conditioned diffusion models generating frames from both the visual context and the prompt.
- Temporal coherence (maintaining consistent shape, style, position across frames) is a key challenge.
- They often reuse methods similar to image generation tools but add temporal awareness and motion smoothing across frames.
- Some tools accept storyboard-style descriptions for specifying scene sequences.

Module 7.2: Business Lens

- Early adopters are already integrating text-to-video transformation into advertising, education, and previsualization processes.
- The tools reduce production costs and accelerate creative timelines significantly.
- The outputs are typically short and sometimes lack polish, so they are best for drafts or idea testing.
- Quality control and human review remain essential for professional applications.

Module 7.3: Generating Avatars and Objects

- AI avatars are digital representations.
- They can be:
 - Photorealistic
 - Cartoonish
 - Fantasy
 - Animated
- There are two avatar types: static (profile photos) or animated (talking heads).
- Use cases:
 - Explainer videos
 - Customer service bots
 - Visual storytelling
 - Gaming
- Video inpainting adds or replaces objects inside video using generative fill.
- Inpainting requires consistent changes across many frames, so it's computationally expensive.
- Free tools often struggle with inpainting, have long wait times, have credit limits, and have low-resolution outputs.

Module 7.3: Ethics in Action

- Avatars raise concerns related to realism, privacy, consent, impersonation, and deepfakes.
- Always disclose when avatars are AI-generated and use ethical upload tools.
- Avoid cultural stereotyping; consider skin tone, attire, and inclusivity.
- Do not use avatars deceptively by creating representations that impersonate others.

Module 7.3: Techie Dive

- Avatar generation uses GANs (generative adversarial networks) and diffusion models.
- Some tools animate avatars using audio-based lip-syncing or motion capture.
- Object insertion combines semantic segmentation, depth mapping, and motion tracking.
- Video inpainting processes hundreds or thousands of individual frames for consistency.

Module 7.3: Business Lens

- AI avatars reduce production costs, as they enabled the creation of videos without the need for actors, props, or crews.
- Businesses must prioritize brand authenticity, visual consistency, and ethical standards.
- Production-ready results typically require paid subscriptions for quality and control.
- Free tools are useful for learning but aren't good for obtaining professional-grade outputs.

Module 7.4: Subtitles and Narration

- AI-generated subtitles automatically transcribe speech, sync with video, and can translate into other languages.
- Hardcoded subtitles are burned-in, always visible, and cannot be toggled off.
- Softcoded subtitles are separate files (.srt, .vtt) that can be toggled on/off.
- AI narration uses text-to-speech (TTS) systems to convert typed words into speech.
- What can be customized?
 - Voice gender
 - Accent
 - Tone
 - Speed
 - Inflection
 - Language
- The benefits of using subtitles are improved accessibility, engagement, and global reach.

Module 7.4: Ethics in Action

- Auto-subtitles and AI narration can misinterpret accents, dialects, or nonstandard speech.
- Inaccuracies frustrate or misinform viewers, so users must review and edit outputs before publishing them.
- It's essential to validate output accuracy in educational and professional contexts.

Module 7.4: Techie Dive

- Subtitles use automatic speech recognition (ASR) to detect words and insert timestamps.
- Narration is generated using TTS models trained on human speech patterns.
- Some TTS systems use deep learning to match emotional tone and natural rhythm.
- Subtitle formats store dialogue and timing information in structured files.

Module 7.4: Business Lens

- Subtitles improve SEO and viewer retention, especially for mobile or silent viewing.
- Narration improves accessibility for people with visual impairments or reading difficulties.
- Together, they enable global audience reach quickly and cost effectively.
- Multilingual content expands market reach without proportional cost increases.

Module 7.5: Audio – Text-to-Speech Transformations and Cloning

- TTS tools convert written text into natural-sounding speech with customizable options.
- TTS use cases:
 - Voiceovers
 - Accessibility
 - Synthetic announcements
- Voice cloning creates speech from a small sample of a specific person's voice.
- Cloning applications:
 - Personalized assistants
 - Branding continuity
 - Historical voiceovers
- Common pitfalls:
 - Consent violations
 - Robotic tone
 - Assuming commercial rights
 - Mismatched voices

Module 7.5: Ethics in Action

- Voice cloning raises concerns related to consent, privacy, deception, and the labor impact on voice actors.
- Using someone's voice without permission can be invasive or illegal.
- Some platforms offer ethical voice licensing with consent and royalties.
- Disclosing when voices are AI-generated maintains user confidence.

Module 7.5: Techie Dive

- Modern TTS uses deep learning models (for example, Tacotron, FastSpeech, and VITS).
- Models map phonemes (the smallest sound units) to realistic speech patterns.
- Voice cloning uses few-shot learning to mimic specific voices from small samples.
- Systems replicate the pitch, tone, inflection, and natural rhythm of speech.

Module 7.5: Business Lens

- TTS helps maintain a consistent brand voice across platforms.
- Benefits:
 - Faster production
 - Cost savings on voice actors and studios
- Trust requires transparency about AI-generated voices in customer-facing roles.
- Always check the licensing terms before using a cloned voice in monetized or public projects.

Module 7.6: Audio – SFXs, Songs, and Mixes

AI can generate:

- Sound effects (SFX): short, intentional audio cues (creaky doors, cash registers, sci-fi beeps).
- Ambient soundscapes: background texture evoking mood (ocean waves, rain, city traffic).
- Songs: AI music generators can create full compositions with customizable tempo, emotion, and instruments.
- Polished audio mixes: AI can use compression, ducking, EQ adjustment, mastering, noise reduction, peak normalization, and sound leveling and volume normalization (ensure consistent volume across tracks).

Module 7.6: Ethics in Action

- Auto-generated music could replace human musicians and engineers.
- Voice styles and music genres copied without consent raise legal concerns.
- Licensing agreements and artist protection policies are still evolving.
- Creators must act responsibly when using cloned voices resembling those of real artists.

Module 7.6: Techie Dive

- AI audio tools use diffusion, GANs, and transformer models trained on labeled audio.
- Tools analyze spectral features and patterns across genres and environmental sounds.
- Audio mastering uses spectral analysis and ML for studio-quality finishing.
- Processes include limiting, EQ adjustment, and background noise suppression.

Module 7.6: Business Lens

- AI-generated sound accelerates production times and cuts costs significantly.
- Entrepreneurs can create custom jingles, music, and SFX without hiring composers.
- For polished releases, combine AI tools with production knowledge or expert consultation.
- Many platforms restrict commercial use, so review licensing terms carefully.

Key Takeaways

- AI tools transform static images into dynamic multimedia through colorization and animation.
- Text-to-video transformation creates clips from descriptions but requires careful prompting and quality review.
- AI avatars enable the creation of compelling videos but there are ethical considerations related to consent and authenticity.
- Subtitles and narration improve accessibility and global reach as long as there is human review for accuracy.
- Voice cloning offers powerful capabilities but raises serious ethical concerns around consent.
- AI music generation accelerates production times, but licensing and quality control remain essential.
- All AI-generated content requires transparency, ethical use, and human oversight.
- Business applications offer cost savings but must balance efficiency against responsible use.