AMCC 5160 · LECTURE 01

When does an image become a world?

Slide-by-slide transcript · 60-slide edition

17,421	105.7 min	120.7 min
spoken words	natural-voice narration	lecture with activities

Harry Yang · Fall 2026/27

AI-Driven Animation and Video Generation

How to use this transcript

Each section corresponds to one slide and one independently generated ElevenLabs clip. The recorded lecture advances only when that measured clip ends. Twelve activity slides add explicit 60- or 90-second working periods; those pauses are labeled on screen and omitted from the podcast edition.

Voice settings: cy45 · Eleven Multilingual v2 · speed 0.95 · stability 0.65 · similarity 0.81 · style exaggeration 0 · speaker boost on.

SLIDE 01 · PRESERVE-ONLY OPENING TITLE

From generative media to multimodal simulation

Source frame 01 · 342 words

Welcome to Lecture One of AMCC 5160.

Welcome to AMCC 5160. I want to begin with a question rather than a product demo: when does an image become a world? A still image gives us a surface. Animation gives us time. A world gives us consequences: things persist, collide, remember, respond, and surprise us. Generative systems are moving across that spectrum, but the artistic problem is not solved by technical capability alone. Our job in this course is to connect two forms of judgment: how the system works, and why a creative choice matters. Look at the threshold in this image. On one side is a bounded frame; on the other is a space that seems explorable. For the next two hours, we will use that threshold as our organizing metaphor. We will build a simple mental model of generative systems, practice reading their failures, and end by framing a studio experiment of your own.

The opening question establishes the course's central claim: a generated frame becomes a world only when change has consequences. Persistence, causality, memory, and point of view matter more than surface polish. Keep this distinction active whenever a model demo appears impressive.

Read the slide as a set of decisions: 1, From generative media to multimodal simulation; 2, Harry Yang · Fall 2026/27. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. Imagine a doorway that glows beautifully. It becomes a world only when entering changes what the character can see, remember, or do.

A useful way to test this is to ask four questions. What is the intended audience experience? Which system decision most directly shapes it? What visible or audible evidence would count as success? And what single comparison would teach us the most if the first attempt fails? Those questions keep the work specific without reducing creative practice to a recipe.

Keep that distinction in mind as we move to the next slide.

SLIDE 02 · COURSE LOGISTICS

Welcome to AMCC 5160

Source frame 02 · 315 words

Now we turn to Welcome to AMCC 5160.

Before we begin, here are the practical details. AMCC 5160 is a three-credit course meeting on Tuesdays from 18:30 to 21:20 in MIL13002. The class quota is fifty students. Canvas is the official channel for announcements, submissions, and any time-sensitive changes; this public course site is the readable front door for the syllabus, schedule, sign-ups, and teaching materials. This course combines visual examples, technical concepts explained from first principles, guided exercises, discussion, and project critique. You are not expected to enter with equal confidence in art and computing. You are expected to test ideas, document your process, and become increasingly fluent across both lenses. Please keep questions active during class. If a concept feels opaque, ask for a different explanation. If a creative example feels culturally narrow or ethically incomplete, name that. The quality of this course depends on building a room where technical curiosity and artistic disagreement can coexist productively.

The logistics are also a contract about how the class will work. Canvas carries official changes and submissions; the public site makes the learning materials easy to revisit. The room is a studio-seminar, so questions, tests, and critique are part of the course rather than interruptions to it.

Read the slide as a set of decisions: 1, Fall 2026/27 · 3 credits; 2, Tuesdays · 18:30–21:20; 3, Room MIL13002 · class quota 50; 4, Canvas is the official channel for materials and announcements. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. If a room or deadline changes, Canvas is the authoritative record; if a student wants to revisit today's concepts, the public site should be the clearer reading experience.

The next step is to turn this idea into a more precise production decision.

SLIDE 03 · TEACHING TEAM

Your teaching team is here to help you connect the lenses

Source frame 03 · 350 words

Now we turn to Your teaching team is here to help you connect the lenses.

I am Harry Yang, the instructor for the course, and Yaofu Liu is our teaching assistant. My work sits across generative media, computer vision, and creative AI. You can find my research and projects at hyang. org. Yaofu will support technical setup, studio questions, and course logistics. Use email for individual matters and Canvas for course-wide information. For project help, the most useful message is specific: include what you are trying to make, what you attempted, one representative result or error, and the decision you are facing. That lets us help you move the project rather than merely diagnose a tool. We will announce consultation arrangements on Canvas. You do not need to wait until a project is “good” to ask for feedback. Early questions about intention, feasibility, reference, or workflow often prevent wasted effort later.

The teaching team connects artistic intention with technical diagnosis. Bring us unfinished evidence: a prompt, a failed shot, a comparison, or an edit. Specific artifacts let us help you much more effectively than a general statement that the model did not work.

Read the slide as a set of decisions: 1, Instructor; 2, Teaching assistant; 3, Harry Yang; 4, Assistant Professor, AMC; 5, yangharry@ust.hk; 6, hyang.org; 7, Yaofu Liu. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. Bring two ten-second clips that differ only in camera motion; the team can then diagnose whether the problem is prompt ambiguity, temporal control, or editing.

A useful way to test this is to ask four questions. What is the intended audience experience? Which system decision most directly shapes it? What visible or audible evidence would count as success? And what single comparison would teach us the most if the first attempt fails? Those questions keep the work specific without reducing creative practice to a recipe.

Use this as a question you can return to when an output surprises you.

SLIDE 04 · ASSESSMENT STRUCTURE

Assessment rewards process, craft, judgment, and explanation

Source frame 04 · 299 words

Now we turn to Assessment rewards process, craft, judgment, and explanation.

The assessment structure has four parts. The weekly presentation is ten percent. Each of the three assignments is worth twenty percent, for sixty percent total. Their due dates are fixed on the syllabus, while the project content remains to be announced. The final presentation is thirty percent and asks you to select one assignment, refine it, and present it as your strongest showcase. Across components, we value four qualities: conceptual clarity, formal and technical craft, evidence of iteration, and the ability to explain decisions responsibly. Polish matters, but polish alone is not enough. A visually impressive clip with no clear intention or process can be weaker than a smaller experiment that asks a sharp question and learns from failure. The final presentation should make the chain of decisions visible: problem or artistic intention, audience, methods, evidence, one representative failure or unresolved limitation, your own contribution, use of AI tools, and what changed through iteration.

The assessment weights reward sustained practice. Three projects contribute twenty percent each, the weekly presentation contributes ten percent, and the final presentation contributes thirty percent. Project content remains to be announced, but the due dates and the standard of evidence are clear.

Read the slide as a set of decisions: 1, Weekly presentation · 10%; 2, Assignments 1–3 · 20% each; 3, Final presentation · 30%. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. A visually ambitious project with no comparison evidence may be less convincing than a smaller project that demonstrates why each iteration changed.

What matters is not the label but the quality of the next decision it enables.

SLIDE 05 · ASSESSMENT CRITERIA

Strong projects connect intention, evidence, craft, and responsibility

Source frame 04 · 276 words

Now we turn to Strong projects connect intention, evidence, craft, and responsibility.

The assessment structure has four parts. The weekly presentation is ten percent. Each of the three assignments is worth twenty percent, for sixty percent total. Their due dates are fixed on the syllabus, while the project content remains to be announced. The final presentation is thirty percent and asks you to select one assignment, refine it, and present it as your strongest showcase.

A strong project is not simply attractive or technically difficult. It proposes an intention, demonstrates control through comparisons, and accounts for sources, consent, and context. These dimensions reinforce one another: responsible documentation often makes the creative reasoning sharper and easier to defend.

Read the slide as a set of decisions: 1, Intention · a clear experience or question; 2, Evidence · iteration, comparison, and technical control; 3, Responsibility · sources, consent, context, and explanation. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. For a character study, intention may be unease, evidence may be three controlled lighting tests, and responsibility may include permission for the face reference.

A useful way to test this is to ask four questions. What is the intended audience experience? Which system decision most directly shapes it? What visible or audible evidence would count as success? And what single comparison would teach us the most if the first attempt fails? Those questions keep the work specific without reducing creative practice to a recipe.

This gives us a practical bridge between creative judgment and technical control.

SLIDE 06 · WEEKLY PRESENTATION BRIEF

The weekly presentation turns your past work into shared course knowledge

Source frame 05 · 333 words

Now we turn to The weekly presentation turns your past work into shared course knowledge.

In Weeks 2 through 11, students give a short weekly presentation worth ten percent. Share one current or past artistic, design, research, or technical project. The work does not need to use generative AI. In fact, projects from photography, animation, installation, performance, design, coding, research, or another medium can give the class a richer foundation. Use roughly ten minutes. Begin with context: what was the situation and why did you make it? State the creative intention or research question. Explain the technical approach at the level necessary for us to understand the choices. Show the work rather than describing it abstractly. End with one lesson, limitation, or open question that connects to this course. Sign up for an empty weekly slot in the Google Sheet. Enter your name, student ID, and topic. The aim is not to prove mastery; it is to give the class a concrete practice to discuss and to help us learn what each person brings into the room.

The weekly presentation is a compact form of peer teaching. Share past work, explain one decision deeply, and end with a useful lesson or question. Ten focused minutes are enough when the example is concrete and the audience can see the evidence.

Read the slide as a set of decisions: 1, Around 10 minutes; 2, Share one current or past artistic, design, research, or technical project; 3, Explain context, creative intention, and technical approach; 4, End with one lesson, limitation, or question; 5, Choose a slot in the sign-up sheet. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. A student might show an earlier game prototype, explain how one interaction rule shaped player behavior, and connect that lesson to generated video.

Keep that distinction in mind as we move to the next slide.

SLIDE 07 · WEEKLY PRESENTATION SIGNUP

Choose an empty slot and make your plan legible

Source frame 05 · 306 words

Now we turn to Choose an empty slot and make your plan legible.

In Weeks 2 through 11, students give a short weekly presentation worth ten percent. Share one current or past artistic, design, research, or technical project. The work does not need to use generative AI. In fact, projects from photography, animation, installation, performance, design, coding, research, or another medium can give the class a richer foundation. Use roughly ten minutes. Begin with context: what was the situation and why did you make it?

The sign-up sheet is intentionally open-ended. Choose an empty slot, enter your own name, student ID, and topic, and never replace another student's entry. If the sheet is view-only, request access from the file rather than creating a duplicate schedule.

Read the slide as a set of decisions: 1, Open the shared Google Sheet from the course site; 2, Use the Weekly presentation sign-up tab; 3, Enter your name, student ID, and topic; 4, Do not overwrite another student’s row; 5, If access is blocked, request access from the sheet. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. If week six has three empty rows, choose one row and enter your information there; the schedule becomes reliable because every student follows the same rule.

A useful way to test this is to ask four questions. What is the intended audience experience? Which system decision most directly shapes it? What visible or audible evidence would count as success? And what single comparison would teach us the most if the first attempt fails? Those questions keep the work specific without reducing creative practice to a recipe.

The next step is to turn this idea into a more precise production decision.

SLIDE 08 · LEARNING RHYTHM

Each class moves through four modes of attention

Source frame 06 · 292 words

Now we turn to Each class moves through four modes of attention.

Our weekly rhythm has four modes. We look closely at art, media, and contemporary systems. We build a technical mental model that is accurate enough to guide action. We make a controlled experiment rather than chasing random outputs. Then we critique and document what happened. This structure draws from strong art-and-technology teaching. CMU’s studio courses combine computational craft, cultural context, looking outward, and critique. NYU ITP presents machine learning through accessible creative applications and prototypes. MIT’s Creative AI work pairs hands-on making with questions about authorship, bias, and ownership. Stanford graphics teaching emphasizes connecting generative controls to real artist workflows. The result is a course where reading, coding, directing, editing, discussion, and presentation all count as forms of inquiry. You do not learn a model by memorizing its name. You learn it by forming a hypothesis, making something, observing what the system permits or resists, and explaining why that matters.

Each meeting cycles through four modes: seeing, understanding, testing, and reflecting. The movement matters. Looking without testing produces taste without evidence; testing without reflection produces outputs without learning. Your documentation connects the four modes into a repeatable practice.

Read the slide as a set of decisions: 1, Look closely; 2, Understand the system; 3, Make a controlled experiment; 4, Critique and document. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. A continuity failure first becomes an observation, then a model-level explanation, then a controlled regeneration, and finally a note in the process log.

Use this as a question you can return to when an output surprises you.

SLIDE 09 · OPENING POLL

You already know more than you think

Source frame 07 · 371 words

Now we turn to You already know more than you think.

Before we define any technical term, let us start from perception. Imagine three short generated clips: one is visually polished but emotionally empty, one contains artifacts but has a strong point of view, and one is technically consistent but feels generic. Which one would you remember tomorrow? Take thirty seconds and choose the evidence you would use: composition, continuity, material logic, point of view, or something else. Then compare with the person next to you. The aim is not to agree. The aim is to notice that you already possess criteria. Technical vocabulary helps us name causes; artistic vocabulary helps us name effects. Neither is sufficient alone. Ask for three answers and write only the nouns on the board. We will return to them near the end. If students say “quality,” ask: what visible property makes you call it quality? If they say “realistic,” ask whether realism is always the goal. This establishes the course habit of moving from vague approval to observable evidence.

The opening poll is diagnostic, not evaluative. Choose one result that feels intentional and identify visible evidence for that claim. Avoid naming a tool or guessing a prompt; begin with composition, continuity, material behavior, rhythm, or point of view.

Read the slide as a set of decisions: 1, Which result feels intentional—and what evidence tells you that?; 2, Composition; 3, Continuity; 4, Material logic; 5, Point of view. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. One clip may feel intentional because the camera waits before moving, while another feels accidental because the subject drifts without narrative cause.

A useful way to test this is to ask four questions. What is the intended audience experience? Which system decision most directly shapes it? What visible or audible evidence would count as success? And what single comparison would teach us the most if the first attempt fails? Those questions keep the work specific without reducing creative practice to a recipe.

What matters is not the label but the quality of the next decision it enables.

SLIDE 10 · LECTURE OUTCOMES

By the end, you can see the system and the artwork

Source frame 08 · 335 words

Now we turn to By the end, you can see the system and the artwork.

These are our four outcomes. First, you will be able to explain a generative pipeline without advanced mathematics. The purpose is not simplification for its own sake; it is to give you a durable map that survives changing products. Second, you will learn to critique an output with two lenses. “I like it” is honest but incomplete. “The motion prior collapses at the cut, which weakens the intended unease” connects mechanism and meaning. Third, we will build a failure vocabulary. Failure is not only a defect. Sometimes it is evidence about the model; sometimes it becomes the material of the artwork. Finally, you will leave with a small studio experiment, not merely notes. This learning design follows a studio pattern used across leading art-and-technology courses: accessible technical explanation, hands-on making, looking outward to prior work, and critique. CMU’s digital-media courses emphasize computational craft within cultural practice; NYU ITP combines accessible ML concepts with creative experimentation; MIT’s Creative AI curriculum pairs making with authorship and ethics.

By the end of the lecture, you should be able to move between system and artwork. The goal is not memorizing every architecture. It is learning enough structure to locate failure, enough critical language to explain an intention, and enough experimental discipline to make the next iteration informative.

Read the slide as a set of decisions: 1, Explain generation as data → representation → model → sampling; 2, Critique an output in artistic and technical language; 3, Identify failures without reducing them to “bad AI”; 4, Frame one responsible studio experiment. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. After today, a flickering face should prompt a temporal-consistency test rather than an immediate switch to a random new style prompt.

This gives us a practical bridge between creative judgment and technical control.

SLIDE 11 · TWO COURSE LENSES

This is one course with two inseparable lenses

Source frame 09 · 371 words

Now we turn to This is one course with two inseparable lenses.

We are not alternating between an art course and a technology course. We are using both lenses on the same object. Consider a camera move generated from a still image. The art lens asks: why this move, at this speed, from this point of view? What does it do to attention? The technology lens asks: how was motion conditioned, what remains consistent, and where does the model invent unseen geometry? The strongest projects make these questions interdependent. A technical constraint can become an aesthetic rule. A creative intention can determine which control method is necessary. The point is not to become equally expert in every domain. The point is to make choices that are legible. CMU describes creative coding as a studio art practice whose medium is software, while NYU's Machine Learning for Artists emphasizes integrating available tools into artistic practice without requiring deep mathematical prerequisites. We adopt that same bridge, but extend it from interactive media into contemporary image, animation, video, and world models.

Art and technology are two readings of the same decision chain. An artistic choice has technical conditions, while a technical setting creates an aesthetic and political effect. Good direction asks both what experience is intended and what system behavior makes that experience possible.

Read the slide as a set of decisions: 1, Art; 2, Technology; 3, Intention; 4, Form & rhythm; 5, Reference & context; 6, Audience experience; 7, Data & representation. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. Choosing a low camera, narrow lens, and slow approach is simultaneously a technical configuration and a statement about power.

A useful way to test this is to ask four questions. What is the intended audience experience? Which system decision most directly shapes it? What visible or audible evidence would count as success? And what single comparison would teach us the most if the first attempt fails? Those questions keep the work specific without reducing creative practice to a recipe.

Keep that distinction in mind as we move to the next slide.

SLIDE 12 · MINI CRITIQUE

Intent is visible in the decisions you can defend

Source frame 10 · 308 words

Now we turn to Intent is visible in the decisions you can defend.

A critique begins by reconstructing intention, not by guessing whether a tool was difficult. Ask what the work is trying to do. Then identify the formal choice that carries that intention: framing, pacing, repetition, texture, sound, scale, or interaction. Only then connect it to the technical decision. For example: “The slow lateral camera motion turns the generated architecture into a memory rather than a location. The image-to-video model preserves the silhouette but invents unstable windows, so the space feels dreamlike. That instability supports the concept until the identity of the central figure also drifts; then the focus becomes accidental. ” This is not praise or blame. It is a causal account. In pairs, take one work you remember—your own or someone else’s—and complete the three questions on screen. You do not need to name the artist. The exercise is about moving from taste to evidence.

Critique traces decisions rather than issuing a taste verdict. Replace 'it looks good' with evidence: where does attention go, what persists across time, what relation feels coherent, and what breaks the proposed experience? This makes feedback actionable without pretending that interpretation is objective.

Read the slide as a set of decisions: 1, “It looks good” is a reaction. A critique traces choices; 2, What is the work trying to do?; 3, What formal choice carries that intention?; 4, What technical choice enables or weakens it?. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. Instead of saying a shot is cinematic, point to the delayed reveal, stable silhouette, and sound cue that redirect attention.

The next step is to turn this idea into a more precise production decision.

SLIDE 13 · MEDIA LINEAGE

Generative media belongs to a longer history of rule-making

Source frame 11 · 369 words

Now we turn to Generative media belongs to a longer history of rule-making.

Generative media did not appear from nowhere. The camera automated part of image capture but did not remove authorship; it reorganized it around framing, timing, selection, printing, and circulation. Animation constructed time frame by frame. Procedural and generative artists wrote rules that could produce families of forms. Machine learning shifts part of the rule-making into learned statistical regularities. World models aim to predict how scenes change under actions. Each shift creates a familiar anxiety: if the machine performs more of the visible labor, where does the artist go? Historically, authorship moves rather than disappears. It can move into system design, dataset choice, constraint, selection, sequencing, interaction, context, or critique. We should also resist a story of smooth progress. New media do not replace old ones; they coexist. A hand-drawn mark, a procedural system, and a diffusion model can all appear in one work. The useful question is not “which technology wins? ” but “what relation between technologies produces the intended experience?

Generative media extends a history of rule-making—from camera capture and animation constraints to procedural systems and machine learning. New tools change where rules are learned and where authors intervene, but they do not eliminate framing, editing, or cultural context.

Read the slide as a set of decisions: 1, Camera: capture; 2, Animation: construct time; 3, Procedural art: encode rules; 4, Machine learning: learn regularities; 5, World models: predict consequences. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. Procedural animation also separated rules from individual frames; learning systems change how those rules are acquired and how predictable they remain.

A useful way to test this is to ask four questions. What is the intended audience experience? Which system decision most directly shapes it? What visible or audible evidence would count as success? And what single comparison would teach us the most if the first attempt fails? Those questions keep the work specific without reducing creative practice to a recipe.

Use this as a question you can return to when an output surprises you.

SLIDE 14 · AUTHORSHIP ACROSS TECHNICAL SYSTEMS

Learning systems move authorship into curation and direction

Source frame 11 · 224 words

Now we turn to Learning systems move authorship into curation and direction.

Generative media did not appear from nowhere. The camera automated part of image capture but did not remove authorship; it reorganized it around framing, timing, selection, printing, and circulation. Animation constructed time frame by frame. Procedural and generative artists wrote rules that could produce families of forms. Machine learning shifts part of the rule-making into learned statistical regularities. World models aim to predict how scenes change under actions.

Authorship moves across the pipeline. Choosing an archive, objective, architecture, interface, and final selection all determines what can be expressed. The visible output is therefore the end of many authored constraints, including decisions made by people the artist may never meet.

Read the slide as a set of decisions: 1, Training data: define the archive; 2, Objective: define what counts; 3, Architecture: define the relations; 4, Interface: define available actions; 5, Selection: define what enters culture. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. A product may hide the seed, force a default safety filter, and rank outputs automatically—each decision affects what the artist can select.

What matters is not the label but the quality of the next decision it enables.

SLIDE 15 · VIDEO GENERATION CHALLENGE

Video is not an image with more frames

Source frame 12 · 355 words

Now we turn to Video is not an image with more frames.

A convincing still image can be judged at one instant. Video requires a relation across instants. The system must preserve identity and geometry while allowing meaningful change. It must decide what is rigid, what deforms, what enters or leaves the frame, how the camera moves, and how sound relates to action. This is why a beautiful first frame does not guarantee a useful clip. A model can optimize local plausibility while losing the larger event. A hand may look correct in each frame but attach differently over time. A shadow may be plausible moment to moment but inconsistent with a moving light source. For artists, this means direction must include time. “A blue room” is an image description. “The room slowly forgets its geometry as the speaker remembers a childhood story” is a temporal proposition. It contains an arc, a rate of change, and a relation between form and meaning. Throughout the course, we will treat time as material—not just duration.

Video generation must coordinate appearance, time, and action. A sharp single frame can still fail as video if identity drifts, motion has no cause, or the cut carries no meaningful change. Evaluate the temporal proposition, not a collection of attractive stills.

Read the slide as a set of decisions: 1, Image quality + temporal consistency + meaningful change + sound + control. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. A perfect portrait that changes identity after two seconds is a strong image and a weak shot.

A useful way to test this is to ask four questions. What is the intended audience experience? Which system decision most directly shapes it? What visible or audible evidence would count as success? And what single comparison would teach us the most if the first attempt fails? Those questions keep the work specific without reducing creative practice to a recipe.

This gives us a practical bridge between creative judgment and technical control.

SLIDE 16 · TEMPORAL PROPOSITION

A temporal proposition gives change a reason

Source frame 12 · 206 words

Now we turn to A temporal proposition gives change a reason.

A convincing still image can be judged at one instant. Video requires a relation across instants. The system must preserve identity and geometry while allowing meaningful change. It must decide what is rigid, what deforms, what enters or leaves the frame, how the camera moves, and how sound relates to action. This is why a beautiful first frame does not guarantee a useful clip.

A temporal proposition names what changes, how quickly, from whose point of view, and why the change matters. This sentence is a directing tool. If it is vague, the generation may move, but the movement will not necessarily create a scene.

Read the slide as a set of decisions: 1, What changes + at what rate + under whose point of view + why. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. 'The room slowly forgets the visitor from the visitor's point of view, until the exit disappears' is a temporal proposition; 'surreal room' is not.

Keep that distinction in mind as we move to the next slide.

SLIDE 17 · TECHNICAL PIPELINE MAP

A durable map: data → representation → model → sampling

Source frame 13 · 351 words

Now we turn to A durable map: data → representation → model → sampling.

Here is the most important technical map for today. Data supplies examples and therefore boundaries. Representation determines how those examples become manageable signals or tokens. The model learns regularities over those representations. Sampling turns a space of possibilities into one particular output. Products hide these layers, but creative control often means intervening in one of them. Curating a dataset changes the learned world. Choosing a latent or token representation changes what variations are easy. Conditioning and architecture change what relationships can be controlled. Sampling settings and iteration change which path is taken through possibility. This map is deliberately broad. It can describe a text-to-image diffusion model, an autoregressive multimodal model, or a video generator with additional temporal machinery. It also gives us a diagnostic habit: when something fails, ask which layer could plausibly be responsible. Do not jump directly from visible artifact to a mystical claim that “the AI is confused.

The technical map separates data, representation, model, and sampling. These layers interact, yet the separation is useful because each produces different symptoms and different remedies. Diagnose the layer before rewriting everything at once.

Read the slide as a set of decisions: 1, Data shapes the examples; 2, Representation compresses what matters; 3, The model learns regularities; 4, Sampling turns possibility into one output. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. If anatomy fails in every frame, investigate representation; if anatomy is correct but changes identity, investigate temporal persistence.

A useful way to test this is to ask four questions. What is the intended audience experience? Which system decision most directly shapes it? What visible or audible evidence would count as success? And what single comparison would teach us the most if the first attempt fails? Those questions keep the work specific without reducing creative practice to a recipe.

The next step is to turn this idea into a more precise production decision.

SLIDE 18 · DATA AS CULTURAL ARCHIVE

A dataset is a cultural archive with edges

Source frame 14 · 306 words

Now we turn to A dataset is a cultural archive with edges.

Data is not neutral fuel. It is a cultural archive assembled through institutions, platforms, labor, permissions, and historical inequality. A model learns what appears often, what appears together, and what is labeled or described. That makes some visual languages easy to produce and others difficult. When a prompt seems to “work,” it may be because the requested subject and style are densely represented. When it fails, the problem may not be your wording. The concept may be rare, poorly labeled, or represented through stereotypes. This creates an artistic opportunity and an ethical responsibility. You can work with the grain of the model, against it, or expose the grain itself—but you should know which you are doing. Ask of any system: whose images, whose language, and whose labor made this output statistically likely? You may not always obtain complete documentation, but the absence of transparency is itself relevant evidence. MIT’s Creative AI curriculum explicitly pairs creative making with questions of ownership, bias, and the consequences of generated media.

A dataset is a cultural archive with uneven coverage. What appears frequently becomes cheap for a model to reproduce; what is absent, mislabeled, or stereotyped becomes expensive to express. Technical fluency includes recognizing those edges rather than blaming every resistance on prompt wording.

Read the slide as a set of decisions: 1, What is common becomes easy. What is absent becomes expensive—or impossible. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. A dataset rich in glossy product photography may render plastic packaging easily but struggle with undocumented local craft techniques.

Use this as a question you can return to when an output surprises you.

SLIDE 19 · DIAGNOSING DATA BOUNDARIES

When a concept resists, diagnose the archive—not only the prompt

Source frame 14 · 266 words

Now we turn to When a concept resists, diagnose the archive—not only the prompt.

Data is not neutral fuel. It is a cultural archive assembled through institutions, platforms, labor, permissions, and historical inequality. A model learns what appears often, what appears together, and what is labeled or described. That makes some visual languages easy to produce and others difficult. When a prompt seems to “work,” it may be because the requested subject and style are densely represented.

When a concept resists generation, inspect frequency, labels, consent, stereotype, and absence. A prompt can only condition what the system has learned to make plausible. Sometimes the right response is a different reference strategy, a composite workflow, or a conscious reframing of the idea.

Read the slide as a set of decisions: 1, Frequency, labeling, consent, stereotype, and absence shape the cost of expression. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. When a ritual object is repeatedly stereotyped, a reference-controlled composite may be more responsible than adding increasingly exotic adjectives.

A useful way to test this is to ask four questions. What is the intended audience experience? Which system decision most directly shapes it? What visible or audible evidence would count as success? And what single comparison would teach us the most if the first attempt fails? Those questions keep the work specific without reducing creative practice to a recipe.

What matters is not the label but the quality of the next decision it enables.

SLIDE 20 · REPRESENTATION CHOICES

Representation decides what the system can change smoothly

Source frame 15 · 287 words

Now we turn to Representation decides what the system can change smoothly.

A representation is the form in which the system handles media. Pixels are direct, but high-resolution video contains enormous redundancy. Latent representations compress visual information into a smaller space. Discrete visual tokens turn image or video content into sequences that can be modeled alongside language. The creative consequence is that representation creates neighborhoods. In a useful latent space, nearby points often correspond to related appearances or concepts. That makes interpolation, editing, and controlled variation possible. But compression also discards information. Fine texture, exact typography, identity details, or unusual structures may be weakly represented. Think of a representation as a set of handles. If a handle exists for camera angle, material, pose, or semantic content, the system can vary it coherently. If multiple properties are entangled, changing one may disturb another. This is why a request to “keep everything identical but change only the reflection” can be much harder than it sounds.

Representation determines what the system can change smoothly. Pixels preserve direct detail, latents compress visual structure, and tokens describe sequences and relations. Each choice creates different strengths, blind spots, and costs for motion and control.

Read the slide as a set of decisions: 1, Pixels: direct but expensive; 2, Latents: compressed visual factors; 3, Tokens: sequences that models can predict. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. A latent representation can interpolate appearance efficiently, while a token sequence may better expose ordered relations such as action and dialogue.

This gives us a practical bridge between creative judgment and technical control.

SLIDE 21 · LATENT SPACE

Latent space is less a library than a landscape

Source frame 16 · 342 words

Now we turn to Latent space is less a library than a landscape.

This image gives us a useful metaphor. A library suggests fixed items waiting to be retrieved. A landscape suggests continuous movement, neighborhoods, paths, and inaccessible regions. A latent representation does not usually store a complete image at each point. It organizes learned features so that the model can traverse combinations. Prompting is therefore not a database query with one correct result. It is an act of navigation under uncertainty. Seed, conditioning, guidance, reference images, masks, and sampling steps all influence the route. Artists can treat this uncertainty as noise to eliminate or as material to choreograph. Notice that navigation still needs direction. More samples are not automatically more authorship. A strong process defines what counts as progress. You might search for a specific emotional rhythm, a recurring visual motif, or a productive kind of failure. Without criteria, exploration becomes scrolling. With criteria, it becomes research.

Latent space is a learned coordinate system rather than a literal landscape. Nearby points are useful only because the model has organized patterns of appearance there. Moving through that space can produce meaningful variation, but it can also expose the assumptions of the training archive.

Read the slide as a set of decisions: 1, Creation becomes navigation, not retrieval. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. Moving between two latent points may smoothly blend faces, yet the path can also reveal which identities the archive has compressed together.

A useful way to test this is to ask four questions. What is the intended audience experience? Which system decision most directly shapes it? What visible or audible evidence would count as success? And what single comparison would teach us the most if the first attempt fails? Those questions keep the work specific without reducing creative practice to a recipe.

Keep that distinction in mind as we move to the next slide.

SLIDE 22 · DIFFUSION MODEL

Diffusion learns to reverse corruption

Source frame 17 · 291 words

Now we turn to Diffusion learns to reverse corruption.

The diffusion idea can be understood without equations. During training, examples are progressively corrupted with noise. The model learns how to predict a direction back toward a plausible clean example, conditioned by text or other signals. At generation time, we begin with noise and repeatedly apply those learned corrections. The important word is repeatedly. Generation is a trajectory, not a single lookup. Early steps establish broad composition; later steps refine texture and detail, though implementations vary. Guidance can push the trajectory toward a condition, but excessive guidance may reduce diversity or create harsh artifacts. For creative practice, this suggests several controls: the initial noise, the number and schedule of steps, the strength of conditioning, masks, reference images, and interventions during denoising. It also explains why small changes can lead to substantially different outputs. We are navigating a probabilistic process whose early decisions cascade.

Diffusion learns a denoising direction by repeatedly corrupting examples and predicting how to reverse that corruption. At generation time, an initially noisy state becomes structured through many conditional updates. The process explains both flexibility and the possibility of instability across steps or frames.

Read the slide as a set of decisions: 1, Add noise during training; 2, Learn a denoising direction; 3, Start from noise at generation time; 4, Iteratively reveal structure. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. The denoising trajectory can preserve a broad composition while gradually resolving material and texture, but small errors may compound across video frames.

The next step is to turn this idea into a more precise production decision.

SLIDE 23 · SAMPLING DECISIONS

Sampling turns a learned possibility into one particular image

Source frame 17 · 269 words

Now we turn to Sampling turns a learned possibility into one particular image.

The diffusion idea can be understood without equations. During training, examples are progressively corrupted with noise. The model learns how to predict a direction back toward a plausible clean example, conditioned by text or other signals. At generation time, we begin with noise and repeatedly apply those learned corrections. The important word is repeatedly. Generation is a trajectory, not a single lookup.

Sampling is where a learned distribution becomes one artifact. Guidance, randomness, schedule, and step count alter the path. Record these choices because an output cannot be understood or reproduced from the text prompt alone.

Read the slide as a set of decisions: 1, Begin with a distribution of possibilities; 2, Apply conditions and guidance; 3, Choose randomness, steps, and schedule; 4, Select one path—and keep the evidence. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. Keeping the prompt fixed while changing only the random seed tests variability; changing seed, prompt, duration, and reference at once does not.

A useful way to test this is to ask four questions. What is the intended audience experience? Which system decision most directly shapes it? What visible or audible evidence would count as success? And what single comparison would teach us the most if the first attempt fails? Those questions keep the work specific without reducing creative practice to a recipe.

Use this as a question you can return to when an output surprises you.

SLIDE 24 · FLOW MATCHING

Flow matching reframes generation as a learned motion

Source frame 18 · 280 words

Now we turn to Flow matching reframes generation as a learned motion.

Flow matching offers another useful mental picture. Imagine probability mass moving from a simple distribution toward the complex distribution of real images or videos. The model learns a velocity field: at any point and time, which direction should a sample move? You do not need the mathematics today. The practical takeaway is that modern generators may formulate sampling as solving a continuous path. This can support efficient training and generation, and it links to diffusion through related continuous-time views. Artistically, the metaphor matters because it emphasizes trajectory. An output is the endpoint of a path through possibility. Interventions can alter direction, speed, or constraints. Later in the course, when we compare schedulers and sampling methods, we will ask not only which is faster or sharper, but which yields the kinds of variation and control our work needs.

Flow matching reframes generation as learning a velocity field that transports simple noise toward data. The practical intuition is continuous direction rather than a long sequence of independent corrections. This helps us discuss speed and trajectories without turning the lecture into equations.

Read the slide as a set of decisions: 1, Instead of many noisy corrections, learn a velocity field through possibility. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. Think of many particles following a learned flow toward plausible images; the field guides direction even though any single path remains a sample.

What matters is not the label but the quality of the next decision it enables.

SLIDE 25 · TRANSFORMER RELATIONSHIPS

Transformers learn relationships across a sequence

Source frame 19 · 346 words

Now we turn to Transformers learn relationships across a sequence.

Transformers model relationships across sequences. Tokens may represent words, patches, compressed visual units, audio codes, or video elements. Attention allows the model to weigh which other tokens matter when predicting or transforming the current one. This helps explain the rise of unified multimodal systems. If different media can be represented as compatible sequences, one model can learn relationships among text, images, sound, and time. But “unified” does not mean every modality is equally represented or controlled. The tokenization, training mixture, context length, and objectives still shape performance. For creators, transformer language gives us a way to think about context. A reference image, previous frame, camera instruction, dialogue track, or edit history can all become conditions within a larger sequence. The challenge is preserving the right relationships over long durations without allowing irrelevant details to dominate.

Transformers model relationships across a sequence. In multimodal systems, the sequence may include words, image regions, frames, or audio events. Attention is powerful because it can connect distant elements, but the model still needs representations and training signals that make the desired relation learnable.

Read the slide as a set of decisions: 1, Tokens can represent words, images, audio, or video; 2, Attention connects relevant parts; 3, Prediction builds structure one relation at a time. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. A token representing a red scarf in an early frame must remain related to the same character after a camera cut or occlusion.

A useful way to test this is to ask four questions. What is the intended audience experience? Which system decision most directly shapes it? What visible or audible evidence would count as success? And what single comparison would teach us the most if the first attempt fails? Those questions keep the work specific without reducing creative practice to a recipe.

This gives us a practical bridge between creative judgment and technical control.

SLIDE 26 · MULTIMODAL CONTEXT

Context is a sequence of relations—not a pile of files

Source frame 19 · 209 words

Now we turn to Context is a sequence of relations—not a pile of files.

Transformers model relationships across sequences. Tokens may represent words, patches, compressed visual units, audio codes, or video elements. Attention allows the model to weigh which other tokens matter when predicting or transforming the current one. This helps explain the rise of unified multimodal systems. If different media can be represented as compatible sequences, one model can learn relationships among text, images, sound, and time.

Multimodal context is not a pile of references. Text may state action, images may anchor identity and space, and audio may define rhythm or an event. Direction improves when each input has a clear job and contradictions between inputs are resolved deliberately.

Read the slide as a set of decisions: 1, Text names intention and action; 2, Images anchor appearance and space; 3, Audio carries rhythm, voice, and event. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. If the text requests panic but the audio reference is calm and slow, the system receives incompatible temporal directions.

Keep that distinction in mind as we move to the next slide.

SLIDE 27 · VIDEO MODEL STACK

A video model must bind appearance, time, and action

Source frame 20 · 352 words

Now we turn to A video model must bind appearance, time, and action.

A useful video stack has at least four concerns. Spatial structure represents objects, surfaces, lighting, and composition. Temporal structure represents persistence, motion, causality, and rhythm. Conditioning states what should happen through text, images, trajectories, poses, audio, depth, or previous frames. Decoding turns the internal representation into visible frames and perhaps sound. When a clip fails, locate the symptom. If a face drifts, identity may not be preserved across temporal context. If the camera direction is ignored, conditioning may be weak or ambiguous. If motion is smooth but meaningless, the model may follow common dynamics without a strong event structure. If details flicker, decoding or high-frequency consistency may be involved. This layered diagnosis is more productive than changing random prompt words. It points toward a control: stronger reference conditioning, shorter shot length, explicit blocking, segmented generation, compositing, or accepting the instability as the conceptual focus.

A video model combines spatial structure, temporal structure, conditioning, and decoding. Failure can originate in any layer. A useful mental model lets you ask whether the problem is geometry, persistence, ignored conditioning, or the final reconstruction of detail.

Read the slide as a set of decisions: 1, Spatial structure: what exists; 2, Temporal structure: what persists and changes; 3, Conditioning: what should happen; 4, Decoding: how it becomes frames and sound. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. A clip may have stable composition, weak object persistence, correct text conditioning, and noisy decoding—all at the same time.

A useful way to test this is to ask four questions. What is the intended audience experience? Which system decision most directly shapes it? What visible or audible evidence would count as success? And what single comparison would teach us the most if the first attempt fails? Those questions keep the work specific without reducing creative practice to a recipe.

The next step is to turn this idea into a more precise production decision.

SLIDE 28 · LAYERED DIAGNOSIS

Locate the symptom before changing the workflow

Source frame 20 · 220 words

Now we turn to Locate the symptom before changing the workflow.

A useful video stack has at least four concerns. Spatial structure represents objects, surfaces, lighting, and composition. Temporal structure represents persistence, motion, causality, and rhythm. Conditioning states what should happen through text, images, trajectories, poses, audio, depth, or previous frames. Decoding turns the internal representation into visible frames and perhaps sound. When a clip fails, locate the symptom. If a face drifts, identity may not be preserved across temporal context.

Layered diagnosis prevents random iteration. A warped hand suggests spatial representation; an identity change suggests temporal persistence; a missing action may be conditioning; shimmer may come from decoding or compression. The symptom does not prove the cause, but it narrows the next test.

Read the slide as a set of decisions: 1, Spatial: geometry, contact, or material; 2, Temporal: persistence, motion, or causality; 3, Conditioning: an ignored or ambiguous instruction; 4, Decoding: flicker, detail, or compression. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. For facial shimmer, compare a higher-quality decode while holding motion and conditioning fixed before rewriting the narrative prompt.

Use this as a question you can return to when an output surprises you.

SLIDE 29 · MODEL VERSUS PRODUCT

A model is a capability; a tool is a set of opinions

Source frame 21 · 353 words

Now we turn to A model is a capability; a tool is a set of opinions.

We need to distinguish a model from a tool. A model is a learned capability shaped by data, objectives, architecture, and weights. A tool wraps that capability in an interface with defaults, filters, pricing, resolution limits, export formats, and a particular workflow. Two tools can expose the same underlying model very differently. One encourages fast prompting; another reveals seeds, masks, control maps, or node graphs. The interface shapes what creators imagine is possible. Defaults can quietly become style. Safety systems and platform policies also affect which subjects or transformations are available. This is why our course does not organize itself around brand loyalty. We will learn transferable concepts, then compare how tools expose them. When evaluating a new product, ask: what model family or capability is underneath; what controls are exposed; what is hidden; what can be reproduced; and what does the platform require you to give up?

A model is a capability; a product is a system of interface defaults, safety rules, versioning, and policy. Two products using related models can behave differently because the surrounding system changes the available actions and the meaning of a result.

Read the slide as a set of decisions: 1, Model; 2, Tool; 3, Training data; 4, Objective; 5, Architecture; 6, Weights; 7, Interface. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. A web interface that silently rewrites prompts may produce different behavior from an API call to a related underlying model.

A useful way to test this is to ask four questions. What is the intended audience experience? Which system decision most directly shapes it? What visible or audible evidence would count as success? And what single comparison would teach us the most if the first attempt fails? Those questions keep the work specific without reducing creative practice to a recipe.

What matters is not the label but the quality of the next decision it enables.

SLIDE 30 · CURRENT VIDEO TOOLS IN PRODUCTION

MiniMax V3 and Seedance 2.5 are tools—not complete filmmaking systems

Source frame 21 · 219 words

Now we turn to MiniMax V3 and Seedance 2.5 are tools—not complete filmmaking systems.

We need to distinguish a model from a tool. A model is a learned capability shaped by data, objectives, architecture, and weights. A tool wraps that capability in an interface with defaults, filters, pricing, resolution limits, export formats, and a particular workflow. Two tools can expose the same underlying model very differently. One encourages fast prompting; another reveals seeds, masks, control maps, or node graphs.

MiniMax V3 and Seedance 2.5 can be treated as current shot-generation tools, but neither replaces story development, continuity planning, editing, sound, provenance, or release decisions. Product literacy means evaluating the complete production chain rather than confusing a strong shot with a finished film.

Read the slide as a set of decisions: 1, Generator; 2, Production workflow; 3, Shot capability; 4, Motion and fidelity; 5, Conditioning; 6, Temporal limits; 7, Story bible. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. Use either tool to generate a controlled shot, then use an edit, sound design, and continuity plan to turn that shot into part of a film.

This gives us a practical bridge between creative judgment and technical control.

SLIDE 31 · AI MOVIE WORKFLOW OVERVIEW

A durable workflow separates decisions into stages

Source frame 13 · 271 words

Now we turn to A durable workflow separates decisions into stages.

Here is the most important technical map for today. Data supplies examples and therefore boundaries. Representation determines how those examples become manageable signals or tokens. The model learns regularities over those representations. Sampling turns a space of possibilities into one particular output. Products hide these layers, but creative control often means intervening in one of them. Curating a dataset changes the learned world.

A durable AI-movie workflow separates development, previsualization, generation, selection, and post-production. Separation creates checkpoints. At each checkpoint, the team can decide what must remain stable, what can vary, and what evidence is needed before spending more generations.

Read the slide as a set of decisions: 1, Develop the dramatic and visual system; 2, Previsualize the edit and continuity; 3, Generate, select, and repair shots; 4, Edit, design sound, disclose, and release. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. A team can approve the logline before building the world bible, approve the animatic before generation, and approve selected shots before final sound.

A useful way to test this is to ask four questions. What is the intended audience experience? Which system decision most directly shapes it? What visible or audible evidence would count as success? And what single comparison would teach us the most if the first attempt fails? Those questions keep the work specific without reducing creative practice to a recipe.

Keep that distinction in mind as we move to the next slide.

SLIDE 32 · DEVELOPMENT LOGLINE

Begin with a dramatic unit—not a prompt

Source frame 23 · 193 words

Now we turn to Begin with a dramatic unit—not a prompt.

Many prompts accumulate adjectives: cinematic, beautiful, surreal, highly detailed. These can influence surface style, but they rarely define a coherent work. Art direction is relational. Who is looking? From where? What is stable? What changes? What does the light reveal? What does the sound contradict? How does the work position the audience?

Begin with a dramatic unit: who wants what, from whose point of view, what changes, and what consequence follows. A prompt describes an input to a tool; a logline describes an experience for an audience. The production should serve the dramatic unit.

Read the slide as a set of decisions: 1, Logline + beat + point of view + change + consequence. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. 'A courier must deliver a memory before the city erases her identity' directs a scene; 'woman walking in futuristic city' mainly describes imagery.

The next step is to turn this idea into a more precise production decision.

SLIDE 33 · WORLD BIBLE

Continuity begins before the first generation

Source frame 23 · 253 words

Now we turn to Continuity begins before the first generation.

Many prompts accumulate adjectives: cinematic, beautiful, surreal, highly detailed. These can influence surface style, but they rarely define a coherent work. Art direction is relational. Who is looking? From where? What is stable? What changes? What does the light reveal? What does the sound contradict? How does the work position the audience?

A world bible translates dramatic intention into continuity anchors. Character proportions, costume logic, setting, palette, lens, light, motion, and sound should agree. The bible need not freeze every detail, but it must define which variations still belong to the same world.

Read the slide as a set of decisions: 1, Character + setting + palette + lens + light + motion + sound. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. If the character's coat may vary in texture but not silhouette or color, the bible defines freedom and continuity at the same time.

A useful way to test this is to ask four questions. What is the intended audience experience? Which system decision most directly shapes it? What visible or audible evidence would count as success? And what single comparison would teach us the most if the first attempt fails? Those questions keep the work specific without reducing creative practice to a recipe.

Use this as a question you can return to when an output surprises you.

SLIDE 34 · STORYBOARD AND ANIMATIC

Storyboard the edit before spending generations

Source frame 24 · 199 words

Now we turn to Storyboard the edit before spending generations.

Style is often the memory of repeated constraints. Instead of asking for a vague look, define a small rule set. One camera rule: every shot is a slow clockwise orbit at eye level. One transformation rule: objects change only when they leave the frame. One material rule: every surface behaves like folded paper. These constraints do three things. They reduce the search space, making iteration more legible.

Storyboard the edit before expensive generation. An animatic exposes missing coverage, weak causality, and impossible continuity while changes are still cheap. Give each shot one purpose, one anchor to preserve, and one clear condition for leaving the shot.

Read the slide as a set of decisions: 1, One shot purpose; 2, One continuity anchor; 3, One exit condition. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. A three-shot animatic may reveal that a close-up has no establishing geography, saving dozens of failed generation attempts.

What matters is not the label but the quality of the next decision it enables.

SLIDE 35 · SHOT BASED GENERATION

Generate shots—not an entire movie

Source frame 24 · 262 words

Now we turn to Generate shots—not an entire movie.

Style is often the memory of repeated constraints. Instead of asking for a vague look, define a small rule set. One camera rule: every shot is a slow clockwise orbit at eye level. One transformation rule: objects change only when they leave the frame. One material rule: every surface behaves like folded paper. These constraints do three things. They reduce the search space, making iteration more legible.

Generate shots rather than asking one model output to solve an entire movie. Limit each shot to one camera behavior and one action beat. Short controlled units are easier to compare, repair, edit, and regenerate without damaging the rest of the sequence.

Read the slide as a set of decisions: 1, One camera move; 2, One action beat; 3, One continuity test. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. Generate a five-second handoff, a five-second reaction, and a five-second exit as separate shots, then construct the exchange in the edit.

A useful way to test this is to ask four questions. What is the intended audience experience? Which system decision most directly shapes it? What visible or audible evidence would count as success? And what single comparison would teach us the most if the first attempt fails? Those questions keep the work specific without reducing creative practice to a recipe.

This gives us a practical bridge between creative judgment and technical control.

SLIDE 36 · SELECTION AS FILMMAKING

Selection creates the film before post-production begins

Source frame 25 · 197 words

Now we turn to Selection creates the film before post-production begins.

Iteration is not generating many options. It is a learning loop. Begin with a hypothesis: “A static camera will make the changing body feel more vulnerable. ” Create controlled variations. Select using criteria derived from the intention. Reflect on what the results teach about both the model and the artwork. Then document the next question.

Selection is a filmmaking act. Generate a small comparable batch, judge against declared criteria, cut the usable interval, and record why it stays. The rejected outputs are also evidence: they reveal the boundary of the current instruction and model.

Read the slide as a set of decisions: 1, Generate a small comparable batch; 2, Compare against declared criteria; 3, Cut the strongest usable moment; 4, Record why the shot stays. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. From six candidates, the most beautiful clip may be rejected because the hand never completes the story-critical gesture.

Keep that distinction in mind as we move to the next slide.

SLIDE 37 · POST PRODUCTION

Post-production restores authorship across the sequence

Source frame 25 · 258 words

Now we turn to Post-production restores authorship across the sequence.

Iteration is not generating many options. It is a learning loop. Begin with a hypothesis: “A static camera will make the changing body feel more vulnerable. ” Create controlled variations. Select using criteria derived from the intention. Reflect on what the results teach about both the model and the artwork. Then document the next question.

Post-production restores sequence-level authorship. Editing creates rhythm and causality; compositing and stabilization repair local artifacts; sound gives events weight and space; color and titles unify the work. Disclosure and delivery are creative responsibilities, not administrative leftovers.

Read the slide as a set of decisions: 1, Edit rhythm and narrative causality; 2, Composite, repair, and stabilize; 3, Design voice, ambience, music, and Foley; 4, Color, title, disclose, and deliver. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. A simple room tone and precisely timed Foley can make a modest image sequence feel spatially coherent and emotionally intentional.

A useful way to test this is to ask four questions. What is the intended audience experience? Which system decision most directly shapes it? What visible or audible evidence would count as success? And what single comparison would teach us the most if the first attempt fails? Those questions keep the work specific without reducing creative practice to a recipe.

The next step is to turn this idea into a more precise production decision.

SLIDE 38 · WORKFLOW PROVENANCE

A professional AI-movie workflow leaves inspectable evidence

Source frame 30 · 222 words

Now we turn to A professional AI-movie workflow leaves inspectable evidence.

Provenance means the history of how a media object came to be. At the studio level, it starts with disciplined documentation: source media, permissions, prompts, settings, model or tool version, edits, and release context. At the ecosystem level, standards and metadata can help attach assertions about origin and modification, though no mechanism is complete or universally adopted. Why does this matter artistically? A documented process can become part of the work’s meaning.

A professional workflow leaves inspectable evidence: rights and permissions, prompts and references, seeds and settings, model version and date, selected outputs, and edit decisions. This archive supports collaboration, critique, troubleshooting, and responsible release.

Read the slide as a set of decisions: 1, Archive sources, rights, and permissions; 2, Keep prompts, references, seeds, and selected outputs; 3, Record model, version, date, and material settings; 4, Document edit decisions, synthetic media, and release context. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. If a collaborator asks why shot four was chosen, the archive should show the comparison, criterion, source, model version, and edit decision.

Use this as a question you can return to when an output surprises you.

SLIDE 39 · PROMPT ARCHAEOLOGY OBSERVATION

Prompt archaeology: observe before you infer

Source frame 22 · 281 words

Now we turn to Prompt archaeology: observe before you infer.

This is our first studio activity. Work in groups of three. Choose one generated image or short clip available on a group member’s device. You are not trying to guess the exact prompt. You are reconstructing the decision structure. Minute one: observe. List only what is visible or audible—camera position, lighting, motion, material, pacing, sound, repeated motifs, and artifacts. Avoid interpretation for a moment. Minutes two and three: infer.

Prompt archaeology begins with observation. Describe only what is visible or audible: camera, light, motion, material, sound, continuity breaks, and repeated motifs. Delay the guess about the prompt or tool until the evidence is shared.

Read the slide as a set of decisions: 1, List only what is visible or audible; 2, Name camera, light, motion, material, and sound; 3, Mark continuity breaks and repeated motifs; 4, Do not guess the tool or prompt yet. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. Describe 'the camera remains static while reflected light moves left to right' before guessing that the creator used a motion brush.

A useful way to test this is to ask four questions. What is the intended audience experience? Which system decision most directly shapes it? What visible or audible evidence would count as success? And what single comparison would teach us the most if the first attempt fails? Those questions keep the work specific without reducing creative practice to a recipe.

What matters is not the label but the quality of the next decision it enables.

SLIDE 40 · PROMPT ARCHAEOLOGY HYPOTHESIS

Turn observation into one testable production hypothesis

Source frame 22 · 220 words

Now we turn to Turn observation into one testable production hypothesis.

This is our first studio activity. Work in groups of three. Choose one generated image or short clip available on a group member’s device. You are not trying to guess the exact prompt. You are reconstructing the decision structure. Minute one: observe. List only what is visible or audible—camera position, lighting, motion, material, pacing, sound, repeated motifs, and artifacts. Avoid interpretation for a moment. Minutes two and three: infer.

The second phase turns observation into a hypothesis. Infer likely conditions, choose one uncertain claim, change one variable, and predict the result. The point is not to reverse-engineer a secret prompt perfectly; it is to practice building testable explanations.

Read the slide as a set of decisions: 1, Infer the likely conditions and references; 2, Choose one uncertain claim that matters; 3, Change one variable and predict the result; 4, Record the input, output, observation, and decision. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. Test whether a reference image controls identity by keeping the prompt and seed strategy stable while swapping only that reference.

This gives us a practical bridge between creative judgment and technical control.

SLIDE 41 · RELATIONAL ART DIRECTION

Art direction is a system of relations, not a bag of adjectives

Source frame 23 · 337 words

Now we turn to Art direction is a system of relations, not a bag of adjectives.

Many prompts accumulate adjectives: cinematic, beautiful, surreal, highly detailed. These can influence surface style, but they rarely define a coherent work. Art direction is relational. Who is looking? From where? What is stable? What changes? What does the light reveal? What does the sound contradict? How does the work position the audience? A useful brief might say: “A fixed waist-height camera watches a temporary wall of ice slowly become transparent as overheard conversations accumulate. The sound becomes clearer as the image disappears. ” This defines point of view, material behavior, temporal arc, and audiovisual relation. It can guide prompting, shooting, compositing, or model choice. Treat the model as one collaborator inside a larger direction system. Your brief should survive if the exact generator changes. That is how you preserve authorship amid rapidly changing products.

Art direction is relational. Instead of stacking adjectives, define subject, point of view, space, light, time, transformation, and sound as a connected system. A coherent relation gives the model fewer contradictory ways to satisfy the request.

Read the slide as a set of decisions: 1, Subject + point of view + space + light + time + transformation + sound. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. 'Cold blue cinematic mysterious' is a bag of adjectives; 'blue window light isolates the subject as the hallway darkens behind her' defines a relation.

A useful way to test this is to ask four questions. What is the intended audience experience? Which system decision most directly shapes it? What visible or audible evidence would count as success? And what single comparison would teach us the most if the first attempt fails? Those questions keep the work specific without reducing creative practice to a recipe.

Keep that distinction in mind as we move to the next slide.

SLIDE 42 · CONSTRAINTS AS STYLE

Constraints produce style when they are repeated deliberately

Source frame 24 · 278 words

Now we turn to Constraints produce style when they are repeated deliberately.

Style is often the memory of repeated constraints. Instead of asking for a vague look, define a small rule set. One camera rule: every shot is a slow clockwise orbit at eye level. One transformation rule: objects change only when they leave the frame. One material rule: every surface behaves like folded paper. These constraints do three things. They reduce the search space, making iteration more legible. They create continuity across separate generations. And they reveal exceptions—moments when breaking the rule has meaning. This is also a bridge between art and technology. A rule can be encoded in prompts, camera trajectories, node graphs, masks, control signals, or post-production. If the tool cannot support the rule directly, that limitation tells you what additional technique or different model you need. The question shifts from “How do I get a good result? ” to “How do I maintain a coherent system of decisions?

Constraints become style when repeated deliberately. A fixed camera rule, transformation rule, or material rule can create a recognizable world. Constraints also make comparison possible because variation has a stable background.

Read the slide as a set of decisions: 1, One camera rule; 2, One transformation rule; 3, One material rule. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. A film that always uses a locked camera and allows only shadows to transform can develop a distinctive visual grammar.

The next step is to turn this idea into a more precise production decision.

SLIDE 43 · CREATIVE ITERATION LOOP

The creative loop is hypothesis, variation, selection, reflection

Source frame 25 · 336 words

Now we turn to The creative loop is hypothesis, variation, selection, reflection.

Iteration is not generating many options. It is a learning loop. Begin with a hypothesis: “A static camera will make the changing body feel more vulnerable. ” Create controlled variations. Select using criteria derived from the intention. Reflect on what the results teach about both the model and the artwork. Then document the next question. If every generation changes subject, framing, model, prompt, seed, and settings, you have produced variety but little knowledge. Controlled variation gives you evidence. Conversely, do not over-control too early. Exploratory phases can be broad, but once you recognize a promising direction, narrow the variables. Keep a process log. For each meaningful iteration, record the input, settings, output, observation, and decision. This becomes material for critique and for the final presentation, where you will show not only a polished result but also one representative failure and what changed after iteration.

The creative loop is hypothesis, controlled variation, selection with criteria, and reflection. The final step is essential: write what changed and what you learned before generating again. Otherwise iteration becomes expensive browsing.

Read the slide as a set of decisions: 1, Hypothesis; 2, Controlled variation; 3, Selection with criteria; 4, Reflection and documentation. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. After a failed run, write 'identity held for three seconds, then changed after occlusion' before deciding whether to shorten, repair, or regenerate.

A useful way to test this is to ask four questions. What is the intended audience experience? Which system decision most directly shapes it? What visible or audible evidence would count as success? And what single comparison would teach us the most if the first attempt fails? Those questions keep the work specific without reducing creative practice to a recipe.

Use this as a question you can return to when an output surprises you.

SLIDE 44 · CONTROLLED COMPARISON

One comparison can teach more than twenty unrelated outputs

Source frame 25 · 202 words

Now we turn to One comparison can teach more than twenty unrelated outputs.

Iteration is not generating many options. It is a learning loop. Begin with a hypothesis: “A static camera will make the changing body feel more vulnerable. ” Create controlled variations. Select using criteria derived from the intention. Reflect on what the results teach about both the model and the artwork. Then document the next question.

A controlled comparison holds important conditions constant and changes one variable deliberately. State the hypothesis and prediction first. After generation, write the observation before choosing the next move so preference does not rewrite the experiment after the fact.

Read the slide as a set of decisions: 1, State the hypothesis; 2, Hold the important conditions constant; 3, Change one variable deliberately; 4, Write the observation before the next generation. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. Compare two clips with identical references and duration, changing only camera-motion strength; measure which better preserves the intended stillness.

What matters is not the label but the quality of the next decision it enables.

SLIDE 45 · FAILURE MUSEUM

Failures reveal the model’s world—and sometimes your own

Source frame 26 · 348 words

Now we turn to Failures reveal the model’s world—and sometimes your own.

Welcome to the failure museum. Each frame is almost plausible. That “almost” is valuable. A broken reflection may reveal that the system learned appearance without geometric causality. Identity drift may reveal weak persistence. A hand may expose training bias and the difficulty of articulated structure. An impossible material may become the central aesthetic of a work. There are at least three responses to failure. Repair it when it distracts from the intention. Constrain it when it teaches you how to direct the system. Reframe it when the artifact creates a meaning worth pursuing. The choice is artistic, but it should be informed by technical diagnosis. Take thirty seconds to choose one frame. Is the anomaly merely an error, or could it become a rule? What context would make it legible as deliberate? This is not permission to romanticize poor craft. It is an invitation to distinguish accidental failure from consciously composed instability.

A failure museum treats broken outputs as diagnostic artifacts. Name what failed and ask what the failure reveals about representation, conditioning, temporal memory, or cultural defaults. Some failures should be repaired; others can become material for the work.

Read the slide as a set of decisions: 1, Do not hide every artifact before asking what it means. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. A melting face may expose weak identity persistence; it may also become meaningful in a story about unstable memory if the choice is intentional.

A useful way to test this is to ask four questions. What is the intended audience experience? Which system decision most directly shapes it? What visible or audible evidence would count as success? And what single comparison would teach us the most if the first attempt fails? Those questions keep the work specific without reducing creative practice to a recipe.

This gives us a practical bridge between creative judgment and technical control.

SLIDE 46 · FAILURE RESPONSE STRATEGY

Repair, constrain, or reframe—the choice must serve the work

Source frame 26 · 201 words

Now we turn to Repair, constrain, or reframe—the choice must serve the work.

Welcome to the failure museum. Each frame is almost plausible. That “almost” is valuable. A broken reflection may reveal that the system learned appearance without geometric causality. Identity drift may reveal weak persistence. A hand may expose training bias and the difficulty of articulated structure. An impossible material may become the central aesthetic of a work. There are at least three responses to failure. Repair it when it distracts from the intention.

Repair distractions, constrain recurring instability, and reframe artifacts that create meaning. The decision must serve the intention. An artifact is not automatically interesting, and a polished repair is not automatically the best artistic choice.

Read the slide as a set of decisions: 1, Repair distractions. Constrain recurring instability. Reframe artifacts that create meaning. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. Repair a distracting extra finger, constrain a camera that repeatedly drifts, or reframe impossible gravity as the world's central rule.

Keep that distinction in mind as we move to the next slide.

SLIDE 47 · FAILURE TAXONOMY

Name the failure before you fix it

Source frame 27 · 375 words

Now we turn to Name the failure before you fix it.

A useful failure taxonomy prevents random troubleshooting. Semantic failure means the content or relation is wrong: the subject performs a different action, or cause and effect are reversed. Spatial failure involves geometry, depth, lighting, reflections, or contact. Temporal failure includes flicker, motion discontinuity, and changing object topology. Identity failure means a person, character, or designed object does not persist. Aesthetic failure means the result may be technically plausible but does not serve the intended visual language or emotional effect. Multiple categories can overlap. A drifting face is identity and temporal. A reflection that moves independently is spatial and temporal. The value of naming is that each category suggests different interventions: reference conditioning, shorter shots, explicit blocking, segmentation, compositing, a different model, or a changed concept. Use the narrowest description you can. “The left hand changes from five to six fingers between seconds two and three while crossing the face” is actionable evidence. “The AI looks weird” is not.

A failure taxonomy gives the studio shared language: semantic, spatial, temporal, physical, identity, and aesthetic. Categories can overlap. Their value is in making the next test more precise and helping collaborators discuss evidence without vague labels.

Read the slide as a set of decisions: 1, Semantic: wrong object, action, or relation; 2, Spatial: broken geometry, depth, or reflection; 3, Temporal: flicker, drift, or implausible motion; 4, Identity: subject changes across frames; 5, Aesthetic: generic, incoherent, or off-intent. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. A floating cup may be spatial, physical, and semantic at once; the taxonomy helps the team decide which aspect the next test will isolate.

A useful way to test this is to ask four questions. What is the intended audience experience? Which system decision most directly shapes it? What visible or audible evidence would count as success? And what single comparison would teach us the most if the first attempt fails? Those questions keep the work specific without reducing creative practice to a recipe.

The next step is to turn this idea into a more precise production decision.

SLIDE 48 · SEMANTIC AND SPATIAL FAILURES

Semantic and spatial failures break what the scene claims

Source frame 27 · 210 words

Now we turn to Semantic and spatial failures break what the scene claims.

A useful failure taxonomy prevents random troubleshooting. Semantic failure means the content or relation is wrong: the subject performs a different action, or cause and effect are reversed. Spatial failure involves geometry, depth, lighting, reflections, or contact. Temporal failure includes flicker, motion discontinuity, and changing object topology. Identity failure means a person, character, or designed object does not persist.

Semantic and spatial failures break what the scene claims. Look for wrong entities or actions, impossible relations, broken contact and scale, contradictory light, and inconsistent depth. These are observable symptoms that can be compared across outputs.

Read the slide as a set of decisions: 1, Wrong entity, action, or attribute; 2, Impossible relation or reversed causality; 3, Broken geometry, contact, or scale; 4, Inconsistent light, reflection, or shadow; 5, Contradictory depth or occlusion. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. If a reflection disagrees with the body position, name the contradiction before proposing inpainting, compositing, or a different camera angle.

Use this as a question you can return to when an output surprises you.

SLIDE 49 · TEMPORAL IDENTITY AND AESTHETIC FAILURES

Temporal, identity, and aesthetic failures break the experience

Source frame 27 · 279 words

Now we turn to Temporal, identity, and aesthetic failures break the experience.

A useful failure taxonomy prevents random troubleshooting. Semantic failure means the content or relation is wrong: the subject performs a different action, or cause and effect are reversed. Spatial failure involves geometry, depth, lighting, reflections, or contact. Temporal failure includes flicker, motion discontinuity, and changing object topology. Identity failure means a person, character, or designed object does not persist.

Temporal, identity, and aesthetic failures break the experience across time. Flicker, topology change, implausible timing, character drift, generic style, or mismatched rhythm may each require a different intervention. Do not collapse them into the single phrase 'bad consistency.'

Read the slide as a set of decisions: 1, Flicker, topology change, or continuity loss; 2, Implausible motion, timing, or physical response; 3, Character, object, or voice identity drift; 4, Generic style or accidental visual language; 5, Rhythm or affect that contradicts the intention. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. If a face remains sharp but gradually becomes another person, the issue is identity drift, not simply low image quality.

A useful way to test this is to ask four questions. What is the intended audience experience? Which system decision most directly shapes it? What visible or audible evidence would count as success? And what single comparison would teach us the most if the first attempt fails? Those questions keep the work specific without reducing creative practice to a recipe.

What matters is not the label but the quality of the next decision it enables.

SLIDE 50 · CRITIQUE PROTOCOL

Critique the intention, evidence, system, and next move

Source frame 28 · 299 words

Now we turn to Critique the intention, evidence, system, and next move.

We will use a four-part critique protocol throughout the course. Intention: what experience or question does the work propose? Evidence: what specific formal feature supports or contradicts that intention? System: what technical and production choices likely produced the result? Next move: what single experiment would teach the most? The protocol is designed to keep critique generous and precise. We do not need to solve the entire project. We need to identify the most informative next move. That might be technical—lock camera motion with a trajectory—or artistic—remove narration and let the visual contradiction carry the idea. When giving feedback, separate observation from interpretation. Say what you saw before what you think it means. When receiving feedback, listen for patterns across viewers rather than defending every decision. Critique is not a judgment ceremony; it is a collective instrument for seeing more clearly.

The critique protocol moves through intention, evidence, system, and next move. This order keeps discussion connected to the work. It also makes critique generous: the group helps the maker learn which decision to test rather than merely voting on taste.

Read the slide as a set of decisions: 1, Intention — what experience is proposed?; 2, Evidence — what in the work supports it?; 3, System — which choices produced the result?; 4, Next move — what single test teaches the most?. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. A critique might state the intended dread, point to the delayed doorway reveal, locate a continuity break, and propose one locked-camera comparison.

This gives us a practical bridge between creative judgment and technical control.

SLIDE 51 · OBSERVATION BEFORE INTERPRETATION

Separate observation from interpretation before judging success

Source frame 28 · 279 words

Now we turn to Separate observation from interpretation before judging success.

We will use a four-part critique protocol throughout the course. Intention: what experience or question does the work propose? Evidence: what specific formal feature supports or contradicts that intention? System: what technical and production choices likely produced the result? Next move: what single experiment would teach the most? The protocol is designed to keep critique generous and precise. We do not need to solve the entire project.

Observation and interpretation are both necessary, but they should not be confused. First establish what can be pointed to. Then discuss the tension or audience position it creates. Shared observation gives disagreement a productive ground.

Read the slide as a set of decisions: 1, Intention — what experience is proposed?; 2, Evidence — what can everyone point to?; 3, Tension — where does the work contradict itself?; 4, Audience — what position or feeling is produced?. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. Everyone can observe that the light changes direction at the cut; interpretations may differ about whether that rupture feels dreamlike or careless.

A useful way to test this is to ask four questions. What is the intended audience experience? Which system decision most directly shapes it? What visible or audible evidence would count as success? And what single comparison would teach us the most if the first attempt fails? Those questions keep the work specific without reducing creative practice to a recipe.

Keep that distinction in mind as we move to the next slide.

SLIDE 52 · MOST INFORMATIVE NEXT MOVE

Choose one next move with maximum information

Source frame 28 · 225 words

Now we turn to Choose one next move with maximum information.

We will use a four-part critique protocol throughout the course. Intention: what experience or question does the work propose? Evidence: what specific formal feature supports or contradicts that intention? System: what technical and production choices likely produced the result? Next move: what single experiment would teach the most? The protocol is designed to keep critique generous and precise. We do not need to solve the entire project.

Choose the next move with maximum information. Identify which decision could explain the result, what uncertainty remains, one test that isolates it, and a threshold that would change the decision. A good next test is small enough to finish and strong enough to matter.

Read the slide as a set of decisions: 1, System — which decision produced the result?; 2, Uncertainty — what claim remains unproven?; 3, Test — what single comparison isolates it?; 4, Threshold — what result changes the decision?. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. If two causes are plausible, choose the comparison whose result would eliminate one cause rather than adding more decoration.

The next step is to turn this idea into a more precise production decision.

SLIDE 53 · AUTHORSHIP CHAIN

Authorship lives in the chain of decisions

Source frame 29 · 334 words

Now we turn to Authorship lives in the chain of decisions.

Authorship in generative practice is not proven by typing every pixel into existence. It appears in the chain of decisions: the brief, references, data, conditions, generations, selections, edits, sequencing, sound, context, and release. The more clearly you can explain that chain, the more legible your contribution becomes. This does not mean every chain is ethically equivalent. Using a living artist's name as a shortcut, reproducing a person's likeness without consent, or hiding automated production in a context that expects disclosure raises different issues. Authorship includes responsibility for inputs and circulation, not just pride in outputs. For each assignment, document what you contributed, what the system contributed, what external work influenced you, and where uncertainty remains. This is not bureaucratic overhead. It strengthens your ability to critique your own process and helps audiences understand the work's conditions.

Authorship lives in a chain: brief, references and data, generation, selection, editing, and context of release. Different people and systems may control different links. Mapping the chain makes credit, responsibility, and creative agency visible.

Read the slide as a set of decisions: 1, Brief; 2, References & data; 3, Generation; 4, Selection; 5, Editing; 6, Context & release. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. An artist may write the brief, a model provider shapes generation, an editor determines rhythm, and a platform frames release; all participate in authorship.

A useful way to test this is to ask four questions. What is the intended audience experience? Which system decision most directly shapes it? What visible or audible evidence would count as success? And what single comparison would teach us the most if the first attempt fails? Those questions keep the work specific without reducing creative practice to a recipe.

Use this as a question you can return to when an output surprises you.

SLIDE 54 · PROVENANCE AND DISCLOSURE

Provenance turns process into inspectable evidence

Source frame 30 · 295 words

Now we turn to Provenance turns process into inspectable evidence.

Provenance means the history of how a media object came to be. At the studio level, it starts with disciplined documentation: source media, permissions, prompts, settings, model or tool version, edits, and release context. At the ecosystem level, standards and metadata can help attach assertions about origin and modification, though no mechanism is complete or universally adopted. Why does this matter artistically? A documented process can become part of the work's meaning. It allows comparison across versions, supports reproducibility where appropriate, and makes your claims inspectable. It also protects collaborators by clarifying consent and contribution. Disclosure is contextual. A speculative gallery piece, a documentary image, an advertisement, and a private sketch do not carry the same audience expectations. The ethical question is not simply "Was AI used? " It is "What belief could this artifact reasonably create, who could be affected, and what information does the audience need?

Provenance turns process into inspectable evidence. Record sources and permissions, model and version, prompts and meaningful settings, selected and rejected media, post-production changes, and the context in which synthetic material is disclosed.

Read the slide as a set of decisions: 1, Keep prompts and meaningful settings; 2, Archive source media and permissions; 3, Record model/tool/version when possible; 4, Disclose synthetic or altered media when context requires it. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. A release note can state which shots are synthetic, whose face or voice was used with permission, and which edits materially altered the output.

What matters is not the label but the quality of the next decision it enables.

SLIDE 55 · BIAS AND UNEVEN FAILURE

Bias appears as defaults, absences, and uneven failure

Source frame 31 · 351 words

Now we turn to Bias appears as defaults, absences, and uneven failure.

Bias is not only an offensive output. It can appear as a default—the kinds of bodies, homes, professions, beauty, or family the model produces without specification. It can appear as absence—cultural forms that are poorly represented or flattened. It can appear as uneven failure—identity preservation or speech alignment working better for some groups than others. Creative practitioners have choices. You can test deliberately across relevant populations, avoid claims your evidence cannot support, use consented and contextualized references, and acknowledge limits. You can also make the bias itself visible, but doing so does not remove the risk of reproducing harm. Critical art still needs careful framing and care for affected people. MIT’s Creative AI curriculum makes ethics part of the making process through questions about ownership, creative partnership, bias, and synthetic media. We follow that principle: ethics is not a final warning slide. It is a design constraint that shapes the project from the beginning.

Bias appears as defaults, absences, and uneven failure. A system may look impressive overall while repeatedly failing particular people, bodies, languages, places, or cultural forms. Responsible evaluation asks who pays the cost of the model's convenience.

Read the slide as a set of decisions: 1, A system can be impressive overall and still fail particular people or cultures predictably. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. If a model repeatedly lightens a skin tone under beauty prompting, the failure reveals a cultural default rather than a neutral quality setting.

A useful way to test this is to ask four questions. What is the intended audience experience? Which system decision most directly shapes it? What visible or audible evidence would count as success? And what single comparison would teach us the most if the first attempt fails? Those questions keep the work specific without reducing creative practice to a recipe.

This gives us a practical bridge between creative judgment and technical control.

SLIDE 56 · WORLD MODELS AND CONSEQUENCE

A world model predicts what happens next under action

Source frame 32 · 304 words

Now we turn to A world model predicts what happens next under action.

A video generator can produce plausible motion without representing a stable, actionable world. A world model aims to predict how a state changes under actions or conditions. The stronger claim is not “it makes video” but “it supports consequence. ” If I turn left, the environment changes consistently. If I pick up an object, its state persists. This is difficult because visual plausibility can hide causal weakness. A system may generate a convincing few seconds while forgetting objects, violating geometry, or responding inconsistently to the same action. We should therefore treat “world model” as a hypothesis to test, not a marketing label to repeat. For artists, even imperfect world models open new forms: explorable cinema, responsive installations, characters that improvise, and environments that change with audience behavior. The aesthetic question becomes how much consistency the work needs. A dream world may productively mutate; a training simulation may require strict reliability.

A world model predicts what happens next under action. The key is conditional consequence: the system receives an observation and action, predicts the next state, and preserves enough continuity for interaction to remain meaningful. This is a useful horizon even when today's tools fall short.

Read the slide as a set of decisions: 1, Observe a state; 2, Receive an action or condition; 3, Predict the next state; 4, Maintain enough consistency to continue. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. In an interactive room, opening a door should reveal a state consistent with earlier actions; otherwise the system predicts frames without maintaining a world.

Keep that distinction in mind as we move to the next slide.

SLIDE 57 · STUDIO CHALLENGE

Studio challenge: direct one meaningful transformation

Source frame 33 · 368 words

Now we turn to Studio challenge: direct one meaningful transformation.

Your first studio challenge is intentionally small. Start with one image that you created or have the right to use. Define one temporal proposition: what changes, at what rate, and why? Add three repeatable constraints—a camera rule, transformation rule, and material or sound rule. Produce three variations, or storyboard them if the required tool is unavailable. Selection is part of the challenge. Choose the variation that best supports the intention, not automatically the cleanest result. Save one failure that teaches you something. Record the input, model or tool, relevant settings, and your observation. A strong example could be: “A still portrait gradually loses all features except eye contact; the camera never moves; skin becomes paper only when the subject exhales; sound becomes drier as identity disappears. ” The point is not spectacle. It is a clear relation between form, system, and meaning. Bring the process—not just the final clip—to our next discussion.

The studio challenge is to direct one meaningful transformation. Start from one image you have the right to use, define one temporal proposition, apply three constraints, produce a storyboard of three variations, and select one intention—not polish alone.

Read the slide as a set of decisions: 1, Start with one image you have the right to use; 2, Define one temporal proposition; 3, Apply three repeatable constraints; 4, Generate or storyboard three variations; 5, Choose using intention—not polish alone. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. A still portrait could slowly lose its shadow while the camera stays fixed, using light, duration, and material as the three constraints.

A useful way to test this is to ask four questions. What is the intended audience experience? Which system decision most directly shapes it? What visible or audible evidence would count as success? And what single comparison would teach us the most if the first attempt fails? Those questions keep the work specific without reducing creative practice to a recipe.

The next step is to turn this idea into a more precise production decision.

SLIDE 58 · ONE PAGE EXPERIMENT

Your first experiment should fit on one production page

Source frame 33 · 218 words

Now we turn to Your first experiment should fit on one production page.

Your first studio challenge is intentionally small. Start with one image that you created or have the right to use. Define one temporal proposition: what changes, at what rate, and why? Add three repeatable constraints—a camera rule, transformation rule, and material or sound rule. Produce three variations, or storyboard them if the required tool is unavailable. Selection is part of the challenge.

The one-page experiment keeps ambition testable. Write the intention and audience experience, choose one source image, define three repeatable constraints, create five variations including one instructive failure, and select using declared criteria.

Read the slide as a set of decisions: 1, Write one intention and one audience experience; 2, Choose one source image you have the right to use; 3, Define three repeatable constraints; 4, Make three variations and keep one instructive failure; 5, Select with one declared criterion. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. Five versions of the shadow experiment can vary only disappearance rate, making the final selection explainable rather than arbitrary.

Use this as a question you can return to when an output surprises you.

SLIDE 59 · SEMESTER ROADMAP

The semester moves from seeing to making to directing worlds

Source frame 34 · 369 words

Now we turn to The semester moves from seeing to making to directing worlds.

The semester follows a learning progression. We begin by seeing: building the vocabulary to read images, systems, representations, and cultural context. Then we make: image generation and editing, adaptation, and video foundations. We move into direction: motion, sound, characters, and interactive worlds. Evaluation and provenance run through every phase rather than appearing at the end. Finally, you select and refine one assignment into a public presentation. The structure borrows three useful patterns from leading courses. From CMU, frequent looking outward, computational craft, and studio critique. From NYU ITP, accessible technical explanations tied directly to creative prototypes. From MIT, hands-on exploration combined with open questions about authorship and ethics. Stanford's graphics teaching also provides a useful reminder that generative controls should connect to established artist workflows rather than forcing artists to abandon them. You do not need to enter as both an artist and an engineer. You do need to become curious about the other lens.

The semester roadmap moves from seeing and describing to making controlled experiments, directing shots and sequences, accounting for provenance, and presenting an authored world. Each stage adds responsibility rather than replacing earlier skills.

Read the slide as a set of decisions: 1, See: representation, models, visual language; 2, Make: image, edit, adaptation, video; 3, Direct: motion, sound, characters, interaction; 4, Account: evaluation, provenance, critique; 5, Show: refined project and public explanation. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. Early projects train observation and control; later projects ask you to coordinate multiple shots, document responsibility, and defend a world to an audience.

A useful way to test this is to ask four questions. What is the intended audience experience? Which system decision most directly shapes it? What visible or audible evidence would count as success? And what single comparison would teach us the most if the first attempt fails? Those questions keep the work specific without reducing creative practice to a recipe.

What matters is not the label but the quality of the next decision it enables.

SLIDE 60 · EXIT TICKET

Your exit ticket is one sentence with a decision inside it

Source frame 35 · 314 words

Now we turn to Your exit ticket is one sentence with a decision inside it.

Return to the opening question: when does an image become a world? Our provisional answer is that a world begins when visual possibility acquires relations, persistence, consequence, and an audience position. Technical systems can provide some of that structure, but an artwork still needs decisions about what matters and why. Write one sentence: “I want to explore blank by constraining blank because blank. ” The first blank is an experience or question, not a product. The second is a system choice: camera, time, material, sound, dataset, reference, interaction, or another constraint. The final blank explains why they belong together. Keep the sentence. It is not a commitment to a final project. It is a first hypothesis. Next week we will examine visual representations and unified multimodal models, then test how the structure of representation affects what can be controlled. Before you leave, compare your exit sentence with the perception criterion you named at the beginning. Has your language become more precise?

The exit ticket compresses learning into a decision: 'I want to explore this experience by constraining this variable because this evidence suggests a direction.' Naming the experience, system choice, and reason turns curiosity into a first experiment.

Read the slide as a set of decisions: 1, “I want to explore ___ by constraining ___ because ___.”; 2, Name the experience; 3, Name the system choice; 4, Name why they belong together. These are not independent checklist items. Their value comes from how they work together and from whether the resulting choice can be observed in the artifact.

Here is a concrete example. 'I want to explore fragile memory by constraining the camera to a locked frame because identity drift becomes meaningful when everything else is stable.'

This gives us a practical bridge between creative judgment and technical control.