TEACHER-FACING FRIDAY DESIGN

Leadership, Responsibility, and Legitimation in Human-AI Systems

CAS Human-AI Collaboration | Module 1 | Unit 7

LITERACY	COLLABORATION MOVE	CONTACT TIME	CHALLENGE
Human-Context	**Reflection + responsible co-construction**	**365 minutes**	**Responsible Decision Logic Experiment**

Position in Module

- Unit: 7
- Topic: Leadership, Responsibility, and Legitimation in Human-AI Systems
- Literacy perspective: Human-Context
- Collaboration level: Reflection, with responsible co-construction
- Fundamentals focus: Fluent generation versus legitimate decision support
- Why this Friday matters: Participants learn that a plausible AI recommendation does not become a legitimate decision merely because a human approves it. They learn to design and test the decision rights, evidence requirements, escalation routes, and opportunities for contestation that make human responsibility substantive rather than symbolic.

Literacy Perspective

- Primary literacy: Human-Context
- Secondary literacy, if any: None
- What this literacy explains: Human-Context Literacy explains why technically useful AI output may still fail when it conflicts with organizational values, decision rights, stakeholder expectations, cultural conditions, or accountability arrangements.
- What participants should manipulate, observe, and interpret: Participants manipulate the decision arrangement around a stable AI-supported recommendation. They observe accountability clarity, evidence use, escalation, contestability, appropriate reliance, and perceived legitimacy, then interpret how organizational context and values shape whether the arrangement works.

Collaboration Level

- Level relation: Reflection, with responsible co-construction
- Collaboration move being learned: Participants move from reflecting on the quality of an AI output to co-constructing the human decision conditions around it. They decide who may recommend, challenge, approve, reject, escalate, and remain accountable.

- What participants should be able to do differently after this Friday: They should be able to diagnose symbolic human oversight, design explicit decision logic around an AI-supported recommendation, and test whether people can understand and exercise their responsibility in practice.

Human-AI Decision Support Fundamentals

This Friday introduces the difference between fluent generation and legitimate decision support. The goal is for participants to understand why confidence, plausibility, and usefulness are not the same as reliability, accountability, or organizational legitimacy.

- Explain that generative AI systems often produce fluent outputs even when the underlying support is weak, incomplete, or uncertain. Fluency is therefore not sufficient evidence that an output should be trusted.
- Introduce calibration as a leadership issue: people need to know when to rely on an output, when to inspect it, when to ask for evidence, when to escalate, and when human judgment must remain decisive.
- Explain decision logic as part of the human-AI system. The system should clarify whether AI informs, recommends, drafts, evaluates, or acts, and which human role remains accountable for the resulting decision.
- Show how uncertainty framing, disclosure, source visibility, confidence language, and explanation affect human interpretation. These elements can support responsible use, but they can also create false assurance if they are not tied to evidence and decision rights.
- Explain that human oversight is not only a final approval step. It can be designed into the collaboration through boundaries, escalation rules, review checkpoints, contestability, and clear responsibility for decisions.
- Treat culture, values, and organizational context as active design variables. The same decision arrangement may be perceived and used differently when norms about authority, participation, evidence, or responsibility differ.
- Introduce measurement before intervention. Broad constructs such as trust or legitimacy must be translated into observable indicators before participants can claim that a redesigned arrangement improved them.
- Caveat to demonstrate: transparency or human sign-off can create legitimacy theatre if responsibility is only symbolic. Participants can compare a workflow where a human merely rubber-stamps an AI recommendation with one where the human has evidence, authority, time, and criteria to contest it.
- Emphasize that greater trust or acceptance is not automatically better. The aim is calibrated trust and appropriate reliance: people should rely on the system only to the extent justified by evidence, context, and the consequences of error.
- Emphasize the central learning point for Unit 7: responsible generative AI use requires leadership design of the conditions under which humans can judge, contest, approve, reject, or be accountable for AI-supported outcomes, plus evidence that those conditions work for the people involved.

Recurring Agentic Coding Demonstration

Use a coding agent that proposes a consequential but non-emergency change, such as replacing a shared dependency and removing the corresponding legacy code. Preserve the same proposal, diff, test results, and known uncertainties across two conditions. In the baseline, the instruction merely says that a human must review the change. In the target condition, a decision protocol identifies the change owner, evidence reviewer, affected system owner, approval authority, evidence threshold, and escalation trigger. Ask participants to determine whether the change may proceed under each condition and record where responsibility becomes clearer or remains ambiguous. The demonstration makes Human-Context Literacy visible through authority and accountability; it does not introduce Unit 9's malicious-content or security-boundary experiment.

Shared Classroom Case

Use one common, fictional, low-risk case so that groups can compare decision arrangements without exposing sensitive organizational information.

Default case:

> **A fictional organization is considering whether to standardize a locally adapted customer-service process. An AI system has reviewed performance data, policy excerpts, employee comments, and a short process description. It recommends adopting the standard process, while noting incomplete evidence about local customer needs and employee workload.**

Prepare one fixed case pack containing:

- the AI-supported recommendation
- the evidence excerpts available to the AI
- known uncertainties and missing perspectives
- the organizational decision to be made
- a baseline instruction stating only that human review is required
- a target decision protocol with explicit roles, evidence requirements, escalation triggers, and a contestation route

Participants should work with the same recommendation and evidence in both conditions. The primary intervention is the decision arrangement, not a better prompt or a newly generated recommendation.

Suggested roles for groups of four or five:

- decision owner: accountable for the final organizational decision
- process owner: responsible for operational consequences
- evidence reviewer: checks whether the recommendation is adequately supported
- affected-stakeholder representative: identifies impacts, values, and grounds for contestation
- AI or process steward, if group size permits: documents system limits, provenance, and use conditions

The baseline should leave authority, evidence thresholds, and escalation largely implicit. The target protocol should make them explicit and exercisable. Participants should not be asked to make real employment, care, financial, or similarly consequential decisions during the class.

Friday Activity Plan

Indicative contact time is 365 minutes, excluding breaks and lunch. Teachers may adjust durations while preserving the sequence from prediction to comparison, measurement, redesign, replay, and transfer.

Time	Activity	Concrete output
45 min	Fluency is not legitimacy	Initial list of conditions for meaningful human responsibility
50 min	Baseline decision simulation	Baseline decision log and individual observation ratings
45 min	From abstract values to observable constructs	Agreed observation instrument with anchored criteria
60 min	Designing substantive decision logic	Responsibility matrix and target decision protocol
60 min	Replay under the target condition	Target decision log and second observation set
45 min	Comparison and interpretation	Comparison table and provisional leadership rule
60 min	Challenge kickoff and contextual transfer	Challenge protocol, baseline/target sketch, first prediction, and evidence plan

Detailed activity blocks

1

45 min Fluency is not legitimacy

Content and manipulation. Distinguish useful output, reliable evidence, justified reliance, accountability, and legitimacy. Introduce symbolic versus substantive human oversight.

Method and expected observation. Impulse with two contrasted decision arrangements; participants identify what a generic human sign-off fails to control.

Concrete output. Initial list of conditions for meaningful human responsibility

Tools and materials. Slides, contrasted example, shared board

2

50 min Baseline decision simulation

Content and manipulation. Groups receive the common case, recommendation, evidence, and generic human-review instruction.

Method and expected observation. Role-based simulation; groups predict where ambiguity will matter, make a provisional decision, and record unresolved responsibility questions.

Concrete output. Baseline decision log and individual observation ratings

Tools and materials. Case pack, baseline instruction, role cards, observation sheet

3 **45 min** **From abstract values to observable constructs**

Content and manipulation. Introduce accountability clarity, evidence use, escalation, contestability, appropriate reliance, and perceived legitimacy.

Method and expected observation. Measurement mini-lab; groups turn each construct into a question, rating, or observable action and compare interpretations.

Concrete output. Agreed observation instrument with anchored criteria

Tools and materials. Construct cards, instrument template, shared board

4 **60 min** **Designing substantive decision logic**

Content and manipulation. Add explicit authority, consultation, approval, evidence thresholds, escalation, and contestation while keeping the AI recommendation stable.

Method and expected observation. Hands-on redesign; peer critique checks whether each role has enough information, authority, time, and criteria to act.

Concrete output. Responsibility matrix and target decision protocol

Tools and materials. Decision-protocol template, responsibility matrix, case pack

5 **60 min** **Replay under the target condition**

Content and manipulation. Process the same recommendation and evidence using the redesigned arrangement.

Method and expected observation. Repeat the simulation with the same participants and comparable time; record decisions, questions, evidence use, escalation, contestation, and ratings.

Concrete output. Target decision log and second observation set

Tools and materials. Target protocol, observation sheet, case pack

6 **45 min** **Comparison and interpretation**

Content and manipulation. Compare baseline and target evidence. Examine whether responsibility became substantive and whether trust became better calibrated rather than simply higher.

Method and expected observation. Small-group analysis followed by cohort comparison; require one claim supported by evidence and one residual limitation.

Concrete output. Comparison table and provisional leadership rule

Tools and materials. Comparison worksheet, shared board

7

60 min **Challenge kickoff and contextual transfer**

Content and manipulation. Select a low-risk AI-supported recommendation or decision process from the participant's context, or retain the common case. Define one primary decision-logic manipulation.

Method and expected observation. Individual or paired coaching; participants produce the first challenge protocol and receive a peer feasibility check.

Concrete output. Challenge protocol, baseline/target sketch, first prediction, and evidence plan

Tools and materials. Challenge template, protocol template, observation instrument

Experiment Focus

- Experiment focus: responsibility and decision-logic variables
- Primary manipulated condition: implicit versus explicit decision logic around the same AI-supported recommendation. The target condition should define decision ownership, review responsibility, evidence requirements, approval rights, escalation triggers, and a contestation route.
- Optional secondary manipulation: one Human-Context variable such as disclosure, stakeholder participation, or explanation may be varied only when the participant can explain why it does not make the comparison uninterpretable.
- What remains constant across comparison: Keep the AI recommendation, evidence packet, decision question, participants or role composition, available time, and observation instrument as stable as practicable.
- Prediction requirement: Before the baseline, participants predict which ambiguities will reduce accountability or appropriate reliance. Before the replay, they predict which indicators should change under explicit decision logic.

Observation Focus

- Accountability clarity: Can participants identify who owns the decision and why that person has authority?
- Evidence use: Can participants identify which evidence is required, which evidence was inspected, and which gaps remain?
- Escalation: Can participants identify a condition that prevents routine approval and the role to which it must be escalated?
- Contestability: Can an affected person identify and exercise a route for questioning or appealing the recommendation?
- Appropriate reliance: Is reliance on the AI proportionate to the evidence, uncertainty, and consequences rather than to fluency alone?
- Perceived legitimacy: Do reviewers judge the process as procedurally justified, and can they connect that judgment to participation, values, evidence, and decision rights?
- Evidence participants should capture: baseline and target decision logs; completed responsibility matrix; observation ratings using a shared four-point scale; short written reasons; evidence or escalation actions; peer or stakeholder comments; and unexpected results.
- Interpretation participants should be prepared to make: Explain which elements of decision logic changed observed behaviour or perception, where causal attribution remains uncertain, and why increased acceptance alone would not demonstrate responsible control.

Challenge Kickoff Deliverables

Participants must leave Friday with:

- selected low-risk AI-supported recommendation or decision process, or explicit choice to retain the common case
- baseline decision arrangement with the main ambiguity identified
- target decision arrangement with one primary manipulation
- first responsibility matrix identifying who recommends, reviews evidence, must be consulted, decides, approves, and receives escalation
- first protocol draft covering the decision and intended outcome, manipulated decision-logic condition, constants, predicted effects, observation indicators, and evaluation criteria
- evidence plan identifying who will provide structured observations and how the two conditions will be compared
- at least one first captured artefact: a baseline decision log, completed observation sheet, decision-protocol sketch, or peer feasibility review

The kickoff should already reflect the challenge-focus logic:

- Difficulty: Technically plausible outputs can still fail because decision rights, responsibility, escalation, evidence requirements, and legitimacy remain unclear.
- Approach: Compare the same AI-supported recommendation and evidence under an implicit baseline and an explicit decision protocol. Observe whether ownership, evidence use, escalation, contestability, appropriate reliance, and perceived legitimacy become clearer or more substantive.
- Artefacts: protocol document; fixed recommendation and evidence pack; prediction table; baseline and target decision logs; responsibility matrix; decision protocol; structured observation evidence; concise interpretation; residual-risk statement; and one practical leadership rule.

Instructor Adaptation and Expert Contribution

The common experiment is the stable spine of the Friday. The instructor should be free to supply the conceptual framing, domain examples, and measurement refinements that best fit their expertise.

If Theresa Schmiedel teaches the unit, her publication-derived background can enter at four deliberate points:

- Culture and values as design variables: use BPM-culture, organizational-change, or value-sensitive-design examples to show that authority and legitimacy depend on context rather than on a universal governance recipe.
- Measurement before intervention: deepen the measurement mini-lab by showing how broad constructs become indicators, scales, and observable evidence without requiring participants to conduct a full validation study.
- Context-aware process thinking: challenge participants to ask whether the decision arrangement fits process variety, uncertainty, local conditions, and the consequences of standardization.
- Human perception in interactive systems: use social-robot or generative-agent examples to examine how explanation, response behaviour, knowledge claims, persuasion, and visible oversight affect perception and reliance.

She may replace the default process-standardization case with a social-robot, organizational-transformation, or generative-agent case provided that the baseline/target comparison, low-risk fictional framing, and common observation logic remain intact. Relevant publications may be used as optional conceptual anchors, but the Friday should not become a biography or literature survey.

Teacher Setup

Mandatory:

- one common fictional case pack with a fixed AI recommendation, evidence excerpts, known uncertainty, and a clear decision question
- separate baseline and target instructions; do not reveal the target protocol before the baseline observation is captured
- role cards for decision owner, process owner, evidence reviewer, affected-stakeholder representative, and optional AI/process steward
- responsibility-matrix and decision-protocol templates covering recommendation, consultation, evidence review, decision, approval, escalation, and contestation
- one shared observation instrument using a four-point anchored scale plus short reasons; anchors should distinguish symbolic, partial, and substantive control
- one worked comparison showing generic human sign-off versus meaningful authority, evidence, time, and criteria to contest
- protocol, prediction, decision-log, comparison, and evidence-capture templates for the challenge kickoff
- demonstration setup for the recurring agentic coding case, if used
- paper-based or slide-based fallback so the core experiment does not depend on live model or platform access
- explicit instruction that class cases must remain fictional or low risk and must not expose confidential organizational information or involve real consequential decisions

Configurable:

- case domain: process standardization, organizational transformation, social robot or service agent, generative AI adoption, or another low-risk socio-technical decision
- exact examples, readings, and research connections used in the impulse
- work mode and group size; groups of four or five allow the suggested role distribution, while smaller groups may combine roles
- whether observation evidence comes from classmates, invited stakeholders, or a teacher-configured equivalent
- whether participants use the shared case throughout or transfer to their own context during the final block
- exact tools; the day can be run with printed materials and a markdown editor, while a chat interface may be added for demonstrations
- time allocation, scaffolding depth, and peer-feedback format, provided the baseline is observed before the target protocol is introduced