

Lesson 3: Reverse Engineering

iste+ascd

The background features several large, overlapping geometric shapes. A large, light blue semi-circle is positioned in the upper right quadrant. Below it, a dark blue semi-circle is visible. In the lower right, there is a light blue rounded rectangle and a large orange circle. The overall design is modern and abstract.

Created with support from:



Lesson Organization

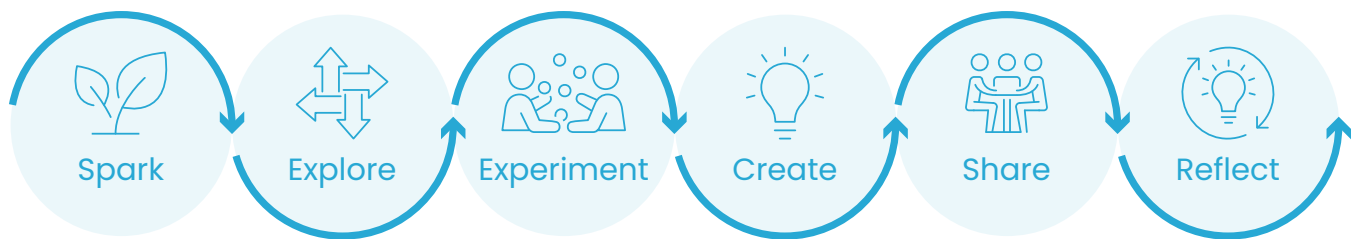
This lesson has three major parts to it:

1. **Lesson Overview – *Get Oriented***

Start here for a quick sense of what the lesson is about, why it matters, and how it connects to the Studio framework. You'll find the learning goals, key standards, and facilitation tips for pacing, setup, and context.

2. **Studio Flow – *Teach and Create***

The main body of the lesson follows the Studio process: **Spark** → **Explore** → **Experiment** → **Create** → **Share** → **Reflect**.



Each phase includes clear student activities, facilitator moves, and checkpoints to guide learning from curiosity to purposeful creation.

3. **Lesson Support – *Deepen and Extend***

Wrap up with practical supports: **Artifacts & Assessment** ideas, **Extensions & Connections** for enrichment or next steps, and a link to the **Resources & References Folder** with lesson materials and examples.

Lesson 3 Overview

Focus

Deconstructing existing AI and digital systems to uncover design choices, hidden assumptions, and user impacts.

Estimated Duration

90–120 minutes (can be split into two sessions)

Stage of the Studio

Process System Analysis and Ethical Reflection

ISTE Standards Alignment

- **1.4 Innovative Designer** – Students apply design thinking in reverse, deconstructing an existing AI or digital system to infer how it was built. They test ideas, gather evidence, and iterate on their understanding as they create a Reverse-Engineered Design Map that visualizes hidden design choices.
- **1.5 Computational Thinker** – Students use decomposition and systems analysis to break down complex technologies into inputs, processing, outputs, and impacts. They identify patterns and make reasoned hypotheses about the algorithms or priorities guiding each system.
- **1.7 Global Collaborator** – Students consider how system design affects different communities, using perspectives like Two-Eyed Seeing to balance technical and human viewpoints. Through reflection and shared analysis, they evaluate fairness, bias, and inclusion in real-world technologies.

Learning Goals

By the end of this lesson, I will be able to:

- Break down an AI or digital system into inputs, processes, outputs, and impacts.
- Identify hidden assumptions and biases in how systems are designed.
- Represent their findings in a Reverse-Engineered Design Map (and optional Model Card).
- Reflect on the ethical responsibilities of designers, using various perspectives.

Facilitator Notes

This lesson works best when students have real systems to take apart. Offer multiple examples (TikTok or YouTube feeds, Quick, Draw!, Most Likely Machine, translate autocorrect features, etc.). Allow students to choose one in small groups. The **As A Matter of Fake Starter Pack** can be run as a capstone “impossible detector” extension, reinforcing why misinformation detection cannot be solved by technical fixes alone.

Big Goal of the Lesson

Students see that AI systems are not “black boxes” but collections of human choices, training data, and algorithms. By reverse engineering them, learners become better critical thinkers, able to ask who benefits, who is excluded, and what responsibilities designers hold.

How the Lesson Flows

- **Spark:** Surprising or flawed AI example provokes curiosity.
- **Explore:** Groups preview systems and learn how to break them into inputs, processes, and outputs.
- **Experiment:** Students test their chosen system to uncover hidden rules.
- **Create:** Build a Reverse-Engineered Design Map and, optional Model Card.
- **Share:** Groups present their maps; facilitator highlights cross-system connections.
- **Reflect:** Students consider ethical responsibilities and alternative perspectives.

Challenge Question

How can we take apart existing AI and digital systems to uncover their hidden design choices, assumptions, and impacts on people?

Framing Concepts

- Every system we use — from recommendation feeds to translation tools — is built on design decisions that shape what we see and experience.
- Reverse engineering helps us look under the hood, breaking a system into inputs, processes, and outputs to see how it really works.
- By mapping hidden assumptions, we uncover issues of bias, transparency, and fairness that affect different users in different ways.
- Designers hold responsibility because their choices influence culture, behavior, and equity. Understanding these choices prepares us to create better systems ourselves.
- Using various perspectives, we can balance technical analysis with empathy and community insight, seeing both what the system does and what it means for people.

Studio Flow

Spark

Purpose: Grab attention and provoke curiosity by showing students that AI systems don't behave randomly, but they follow hidden rules and patterns that can be uncovered.



Intended Outcome: Students begin to question why a system produced a certain output and recognize that there are invisible design choices at play.

Mini-Challenge Activity (10 minutes)

- Facilitator shows one striking system behavior. Options:
 - Quick, Draw! (draw a cat and it guesses "potato").
 - TikTok or YouTube feeds (two side-by-side screenshots look totally different).
 - Translate and Autocorrect (type an idiom, e.g., "it's raining cats and dogs," and show a literal translation).
- Prompt students:
 - "Your job is to be system detectives. Don't just laugh at the weirdness, ask: *Why did the system do that? What do you think it was using behind the scenes?*"
- In pairs, students generate 2–3 quick "suspect factors" (inputs or rules that could explain it).
- Groups stick answers on the board (sticky notes or digital whiteboard).

Artifact: Sticky notes or short list of suspected hidden rule and/or/inputs from class brainstorm.

Facilitator Moves

- Keep energy high — treat it like a puzzle.
- If students say "the AI is dumb," redirect:
- "Instead of calling it dumb, let's think — what choice or data might have made it do that?"
- If groups get stuck, seed them with options: "Could it be the data it collected, the time of day, the way you drew it, the popularity of content?"

Scaffold

- Provide sentence starters:
 - “It used _____ to make its guess.”
 - “It prioritized _____.”
 - “Maybe it ignored _____.”
- Show a pre-captured screenshot in case of technology challenges.

Checkpoint (readiness to move on)

- Students can name at least two possible factors behind the system’s behavior.
- Whole class sees that “black box” systems can be broken into *inputs* → *processing* → *outputs*.

Explore

Purpose: Move beyond curiosity to structured analysis. Students learn a framework for dissecting systems and start applying it to reveal early patterns, design choices, and interdependencies.



Intended Outcome: Students understand how systems connect data, algorithms, and outcomes — and begin identifying the *questions worth testing* in the next phase.

Mini-Challenge Activity

1. Revisit the Spark

“We just saw that systems behave in surprising ways. Now we’ll figure out *how to think like a reverse engineer* — how to take that weirdness apart logically.”

2. Model “Deep Deconstruction”

- Using one shared example (Quick, Draw! or Translate), model the full **system map**:
 - *Inputs*: what data enters (drawings, keystrokes, clicks)
 - *Processing*: what the system prioritizes or filters (pattern matching, engagement scores)
 - *Outputs*: what the user sees (label, recommendation, correction)
 - *Feedback loop*: how output data feeds new inputs (likes, retraining, personalization)
- Emphasize this is not static — it’s dynamic and recursive.

3. Group Investigation Setup

- Each group chooses one system from the menu (TikTok/YouTube, Quick, Draw!, Most Likely To, Translate, or Fake News Starter Pack).

- Students complete a mini “System Blueprint”:
 - List 3 data types the system collects (inputs).
 - Predict 2 things the algorithm probably values (processing).
 - Identify 1 way it could amplify or distort outcomes (impact).
- Groups post their blueprints so others can browse.

4. Synthesize as a Class

- Facilitator leads short debrief:

“What patterns do we notice across systems? What seems to drive most of them?”
- Circle recurring themes: engagement, popularity, accuracy, bias, user retention, etc.

Artifact: Group “System Blueprint” (proto-map outlining early hypotheses).

Facilitator Moves

- Draw explicit connections to **systems thinking**: “Every part influences the others.”
- Push for reasoning, not guessing: “What evidence makes you think that’s an input?”
- Use clear examples for modeling. Don’t skip the Impact step — it primes them for ethics later.
- If students can’t think of systems, suggest by category such as (“social media, translation, drawing, harmful bias/facial recognition”).
- Encourage diversity of choices so the Share later has variety.
- Preview the next step: “We’ll test these assumptions in Experiment to see which ones hold up.”

Scaffolds

- Provide “System Blueprint Template” with guiding prompts:
 - “What does the user give the system?”
 - “What must it be doing behind the scenes to decide?”
 - “How might its goals shape bias or error?”
- Offer one filled-in example for reference.

Checkpoint

- Every group has a System Blueprint with clear hypotheses about inputs, processing, and outputs.

- Class discussion reveals at least two shared system patterns (e.g., personalization or data feedback).
- Students are ready to test these hypotheses in the next phase.

Experiment

Purpose: Give students hands-on time to test their chosen system, collect evidence, and uncover hidden assumptions.



Intended Outcome: Students identify at least one hidden “rule” or assumption in the system and gather evidence to support their claim.

Mini-Challenge Activity

1. Set the frame:

- “Now you’ll run experiments on your system. Take on the role of a system tester: make a hypothesis, test it, record what happens, and then guess what rules the designers built in.”

2. Group work:

- Students investigate their chosen system with at least 2–3 quick tests.

Examples of mini-tests by system:

- **TikTok/YouTube feed:** Watch or like only sports for 5 minutes; what appears next? Compare with a classmate’s feed.
- **Quick, Draw!:** Draw the same object three different ways (stick figure, cultural style, abstract). Which get recognized?
- **Most Likely To Machine:** Try 2–3 different photos (or use demo images). How are people labeled differently?
- **Translate/Autocorrect:** Enter idioms, slang, or names. Does it misinterpret meaning? Does it change based on capitalization?
- *Optional Extension – Fake News Starter Pack:* Follow steps to generate/analyze fake content and track what makes it look “real.”
- Students record results in the **Reverse-Engineered Design Map** template (Inputs, Processing Guesses, Outputs, Impacts).

3. Quick pair-share:

- Each group identifies one hidden “rule” they discovered and shares it with another group nearby.

Artifact: Notes, screenshots, or sketches with evidence of system behavior and one emerging hidden rule or assumption.

Facilitator Moves

- Remind groups of the classic testing principle: “One variable at a time!”.
- Circulate with prompts:
 - “What do you think the system is prioritizing?”
 - “What would happen if you change just one input?”
 - “Who might be left out by how it responded?”
- Push groups to connect tests to possible *designer goals* (e.g., engagement, accuracy, profit, user safety).

Scaffolds

- Provide hypothesis frames:
 - “If I do ____, the system will ____.”
 - “Because it seems to value ____, I predict it will show ____.”
- Encourage evidence capture: screenshots, tallies, short notes.
- Offer pre-filled **Map starters** with Inputs and Outputs filled in so groups just add Processing + Impacts.

Checkpoint (readiness to move on)

- Each group has documented at least one test and its result.
- Each group can name one hidden rule or assumption.
- Evidence is in progress -map partially filled.

Create

Purpose: Synthesize observations into a clear Reverse-Engineered Design Map that shows how the system works — inputs, processing, outputs, and impacts — and highlights hidden assumptions and ethical considerations.



Intended Outcome: Students create a visual representation of their system that demonstrates critical thinking about data, algorithms, and design choices.

Main Challenge Activity

1. Introduce the task:

- “Now you’ll bring your detective work together into a clear map that shows how your system works under the hood. You’re not guessing everything exactly right — you’re building your best evidence-based model.”

2. Group creation:

- Using the **Reverse-Engineered Design Map template**, students:
 - Identify **Inputs** -data collected
 - Describe **Processing** - rules, priorities, or assumptions they observed
 - List **Outputs** -what the system delivers
 - Note **Impacts** -who benefits, who is harmed, biases uncovered
- Add **Designer Assumptions** such as “More watch time = more interest”
- Optional for extension groups: complete a **Model Card** with system purpose, intended users, risks, and limitations.

3. Prep for Share:

- Groups make their map presentation-ready (paper posted on wall, digital diagram in a tool like Canva or Miro, or slide snapshot).

Artifact

- **Reverse-Engineered Design Map** (required)
- **Optional Model Card** (extension groups)

Facilitator Moves

- Model a quick, simple example map at the start (e.g., Quick, Draw!).
- Circulate and push groups with targeted questions:
 - “What data does it *definitely* use? What data might it ignore?”
 - “What rule seems to drive this output?”
 - “Whose perspective might not be represented in the training data?”
- Remind students that clarity matters more than perfection — “Your goal is to make the invisible choices visible.”

Scaffolds

- Provide **Map template** with labeled boxes (Inputs → Processing → Outputs → Impacts).
- Sentence frames for assumptions:
 - “The designers assumed that ____.”
 - “This system values ____ over ____.”
- For groups struggling: give them a **half-filled template** (Inputs/Outputs) so they only add Processing and Impacts.
- For advanced groups: encourage them to brainstorm **possible mitigations** (“What would you change as a designer?”).

Checkpoint

- Each group has a complete map with all four sections (Inputs, Processing, Outputs, Impacts).
- At least one assumption and one ethical impact are explicitly stated.
- Groups are prepared to explain their map to peers in the Share phase.

Share

Purpose: Make student thinking visible, build connections across different systems, and highlight common patterns in how AI systems are designed.



Share

Intended Outcome: Students communicate their reverse-engineered insights clearly and recognize similarities and differences across multiple examples.

Activity

1. **Setup:**
“Now each group will present your Reverse-Engineered Design Map. Remember: focus on what you uncovered about how the system works, not every detail.”
2. **Lightning presentations:**
 - Each group gets **90 seconds** to share:
 - One surprising input or assumption
 - One key output pattern
 - One impact on users (bias, fairness, or exclusion)
 - Other groups can ask one clarifying question (30 sec)
3. **Facilitator synthesis:**
 - Capture insights on the board under 3 headings: **Inputs, Processing Rules, Impacts**
 - Point out cross-example themes (e.g., “Notice that both TikTok and Quick, Draw! rely heavily on training data, but they affect users differently”).

Artifact

- Group Reverse-Engineered Design Maps displayed (on walls, slides, or a shared drive).
- Class board or chart with collected “design assumptions” and “impacts” across systems.

Facilitator Moves

- Keep groups concise by timing them (use a visible countdown).

- Ask probing follow-up questions if needed:
 - “What evidence made you think the system prioritized that?”
 - “Who do you think is most affected by this assumption?”
- Emphasize that *nobody gets it 100% right* — the point is to show plausible explanations based on evidence.

Scaffolds

- Provide a presentation frame:
 - “One input we found was ____.”
 - “The system seemed to process it by ____.”
 - “That led to an output where ____.”
 - “The impact was ____.”
- Allow groups to post their map and do a “gallery walk” instead of live presentations if time is short.

Checkpoint

- Every group has shared or posted their map.
- Class has a visible list of cross-cutting insights (design assumptions and impacts).
- Students can name at least one similarity and one difference across systems.

Reflect

Purpose: Deepen insight by connecting technical observations to ethics, empathy, and responsibility. Use different “lenses” (Two-Eyed Seeing, community impact, designer responsibility) to make sense of the maps together.



Intended Outcome: Students recognize that systems are built on human choices, and they articulate how those choices impact users, communities, and fairness.

Activity

1. **Set the frame:**
“Reverse engineering shows us how systems work. Reflection helps us ask so what? What do these design choices mean for people, communities, and fairness?”
2. **Lens rotation:**
 - Display three lenses on the board:
 - **Designer Responsibility Lens** – What choices were made? Who benefits? Who is harmed?

- **Two-Eyed Seeing Lens** – How does looking through both a technical lens (data/algorithms) and a community lens (stories, lived experience) change what we notice?
- **User Impact Lens** – How does this system shape behavior, identity, or opportunity?
- Groups revisit their maps for 3–4 minutes, highlighting one insight through each lens.
- Then share out with each group contributing one insight under one lens (posted on wall/board).

3. AI mirror activity:

- Groups ask an AI to summarize their map and ethical impacts.
- Then critique: *What did the AI miss? Did it overlook bias, community context, or empathy?*

Artifact

- Maps annotated with lens-based notes (sticky notes or highlights).
- Class board showing 2–3 insights under each lens.

Facilitator Moves

- Keep reflection concrete: always tie back to examples (“What’s one thing TikTok’s designers assumed?” not “What is design responsibility in general?”).
- Draw connections between groups: “Notice how both Translate and Most Likely To show issues when cultural nuance is ignored.”
- Push for empathy: “If this system were deployed in a different community, how would its impacts shift?”

Scaffolds

- Provide lens prompts on handouts or slide:
 - Designer Responsibility: “What choice did the designers make on purpose? What was left out?”
 - Two-Eyed Seeing: “What do we see with data? What do we see with lived experience? How do they differ?”
 - User Impact: “What does this system encourage you to do more/less of?”
- Offer examples to get discussion rolling (e.g., “In Quick, Draw!, diverse drawing styles might be misrecognized → designer responsibility is dataset diversity”).

Checkpoint (lesson closure)

- Students have contributed at least one insight under a lens.
- Class collectively sees that hidden design choices result in ethical consequences.
- Students leave with language for both the technical breakdown and ethical reflection.

Lesson Support

Artifacts & Assessment

Student Artifacts

- Reverse-Engineered Design Map (required)
- Annotated insights from reflection lenses (Designer Responsibility, Two-Eyed Seeing, User Impact)
- Optional Model Card (for advanced groups)

Assessment Focus (aligned to Studio Rubric)

- Responsible & Ethical Use of AI: Attention to bias, inclusivity, and explainability in their analysis.
- Innovation & Creativity: Originality in how students visualize and explain the system.
- Design Quality & Accessibility: Map clarity, usability, and completeness (inputs, processing, outputs, impacts).
- Reflection & Process: Lens-based insights showing empathy, ethical reasoning, and systems thinking.

Extensions & Connections

Extension Challenges

- **Impossible Detector (Starter Pack: As A Matter of Fake):** Students attempt to design a fake news detector, then reflect on why it cannot fully succeed and what other safeguards (policy, education, provenance) are needed.
- **Remix the System:** Redesign the system to fix one of the uncovered biases (e.g., broaden training data for Quick, Draw!, change TikTok's prioritization rule).
- **Apply a Constraint:** Map the system again imagining it was designed with frugal innovation (e.g., limited data or budget). How would assumptions shift?
- **Global Lenses:** Compare the same system across two contexts (e.g., Translate English to Spanish vs. English to Indigenous language). What new impacts show up?

Connections to Future Lessons

- Leads into Lesson 4 (Impossible Tasks) by surfacing the difficulty of explainability, transparency, and fairness.
- Strengthens the foundation for Challenge Projects: students will need to critically analyze systems they design.
- Connects to SDGs: builds awareness of how design choices affect equity, inclusion, and sustainable communities.

References & Resources

To support flexible use and continual improvement, all lesson resources are housed in shared online folders. This approach allows us to update materials as new tools and examples emerge or suggestions are made. This also ensures they align with your school or district's preferred platform. Each link below will take you to a list of the editable templates, example artifacts, and any supporting media referenced in the lesson.

Access the full set of resources through your preferred platform:

- [Google Documents Resource List](#)
- [Microsoft Documents Resource List](#)