



# Understanding how AI Systems Perform Tasks, But Are Not Human

## Competency

Learners describe how AI operates in realistic and accurate terms, avoiding language that suggests AI possesses human feelings or understanding. They understand that their language can either clarify or perpetuate misconceptions about AI.

## Key Method

The educator examines how AI systems are designed to produce humanlike language. The educator examines how training processes and human feedback shape AI responses and explains why those responses may appear humanlike.

The educator analyzes common examples of anthropomorphic language and reflects on their own language use with technology and its impact on

the classroom. They then apply this understanding by developing clear explanations and examples that help students and colleagues distinguish between human reasoning and automated task performance.

## Method Components

AI Predicts. It does not think.

AI tools often produce responses that sound human, but they are not “thinking” or “understanding” anything the way people do. Keeping that distinction clear helps us

teach students how to use AI safely and realistically. This approach allows teachers to model appropriate and ethical use of AI.

- AI systems generate outputs by analyzing vast amounts of data and identifying patterns.
- AI predicts what statistically should come next.
  - Such as the next word, label, or item in a sequence.
- Outputs may appear thoughtful or creative, but they are a result of prediction, not comprehension or feeling.
- AI does not have feelings, intentions, judgment, or life experience.
- When students (or adults!) assume AI “knows” or “AI understands,” it can lead to over-trust. This over-trust leads users to skip reviewing outputs.

To improve modeling for students and our own use, it is helpful to name a few common types of AI tools in education:

- **Machine Learning (ML):** Systems that identify patterns in data to make predictions or classifications.
  - *Education example: A system reviews patterns in student quiz results and highlights which standards students may be struggling with, helping a teacher decide where to reteach or provide additional practice.*
- **Deep Learning:** A subset of ML that uses layered networks to support complex tasks like language or image processing.
  - *Education example: A system converts spoken instruction into written captions or transcribes student presentations by identifying patterns in audio and language.*
- **Generative AI:** Tools that create new content, such as images, music, or text. These are based on patterns or training data.
  - *Education example: An educator uses a tool to generate a draft example, discussion prompt, or explanation that they then review, revise, and adapt for instructional use.*
- **Rule-based (Expert) Systems:** Tools that follow predefined rules to reach outcomes in specialized fields.
  - *Education example: A system checks whether a student meets course prerequisites or flags scheduling conflicts in accordance with established rules.*

Starting to understand these categories helps us understand that AI may be helpful and powerful, but it is still pattern-matching, not human thinking. Accurately naming the system without attributing human qualities to it is an essential step toward avoiding anthropomorphism.

## Why This Creates Anthropomorphism

AI systems are trained on extensive collections of human-written text, such as books, websites, and conversations. During training, the system repeatedly predicts what

word or phrase is likely to come next based on patterns in that data. Because these patterns reflect how humans communicate (including tone, phrasing, and conversational structure), AI-generated outputs can sound empathetic, even though they are generated through statistical prediction.

As a result, people may begin to describe AI behavior using human-centered language. In educational settings, this often sounds like students saying, “AI knows,” “AI understands me,” “AI said,” “AI told me,” or “I couldn’t decide, but AI helped me decide.”

Anthropomorphic framing is the practice of attributing human-like understanding, intention, or decision-making to a non-human system. This framing can obscure the technical realities and limitations of how AI systems operate.

In some tools, this illusion is reinforced by design choices that prioritize personalization. For example, systems may prompt users to share names, preferences, or background details and then reference that information in future responses. While this can feel engaging or even “cool,” it can also blur essential boundaries. The system does not get to know the student; it stores and reuses information to shape outputs.

For educators, this means slowing down to make intentional choices. Before enabling personalization features or modeling their use with students, it is worth asking what message the design sends. Choosing neutral naming, limiting personal data, and explicitly explaining how the tool uses information helps keep the focus on learning rather than perceived connection. Small instructional decisions reinforce that AI is a tool for learning.

## Reinforcement Learning from Human Feedback

After understanding why AI sounds human, it helps to look at how those responses are reinforced during development. One key process behind this is Reinforcement Learning from Human Feedback (RLHF).

As educators, we take the name itself as a useful clue. RLHF does not mean that an AI system understands human values or makes judgments. Instead, it is a process used during development to adjust which types of responses are more likely to appear based on human preferences. In plain terms, humans review AI outputs, compare them, and indicate which responses are more helpful, precise, safe, or appropriate. The system is then tuned to produce more responses like those that were rated higher.

What RLHF looks like in practice:

- **Humans compare responses:** Human feedback comes from two places: reviewers who help tune the system before release, and users whose feedback informs future updates. Classroom use does not retrain the AI in real time, but it does influence how later versions are adjusted.

- **Pattern Type:** The model is tuned to reproduce interaction patterns that humans consistently rate higher, often with a thumbs-up vote or by selecting a “I prefer this answer.” Humans usually select supportive phrasing and cautious explanations that reinforce language patterns.
- **Anthropomorphic Effect:** The output resembles human language patterns. The user may interpret this language emotionally, even though the system is reproducing preferred patterns rather than responding with understanding.

RLHF helps explain why AI systems often sound agreeable or aligned with the user. The system is not deciding to be helpful; it is producing responses that match patterns humans previously preferred. Without this context, it can be easy for students or educators to mistake tone for reasoning or agreement for accuracy.

## Instruction Tuning and Dialogue Training

Building on reinforcement processes such as RLHF, AI systems also undergo instruction tuning and dialogue training during development. In this stage, systems are exposed to extensive collections of example prompts and dialogue exchanges, increasing the likelihood that they will follow directions and produce structured, back-and-forth responses. This tuning occurs during development and does not happen in real time through individual user interactions.

- **Pattern Type:** Question–answer sequences, conversational repair (such as, “I’m not sure what you mean, could you clarify?”), and turn-taking behavior.
- **Anthropomorphic Effect:** Because these interaction patterns resemble common human conversations, such as acknowledging a question or offering clarifications, responses can appear empathetic or thoughtful. This effect reflects reinforced communication patterns, not a lack of understanding or intent.

## Anthropomorphic Emergence

Even when AI responses sound human, the system is not experiencing emotions or understanding meaning. It reproduces language patterns commonly used when people express feelings, giving the impression of emotional awareness.

These patterns include:

- Emotional and Relational Language
  - Responding kindly to sad or frustrated inputs (such as, “I understand that you are confused,” “Those feelings are normal”)
  - Validating the user through supportive or reassuring phrasing
- Conversational Positioning
  - Using “you” and “I” in socially consistent ways

- Assigning itself a role (such as responding as a thought partner, even if not prompted to have that role)
- Following social cues, such as prompting or eliciting a response
- Politeness and Social Conventions
  - Using politeness formulas (“please,” “thank you,” “I would be happy to help”)
  - Including apologies when a response is incorrect
- Narrative Coherence
  - Organizing responses into cause-and-effect explanations that resemble human reasoning

When these patterns appear together, AI can seem humanlike, even though it is simply generating language based on patterns rather than on awareness or understanding.

Educators and students often unintentionally use humanlike language when describing AI behavior, such as saying the model:

- *Understands* the user
- *Knows* information
- *Tries* to help
- *Clarifies* because it senses confusion
- *Thinks or decides* what to say next

In reality, the model is only predicting statistically likely responses—without awareness, intention, or emotion.

Educators who want to explore this topic further may:

- Compare how raw AI models differ from aligned models (such as, pre-RLHF vs. fine-tuned versions).
- Notice linguistic markers such as empathy words, hedges (“maybe,” “I think”), or conversational scaffolds that make AI seem humanlike.
- Read about *anthropomorphic bias*—how people naturally project human intentions onto patterned text.

AI systems are not responding like humans. They generate text through pattern recognition and rapid prediction, but they do not understand intention or emotion.

What are some applications of AI that present with anthropomorphism?

There are many ways AI assists us in education. Examples include drafting emails, outlining lesson ideas, summarizing content, translating languages, and more. Anything produced by AI, however, still requires human review and revision.

Examples of AI use in education:

**Communication:** When drafting emails, family newsletters, classroom updates, or short communication snippets, AI can provide a clear, well-structured starting point. It can also help shorten messages, adapt content for different formats (such as school platforms or websites), or translate communication into other languages. The educator remains responsible for accuracy, tone, and appropriateness.

**Student Feedback and Revisions Support:** AI-supported features embedded within curriculum platforms or instructional tools can provide timely feedback, prompt student thinking, or suggest revisions. These features are often engaging, sometimes gamified, and may adjust support based on student input, helping learners access content more independently and supporting differentiation.

In some educational tools, anthropomorphism is further reinforced through naming, avatars, or “student companion” features (for example, a named chat assistant or a character presented as a learning partner). While these design choices can increase engagement and reduce anxiety, they may also strengthen the impression that the tool is a friend or thinking agent. Educators should be aware of how naming and character design influence student perception. Also, they should discuss with students that these tools support learning and do not replace human relationships or instruction.

Because these tools often use encouraging and responsive language, they can create a cheerleader effect, making the system seem supportive or “on the student’s side.” This mistake can lead students or educators to describe the tool in human terms, such as “understanding” or “knowing”. That tone can feel supportive, but it stems from learned language patterns rather than judgment or understanding. Teachers still need to decide when AI responses are helpful, misleading, or inappropriate.

What are some of the challenges of using AI on vulnerable populations?

In educational settings, anthropomorphic emergence, mimicry, and sycophancy can have uneven impacts on learners. Students who are still developing reasoning skills, academic confidence, or self-regulation may be more likely to over-attribute trust to AI tools—especially when AI outputs sound supportive or aligned with the user.

This is where teacher expertise matters. Educators know their students’ developmental levels, academic readiness, and learning needs, and they use

that knowledge to decide when guidance is appropriate, when struggle is productive, and when support should be adjusted. This knowledge makes teachers' judgment especially important when working with vulnerable learners.

Two interrelated challenges educators should be aware of are mimicry and sycophancy.

**Mimicry** refers to AI's ability to reproduce language, tone, and visual patterns that closely resemble human communication. These outputs are generated through pattern recognition. Because AI systems reflect the data they are trained on, gaps or imbalances in that data can result in misleading outputs, particularly for groups that are underrepresented or misrepresented in training datasets.

In educational contexts, mimicry can surface in several ways.

- AI-generated explanations may sound confident even when they are inaccurate.
- Writing feedback may resemble supportive teacher language without the instructional judgment that normally guides such feedback.
- AI-generated media may appear so realistic that they are mistaken for authentic sources.

One increasingly relevant example is the use of deepfakes. AI-generated images, audio, or video can convincingly mimic real people or events, making it difficult for students to distinguish between authentic and synthetic media. In school settings, this raises important concerns related to misinformation and digital literacy. As AI-generated media becomes more realistic, educators can no longer assume that visual or audio evidence is trustworthy.

Other classroom-relevant impacts of mimicry may include:

- Over-reliance on AI explanations that sound authoritative but lack accuracy
- Reduced questioning when responses feel polished or reassuring
- Increased difficulty distinguishing original student work from AI-generated content

**Sycophancy** refers to an AI system's tendency to align with user input or preferences rather than challenge assumptions or present alternative viewpoints. Because many systems are tuned to prioritize user satisfaction, they may reinforce ideas the user already holds, even when those ideas are incorrect.

In educational use, this can occur when AI tools consistently validate student responses, endorse questionable reasoning, or avoid introducing productive disagreement. When learners repeatedly receive affirming or agreeable feedback, they may develop a distorted sense of correctness or confidence without engaging in deeper analysis.

This pattern can be especially concerning for students who are seeking reassurance, struggling academically, or unsure of their own judgment. AI empathetic-sounding validation can be misleading when it comes from a system that cannot evaluate nuance or context.

Understanding how mimicry and sycophancy function allows educators to make more intentional instructional choices. Rather than treating humanlike AI responses as evidence of understanding or reliability, educators can model how to question tone, evaluate content, and verify information.

By explicitly naming these patterns and discussing their limits, educators help students develop critical awareness of AI-generated outputs. This supports healthier boundaries, reduces over-trust, and reinforces the role of human judgment, instruction, and accountability in learning.

## Human Words, Machine Meaning

The lexicon of AI has grown and altered the meaning of many words. To better understand the process of machine learning and respond to it, it is only natural that humans connect words from our language to these new tools. Many of these are uniquely human-centric, yet they explain non-human processes. The list will always be growing. Here are a few with the non-anthropomorphic meanings.

Anthropomorphic words for AI include "think," "see," "recognize," "learn," and "help," which attribute humanlike cognitive abilities and actions to machines.

Learn, cognition	Pattern sorting and recognition to apply to future processes
Think, decide	Applying the data set information to a current situation
Perspective	How the machine is responding
Opinion	Sorting the possible responses based on an algorithm
Emotional or social words	Applying the use of words such as "Thank you," "I understand," etc.
Personal Identity	Use the words or title of me, my, I, as well as you, your

Understanding these terms helps us describe AI accurately and avoid unintentionally using terms that are accurate but misleading, which in turn helps us and our students remember that AI is not human. It is a tool we guide, not a system that thinks or feels.

### AI in Education

# Supporting Rationale and Research

Alabed, A., Javornik, A., and Gregory-Smith, D. (2022). AI anthropomorphism and its effect on users' self-congruence and self-AI integration: A theoretical framework and research agenda. *Technological Forecasting and Social Change*, 182, 121786. <https://doi.org/10.1016/j.techfore.2022.121786>

<https://doi.org/10.1016/j.techfore.2022.121786>


Paterson, J. M. (2025). AI mimicking and interpreting humans: Legal and ethical reflections. *Bioethical Inquiry*, 22, 539–550. <https://doi.org/10.1007/s11673-025-10424-9>

## Resources

AI Support for Educators

[Microsoft Elevate for Educators](#)

Observing Humanlike Responses in AI Outputs

 Can AI Companions Help Heal Loneliness? | Eugenia Kuyda | TED

[Can You Be Friends with AI? Pros and Cons Explained](#)

Recognizing Anthropomorphism in Everyday AI Conversations

[AI Mimicking and Interpreting Humans: Legal and Ethical Reflections - PMC](#)

[AI Sycophancy: The Downside of a Digital Yes Man](#)

[The Threat Posed by Deepfakes to Marginalized Communities | Brookings](#)

Responsible Use of AI in Educational Contexts

[Resources for Empowering Educators in the Age of AI | NEA](#)

[What is a Chatbot and How Does it Work | Common Sense Media](#)

Understanding Anthropomorphic Language in AI

[AI Glossary of Terms](#)

[Mimicry Argument Against AI | MIT Sloan Teaching and Learning](#)

## Submission Guidelines and Evaluation Criteria

*To earn this micro-credential, you must receive a passing score in Parts 1 and 3 and be proficient in all components in Part 2.*

### Part 1. Overview Questions (Provides Context)

#### **250-300 words**

Please do not include any information that will make you identifiable to your reviewers.

Please answer the following questions:

1. What is your role in education, and how do you currently use AI in your work?
  - a. Include information such as what devices are used, whether AI platforms are allowed or blocked, and how often AI tools realistically show up in your work.
2. How have you noticed the humanlike responses in your use of AI?
  - a. Describe a few examples (tone, emotional language, decision-like language) that illustrate anthropomorphism.
3. How do you currently use AI when planning, problem-solving, or generating ideas?
  - a. Explain what a typical “conversation” with AI looks like and how it progresses—for example: brainstorming → refining → questioning → finalizing.

#### **Passing:**

The response includes specific information explaining how AI is utilized in the applicant’s role, as well as whether and how students have access to AI tools. The applicant clearly describes examples of humanlike or anthropomorphic responses they have observed in AI outputs or conversations, demonstrating an accurate understanding of how these responses differ from human reasoning.

### Part 2. Work Examples/Artifacts/Evidence

To earn this micro-credential, please submit the following four artifacts as evidence of your learning. See the Rubric for the passing score.

\*Please do not include any information that will make you or your students identifiable to your reviewers.

### **Artifact 1: AI Glossary**

Create a glossary of 10 terms from the current AI lexicon that use humanlike or emotion-based language (such as *think, decide, understand, feel*).

For each term, include:

- A brief explanation of how the term is commonly used in an anthropomorphic way when talking about AI.
- A clear description of the actual machine action the AI is performing, instead of the human behavior the word suggests.

Use a consistent format for all entries.

### **Artifact 2: Discussion Points**

Over two weeks, observe or participate in conversations about AI with colleagues, students, friends, or family. Select three to five conversations where people used humanlike language to describe AI.

For each conversation, include:

- *A brief description of the conversation (context only; no identifying details).*
- *The specific anthropomorphic phrase or idea that was used (such as “It’s trying to tell me ...” “It decided ...”).*
- *An explanation of how that phrase assigns human characteristics to the machine.*
- *A short clarification you could offer that explains the actual machine-learning process behind the response (such as, “Instead of saying the AI is thinking, we can say it is predicting based on patterns it has learned.”)*

You may present the conversations in paragraph form or as separate entries.

### **Artifact 3: Prompt Comparison**

Create three AI-generated outcomes, using one tool (such as a poem, short paragraph, image, or other creative output). Use different input prompts for each product, and in each prompt, intentionally include a request that encourages humanlike processing (such as asking the AI to “feel,” “imagine,” “decide,” or respond with empathy).

For each of the three products, include:

1. The exact prompt you used, including the humanlike language.

2. The full AI-generated output.
3. An analysis of anthropomorphism in the output, including:
  - Where emotions, empathy, or other human characteristics appear.
  - How tone, voice, or encouragement show up in the response.
  - Whether the AI is mimicking a particular type of human voice or style.
4. A brief reflection (three to five sentences) explaining how the AI's response demonstrates pattern-based prediction rather than human understanding.

Your final document should clearly include:

- three prompts
- three outputs
- three analyses/reflections

#### **Artifact 4: Tool Comparison**

Create three entries using different AI tools. For example, create an email or document using ChatGPT, Claude, and Gemini. Use the same input prompt for each and note how each one responds with humanlike language. Create a comparison document outlining the differences between these projects and identifying the anthropomorphic elements in each. Reflect on when humanlike responses are helpful and how you might apply this knowledge moving forward.

Using the same prompt, create three AI-generated products using three different AI tools (for example: ChatGPT, Claude, and Gemini).

In your artifact, include:

1. The exact prompt used for all three tools.
2. The full output from each tool (copied or screenshot).
3. A comparison of the three outputs, describing:
  - Where each tool uses humanlike or emotional language
  - Differences in tone, voice, politeness, empathy, or encouragement
  - Any patterns that show how each tool mimics human communication differently
  - Any personalization or familiarity cues that could create the impression the tool “knows” the user (for example: referencing prior context, mirroring preferences, or using relational language).
4. A brief reflection (200-350 words) explaining:
  - When humanlike responses might be helpful
  - When they could be confusing or problematic
  - How personalization or familiarity cues may affect different student populations, including those who may be more vulnerable to anthropomorphic framing
  - How this comparison will influence your future use of AI tools

Your Artifact 4 submission should include:

- one shared prompt
- three outputs
- one comparison section
- one reflection

## Part 2. Rubric

	<b>Proficient</b>	<b>Basic</b>	<b>Developing</b>
<b>Artifact 1: AI Glossary</b>	<p>The glossary includes 10 humanlike or emotion-based AI lexicon terms.</p> <p>Each entry:</p> <p>Clearly identifies the anthropomorphic meaning of the term, and</p> <p>Accurately explains the actual machine action behind the term.</p> <p>Entries are complete, accurate, and consistently formatted.</p>	<p>Glossary includes fewer than 10 terms <i>or</i> includes terms that are not humanlike.</p> <p>Some entries describe anthropomorphism, but the explanations of the machine's actual actions are incomplete, unclear, or inconsistent.</p>	<p>The glossary is missing multiple required elements.</p> <p>Terms do not relate to anthropomorphism, and/or entries do not explain how the machine works.</p> <p>The evidence does not demonstrate that AI understands humanlike language.</p>
<b>Artifact 2: Discussion Points</b>	<p>The artifact includes three to five clearly described conversations that use anthropomorphic language about AI.</p> <p>Each example: Identifies the specific humanlike phrase or assumption,</p> <p>Explains how the phrase assigns human characteristics to AI,</p>	<p>The artifact includes three to five conversations, but the descriptions are general or incomplete.</p> <p>Examples identify some anthropomorphic language, but explanations of how it is humanlike or clarifications of the actual machine's actions are missing or only partially accurate.</p>	<p>The artifact includes fewer than three conversations, lacks examples of anthropomorphic language, or does not explain how AI responses were misinterpreted as humanlike.</p> <p>Clarifications of machine processes are missing or inaccurate, demonstrating a limited understanding of anthropomorphism.</p>

### AI in Education

Understanding how AI systems perform tasks, but are not human

Revised on: Apr 15, 2026

	<p>Provides an accurate clarification of the actual machine-learning process.</p> <p>Entries are specific, accurate, and demonstrate understanding of anthropomorphism.</p>		
<b>Artifact 3: Prompt Comparison</b>	<p>Artifact includes three AI-generated products, each created with a prompt that intentionally uses humanlike language.</p> <p>For each product, the submission:</p> <p>Provides the exact prompt and full output,</p> <p>Identifies where emotions or human characteristics appear,</p> <p>Analyzes tone, voice, or encouragement used in the response, and</p> <p>Explains how the anthropomorphic elements reflect pattern prediction rather than human</p>	<p>The artifact includes three products, but one or more required components are partially developed.</p> <p>Examples may: Include prompts and outputs but have limited analysis,</p> <p>Identify some anthropomorphic elements, but do not fully explain them, or</p> <p>Provide reflections that are general or incomplete.</p> <p>Work shows a developing understanding of anthropomorphism but lacks complete clarity or depth.</p>	<p>The artifact includes fewer than 3 products or is missing major components.</p> <p>Submissions may:</p> <p>Lack clear prompts or outputs,</p> <p>Fail to identify humanlike language,</p> <p>Provide little or no analysis of tone, voice, or anthropomorphism, or</p> <p>Shows an inaccurate or minimal understanding of AI's pattern-based responses.</p> <p>The artifact does not meet the expectations for demonstrating competency.</p>

**AI in Education**

Understanding how AI systems perform tasks, but are not human

Revised on: Apr 15, 2026

	<p>understanding.</p> <p>All components are complete, specific, and accurate.</p>		
<b>Artifact 4: Tool Comparison</b>	<p>The artifact includes the same prompt used with three different AI tools, and all three outputs are provided.</p> <p>The comparison section clearly analyzes:</p> <ul style="list-style-type: none"> <li>Where each tool uses humanlike or emotional language,</li> <li>Differences in tone, voice, politeness, empathy, or encouragement, and</li> <li>Observable patterns in how each tool mimics human communication, including any personalization or familiarity cues that could influence user perception or trust</li> <li>Reflection explains when humanlike responses may be helpful or problematic and describes how this understanding will influence</li> </ul>	<p>The artifact includes the same prompt and outputs from three tools, but the analysis is partially developed.</p> <p>The comparison and/or reflection may:</p> <ul style="list-style-type: none"> <li>Fall short of the suggested word counts,</li> <li>Mention humanlike responses but lack depth or specificity, or</li> <li>Provide general observations without thoroughly examining tone, voice, or anthropomorphic patterns.</li> <li>The work demonstrates a developing understanding of anthropomorphism but lacks complete clarity and thoroughness.</li> </ul>	<p>Artifact is missing major components (e.g., fewer than 3 tools, missing outputs, unclear prompt).</p> <p>The comparison is limited or inaccurate, with minimal attention to humanlike language or tool differences.</p> <p>Reflection is missing, incomplete, or does not address the impact of humanlike responses.</p> <p>The submission does not demonstrate sufficient understanding of anthropomorphism or how different tools generate humanlike text.</p>

**AI in Education**

	<p>the applicant's future use of AI.</p> <p>All components are complete, accurate, and demonstrate a strong understanding of anthropomorphic framing and its impact on user trust and decision-making.</p>		
--	--	--	--

## Part 3 Reflection

**(300-500 words)**

For tips on writing a good reflection, review the following resource:

[How Do I Write a Good Personal Reflection?](#)

*Please do not include any information that will make you identifiable to your reviewers.*

Answer the following questions:

1. Are there any populations that you currently work with for which this new information is particularly relevant?
2. How do humanlike responses encourage users to engage with AI in more personal or emotional ways?
  - a. What is the impact, both positive and negative, of using AI on the time spent, the product produced, and the user's skill?

**Passing:**

Response addresses all reflection questions with specific examples and demonstrates a clear understanding of how anthropomorphic AI responses affect users, including vulnerable populations. The reflection thoughtfully considers both the benefits and challenges of humanlike AI behavior, connecting the educator's learning to future instructional or professional practice.