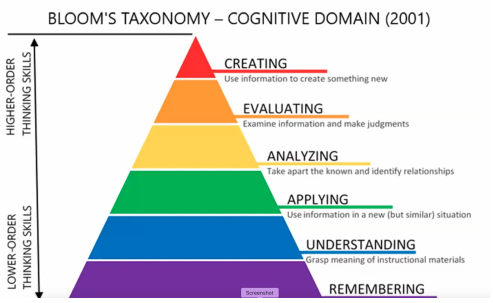
- **What does it mean to say that "trust is a construct" in the context of AI? How does this perspective influence the design of trustworthy AI systems?**
    - Sample Answer: Trust is considered a construct because it's shaped by human values and expectations. This means AI systems must be built to align with societal norms and context to be considered trustworthy.

- **What is one challenge with relying solely on human feedback to align large language models?**
    - Sample Answer: human feedback is often inconsistent, expensive, and hard to scale.

- **In what ways does the Turing Test fall short as a comprehensive measure of artificial intelligence?**
    - *Answer:* The Turing Test assesses a machine's ability to mimic human responses but doesn't evaluate whether the machine possesses genuine understanding or consciousness.
- **Which of the following is a limitation of the Turing Test in evaluating artificial intelligence?**
    - A. It requires physical embodiment of the AI.
      B. It measures only the speed of responses.
      C. It focuses on output indistinguishability rather than actual understanding.
      D. It assesses the AI's ability to perform calculations.
    - **Answer:** C. It focuses on output indistinguishability rather than actual understanding.
- **Which of the following is a key challenge in applying financial-style risk frameworks to AI systems?**
    - A. AI systems are always deterministic
      B. AI risks are harder to quantify and may change over time
      C. Financial systems are not regulated
      D. AI systems do not require governance structures
    - **Answer:** B. AI risks are harder to quantify and may change over time
- **Rebecca Hwa used the example sentence from an advertisement (in a magazine about computer systems from the 80's) "A computer that understands you like your mother." Why is this sentence ambiguous, and what does this tell us about challenges in AI language understanding?**
    - Answer: The sentence is syntactically ambiguous: it can mean either the computer understands you in the way your mother does, or that it understands the fact that you like your mother. This shows how even simple sentences require models to resolve structure and context — something AI still struggles with.

- Alexa Joubin talked about Bloom's Taxonomy of cognitive thought (shown at the right). As LLM's become more integrated in education, how does this taxonomy help think about student assessment?
    - **Answer:** The Taxonomy helps distinguish between surface-level skills (like remembering or understanding) and deeper cognitive engagement (like evaluating or creating). Since LLMs can easily handle tasks at the lower levels, assessment should increasingly focus on higher-order skills that require human judgment, synthesis, and critical thinking.



- Multiple Choice: AI nationalism refers to:
    - A. The use of AI in war
      B. Prioritizing domestic AI development for geopolitical advantage
      C. National-level bans on generative models
      D. Only using AI for public good
    - Answer: B

- How might AI nationalism challenge the development of globally trustworthy AI systems
    - Sample Answer: National priorities may conflict with global standards, leading to fragmented governance and inconsistent trust frameworks.

- "In class, we've discussed various 'guardrails'—technical or procedural measures used to make AI systems more trustworthy. Name two specific guardrails and briefly explain how each contributes to trustworthiness."
    - Sample Answer: Any 2 of the below:
        - **Human oversight** – Ensuring humans remain in the loop for high-stakes decisions.
        - **Model interpretability** – Designing models that are explainable and understandable.
        - **Data transparency** – Providing clarity about where training data comes from and how it's used.
        - **Bias auditing** – Regularly testing for and mitigating algorithmic bias.
        - **Risk management frameworks** – Like NIST's AI Risk Management Framework.
        - **Governance structures** – Oversight by boards or external review panels.
        - **Temperature control in generative models** – Controlling randomness in outputs to match risk level.
        - **Model cards and datasheets** – Documentation that helps communicate model limitations and intended use.

        - **User education and training** – Teaching users how to use AI tools responsibly.
        - **Legal or regulatory compliance** – Adhering to frameworks like GDPR, AI Act, etc
- One current fear of AI is that it will create a crisis of trust/confidence crisis because we no longer trust that words/documents are not written by humans. Please name one or more previously technology changes that generated similar crises.
    - Answer: Any of the following
        - **Printing Press:** *The printing press caused a crisis of trust by enabling mass distribution of unchecked or heretical ideas.*
        - **Photography and Photoshop:** *Photo editing tools undermined trust in visual evidence by making it easy to alter images.*
        - **Radio and Television:** *Mass broadcast media raised fears about propaganda and manipulation of public opinion.*
        - **Word Processors / Typewriters:** *Typewriters and word processors made it easier to forge official-looking documents without handwriting.*