

Beyond the Language Analogy: Lessons for Education from How AI Models Process Images and Video

Dr. Vladimir Shapiro

Northeastern University, Boston, The United States

Abstract

Educational research increasingly uses Artificial Intelligence systems as a metaphor for human learning, but has drawn almost entirely on systems built for text, whose units (words) already carry meaning. Systems built for recognizing pictures and video, we argue, surface questions that the text-based metaphor cannot. An architectural analogy is informative when both systems must solve the same underlying problem, not when their mechanisms coincide; on that basis, we develop three hypotheses through structured analogical reasoning. First, interpreting a picture with text means mapping words onto what they denote. AI needs a dedicated component for this; people integrate words and pictures readily when the two repeat one another, but struggle when each contributes complementary information. Closing that gap is a distinct competence that may benefit from explicit instruction. Second, when trained on too little data, these systems are harmed by varied examples, benefiting only once consistent ones establish a baseline. This mirrors the variability-of-practice and expertise-reversal debates, suggesting varied-context practice may help more after single-context grounding than before. Third, dividing an image into raw units is only the start; a vision model must learn which units carry meaning, attending to some regions over others. Expertise is similar: in radiology, a novice and an expert see the same image, but only the expert attends to what matters. This suggests teaching perceptual attention is a distinct task, separate from teaching reasoning about structures once identified. Each hypothesis is evaluated against established learning-science findings, acknowledging the limits of reasoning from machine to human pedagogy.

Keywords: cognitive load theory; cross-modal alignment; instructional sequencing; multimodal transformers; perceptual chunking