UPSIDE LEARNING

Position Paper

# Artificial Intelligence in Designing Learning

## *Unpacking the hype*

With the sudden burst of enthusiasm sparked by Generative Artificial Intelligence (AI) applications, we're seeing a lot of prognostications. Some are wildly enthusiastic, promising a panacea of benefits. Others are woefully worrisome, forecasting a host of horrors. We want to stake out what we think is a practical take somewhere in the middle. This isn't because we want to straddle a line, but because we have a deep conceptual background in AI and want to convey a realistic exposition of what's possible, practical, and what currently remains ephemeral.

Here we start by providing a historical overview as a means to define Artificial Intelligence. From there, we unpack some of the myths and realities of AI. We close by exploring what we expect are realistic directions to take. If you wish to skip the history and technical bits, jump straight to Documenting Potential Directions section below.

# Defining AI

While AI has several gestations, the most prevalent is that AI emerged as a companion piece to psychology. While there are attempts to get computers to do smart things in *any* way, the main focus was to try to model how humans think as a way to provide a test for hypotheses. The computer models, to the extent that they behaved like people, were considered to be accurate.

The first attempts were symbolic, based upon a view that humans exhibited formal logical reasoning. For example, Newell & Simon's General Problem-Solver, and subsequent iterations (e.g. SOAR), could break problems down into sub-problems, solve the sub-problems, and put the solutions together to create an overall answer. They could even chunk sub-solutions together, effectively learning.

Subsequent models used propositional logic (if-then rules) to build so-called expert systems. This was the promise in the 80s, that we could ask experts what they did, build models and run them, and then have fatigue-less expertise on tap. The massive failure of these initiatives led to an AI winter. It turns out that what experts said they did, and what they actually did, had little correlation.

Continued research in the cognitive sciences struggled to use symbolic models to match human behavior. Efforts such as Zadeh's Fuzzy Logic and Hofstadter's slip nets were attempts to soften symbolic behavior. David Rumelhart, who's schemas echoed Minsky's frames and Schank's scripts, recognized the problem in trying to use such models. With colleague Jay McClelland and their research group, they resurrected and advanced previous attempts to leverage neural networks, using a sub-symbolic approach. They found that such models more closely mimicked human behavior than did the symbolic models. The launch of their two-volume set on Parallel Distributed Processing catalyzed the connectionist work that characterizes much (if not most) of machine learning today.

We still have symbolic models today, in well-defined areas. However, the ability to train systems on historic data to react to new situations has been extremely valuable. As such, companies have raced to create ever more powerful systems as a way to capture the interest, and investment, of venture funding and customers. The so-called 'generative' systems, which use a background of data in particular areas (e.g. images or text) to respond to requests with amalgamated responses, have fueled interest and concerns.

# Detailing What's Real

The first thing to realize, despite the hype, is that we're nowhere near what's termed Artificial General Intelligence (AGI). That would be a system similar to our own mental architecture, capable of reasoning in new domains and acquiring knowledge to become an expert in what's studied. Humans can approach new areas and handle them with what's known as 'graceful degradation', that is, we may not do as well as in areas known to us, but we can still make some reasonable inferences. Traditional symbolic systems fall apart brutally at the edges, being termed 'brittle'. Machine learning systems likewise are limited to what they've been trained on.

There's a fundamental reason that these systems have to fail: they have no real-world knowledge.
They are symbol-processing machines, capable of matching patterns, but they have no idea what the patterns actually mean. Systems using Large Language Models (LLMs), like ChatGPT, get very good at predicting what a reasonable response would look like and providing that. They have no true understanding, however, as they're probabilistic engines. If what they say isn't true, they have no awareness thereof.

The approach to AGI is limited by several things. For one, intelligent systems need to have their patterns matched with visceral experience, what Stevan Harnad termed the "symbol-grounding problem". Another is that the scope of training can't just be in-depth in one area, but has to have experience across areas. A third requirement would be the capacity of an engine to hold sufficient memory and processing power to handle the computation. We already strain resources in these limited areas (at an environmental cost). So, the fear of an AGI takeover, with either benign dictatorship or human-erasing tendencies, is unjustified anywhere in the near future.

A second concern is as to what is considered intelligent. It's always been the unwritten joke that "once it's understood, it's no longer AI". Many things organizations do are captured in rules that run on systems, such as spacing learning. Unless it's adapting, it's not intelligent. Even if it is, it's still only *arguably* AI. Thus, just because something's claiming to be using AI, is it really just clever programming? Of course, we may not care, if it is, in fact, doing something clever and useful.

One of the real issues is in ethics. To do machine learning, vast quantities of data are required. Where does that data come from? For instance, ChatGPT's LLM uses a portion of the internet as a knowledge base. There are several problems with this. For one, the internet is notoriously ungoverned, and there're wrong answers out there. It's been seen in learning tests that it can recommend learning styles, an approach decisively debunked scientifically. Similarly, historical biases emerge in the datasets, such as stereotyping doctors as males and nurses as females. Further, there's an attribution problem. The Washington Post provided a way to investigate Google's C4 database, and our own Upside Learning site is (a very small) part of it! What benefit accrues to us if someone uses C4 to answer a question? The mechanism underlying such systems means that they *can't* answer questions about attribution, so there's no way to identify the actual source of the material.

Ultimately, we need controls. AI should be complementing us, not competing with us. There are things that our brains do that are very hard for technology, even AI, to do. Similarly, there are things technology does brilliantly that we struggle with. The right approach is to actively design AI as a complement and a partner, not a replacement. Except, of course, for things we don't want to do, we shouldn't put people out of jobs they value to be replaced by an AI that has no real comprehension. As the joke has it, we don't want people doing menial tasks while AI gets to do art and music! Let's focus on replacing the rote, not the creative.

# Where Can AI Help?

Which opens the question, what can AI do, and how do we use it in learning? Markus Bernhardt and Clark Quinn wrote an article on the future of L&D in an AI world, providing some answers. In it, we documented how AI could process content to create a topic map and monitor people's progress through it, increasing their knowledge. Moreover, it can adaptively inform what one person should focus on versus another. We also noted that it took well-curated content for this to work, and AI (still) cannot create meaningful practice. Thus, we suggested a partnership between instructional designers and AI.

This approach continues to be a reasonable focus. While LLMs have inspiring abilities in language, they are wholly dependent on the quality of the knowledge model they use as the basis for their language skills. As such, they may make good thought partners, providing potentially lateral inputs on topics, but what they say shouldn't be taken as gospel. If you do control what knowledge model they use, their answers can be more trusted. For instance, chatbots trained on the policies and procedures of a company can provide accurate answers, freeing up HR personnel to deal with other challenges. They can even point out inconsistencies in such policies and procedures.

Increasingly, we will be able to use AI to generate content, images, and the like. However, AI, as yet, can't handle images *and* text, together. That is, it can't have prose that references a particular part of an image, which is a gap for supporting learning. While it can diagnose knowledge gaps, it can't diagnose skill gaps, nor design meaningful practice. It can do a good job of supporting search *if there's a good knowledge model.*

We see AI as a valuable adjunct to what we do. That is, it can help automate rote knowledge tasks that are nice to offload. This frees us up to do more of what we prefer. We should continue to view all technology, including AI, as a complement to our capabilities, and design for the same. We understand context, and emotion, for instance, as well as evaluating the application of knowledge to solve problems. Thus, we need to look at how AI can take on what it can do well, specifically in quantity and rote knowledge, and we need to provide oversight and engagement.

As stated, AI is good at large quantities of tasks around knowledge. For instance, it can take good content and extract the underlying structure, then serve as a knowledge guide. It can also serve as a thought partner in the creation of content. What is needed on our side is several things. For one, we should never take what AI says without vetting it. AI also isn't capable of designing deep learning challenges (at least, not yet, but that's also something we may prefer to undertake). Thus, we have a role to complement the knowledge with skills that are sensitive to context and invoke human engagement in the challenges. Human creativity in the design of practice that successfully integrates meaningful practice across sufficient contexts to support retention and transfer over time becomes the continuing added value.

That, at least, is what we see. We continue to develop our ability to effectively integrate learning science with engagement to create learning experiences. We also continue to explore how we partner with AI to create a unified whole that is as efficient as possible but is similarly as effective as needed. That's the way we think AI should be applied, and consequently, that's what we aim for: to provide the best outcomes for customers and society as a whole.