

A Coach’s Notes¹

Everett Rutan
Connecticut Debate Association
ejrutan3@ctdebate.org

Connecticut Debate Association
November 2025

AI and Debate: A Primer

Contents

Preface	1
Introduction.....	2
What is AI?.....	2
Thought Experiments	2
Associative Indexes.....	3
LLMs and Humans.....	4
Searle and Turing.....	5
Using AI in Debate	5
Problems with AI in debate.....	6
AI in Debate: Best Practices	7
The Debater Is the Weak Link.....	8
Idea Generation and Research	9

Preface

This is a special edition of my Coach’s Notes, the first of two parts. This piece explains “artificial intelligence”, at least in its current incarnation, and how it might be used in debate. The second discusses computer use at CDA tournaments.

¹ Copyright 2025 Everett Rutan, all rights reserved. This document may be freely copied for non-profit, educational purposes. The opinions expressed herein are those of Everett Rutan alone and do not represent the views of nor have they been endorsed by the Connecticut Debate Association or any other party.

I may write some commentary on the motion at a later date.

I appreciate any feedback you have, good and bad. The best comments and suggestions will find their way into subsequent issues. I would also consider publishing signed, reasoned comments or replies from coaches or students. If you would like to reply to my comments or sound off on some aspect of the debate topic or the CDA, I look forward to your email.

Introduction

“Artificial intelligence” is an old term that has been applied to a variety of technologies over the past 70 years. It is currently being applied to Large Language Models (LLMs), a powerful, disruptive new technology that has exploded into the world with amazing capabilities and very little preparation. It has roiled education, as teachers and students struggle to adapt. This note provides background on the technology and on its use in debate,

What is AI?

Key points:

- Intelligence is hard to define
- Turing’s imitation game is well-known but provides little insight into what is being called AI today
- Searle’s Chinese room thought experiment provides a useful way to think LLMs
- LLMs are associative indexes created statistically from the proximity and frequency of items in the training data.
- Both the size of the training data sets and the complexity of the index make full human understanding difficult

Thought Experiments

I don’t know of any good definition of “intelligence”. It’s one of those, we-know-it-when-we-see-it things, usually by comparison with human behavior. This is the basis for the “Turing Test”, which he called “The Imitation Game” and popularized by the movie of the same name.² Basically, if you cannot tell whether the entity you are communicating with is or isn’t human, that entity should be considered intelligent. This is a mathematician’s approach: two things in mathematics are equal if you can’t tell them apart.

² Turing’s paper, “Computing Machinery and Intelligence” (*Mind* 49: 433-460), is quite readable and worth your time. You can download a pdf here, <https://courses.cs.umbc.edu/471/papers/turing.pdf>.

Turing's is an interesting thought experiment, but it is not very useful for understanding the current version of machine intelligence, Large Language Models (LLMs).³ John Searle, a philosophy professor at the University of California, Berkeley, presented a different thought experiment called the "Chinese Room"⁴ as an argument why no computer program can ever achieve consciousness.⁵ Consciousness implies self-awareness, and so is not the same as intelligence, though they are related. But I find Searle's Chinese Room a very useful way to think about large language models (LLMs) like ChatGPT and its competitors.

A person in a room has to answer messages passed to him in a totally unfamiliar language, in Searle's telling Chinese logographic characters. The room has a huge library of instruction books. The person combs through the books looking up the symbols he receives and copying the symbols the books indicate as the proper reply. Searle assumes the books are complete, and the man can reply to any message in perfect Chinese.

The Chinese Room is a thought experiment, like Turing's imitation game, or Einstein's "riding on a beam of light" which he used to explain special relativity. Details like how the books could be created or how cumbersome the process of looking up and matching symbols in real time might be don't matter. Searle's point was that neither the man nor the room has any understanding of the conversation, so while might seem intelligent in Turing's sense, cannot be considered conscious in a human sense.

Associative Indexes

While we hope ChatGPT isn't a little man in a room with a lot of books, the Chinese room model is very much how LLM's work. The system receives a message in the form of a verbal or oral or visual prompt, consults a neural network with billions or trillions of parameters created from an enormous amount of training data, and produces a reply. This neural network is a massive associative index based statistically on:

- proximity, how "close" different snippets of information are to each other in the training data;
- frequency, how often the items appear.

The LLM program responds to the prompt by building a reply piece by piece with what the evaluation of the training data, summarized in the neural network parameters, calculates as what

³ Artificial intelligence has a long history using different approaches. The Wikipedia entry on "[History of artificial intelligence](#)" is a good place to start for those interested.

⁴ Searle picks "Chinese" because most of his audience would know of Chinese writing, but no comprehension of the symbols. It could just as well have been Sanskrit, cuneiform, hieroglyphics, or Greek for most of us.

⁵ Searle's paper, "Minds, Brains, and Programs" (*Behavioral and Brain Sciences* 3 (3): 417-457) is also readable and worthwhile. You can find it here: <https://web-archive.southampton.ac.uk/cogprints.org/7150/1/10.1.1.83.5248.pdf>

should most likely follow.⁶ Obviously this simplifies things considerably, but it’s a useful mental picture to help understand what is going on.

Before printing was invented and paper was cheap, books were rare and expensive. Scholars had to rely much more on their memory. One technique they used to train their mind to remember more was to create a “memory palace”. This mental construct might be a large house or even a palace the individual was familiar with. They would purposely work to associate something they wanted to remember with particular room or object in a room. Related material would be associated with nearby objects. To help recall something, they would mentally walk through the building until they came to the right location or object. The association would help recall what they were searching for.⁷

This should all sound very human to you. Even without that elaborate technique, what we read, hear, or see makes connections with things we have learned and retained over the course of our lives and leads us to formulate a reply. Something we see or hear may bring back a long forgotten memory. Marcel Proust’s *A la Recherche du Temps Perdu* is a fictional memoir brought forth when the protagonist bites into a cookie, a madeleine. Those in the tech industry who believe LLMs will lead to artificial general intelligence (AGI) believe this is all thinking is.

LLMs and Humans

LLMs are very complex, but essentially just large Chinese rooms with enormous libraries and very fast retrieval and response times. The training data sets used—perhaps the “entire internet”, but even if only the corpus of Wikipedia—are much larger than any one person can absorb or comprehend in a lifetime. Unlike humans, LLMs never forget anything they have learned. The generated associations are statistical, based on frequencies found in the training data, more exact, perhaps, but conceptually similar to what happens as the human brain strengthens some connections and weakens others as it learns.

The Chinese room model also helps think about potential weaknesses in LLMs, though again these weaknesses are similar to those in our own brains. The training datasets are so vast that no one can “know” what is in them. It isn’t that these datasets aren’t there to be reviewed, it’s simply that they are so large. The tendency of some LLMs to hallucinate may be due to flaws in the training data, as anything found in the data will feed into the statistical calculations.⁸

Similarly, no one knows all the calculated associations in the trained model—again the problem is the number of calculated parameters and the complexity of how they interrelate, not that these

⁶ Stephen Wolfram provides a readable explanation of the process here:

<https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/>

⁷ Jonathan Spence, Yale historian, discusses this technique in his biography, *The Memory Palace of Matteo Ricci* (Penguin, 1985), a 16th century Jesuit missionary. Benedict Cumberbatch, with the help of television magic, demonstrates the technique in his portrayal of Sherlock Holmes in the PBS series *Sherlock*.

⁸ LLMs were initially trained on vast “sweeps” of data from the internet. Some now screen, or pay other companies, to clean data as they work to improve LLM quality and performance.

aren't available to be examined. One cannot predict performance in all situations. There is a “whack-a-mole” quality to the industry as users find ways to make models misbehave and developers try to adjust their behavior. Researchers are also trying to develop ways for LLMs to “explain” the “reasoning” behind their replies, and to better understand how LLMs capture and store knowledge.

In many ways, LLMs seem quite human. No one knows all the influences you may have been subject to over the course of your life—and you will have forgotten many. A random phrase or song or image may summon memories long buried. Others can't always predict what we will do in any particular situation, and we often can't explain the reasons why did it. And our behavior is often less than perfect, or even acceptable.

Searle and Turing

Reading Turing, I found what I thought was an interesting connection to LLMs. At the end of Turing's article he speculates on how one might program a computer to play his imitation game. Though far beyond the capabilities of the machines of the 1950's, Turing foresaw their likely improvement. He suggests that programming a computer with all it would need to know would be very difficult. Efforts were made to build systems “by hand” this way made progress but ultimately were unsuccessful. Turing speculated that a machine that could “learn” like a child, i.e. teach itself or be taught, was much more likely to succeed. The neural network paradigm behind LLMs seems to be just that self-teaching technology, and it seems to have created Searle's Chinese room.

Using AI in Debate

Key points:

- Conceptually, an LLM is just another research resource for developing arguments
- LLM output quality is highly dependent on the prompt or input provided
- LLMs problems with hallucinations or incorrect answers a known, but similar to those of any other research resource one might use
- The debater must understand their case in order to present and argue it successfully

An LLM is a vast repository of information and its interrelationships. Clearly it “knows” more than you or I or the CDA motion packet. It is easier to use than reading the packet, other outside material, looking something up on Wikipedia, or searching the internet. It is also much faster.

Prompting an LLM with the motion, pro or con, will result in an essay of varying length supporting the position. The output could provide arguments for a Gov or Opp case, or might even be read directly as the PMC or Opp case portion of the LOC.

Many debaters argue cases for and against particular motions or topic areas as part of their team practice. They may write these case files up and store them on paper or more likely on computer. Some leagues permit them to reference this material during case prep before a round. There are also internet sites and services with prepared cases on a variety of topics. Using an LLM prompt to write a case is like using one of these services to download a case.

There is nothing “wrong” or “improper” about a league allowing debaters to access LLMs during case prep. Leagues already differ as to when motions are announced, whether and how often they change. They also differ about which materials and how much time debaters should have before a debate. Various forms of parl—CDA, NPDA, NYDDL, World Schools, British or World University—and LD, PF, and policy all do things a bit differently. All have their proponents and detractors. So long as everyone knows and follows the same rules, the competition will be fair.

Problems with AI in debate

The problems using AI in debate are no different from those a debater faces using any resource to help develop their case.

Good research requires asking good questions. Go to a library, look up a subject in Wikipedia, enter a google search, what you get depends on where you start and how you follow up information you find. Ask good questions, verify the answers you get, organize the material properly and likely you will produce a good result. Fail to do any of those, and you are likely to be embarrassed when others review your work.

LLM output depends on the quality of the prompt as much as the quality of the LLM. Good answers require good questions. Asking good questions is harder than providing good answers. Simply using the motion as a prompt may not generate a particularly good case. A better prompt likely requires some understanding of the topic by the person creating the prompt. Some worry AI will replace workers, but it also is creating new roles, one of which is “prompt writer”, a specialty apparently much in demand these days.

The linkages created when an LLM is trained are statistical. They reflect the most frequent connections in the training data set. Statistical models are known to exhibit “regression to the mean.” LLM output will be the “average” or most common, or most likely answer. Unique or insightful answers are less likely. For most debaters, an “average” case is probably good enough, but that case may be vulnerable to arguments that are less common.

LLMs are known to “hallucinate”, providing arguments that do not make sense backed by authorities that do not exist. Nothing guarantees the accuracy of material you find in a library, on Wikipedia, from an google search, or the CDA motion packet. Humans “hallucinate” too, all of us at some time, because we are mistaken about what we think we know or err in our reasoning. The problem will likely decline as models improve but is unlikely to be eliminated. Unless you have some understanding of the topic, you will be unable to detect mistakes or

hallucinations, any more than you can when the person in front of you spouts reasonable sounding nonsense.

Even if what the LLM generates is sound, that output needs to be reviewed and understood to be presented and argued. Simply reading the output does not confer understanding of the case: the strategy behind it or the tactics needed to present and defend it successfully. You may have time to “learn” the case from LLM output produced prior to a round. But that will not necessarily provide responses to arguments and rebuttal presented by the other team during the debate. Granted, LLMs are improving rapidly, and some now respond to verbal input and can provide verbal responses. One can see the day when a debater could allow an LLM to listen to an opponent’s speech and write a reply in real time. Again, having a response and understanding that response sufficiently to present it are very different things.

The key factor underlying each of these three weaknesses is understanding. A debater making arguments they do not understand is extremely vulnerable. They may win rounds against weaker teams but likely find less success as power matching moves them against teams that “own” their cases. It’s the difference between being given an answer and taking the time to work out the answer yourself, between surface and depth.

I have personal experience of this, debating over 50 years ago, when computers were the size of refrigerators and locked away in secure rooms. Many teams were known to be “plastic sheeted”. They didn’t have LLMs to write their cases but relied on their coaches. These were typed up and held in plastic page protectors, hence the term. These teams would do well in the early rounds but were easy pickings as the day moved on. An argument their coach had not anticipated, or that they could not find the right sheet for, or even an argument presented in a way that did not match the phrasing on those sheets, would usually be fatal. You still see this today with teams in some parliamentary leagues relying on libraries of prepared cases, prepared by someone other than the debater using them.

AI in Debate: Best Practices

Key points:

- AI best practice is to use it as an idea generator and research tool
- This seems to be an emerging consensus among successful adapters in education, business and government
- A debater would have to review, validate and understand LLM output to use it successfully, as they do the motion, packet, or any other resource
- Debate is a spoken word contest, limiting the usefulness of LLMs during the round itself
- Nothing prevents a league from adopting AI as a permitted tool, and it is more likely than not that this will occur in some formats and leagues if it has not happened already

“If you’re gonna play the game, boy, you gotta learn to play it right.” *The Gambler*, song by Don Schlitz, popularized by Kenny Rogers

Leagues differ in types of motions, amount of prep time, and allowable research materials. It is likely that some will experiment with allowing the use of LLMs. They may become an accepted research tool. This is already happening in education, industry and government. Like laptops, tablets, and cell phones, this is a technology that is not going away.

The strength of an LLM is the ability to quickly produce a reasonable answer to any prompt from a vast, interconnected knowledge base as described above. This is what debaters would call a *prima facie* case: an argument that, in the absence of any reply, would cause a reasonable person to agree. If the opponent is also using an LLM, the two cases may be evenly matched. But debaters may differ in their ability to write good prompts, there is some risk the training data set may be biased, one side or the other may better understand what is produced.

The Debater Is the Weak Link

Debate is a contest conducted using the spoken word. One could simply read the LLM output. Many packets hold articles that could almost be read into the round as a Gov or Opp case, but I’ve never seen it done. Many debaters spend time writing out and reading their speech, and some read speeches written by other team members. You can look to your own experience to decide how well this works. A debater has to understand the case well enough to appreciate its weaknesses and respond to attacks. If you write the case and the speech yourself, you likely have a good understanding. Otherwise—LLM output, packet articles, a team mate’s work—you will need to spend some time with it to be successful.

Debate is about reply, or clash. An LLM cannot anticipate fully what an opponent might say. One might feed the LLM prompts based on an opponent’s speech to generate a response “on the fly”, or even let the LLM “listen” to the other team speaking and generate a running reply. But like the original case, the debater would need time to verify and understand the generated output. LLM replies can be wordy, so there might be need to extract the essential points to stay within speaking time. A series of replies to a series of prompts might not come together as a coherent speech. One might generate an extended prompt, submitting it to the LLM only at the conclusion of the opponent’s speech, but there is no prep time between speeches so one would then have to stand and read it “cold” with uncertain results.

It should be clear by now that primary weakness is the debater not the LLM. LLMs with speech and hearing capabilities have already been demonstrated debating in real time. But high school debate requires you stand and speak. Standing to read something you have not had the time to absorb and understand usually falls flat. Reading from a teleprompter—think of material scrolling up on your laptop—or speaking from prompts fed through an ear bud is very difficult if it is unfamiliar unpracticed material. You can try these; you don’t need to take my word.

Idea Generation and Research

It also should be clear to you by now that an LLM is best used as a research substitute. In some ways an LLM can replace using the packet, the library, Wikipedia or google. Google search is now augmented by their own LLM's response to what the user types in, if you want it.

The best articles I have read from journalists, academics, and business people who have successfully adopted AI describe using LLMs to generate ideas. Input a prompt related to the project, review the answer, and refine the prompt to improve the answer. The user in these cases is a professional with experience and subject matter knowledge that allows them to generate good prompts, evaluate the output, and refine the prompt to improve the output. Interestingly, the most common articles describing the failed use of LLMs are about Judges admonishing law firms for turning in LLM-generated briefs and motions that were never edited properly and so contained ridiculous claims and false case citations.

In CDA debate, if use of LLMs were permitted, one would start with prompts based on a Gov and Opp reading of the motion. The LLM would quickly provide prima facie Gov and Opp cases or case outlines of one to two pages. Debaters could review and discuss these to find arguments they find reasonable or prefer. Supporting details would need to be similarly reviewed, and references scrubbed for accuracy. This might be done by further prompts to the LLM, or by reading the motion packet if it was still used. The LLM would be an additional input to aid case development.

Like any resource, allowing it to take over your case development would be risky. CDA debaters arrive with limited background knowledge about the motion they are presented with at 9AM on tournament day. But it is unclear that the mistakes they might make using an LLM are much different from mistakes they already make using the motion package and their own knowledge.

I don't think using the LLM in-round would be very effective. Unless you let the LLM speak for you the intellectual demands of listening, prompting, reviewing, organizing would probably overwhelm most of us in the limited timeframe of the round. The resulting speech would be either very disjointed, or obviously wooden as the speaker read unfamiliar material.

There is a good chance that this "LLM as adviser and guide" will be adopted in education, academia, business, and debate. I don't think it would harm the quality or value of debate as a competition, but I am not advocating in favor of adopting it in CDA. That is a discussion for another time.