

# Harnessing AI for Interoperability

Enhancing Artifacts and Innovation

# SESSION PRESENTERS



**TQ White II**

Member Emeritus,  
A4L Community



**John W. Lovell**

Technology Director,  
A4L Community

# What Possible Use?

## Anything where the answer is possible but hard

- Test data for data model
  - Unity Data Generator
- Align with data dictionary
  - Unity CEDS Matching
- Reverse dictionary search
  - CEDS Element Lookup

# Intro to AI

## Short answer: It's Magic

- Have you ever noticed how much easier brainstorming is when you are with other people?
  - Alone, we struggle to come up with a new idea
  - When someone else says something, we naturally think of a response, the next thing to say
  - Given a series of words, our brains naturally think of the next words that make sense

# Intro to AI

---

## Starts with a lot of learning

- Think about children.
  - They see, feel and hear a ton of stuff and babble and wave their arms at random
  - Sometimes they get positive feedback, a smile, food, etc
  - So whatever they did is embedded in the brain
  - Causes it to be more likely to be used as the response, the next thing that makes sense

# Intro to AI

## Learning and Training

- Fundamentally, these things read the entire internet, treat each sentence as gospel and layer them into its brain
- Like brainstorming, you give it a prompt that inspires it to say the next thing that makes sense or seems likely to be right but is sometimes stupid
- Like a baby learning, it is given examples that help (hold your spoon this way, don't use the word nazi) that also layer into its brain that train it to be better
- There's a ton more but that's the idea. It absorbs all the information and can use that to think of a good thing to say next and is explicitly given lessons on how to be useful.

# Intro to AI

## Recognize Patterns to Guess What's Next

- Technology concept goes back to the sixties and the Perceptron
- Reading the entire internet actually means breaking it into a series of (nearly) arbitrary tokens and using Perceptron idea to capture the likelihood of a pattern existing
- It is true that the basic mechanism is to guess the next best token. Also true that this results in magic
- Nobody knows why it has the clear appearance of strategizing and thinking
- My opinion is that this is what humans really do

# How Does That Help Us?

## A Problem that People Could Do but Won't

- Not having sample data has always been a problem. Drove me insane with my first experience with SIF in 2010.
- In most industries, you'd grab a bunch of production data, obfuscate, and send it to the developers.
  - Education has too many privacy rules for that
- The data is very complicated. It's just too expensive to imagine paying people to do it.

# How Does That Help Us

---

## Same Way as a Person Would

- Unity Data Generator uses GPT and other tools
  - Good instructions in the prompts
  - Multiple steps
- Give it an element and simply ask it to make up data
- Give it made up data and ask it to make sure it's good
- Do technical validation
- Give GPT the error message and ask it to fix
  - repeat until valid

# CEDS Alignment

## 15,000 Element Descriptions 2,000 Definitions

- No human can compare a sentence to 2k definitions even once
- 15k times? Beyond possibility
- GPT is happy to do it.
- But it's slow!!

# Enter: Semantic Distance

## The Coolest Idea

- Worth noting: 'semantic' is a metaphor. It's something far weirder. It's all tokens. But, it acts like semantic distance.
- GPT is willing to encode text into a big sequence of numbers that represents a point in space, super-hyper-space, 1538 dimensional space
  - but geometry doesn't care if it's a 2D Cartesian graph or 3D or 15kD
- Encode all the definitions. Then encode an element description. How far apart?
- Often there are several nearby, so, get a list and, ask GPT to decide which is best.

# CEDS Is At Risk

## If the Feds kill it, it's up to us

- Turns out that the fundamental function of CEDS is 'find a definition'
  - *(all kinds of other stuff but mostly derives from this idea)*
- Also, the whole alignment thing comes down to repeated 'find a definition'
- Happy coincidence. The tool exists. It just needs to be repurposed to keep CEDS afloat.

# Getting to Work

## Software to Execute the Processes

- Bunch of tools...
  - Absorb CEDS and Data Model into a database
    - found 1,905 CEDS, 15k NA Data Model elements
- Calculate semantic vectors for everyone
  - distinctly mixed results
- Need QA mechanism to decide if the AI got it right

# A New Tool

## Introducing the UnityCeds Web Tool

- Would have stuck with batch processes until decision to have interactive QA tool
- Show NA Model definition and CEDS definition together.
  - Thumbs up or down
- If no good, would like to figure out if there's a better one
  - Interactive search tool arrives

# My New Coding Partner, Claude

## A life-changing improvement for programmers

- Anthropic's Claude Code writes software incredibly well.
- This took about ten minutes and would have taken an hour.

# I Gave Claude This:

|          |          |                         |
|----------|----------|-------------------------|
| APPROVE  | GlobalID | @RefID Description text |
| terrible |          | GlobalID CDS Definition |

# And This:

I want to add a new tool in addition to the spreadsheet and outline for namodel.

Here is a wireframe picture: /Users/tqwhite/Downloads/evaluateTool.jpeg

It is called evaluate-definitions and be triggered by a new button labeled "Evaluate".

It should show a list of all the elements in an object and, instead of a collapsable display like the outline, it should show as a static table.

Each row contains a column with buttons first, then the cedsMatchesConfidence and then a structured cell with two lines of text.

The top line should show the element name (@refId or whatever) and the description. The second line should show the CEDS match as GlobalID, and definition (or label or comment, whichever is set).

The list should have controls (table column heads would be best) that sort in order of ceds confidence and namodel element name. The initial sort should be by confidence.

The buttons should be stacked vertically. The top says Approve. The bottom one says Terrible.

Organize the column with text so it is easy for a person to read them both together.

Filter the list to only show elements with a cedsMatchesConfidence defined. If there are none, it is important to show a legend saying so so the user doesn't think something is broken.

# And Got This:

| Actions                                | Confidence ↓ | Definition                                                                                                                                                                                                                                                                                                                                                                                                                          |
|----------------------------------------|--------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <div>APPROVE</div> <div>TERRIBLE</div> | High         | <p><b>GunFreeSchoolsActReportingStatus</b> - An indication of whether the school or local education agency (LEA) submitted a Gun-Free Schools Act (GFSA) of 1994 report to the state, as defined by Title 18, Section 921.</p> <p><b>000134</b> - An indication of whether the school or local education agency (LEA) submitted a Gun-Free Schools Act (GFSA) of 1994 report to the state, as defined by Title 18, Section 921.</p> |

Disclaimer: This is true *but* it took a few tries to flesh out the prompt so it completely worked and it gives a different result every time. If I weren't doing the presentation, I would have done a much less thorough prompt and then asked for corrections.

# Website Demo? Goodbye?