KENEXIS

AI Ready Process Safety Information

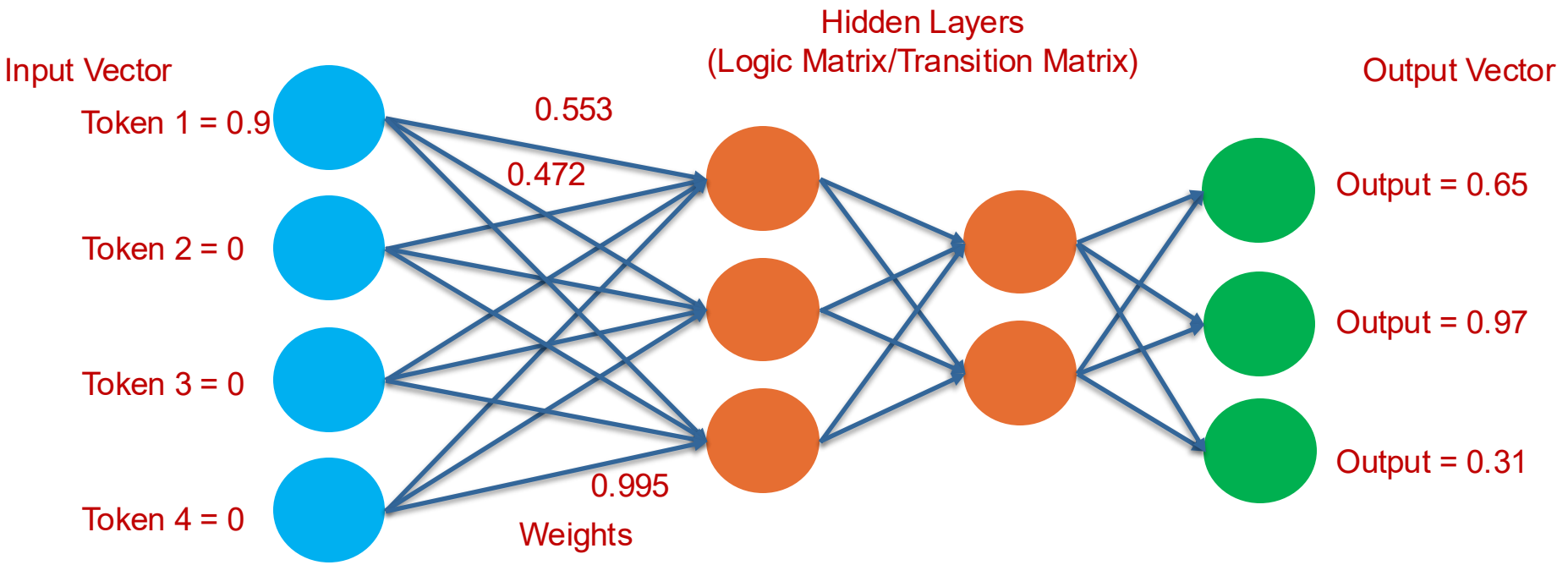
Agenda

- Review of AI Mechanics
  - Systematic Word Guesser
  - Do not romanticize or anthropomorphize
- Why ANN overgeneralize
- Agents – systematic code recursively calling ANN
- Retrieval Augmented Generation
  - Vector Databases vs. Semantic Search
- Context Poisoning
- ANN only understand text – How do we "force" it to look at drawings?

The AI Equation

$$\text{Max}^{1}\ \text{Out where; } \overrightarrow{\text{Out}} = \overrightarrow{\text{In}}\ \text{X}\ \left|\ \text{Logic}\right|$$

(Matrix Multiplication)



[1] Max output token not always taken, randomly lower tokens are selected

ANN – Statistical Next Word (Token) Guesser

- LLM predict statistically likely sequences
  - Higher frequency in training, more likely output
- No inherent truth model or physics model
- Tends to “General” answers, not specifics
  - Decides what is generally the result of a pump failure, not the specific result of pump failure in your plant

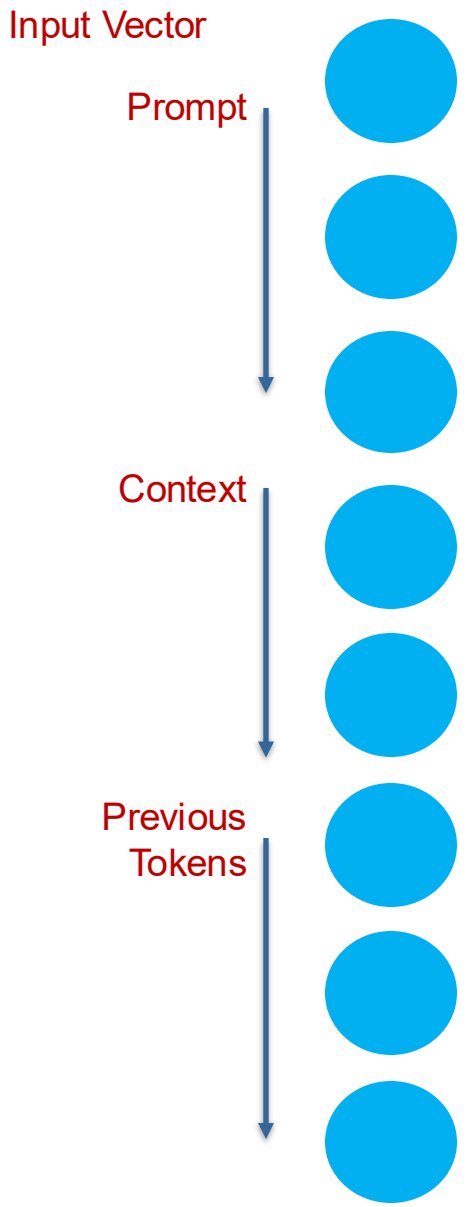
Agents

- Systematic code that calls the ANN
- Usually recursive
- Preprocess input requests – prompts
- Postprocess output results
- Use prior results to generate subsequent steps
- Provide additional context to the ANN

Retrieval Augmented Generation

- "That's not a product, that's a waste stream" – Dr. K Knabel
- Provide additional information that is retrieved from a data store
- Provides context for ANN (more input tokens)
- Combined with system instructions and metadata
- Building the input vector
  - Prompt → Additional Context → All tokens generated to current run

Important: The LLM does NOT differentiate "meaning" vs. "instructions" – it just sees a tokens in a sequence



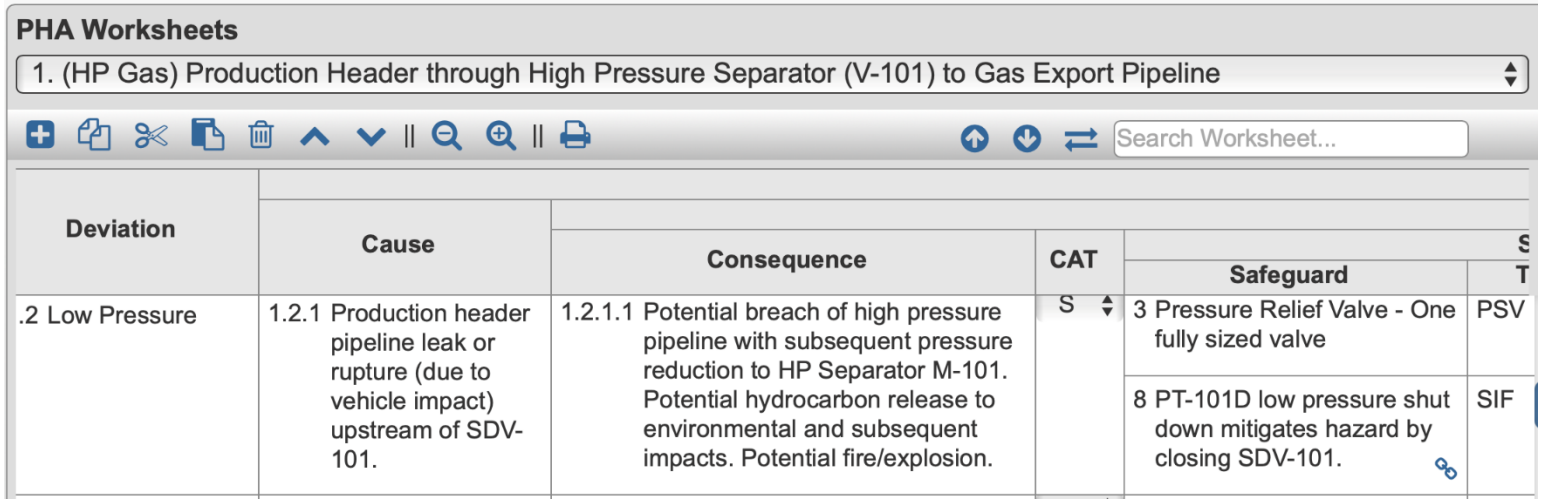
Context Retrieval – Vector vs. Semantic

- Semantic
  - Traditional key-word searching
  - Best for structured data
  - Difficult for LLM, no neighbor cell context
- Vector
  - "Mathematical Nearest Neighbor"
  - Based on embeddings (numeric vectors)
  - Good for images and unstructured data

Img = <yellow=0.9, green=0.3, black=0.02, day=0.9. night=0.05, hot=0.76, autumn=0.35>

Process Safety Information for Context

- PSI needs to be stored so that it can be strategically provided as context to the LLM
  - Chunked
  - Vectorized
  - Semantic Databases
  - Natural Language Rendering



PHA Worksheets

1. (HP Gas) Production Header through High Pressure Separator (V-101) to Gas Export Pipeline

Search Worksheet...

Deviation	Cause	Consequence	CAT	Safeguard	S T
.2 Low Pressure	1.2.1 Production header pipeline leak or rupture (due to vehicle impact) upstream of SDV-101.	1.2.1.1 Potential breach of high pressure pipeline with subsequent pressure reduction to HP Separator M-101. Potential hydrocarbon release to environmental and subsequent impacts. Potential fire/explosion.	S	3 Pressure Relief Valve - One fully sized valve	PSV
				8 PT-101D low pressure shut down mitigates hazard by closing SDV-101.	SIF

Low pressure in the high pressure separator can occur as the result of a leak or rupture upstream of shutdown valve SDV-101. A leak or rupture will result in pressure reduction in the high pressure separator with attendant release of the material in the vessel to atmosphere. Since the material in the vessel is flammable, if there is a source of ignition this might result in a fire or explosion.

What about images (like P&IDs)?

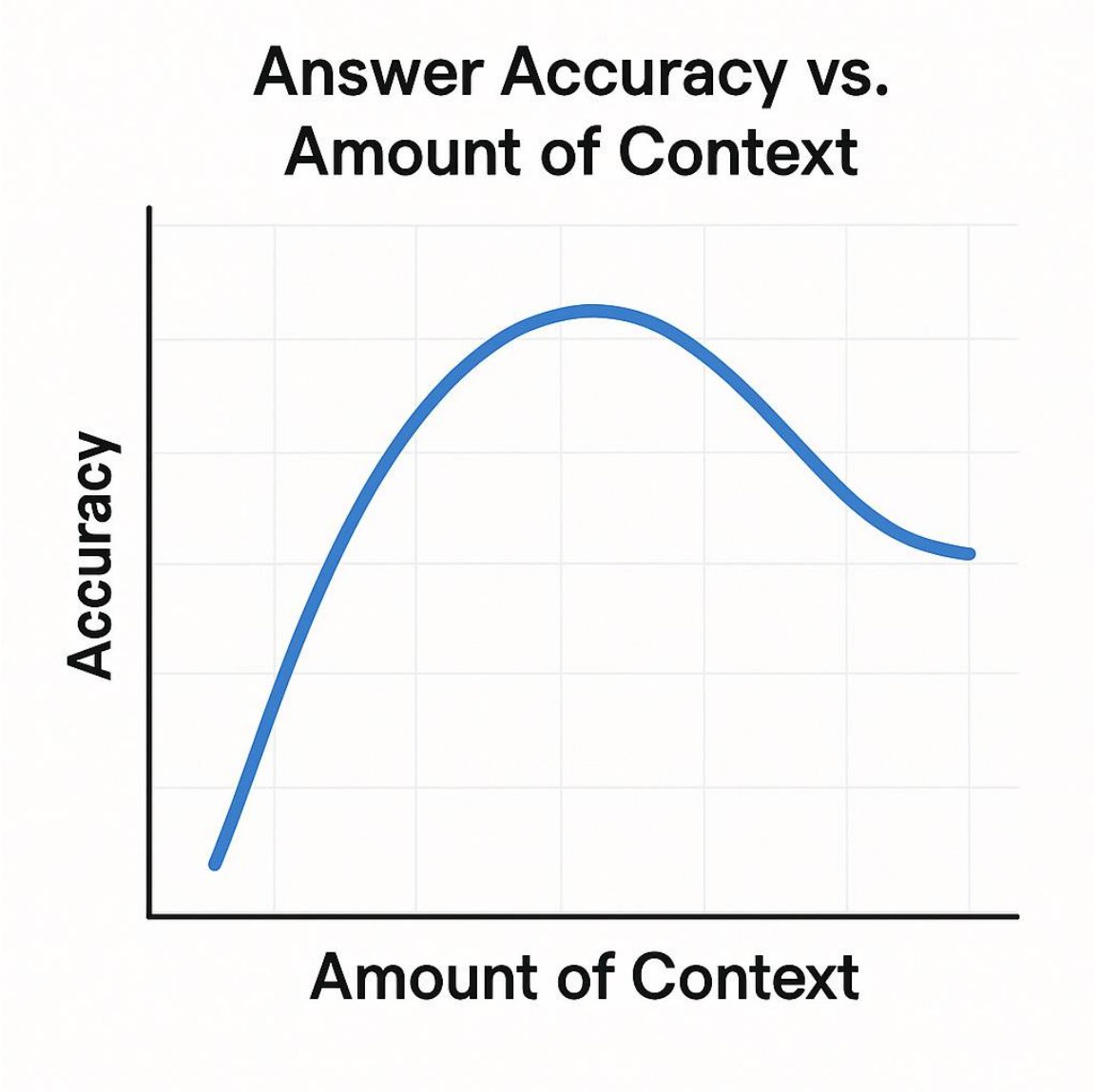
- ANN have no ability to process drawings
- LLM provides use agents to extract data from images
  - PDF text
  - Image recognition
- Image processing only as result of prompt
- Full context of drawing not understood without significant prompting

Solution for Engineering Drawings

“Un bon croquis vaut mieux qu’un long discours.”
“A good sketch is better than a long speech.” – Napoleon Bonaparte --- but not for LLM…

- Convert drawing content to structured data
  - DEXPI standard
  - Image recognition
- Natural Language Rendering
  - Describe what is happening in the drawing as a narrative
  - Chunk/vectorize

The Catch… Context Overload / Context Poisoning



Conclusions

- Without good context, LLM will generate generic results that are not valuable for hazards analysis
- LLM output is improved with good context as part of the input vector
- To provide a sufficient set of highly relevant context
  - Convert images to text
  - Convert tabular data to natural language
  - Chunk and vectorize

KENEXIS

Questions?