Trisotech

Low Code Neuro-Symbolic Agents

Prompt Engineering through
Process and Decision Orchestration.

Denis Gagne
CEO & CTO
dgagne@trisotech.com

Caveats

- This presentation is not about creating Large Language Models (LLMs) but rather about using them.
- This presentation is not about the copyright, safety, and ethical use of AI.
- Presented herein are generalised concepts to simplify understanding.
- Everything presented herein can be achieved using the Trisotech Digital Enterprise Suite (DES). Have fun!

Content inspired by

- Andrew Ng, Yoshua Bengio, Geoffrey Hinton, Gary Marcus, Randy Goebel, David Pool, Alan MacWorth, Tony Seale, James Briggs, Linus Ekenstam, Afaque Umer, Lilian Weng, John Maeda, Mustafa Ispir, Corbus Greyling, Tomer Cohen, Malte Pietsch, Raphaël Mansuy, Kingsley Uyi Idehen, Alex Lenail, Giuseppe Scalamogna, Marcel Pociot, Tomaz Bratanic, Varshita Sher, …

- OpenAi, AssemblyAI, Pinecone, fiddler, a16z, 451 Research, Gartner, IDC, McKinsey, Nvidia, Microsoft, Google, AWS, Meta,HumanFirst, Deeplearning.AI, AlphaSignal, PromptingGuide.ai, Langchain, Weights&Biases, LearnPrompting, NN-SVG, Petal, Abacus.ai, …

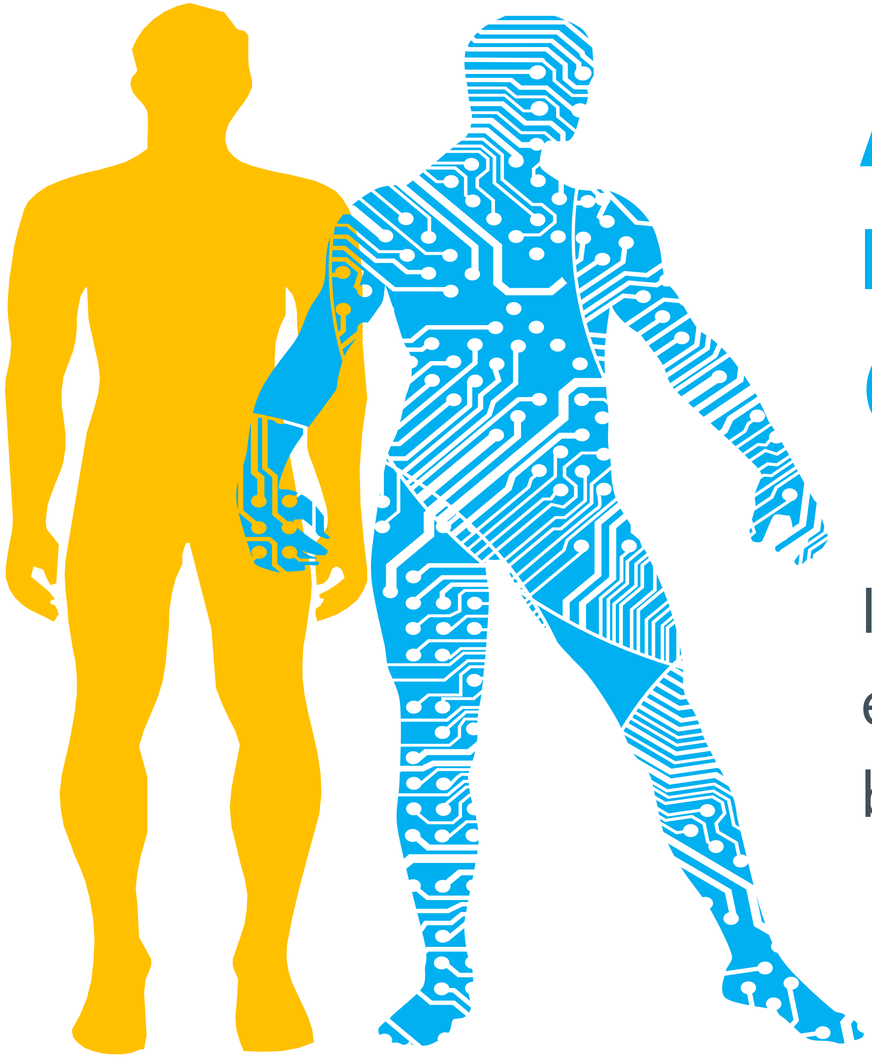
- Tons of research papers on arxiv.org

AI

Demystifying Artificial Intelligence

Natural Intelligence (NI)

Intelligence exhibited by Humans and other Animals.



Artificial Intelligence (AI)

Intelligence exhibited by Machines.

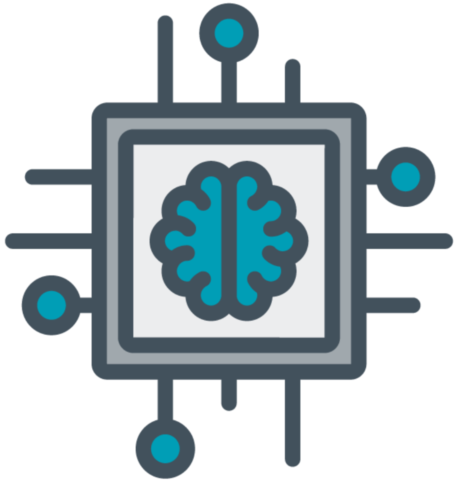
The main approaches to AI

Symbolic AI:

Inspired by computer science

Symbol processing engine (information processing metaphor).

For symbolists, neurons systematically implement the basic operations that are required for symbol processing.



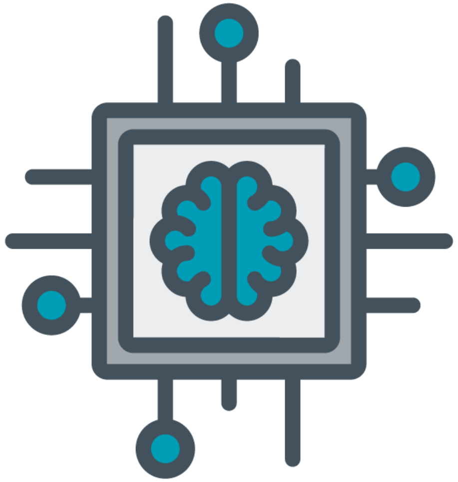
Sub-Symbolic AI:

Inspired by neurobiology

Artificial Neural Networks of computations. (Took many names over the year: Connectionism, Neural Nets, Machine Learning,…)

For sub-symbolists, nodes, links, activation, weights, emulate neural activity.

Prolog to this project



Since my AI Grad Studies in the 80s:

- I have been a nay sayer of sub-symbolic and Neural-Networks AI for Decision Automation.
- Sub-symbolic AI has proven good for some tasks but has been poor for decision automation.
- Symbolic AI has proven better for decision automation.

Symbolic and Sub-Symbolic AI Characteristics

Symbolic AI

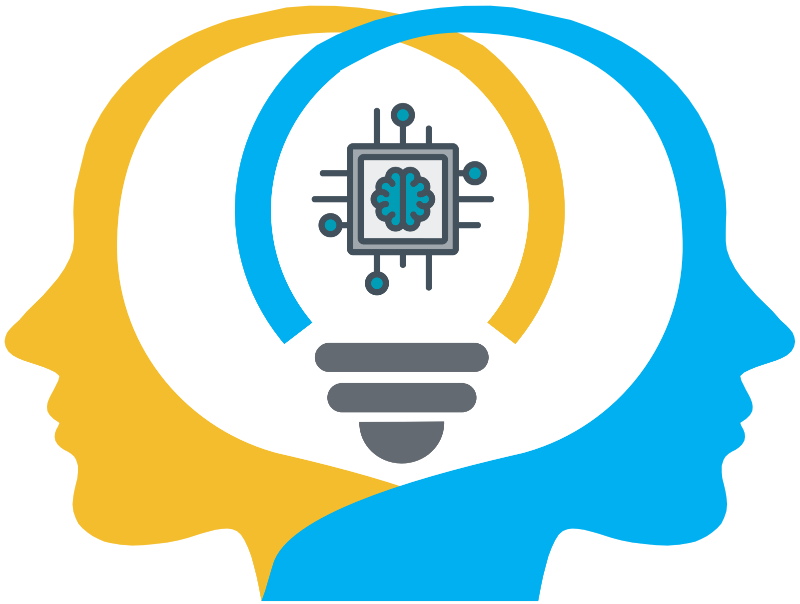
- Better at abstract problems
- More useful for explaining people's thought
- Easier to explain
- Knowledge elicitation needed upfront
- Not so Big Data
- Easier to debug
- Easier to control

Sub-Symbolic AI

- Better at sensing and perceptual problems
- More useful for connecting to neuroscience
- Easier to scale up
- Less knowledge upfront
- Big Data
- More robust against noise
- Better performance

Neuro-Symbolic AI

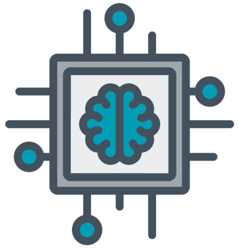
Symbolic AI



Sub-Symbolic AI

Field of artificial intelligence that focuses on the integration of neural networks and symbolic architectures in a manner that addresses strengths and weaknesses of each, in a complementary fashion.

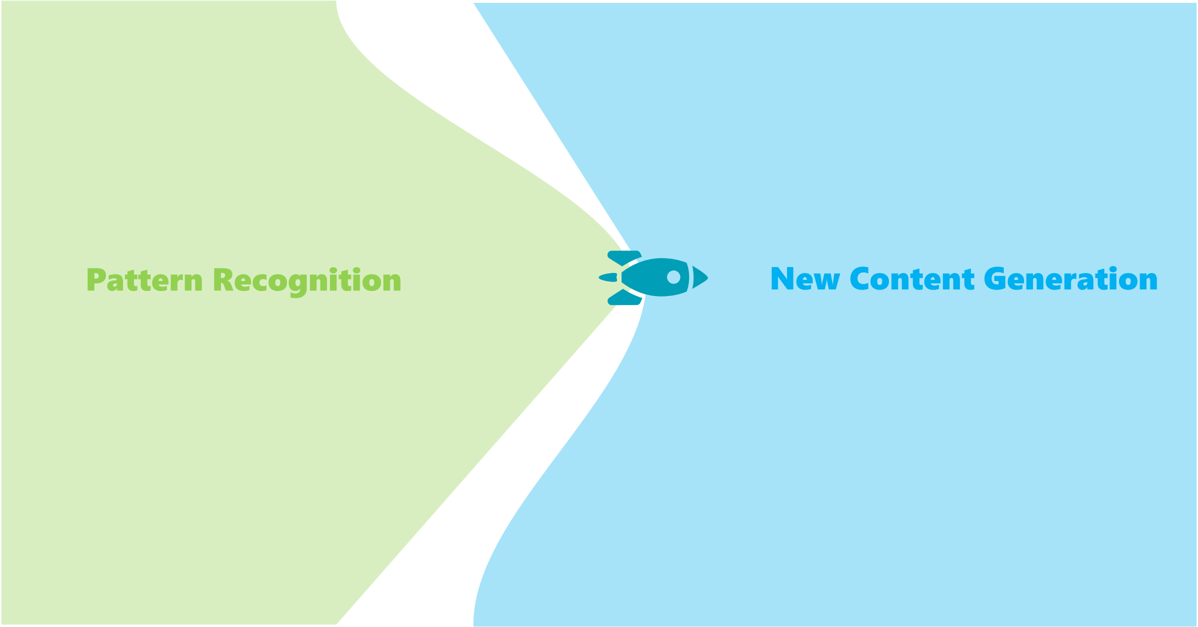
The ChatGPT Revelation



OpenAI

30 Nov 2022

The Advent of Generative AI

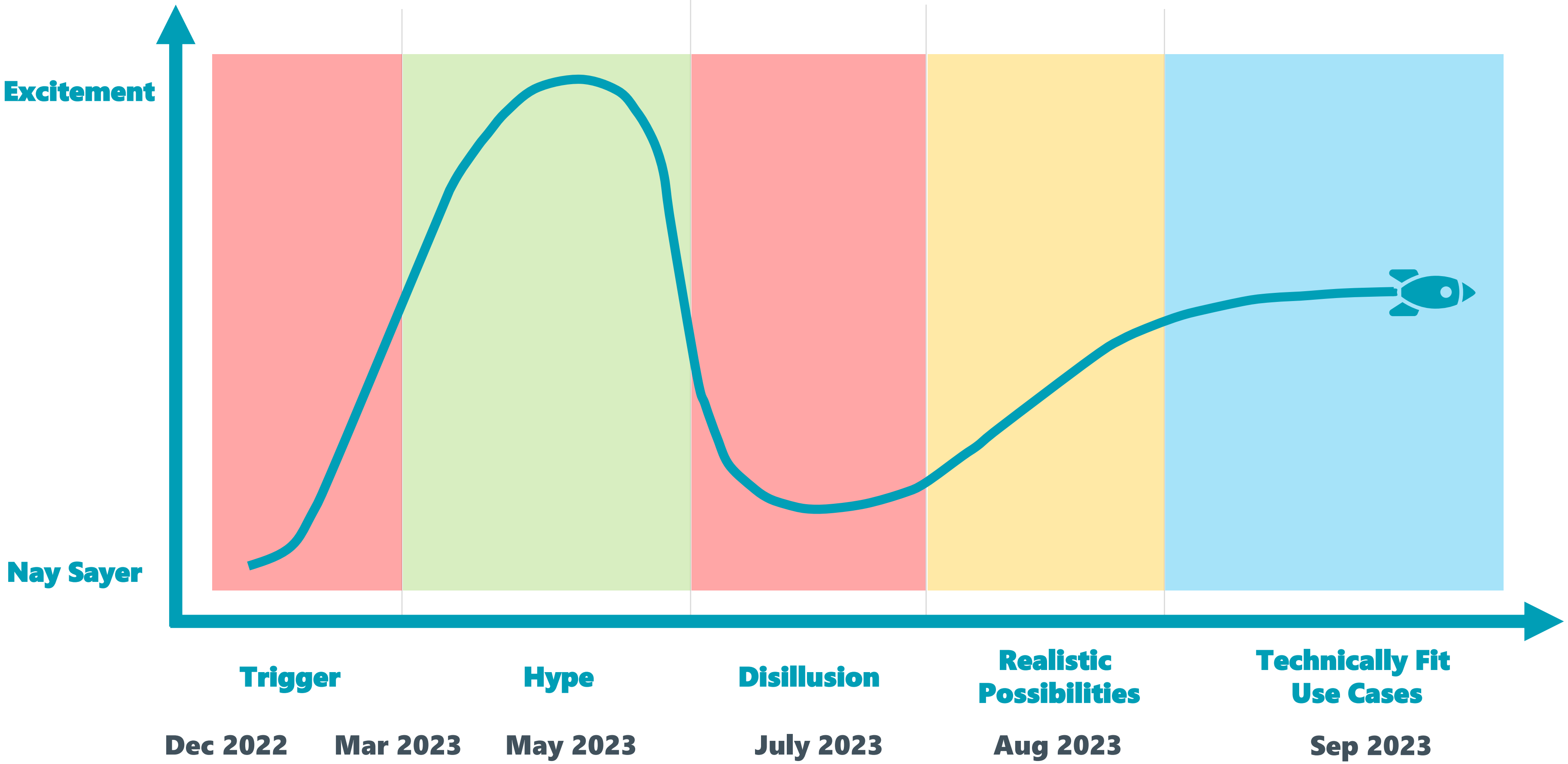


Took the world by surprise:

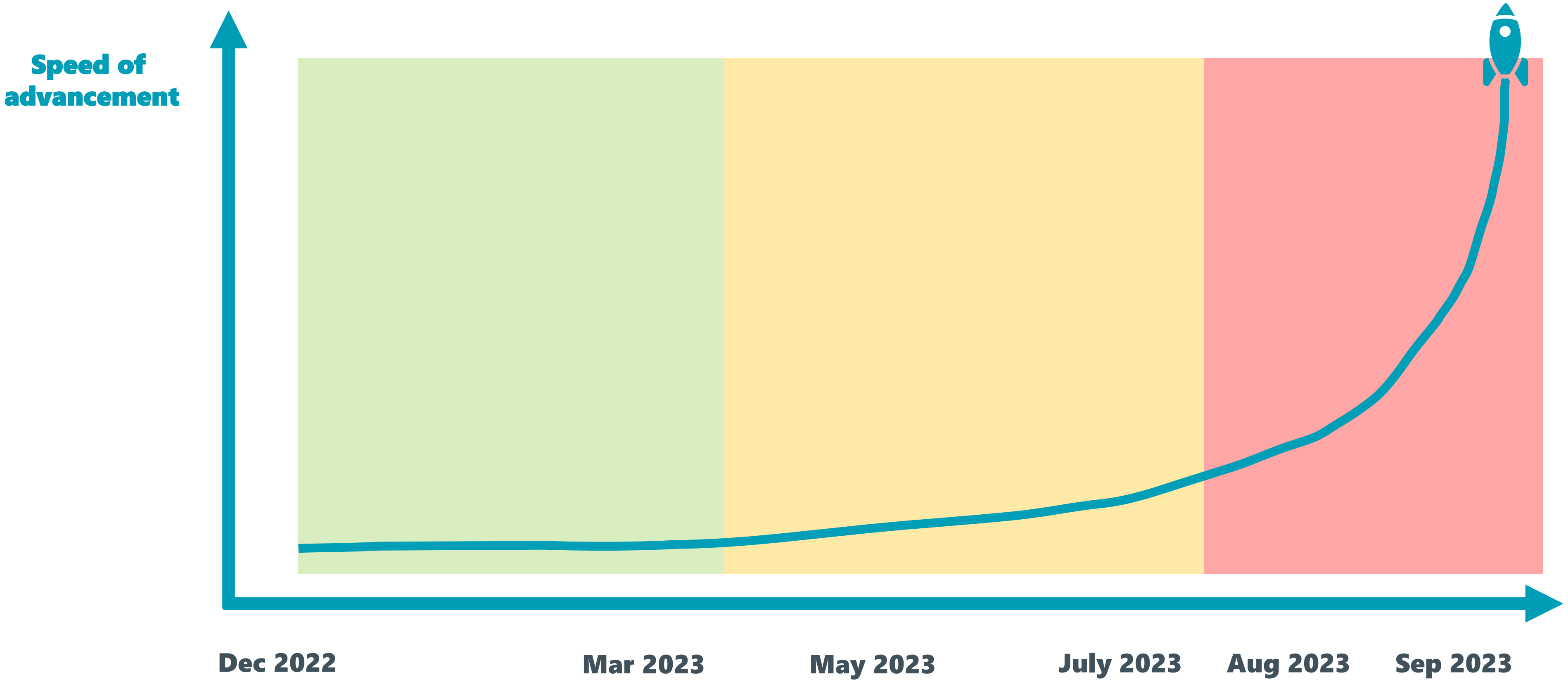
- Barrier to entry is lower [and thus]
- Time-to-value [seems] shorter

Trisotech

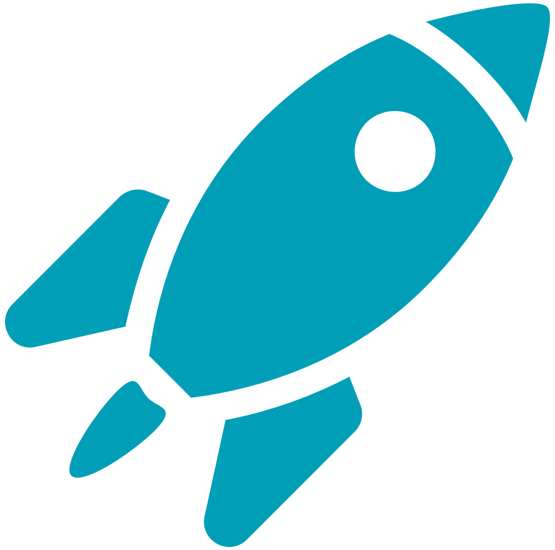
My Emotional Curve during this project



Rate of Publications and Technology Evolution



Now Everywhere



- **Business leaders** are asking how AI will affect their companies.
- **Governments** are wondering how it will affect the labor market, what risks it poses, and how to regulate it.
- **Companies** are trying to figure out how to use it without "giving away" their data to one of the platform vendors.
- **Developers** are experimenting with creative uses of generative AI.

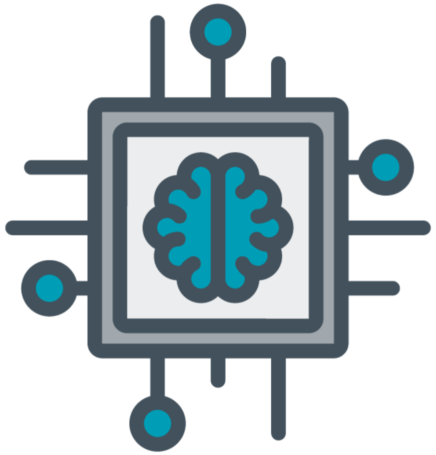
Should we get ready for:
The Future of AI or
The AI Future

AI

Demystifying Generative Artificial Intelligence

Large Language Models (LLMs)

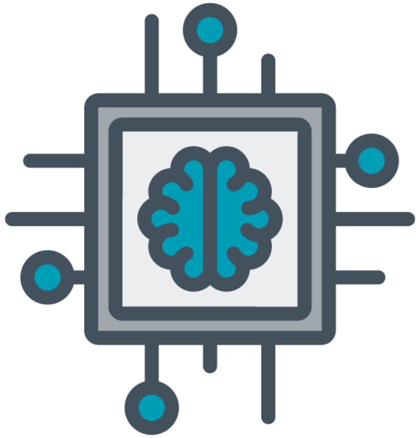
Large Language Model

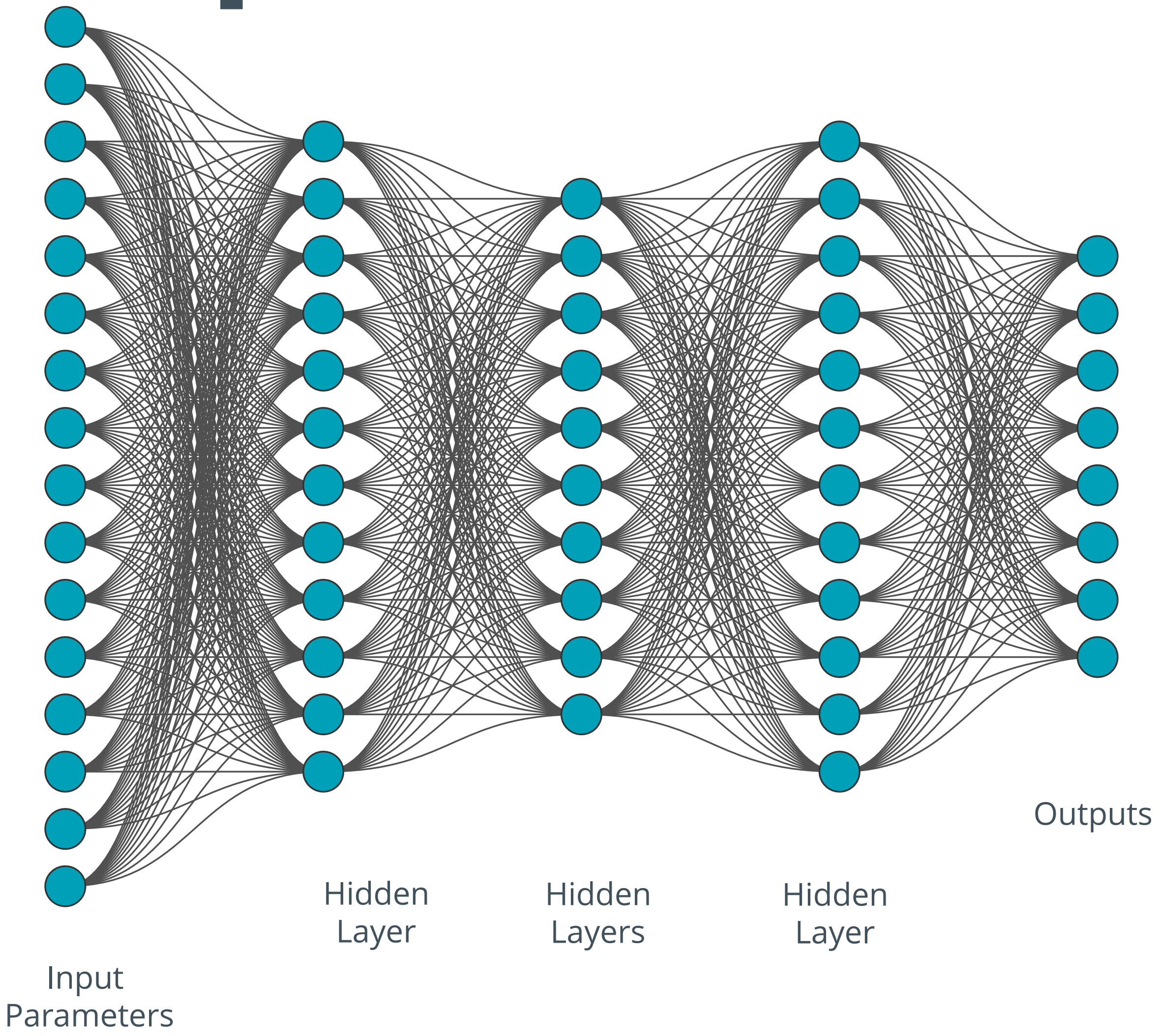


Foundation Model

- LLMs are deep learning models that can generate language outputs = Language Models.
- LLMs have billions of parameters and are trained on billions of words = Large.
- The term Foundation Models generally refers to models that are general purpose models which excel at a wide range of tasks, as opposed to being trained for one specific task.

GPT-3 has 175 billion parameters and 96 layers. GPT-4 has 1.8 trillion parameters.



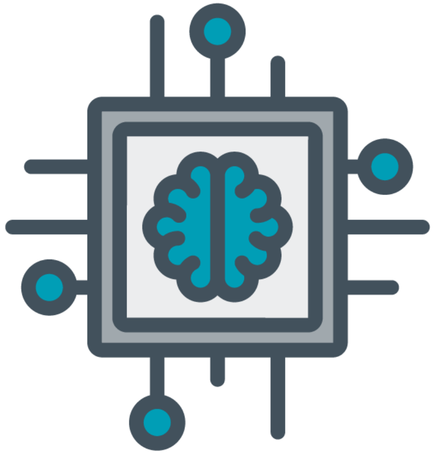


https://alexlenail.me/NN-SVG/

Just feed in lots of data…big data

The Near Future of Multi-Modal Generation

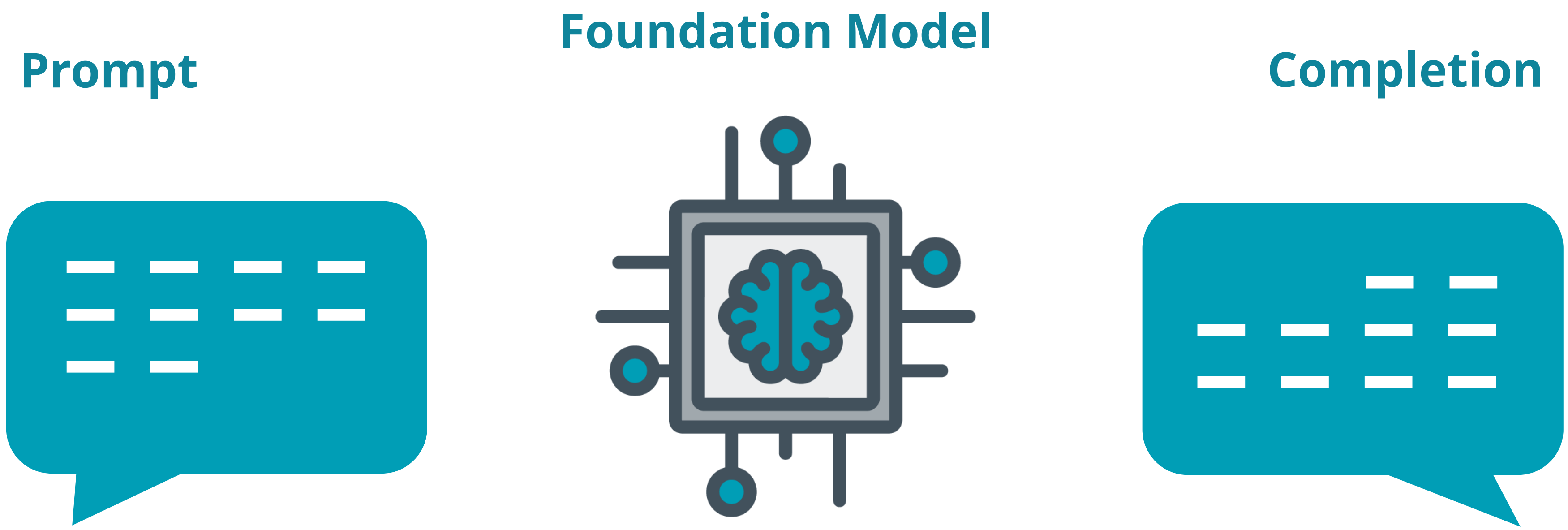
Foundation Model



Generation

- Text Generation
- Code Generation
- Image Generation
- 3D Generation
- Video Generation

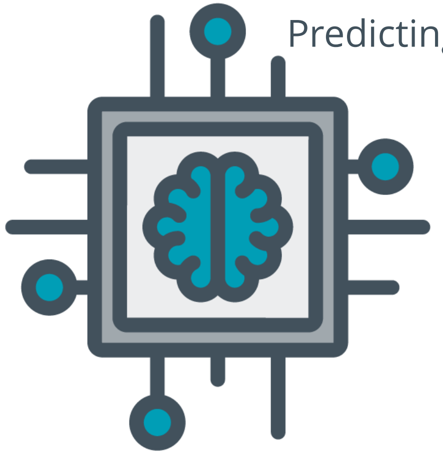
Prompt Completion



Trisotech

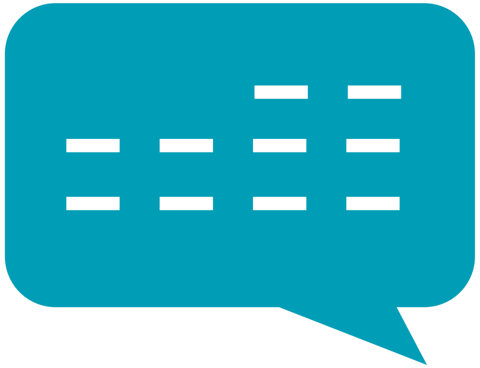
Prompt Completion

Foundation Model



Predicting the next token

Completion



Completion is Stochastic

- The completion is randomly determined.
- Sometimes randomness is desirable (Creative Contexts).
- Sometimes it is not (Factual Contexts).

Trisotech

Foundation Models Main Settings

Tokens

- the basic units of text or code.

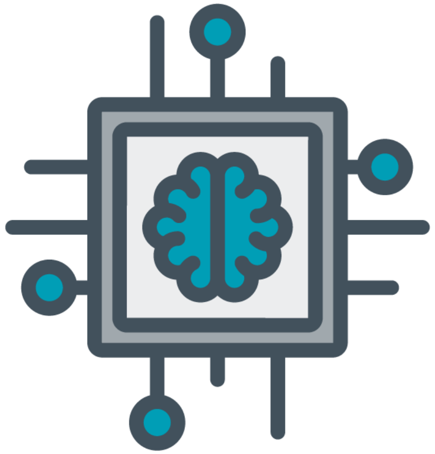
Temperature

- the lower the temperature, the more deterministic the results.

Top p

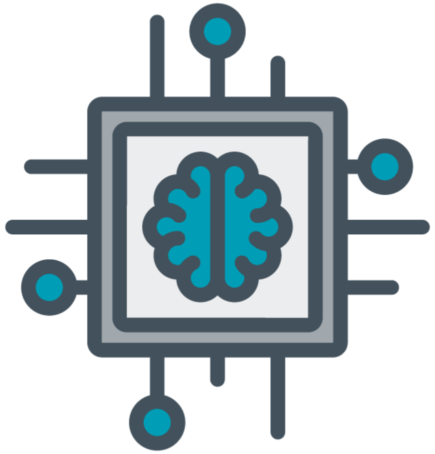
- chooses from the smallest possible set of words whose cumulative probability exceeds the probability p.
- The general recommendation is to alter one, not both.

Foundation Model

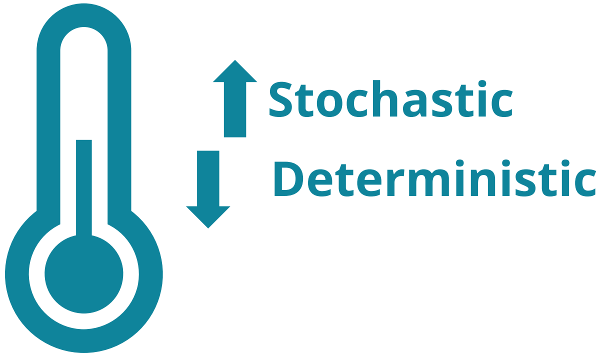


Foundation Models Main Settings

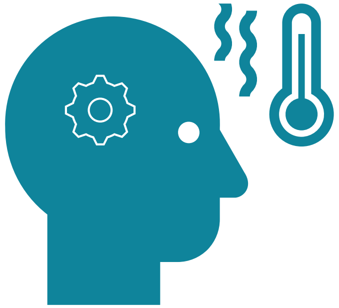
Foundation Model



Temperature



Hallucination



Be mindful that if your foundation model has a high Temperature, it may start Hallucinating.

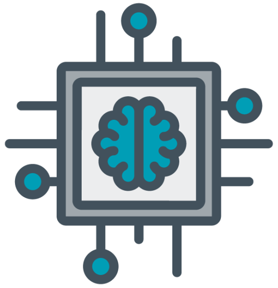
Foundation Models Main Settings

Token Example

Tokens are the basic units of text or code used by large language models.
An example of a multi-token word is demystifying.

Temperature Example

The sky is



More likely

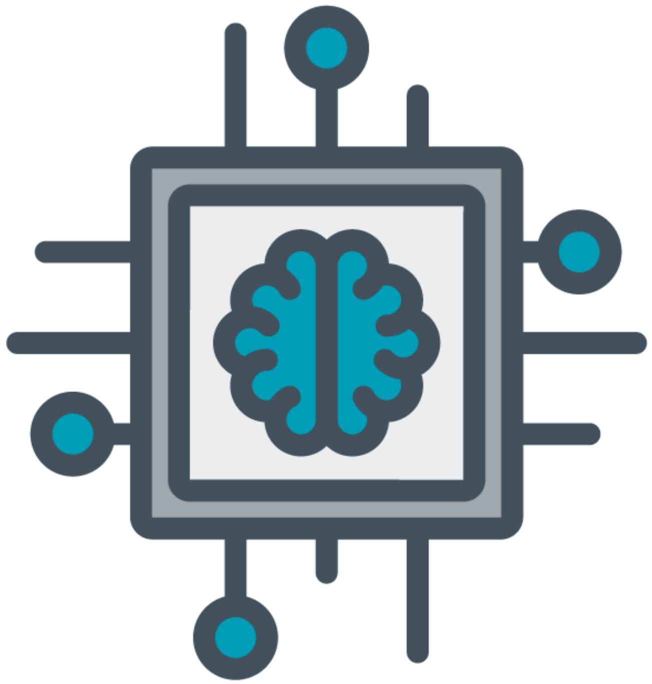
Less likely

- blue
- clear
- cloudy
- dark
- beautiful

The sky is blue

OpenAI Tokenizer

Prompt Completion Challenges

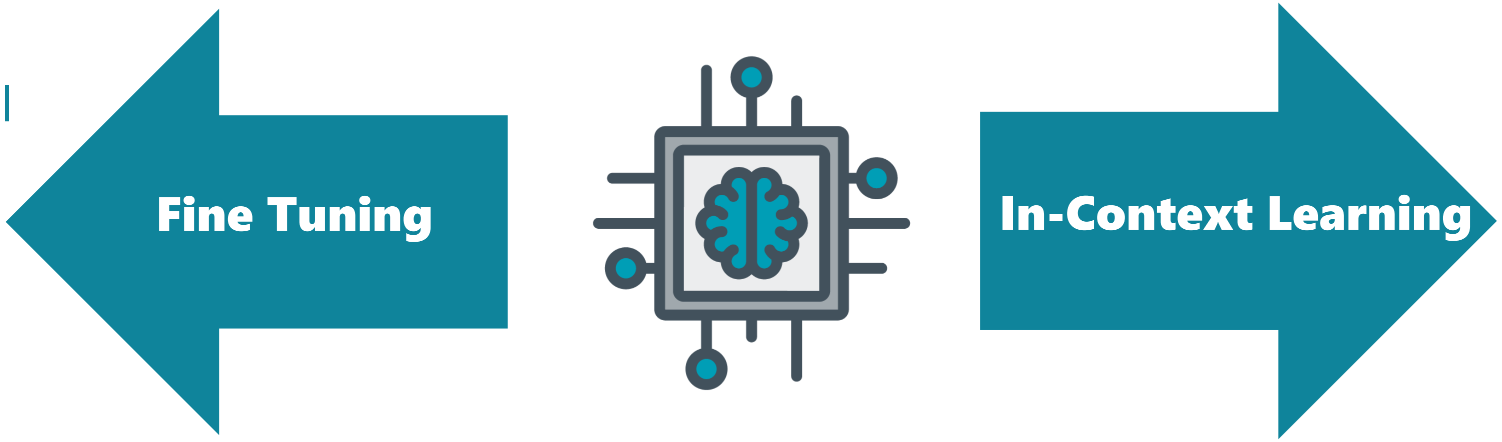


- Hallucination
- Training Window Limitation
- …
- API Cost
- API Latency
- …
- Bias
- Ethics
- Copyrights
- …
- Prompt Injection Attacks
- …

Fit for Purpose Models

Foundation Model

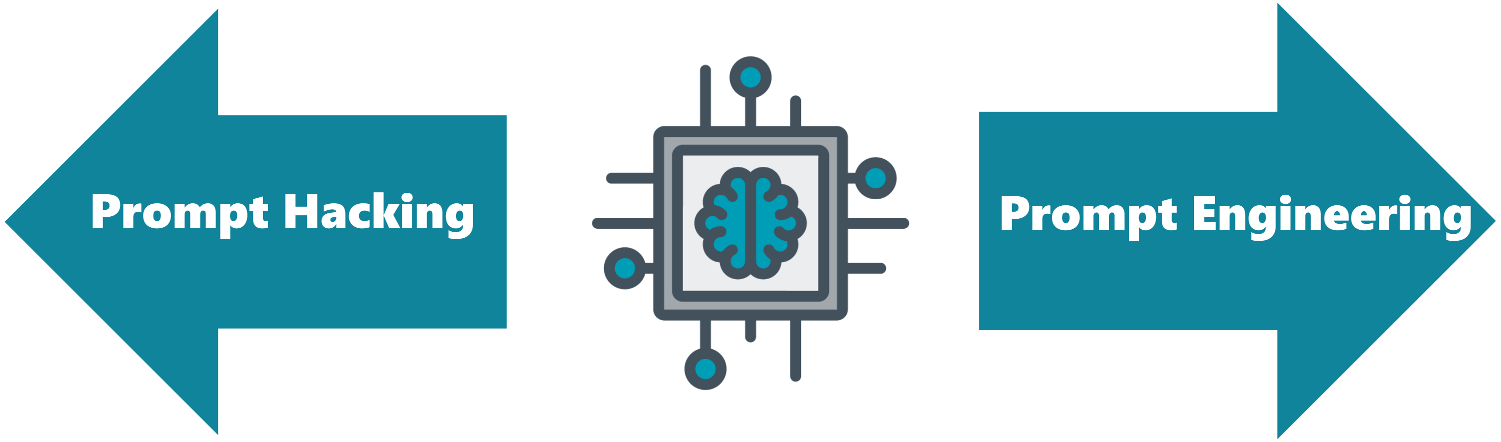
A foundation model that has been retrained for a specific purpose, task, or domain.



Manipulating the generative behavior of foundation models via prompt conditioning using a provided context.

In-Context Learning

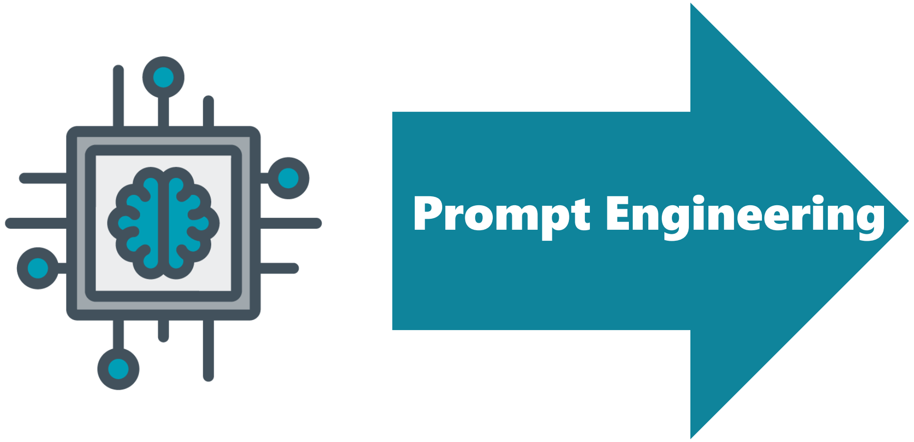
Foundation Model



A trial-and-error technique of crafting prompts.

The process of crafting effective prompts to interact with foundation models, aiming to elicit desired responses.

Prompt Engineering

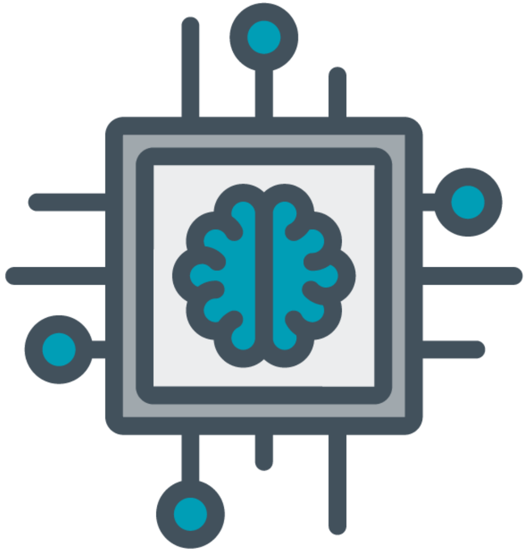


Prompt engineering is emerging as a key methodology to leverage large language models to obtain desired outputs by providing explicit instructions or constraints.

By carefully crafting prompts, we can shape language models behavior and align their outputs with our intentions.

Prompting Techniques

To scaffold the foundation model completion to be as desired (creative/factual/logical?):



- Prompt Template
  - Structured Prompt
  - Role-Based Prompt
- Few Shot Training
- Prompt Pipeline
- Prompt Chaining
- Retrieval Augmented Generation (RAG)
- ReAct
- Chain of Thoughts
- Tree of Thoughts
- Graph of Thoughts
- Program Simulation
- …

AI

Creating Generative AI Applications

Then: “it’s a simple matter of programming”.

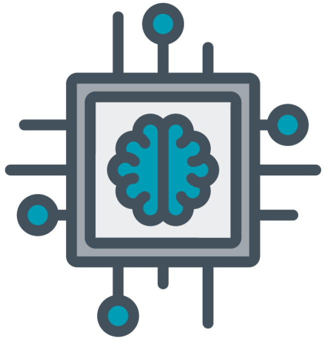
Now: “it’s a simple matter of training”.

It was never simple then; it is not simple even now.

Difference between Software Engineering and Prompt Engineering

Initial Development → Debugging & Improvement

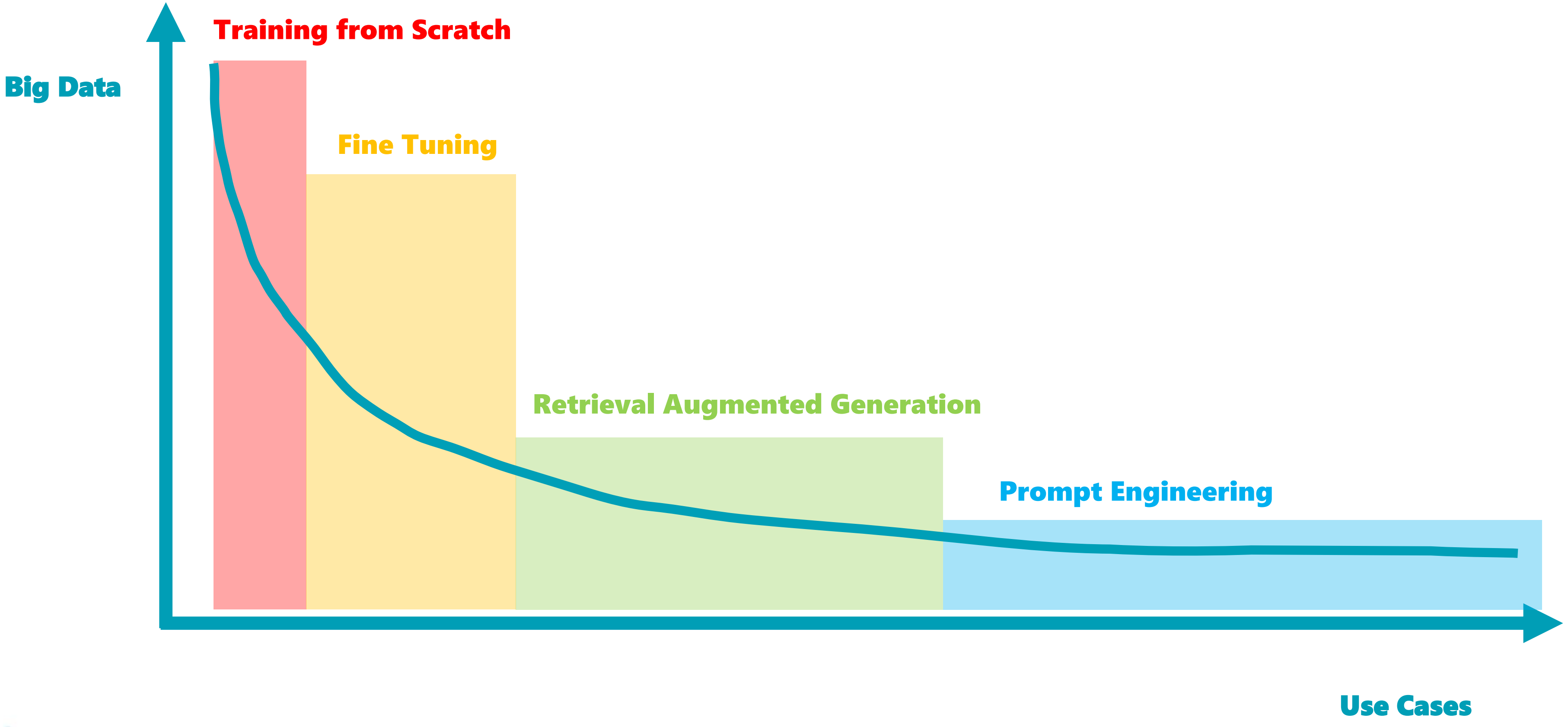
- Implementation
- Coding
- Loose Coupling
- Plan Functionality
- Confirming via Unit Test
- Complicated
- Predictable



Initial Development → Debugging & Improvement

- Experimentation
- Analyzing
- Chaining Dependencies
- Seek Quality
- Insight via Evaluation
- Complexity
- Unpredictable

Data Requirements for Creating GenAI Apps

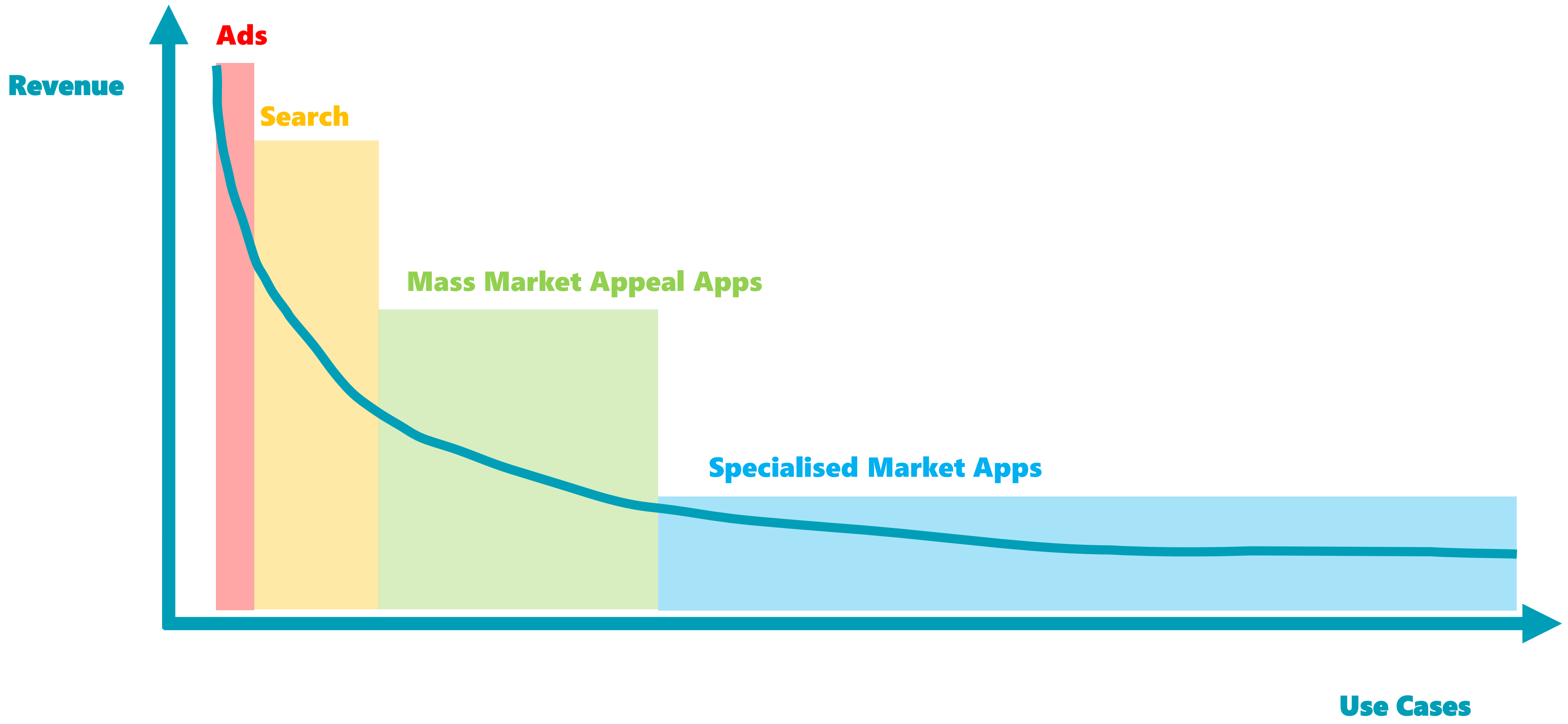


The two most common applications of GenAI remains:

Building customer service chatbots

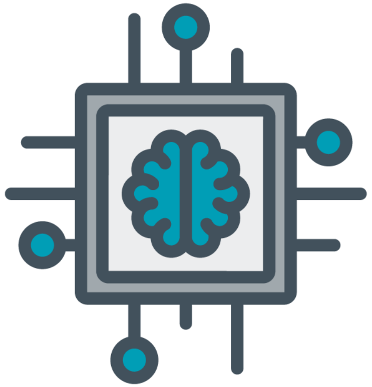
Answering questions based on documents

Potential markets of GenAI Apps



Creating GenAI Apps

BPM+



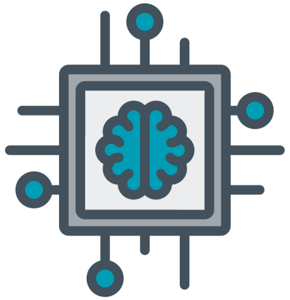
GenAI

Neuro-Symbolic AI

Combining the benefits of Low-Code Model Driven App creation with Generative AI new possibilities.

BPM+ Triple crown of BPMN, DMN and CMMN combined with engineered GenAI prompt components.

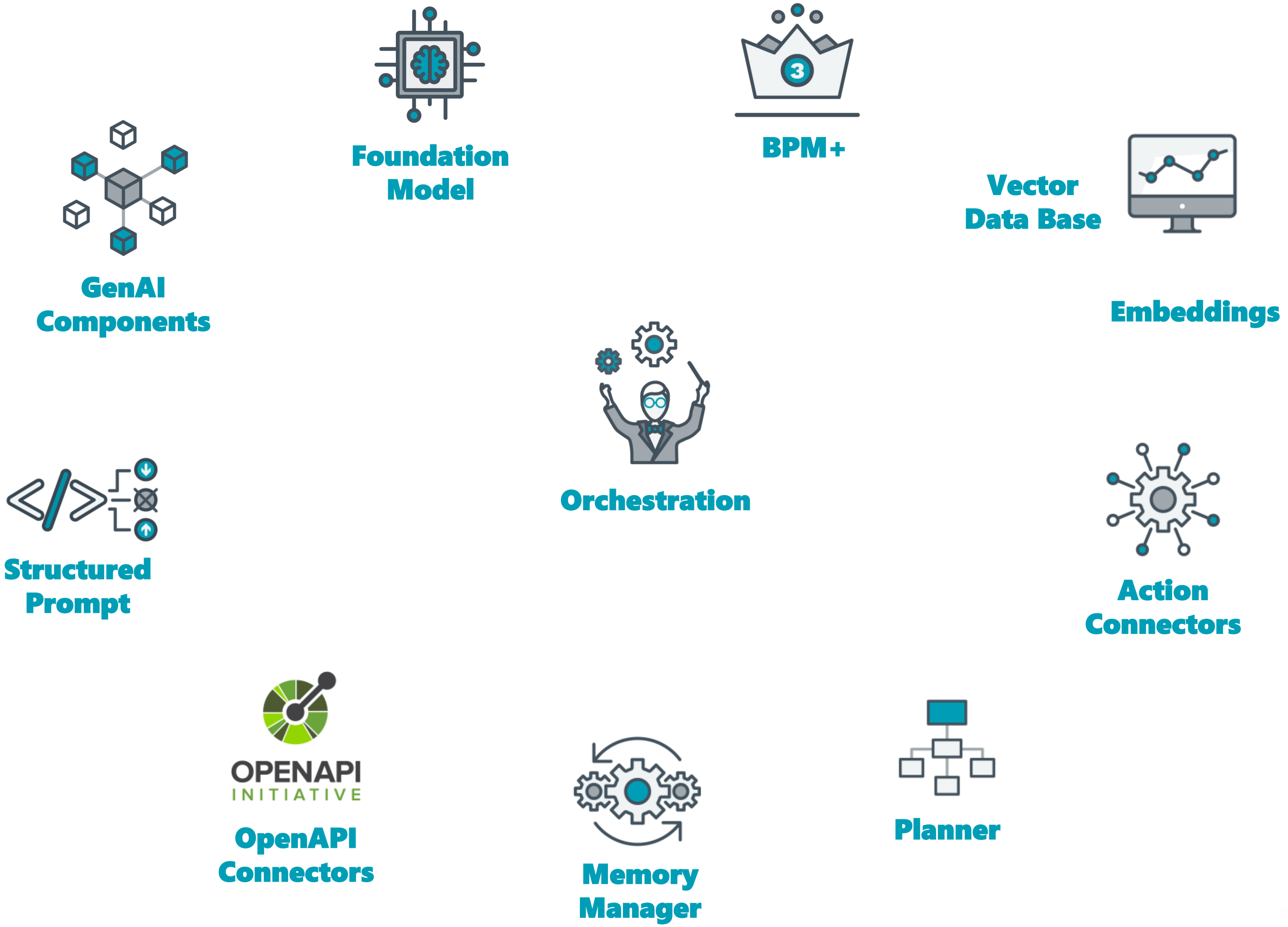
GenAI Feature Set Exploration



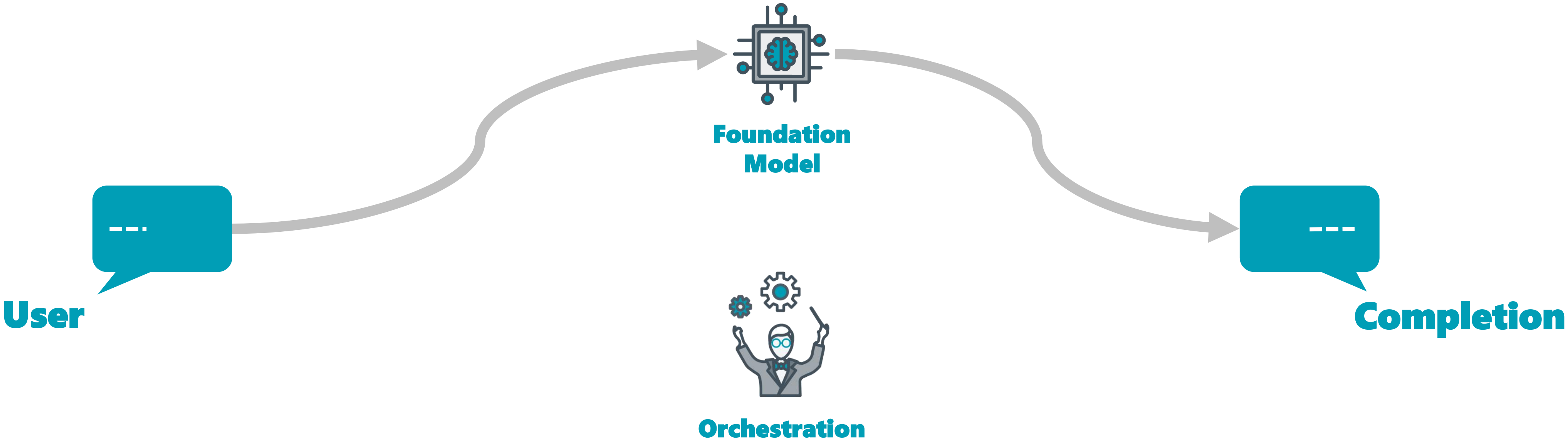
Neuro-Symbolic AI

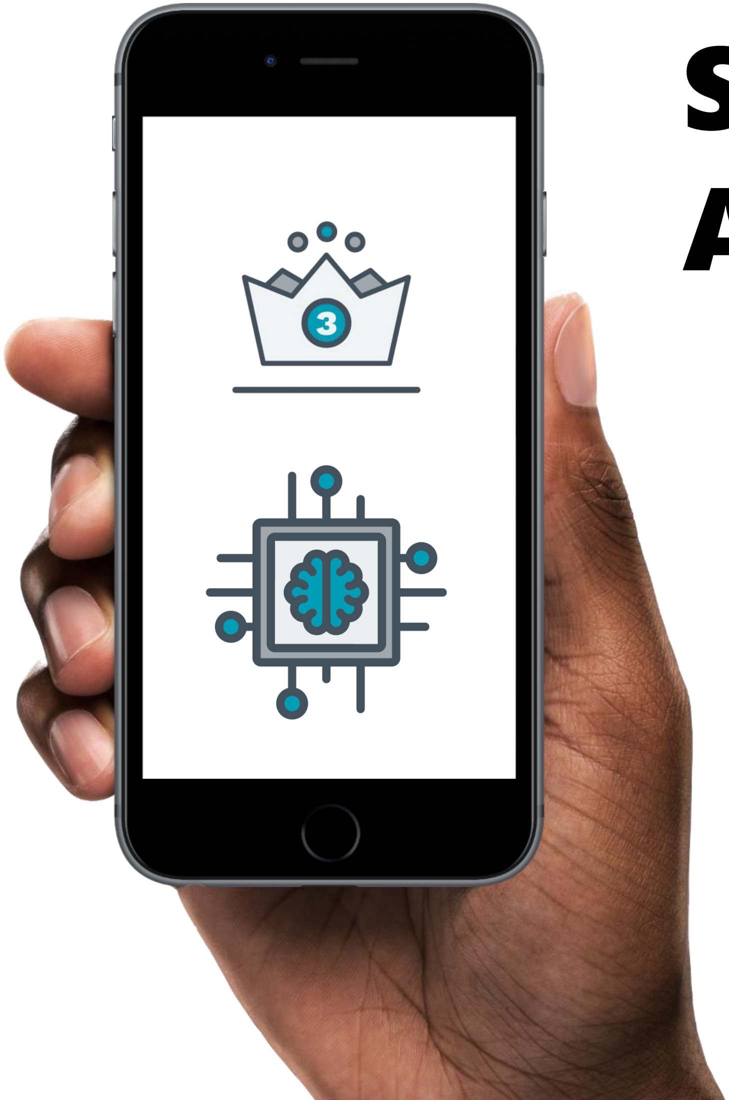
The building blocks of prompt orchestration. A set of features that accelerates the creation of Generative AI applications.

GenAI Orchestration



Prompt Completion





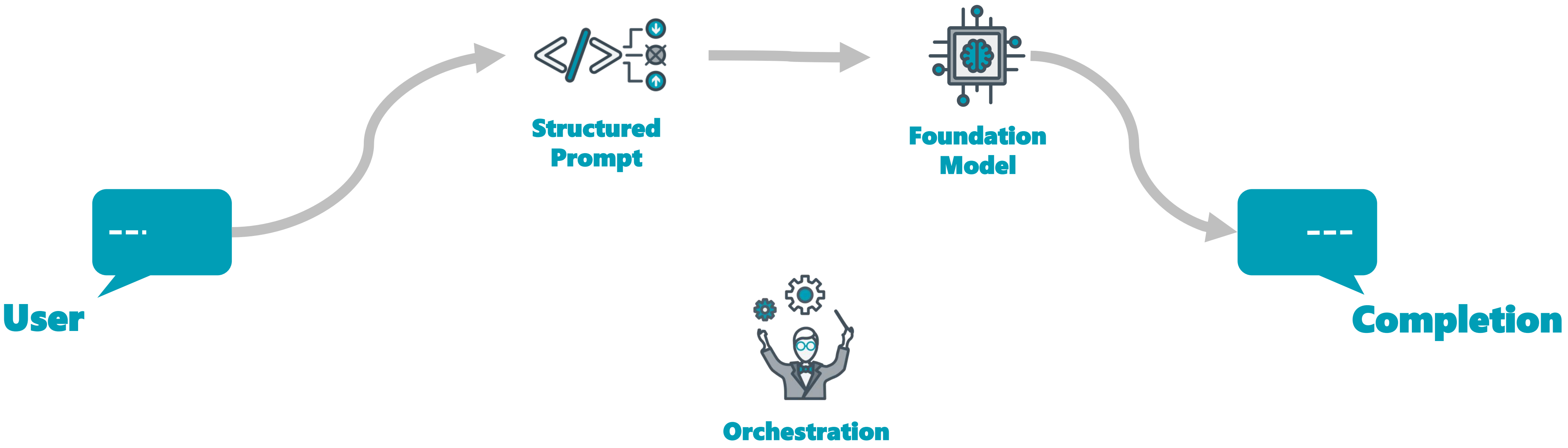
Sentiment Analysis

Trisotech

Reusable GenAI Components

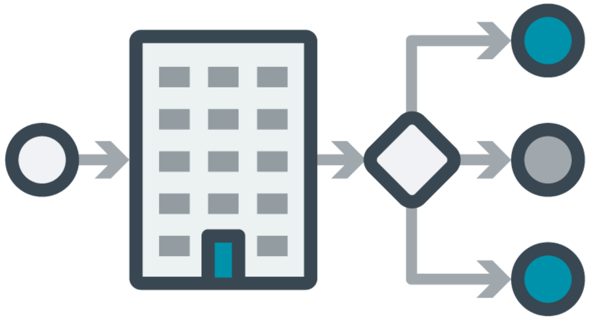
- **Structured Prompt Generator:** Applies prompt engineering best practices.
- **Question Answering:** Generates an answer to a question.
- **Text Summarization:** Generates a shorter version of a text.
- **Sentiment Analysis:** Detects positive or negative sentiment of a text.
- **Text Translation:** Translate a text into a different language.
- **Text Expansion :** Generates a longer text from an excerpt.
- **Text Generation:** Generates a coherent text according to criteria.
- **Image Generation:** Generates an image from a text.
- …

Structured Prompt Completion



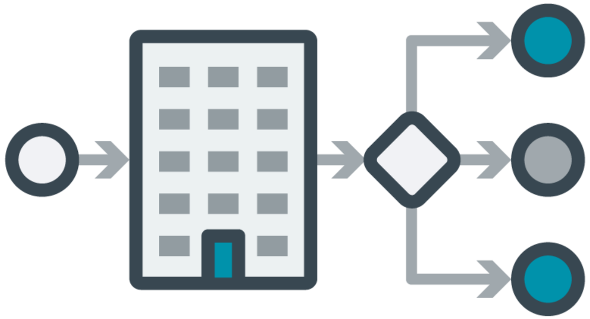
Structured Prompt Completion

Reference Source and Business Data



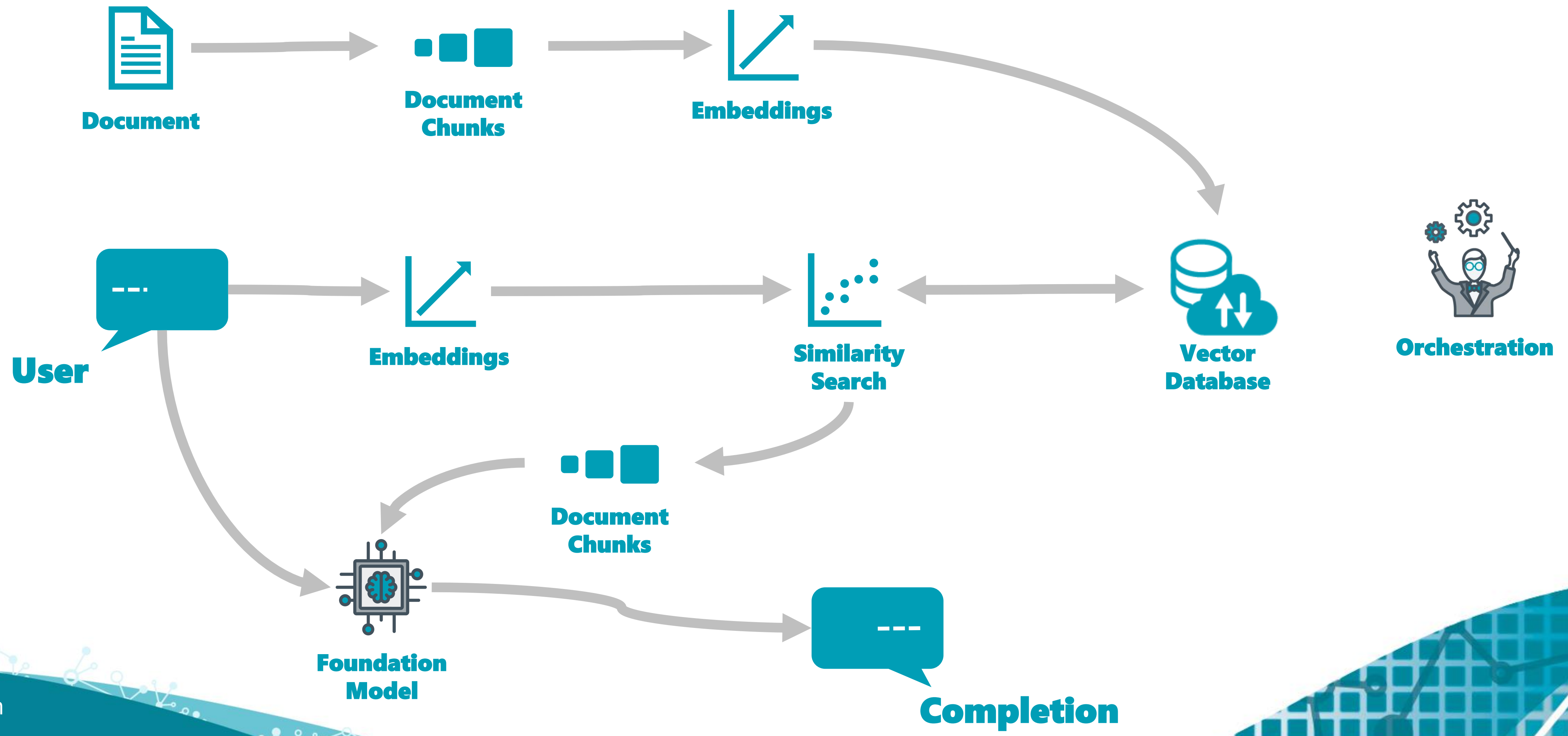
- One of the big complaints about Generative AI is there is no anchoring to sources.
- Fragments of text are returned but no one can say why or its evidence.
- Retrieval Augmented generation (RAG) is proposed as an improvement.

Retrieval Augmented generation (RAG)



- Tackles the limiting Training Window with Task Specific Data/Documents.
- For each user query or question, contextual chunks of text are retrieved to be added into the prompt context.
- These chunks of text are retrieved based on semantic similarity with the user query or question.

Retrieval Augmented Generation (RAG)



Questioning Documents with RAG

Chatbots and Agents

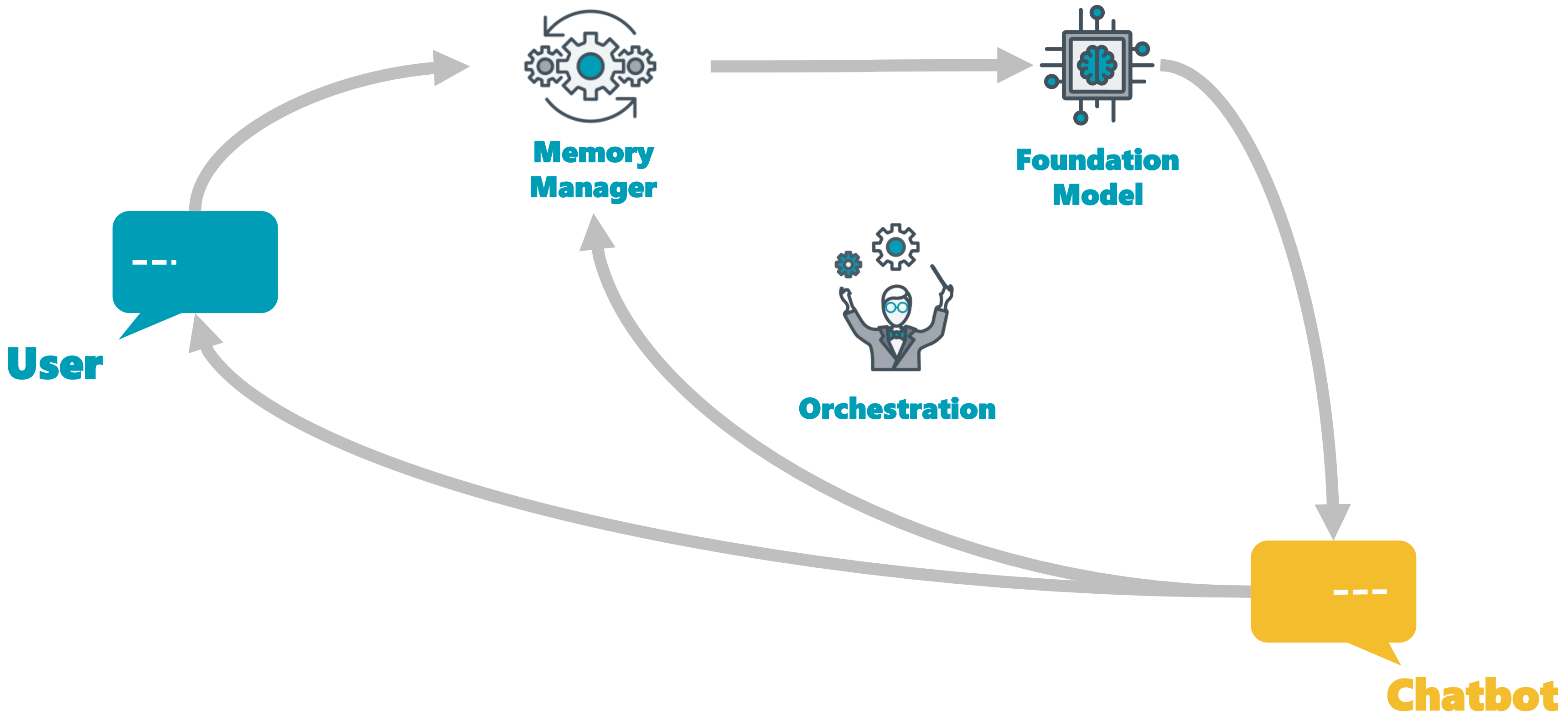
Chatbot

a conversational computer program designed to **simulate conversation** or interaction with human users via natural language.

Agent

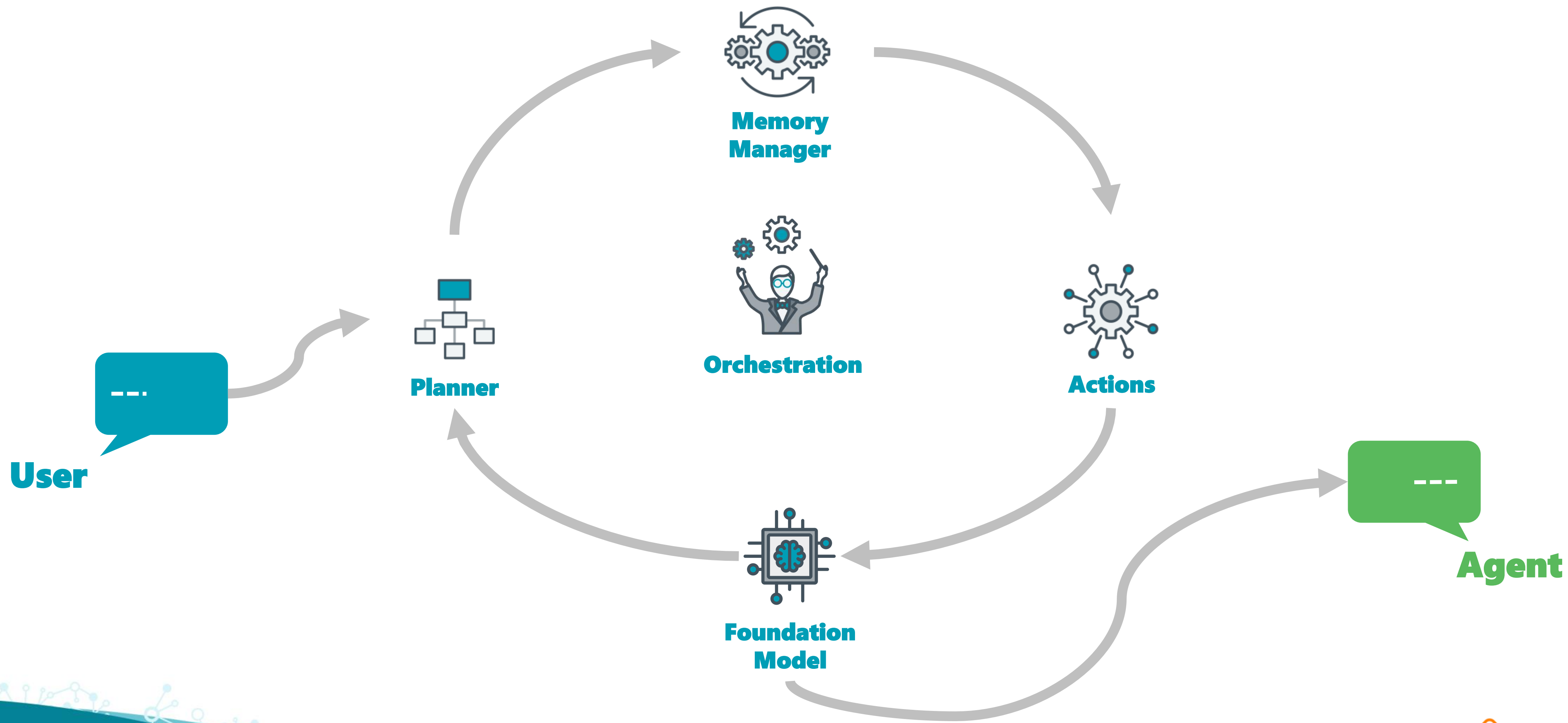
a system aware of its context which **takes actions** to maximize its chances of success.

Chatbot Orchestration



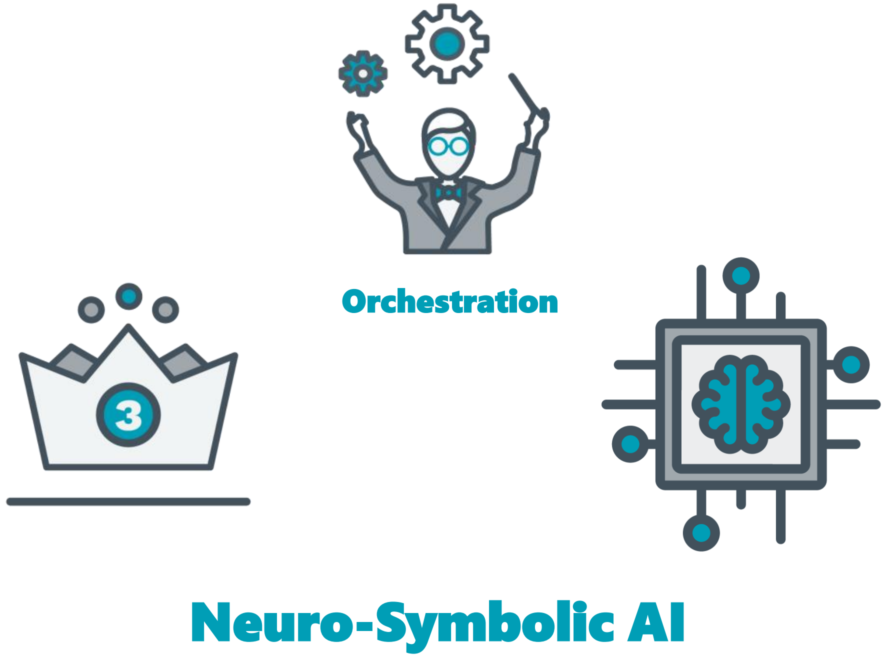
Trisotech

Agent Orchestration



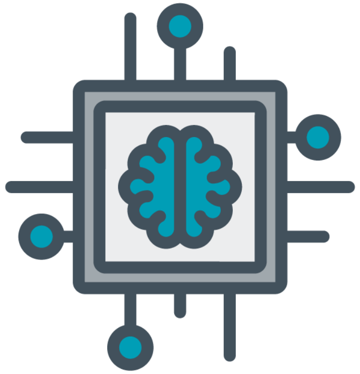
Trisotech

Co-Pilot



I like the term co-pilot as it implies human control. The benefit of engineered GenAI prompt orchestration is that we can add validation Human Task at any step to ensure desired behavior.

Lessons Learned



- We are very early in exploring GenAI.
  - Things are progressing very rapidly.
- GenAI completion is randomly determined.
  - Sometimes randomness is desirable (Creative Context).
  - Sometimes it is not (Factual Contexts).
- Use GenAI where it is efficient otherwise use something else.
  - LLMs are known for their language capabilities not scientific capabilities.
  - Why use tons of computation resources to solve 1+1.
- GenAI is not the App. It is an Interface. At best a Component of an App.
  - As such it needs to be componentized and orchestrated with other programmable components.

Conclusions

Neuro-Symbolic AI

- Should we get ready for “The AI Future” or the “The Future of AI” ?
  - It depends on if we are seeing a “GenAI Revolution” or just a “GenAI Hype Bubble”.
- Use GenAI for Technically Fit Use Cases
  - Use GenAI where it is efficient otherwise use something else.
- From someone who learned AI in the 80s:
  - We sure live in interesting times.
  - I believe Neuro-Symbolic AI time has come.

BPM+

GenAI

Any questions?

Denis Gagne
dgagne@trisotech.com

THANKS!