



Crafting an AI Mixologist Using Semantic Kernel

what's your go-to cocktail?

CHEERS!

Ron Dagdag

R&D Engineering Manager at



7th year Microsoft AI MVP



TABLE OF CONTENTS

01

Generative AI

What are foundation
models

02

Large Language Models

How language models
work?

03

Semantic Kernel

Semantic Completion and
Semantic Similarity

04

AI Agents and Copilots

Plugins, Planners, Personas

"Teaching AI to mix a cocktail is like trying to explain a hangover to a toaster. It understands the ingredients but not the headache!"

-ChatGPT

"AI Headaches"

01

GENERATIVE AI

Foundation Models



Artificial Intelligence

Machine Learning

Deep Learning

Generative AI



Artificial Intelligence

Seeks to create intelligent machines that can replicate/exceed human intelligence



Machine Learning

subset of AI that enables machines to learn from existing data and improve upon that data to make decisions or predictions



Deep Learning

a machine learning technique in which layers of neural networks are used to process data and make decisions

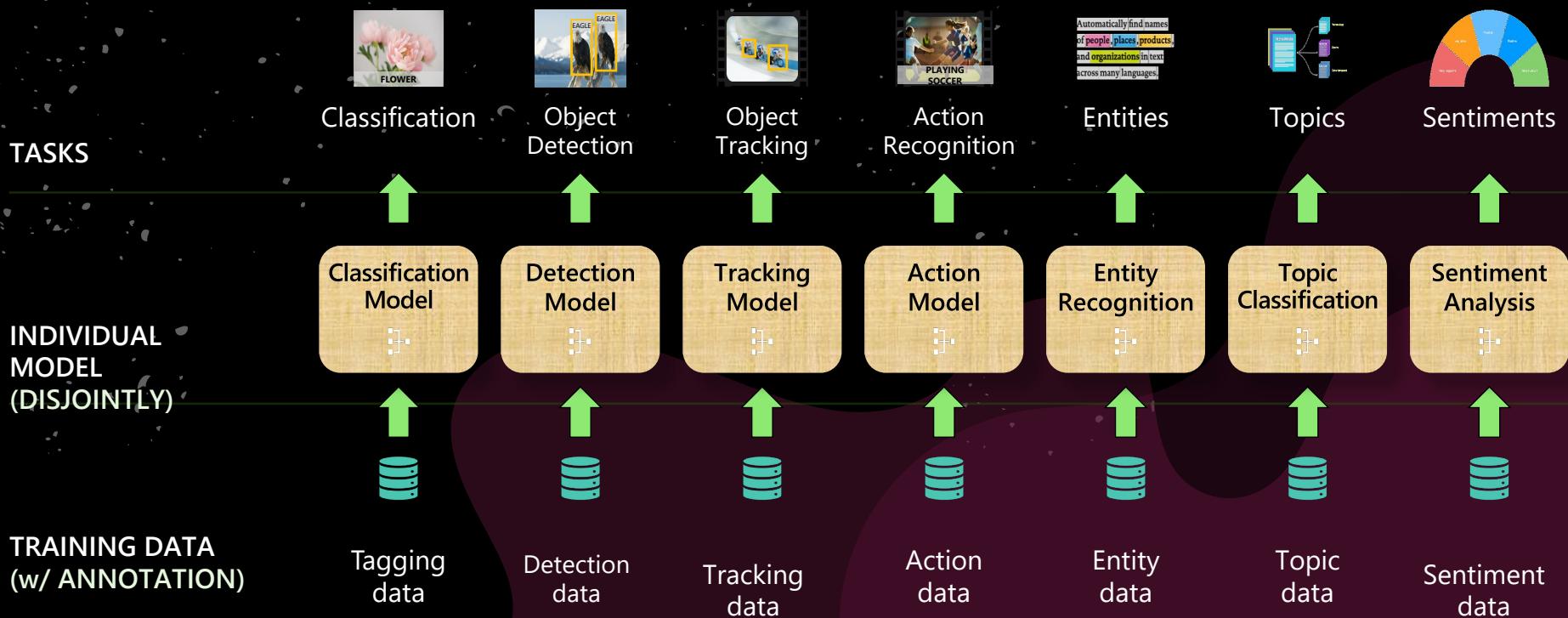


Generative AI

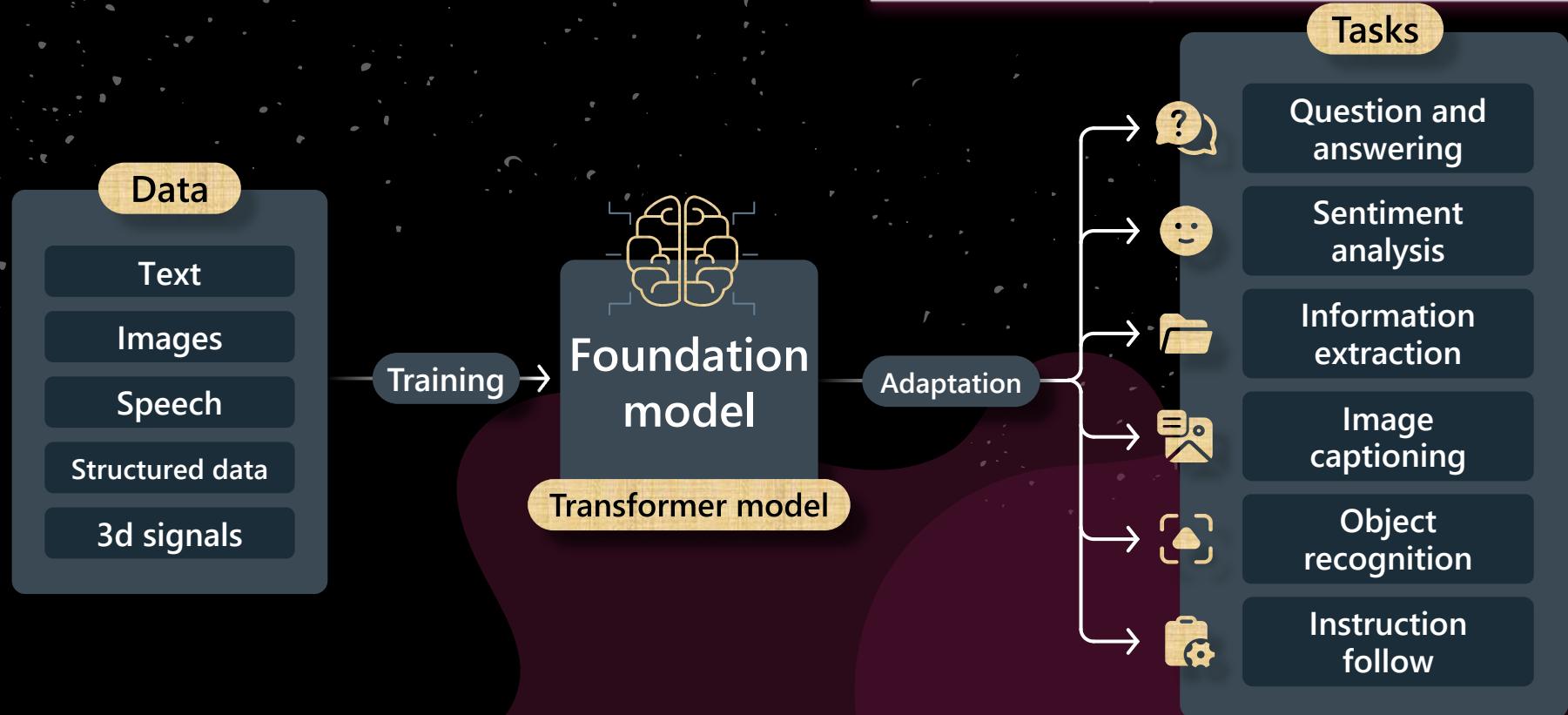
Create new written, visual, and auditory content given prompts or existing data.

Traditional model development

High cost and slow deployment—each service is trained disjointly



Foundation models

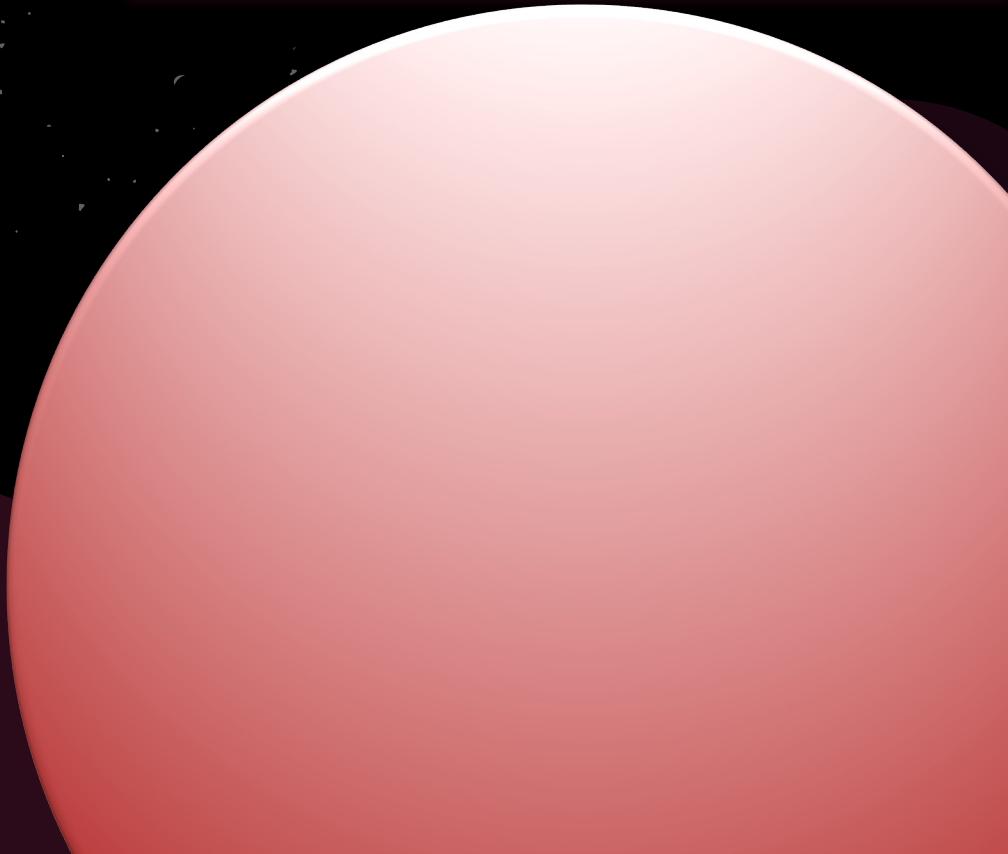


Foundation models

1.76T
GPT-4 (300x)

175B
GPT-3

17B
Turing-NLG

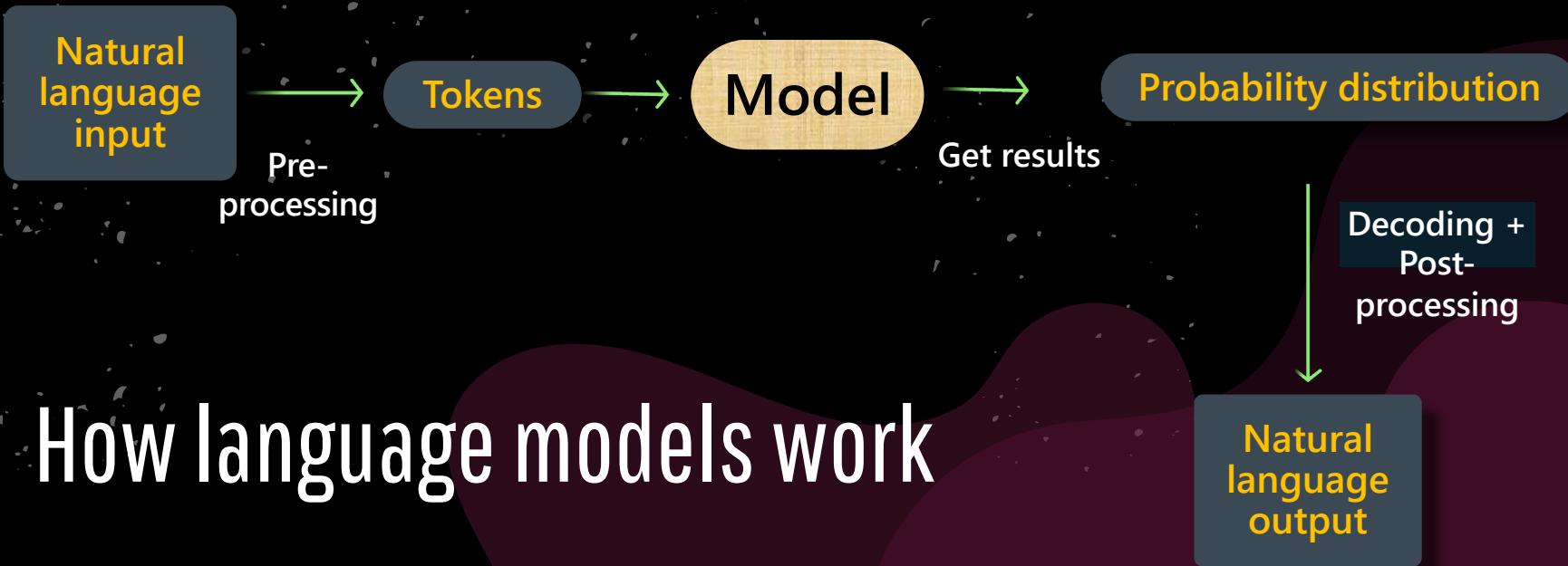




02

LARGE LANGUAGE MODELS

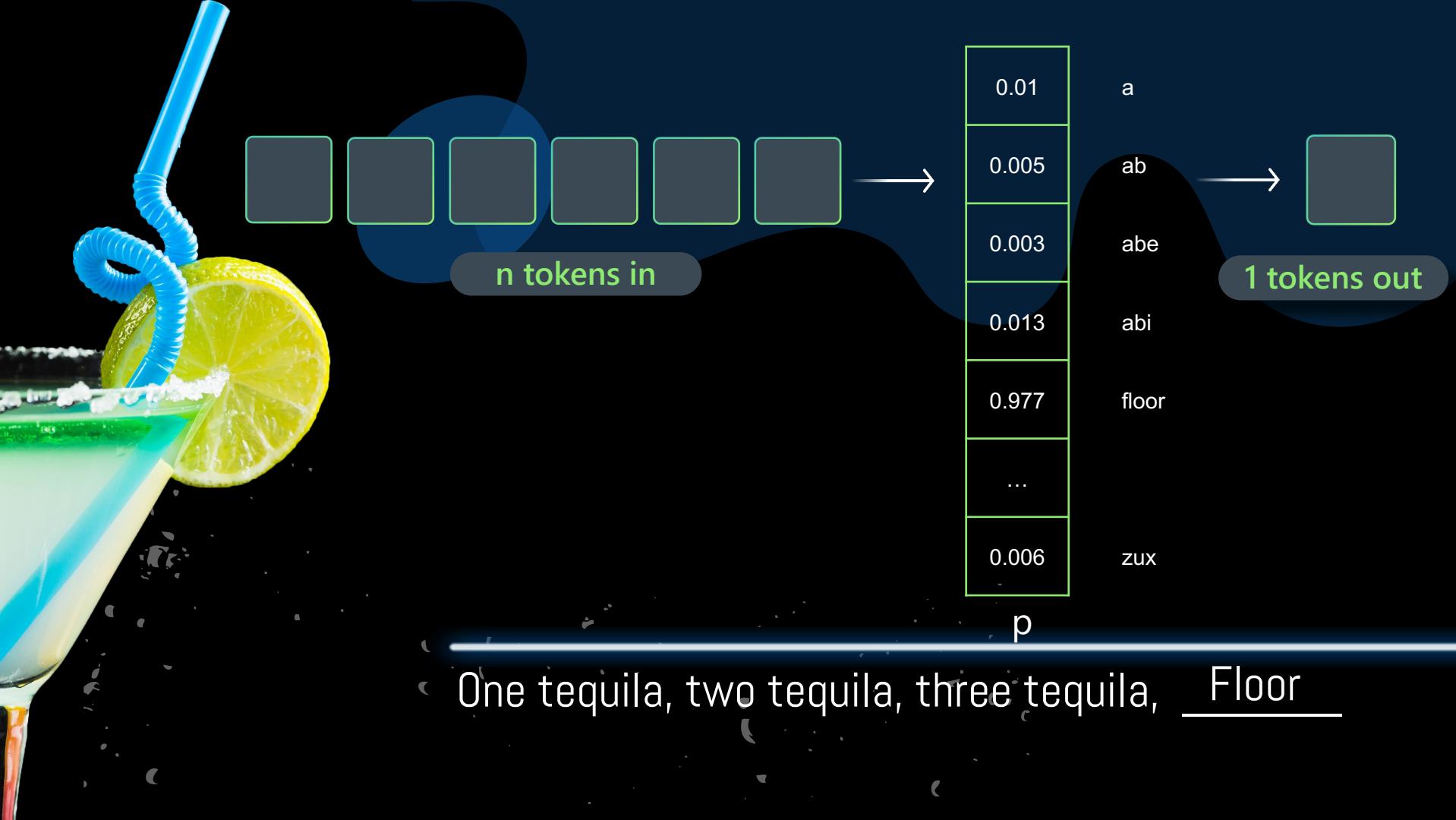
How language models work?



How language models work



One tequila, two tequila, three tequila,





<https://platform.openai.com/tokenizer>

Tokens	Characters
45	154

Wastin' away again in Margaritaville
Searchin' for my lost shaker of salt
Some people claim that there's a woman to blame
But I know, it's nobody's fault



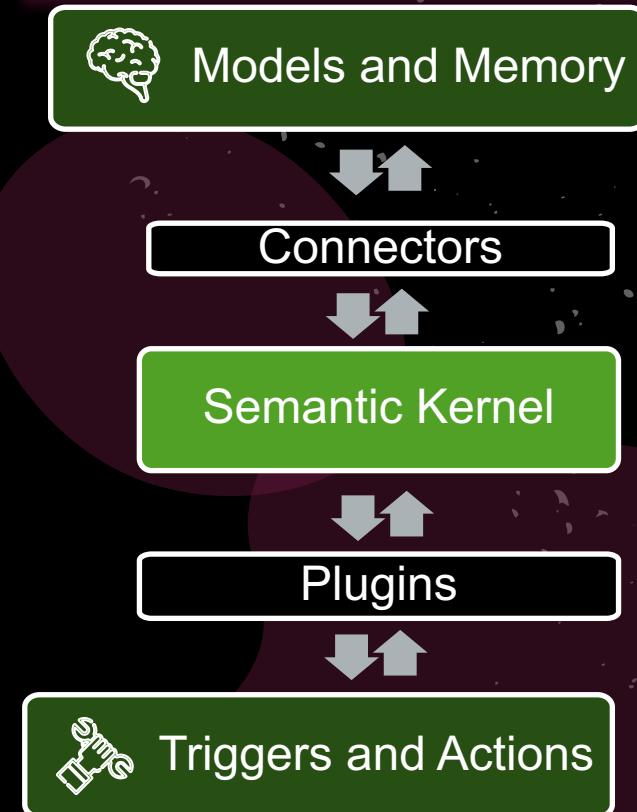
03

SEMANTIC KERNEL



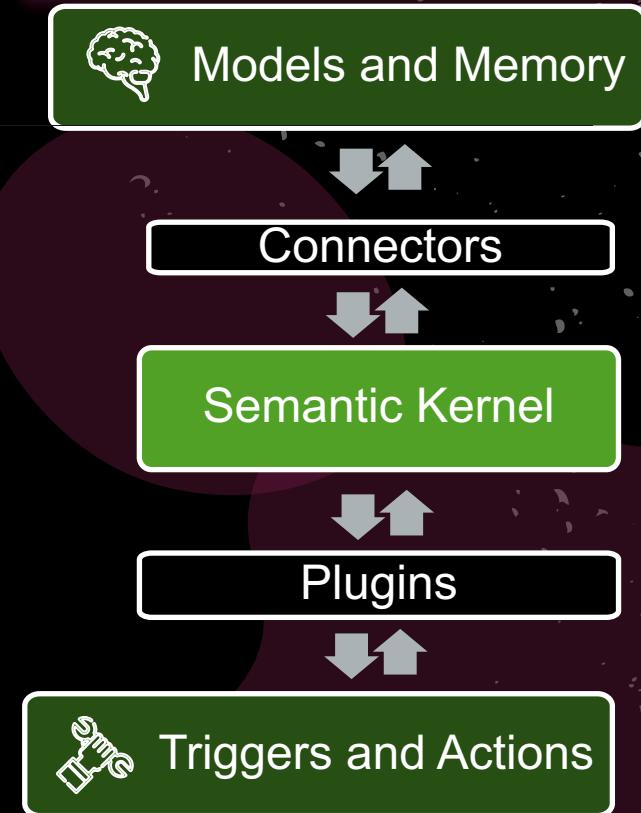
SEMANTIC KERNEL

- Lightweight SDK – mix Code with LLMs
- Supports OpenAI, Azure OpenAI, and Hugging Face
- Languages: C#, Python, and Java
- Templating, chaining, and planning capabilities out-of-the-box.
- Define plugins then chained together or automatically orchestrate plugins with AI.



SEMANTIC KERNEL

Concept	Description
Kernel	Orchestrates a user's ASK expressed as a goal
Planner	Breaks it down into steps based upon resources available
Plugins	Customizable resources - built from LLM AI prompts and native code
Memories	Customizable resources - manage contextual information
Connectors	Customizable resources - enable external data access



SEMANTIC KERNEL



SEMANTIC COMPLETION

Prompt Engineering

Estimate the next best
completion of the prompt



SEMANTIC SIMILARITY

Context

Search and Measure the distance
between the semantic meanings of
words, phrases, sentences, or
documents

04

AI AGENTS and COPILOTS

Plugins, Planners, Personas



SPECTRUM OF AGENTS

SIMPLE to FULLY AUTOMATED



CHATBOT

Simple Back and Forth chats with user



RAG

Retrieval Augmented Generation for conversation grounded in real data



COPILOT

Agents that work side-by-side with user to complete task



FULLY AUTONOMOUS

Agents that can respond to stimuli with minimal human intervention

CORE BUILDING BLOCKS OF AN AGENT



Plugins



Planners

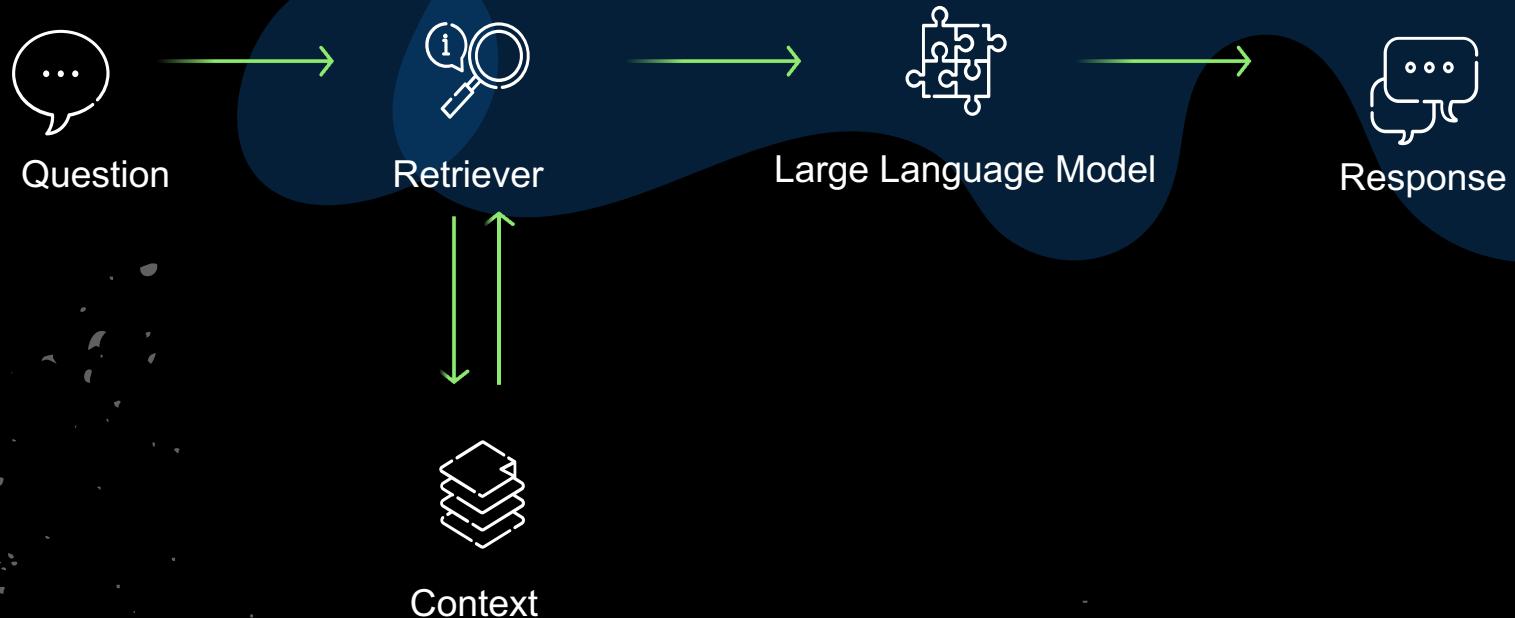


Personas

giving your
agent skills

giving guidance to
your agent

giving your agent
a personality



Retrieval Augmented Generation

How about a little cocktail magic?

<https://github.com/rondagdag/crafting-mixologist-semantic-kernel/>



| Do not humanize AI!

Generative AI

is not:

Intelligent

Deterministic

Grounded

Generative AI

can not:

Understand language or maths

Understand manners or
emotions

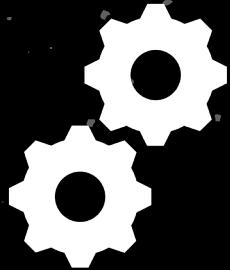
Know facts that are not in its
training dataset



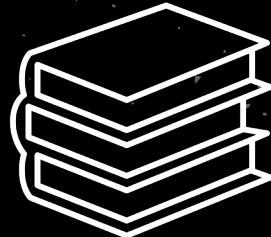
Image created by DALL E2

Always

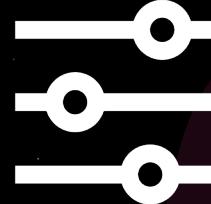
I How to mitigate hallucinations



Effective prompt
engineering



- Provide validated data sources
- Ask for citations



- Tweak parameters (temperature/max length)
- Content filtering



Schillace Laws

Relationships Between Code And Models



Don't write code if the model can do it; the model will get better, but the code won't.



Code is for syntax and process; models are for semantics and intent.



Text is the universal wire protocol.

Schillace Laws

Trade-offs And Challenges When Using LLMs



Trade leverage for precision; use interaction to mitigate.



Uncertainty is an exception throw.



The system will be as brittle as its most brittle part.



Hard for you is hard for the model.

Schillace Laws

Trade-offs And Challenges When Using LLMs



Ask Smart to Get
Smart



Beware “pareidolia of
consciousness”;
the model can be used
against itself.

"Mixing a cocktail is like programming AI: the right combination of ingredients leads to something extraordinary, but one wrong element and you've got a recipe for disaster"

-ChatGPT

"AI Responsibly, Like You
Drink Responsibly."



FINAL WORDS

Generative AI

flexible, reusable models

Semantic Kernel

Connectors and Plugins



Large Language Models

Chat Completion and Tokens

AI Agents and Copilots

Spectrum from Simple to Fully Autonomous
Copilot work side-by-side w/ user

THANKS!

Do you have any questions?

ron@dagdag.net

[in/rondagdag/](https://in/rondagdag)

[@rondagdag](https://twitter.com/rondagdag)



CREDITS: This presentation template was created by [**Slidesgo**](#), including icons by [**Flaticon**](#), infographics & images by [**Freepik**](#)

