Department of Industrial & Systems Engineering

Agentic AI for Smart Mobility

Smart mobility encompasses a broad spectrum of systems, services, and business models that collectively aim to address diverse end-user needs while advancing the broader goals of society and environmental sustainability. It integrates advanced technologies such as softwarization, artificial intelligence, connectivity, and electrification, together with innovative economic paradigms including the digital economy, servitization, sharing economy, gig economy, experience economy, and circular economy. The emergence of agentic AI represents a paradigm shift in how intelligent systems perceive, reason, and act within dynamic, real-world environments. Powered by large reasoning models, structured memory, and integrated toolkits, AI agents are increasingly capable of multi-step reasoning and collaborative behavior, creating transformative opportunities across multiple domains, including smart mobility. This talk highlights potential applications of agentic AI in smart mobility systems and services, such as personalized trip planning, frictionless parking, seamless Umrah experiences, automated driving, adaptive traffic management, road safety, and last-mile delivery. These scenarios illustrate how agentic AI can enable context-aware, personalized, and efficient mobility solutions.

05 November 2025 **2:00 PM** **Building 22, Room 130**

Dr. Alaa Khamis

Dr. Alaa Khamis is an Associate Professor in the Department of Industrial and Systems Engineering and Director of the AI for Smart Mobility Lab at the Interdisciplinary Research Center for Smart Mobility and Logistics, King Fahd University of Petroleum and Minerals (KFUPM). Before joining KFUPM, he was the AI and Smart Mobility Technical Leader at General Motors. He also serves as an Adjunct Professor at the University of Toronto and Ontario Tech University. Dr. Khamis has authored three books and over 190 scientific papers in refereed journals and conferences, and holds 72 U.S. patents, trade secrets, and defensive publications. He is the author of Smart Mobility: Exploring Foundational Technologies and Wider Impacts and Optimization Algorithms: AI Techniques for Design, Planning, and Control Problems. His research focuses on the intersection of AI and mobility systems, services, and business models, addressing challenges such as seamless integrated mobility, contextual observability in software-defined vehicles (SDVs), and optimization in mobility, logistics, and infrastructure. He is the recipient of the 2018 IEEE Member and Geographic Activities (MGA) Achievement Award, the Best Paper Award at the 2023 IEEE International Conference on Smart Mobility, the 2022–2024 GM Critical Talent Award, and first place in the 2025 Sustainable Solutions for Pilgrims Challenge.

King Fahd University of Petroleum and Minerals
ISE Department & IRC for Smart Mobility and Logistics

Agentic AI for Smart Mobility

Alaa Khamis, Ph.D

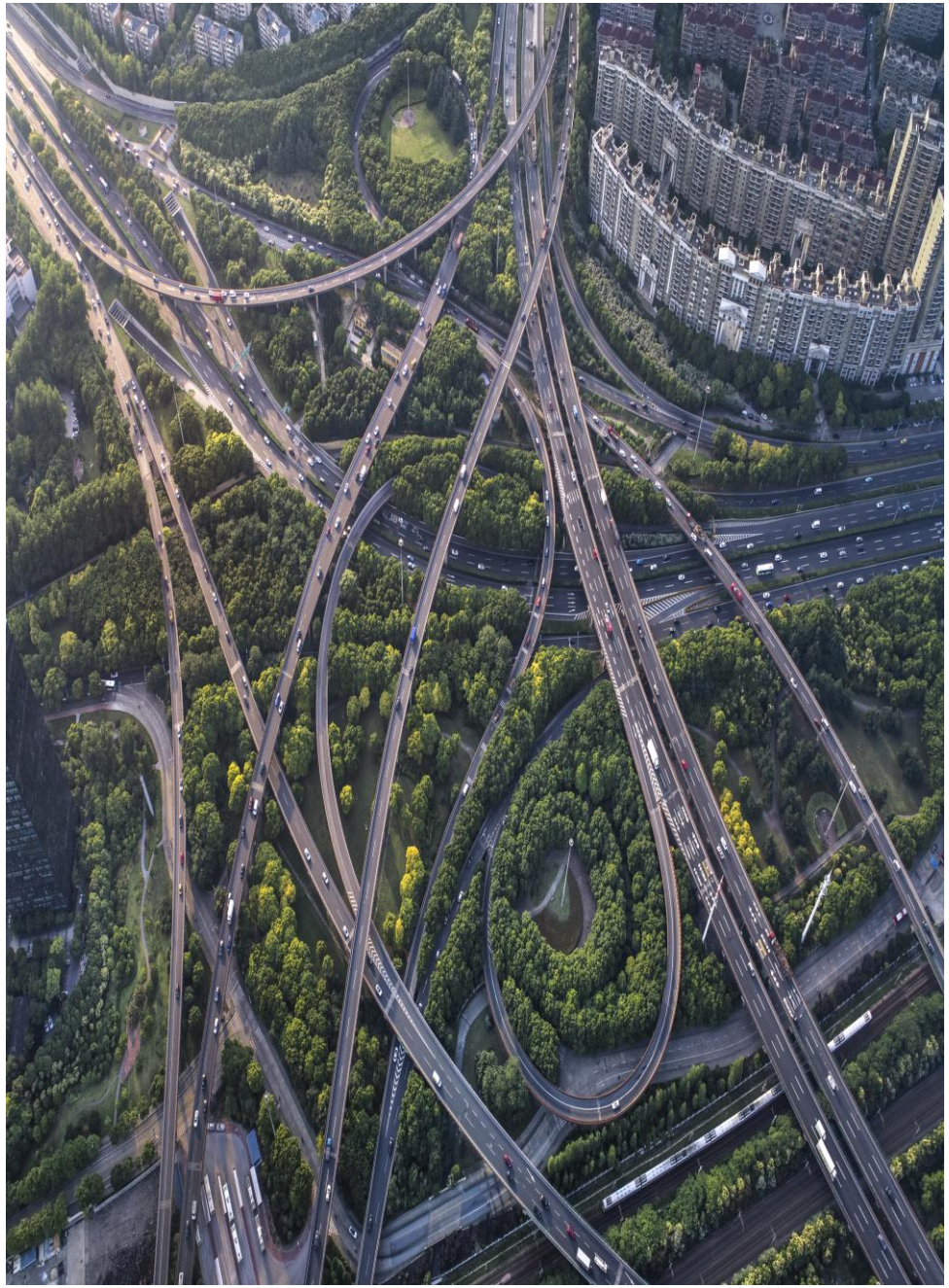
Associate Professor in ISE Department

Director of AI4SM Lab at IRC SML

November 5, 2025

Outline

- About AI for Smart Mobility Lab
- Motivating Scenarios
- What is Agentic AI?
- AI Agent Components
- Use Cases



AI for Smart Mobility Lab at KFUPM

Mission

- Our mission is to advance **smart mobility** as a transformative enabler of **sustainable development**.
- Our research focuses on the **intersection of AI and mobility systems, services and business models**.

Smart Mobility

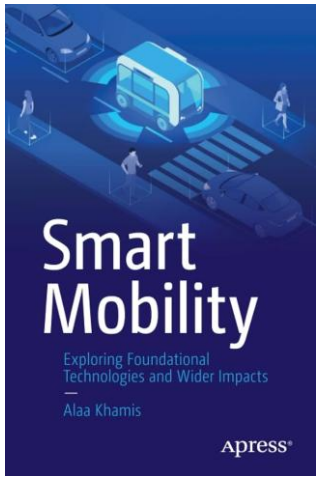
Existing and emerging smart mobility business models

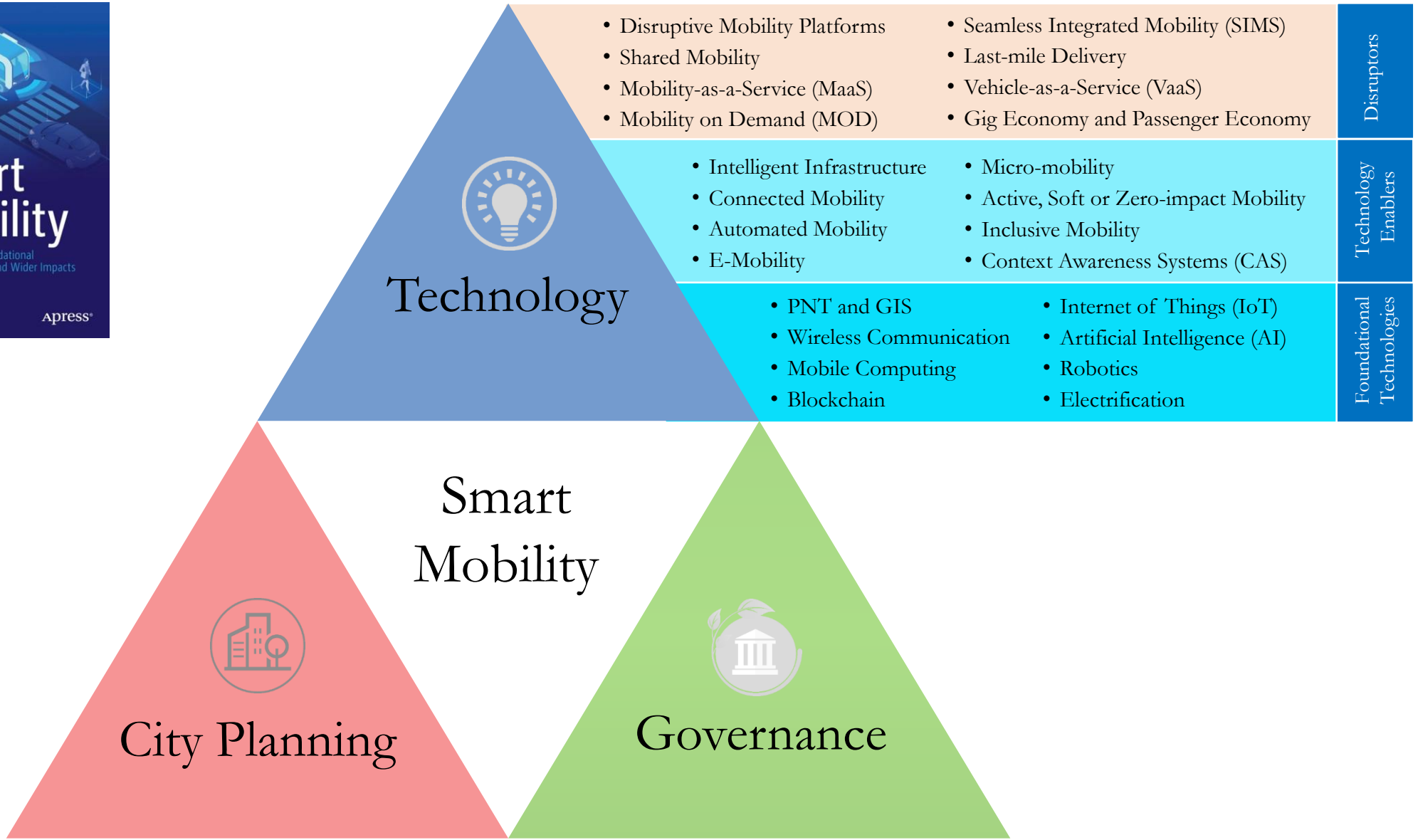
Existing and emerging smart mobility services

Existing and emerging smart mobility systems

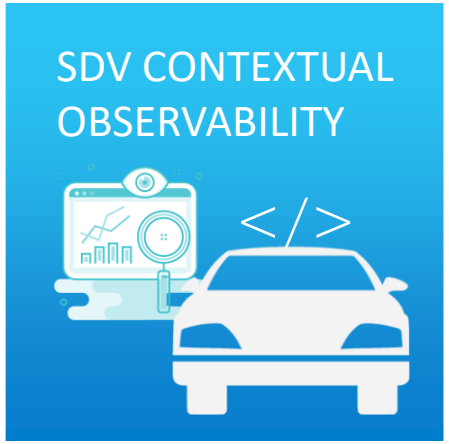
Alaa Khamis and Suzette Malek, "Smart Mobility for Sustainable Development Goals: Enablers and Barriers," The IEEE International Conference on Smart Mobility, 2023.

Smart Mobility





Ongoing Projects



- **Title:** Contextual Observability of Software-Defined Vehicles
- **Objective:** Develop a testbed for software-defined vehicle (SDV) contextual observability.
- **Collaboration:** IRC SML and auto OEMs, NGOs and Suppliers

- **Title:** Agentic AI-based Framework for SIM
- **Objective:** Develop as a unified platform that integrates multimodal transportation options.
- **Collaboration:** RCRC, MIT, VTTI

- **Title:** SmartDispatch: AI-driven Optimization for Eco-Efficient Last-Mile Delivery
- **Objective:** Develop an AI-driven routing model for eco-efficient last-mile delivery.
- **Collaboration:** IRC SML KFUPM

- **Title:** Enabling Cybersecurity Adaptation in Software Architecture
  Objective: What-if architecture analysis of existing software systems with poor or no records of architectural decisions.
- **Collaboration:** UAB, Chile

Collaboration & Facilities

الهيئة الملكية لمدينة الرياض
ROYAL COMMISSION FOR RIYADH CITY

Massachusetts Institute of Technology

JTL Transit Lab
URBAN MOBILITY LAB AT MIT

VIRGINIA TECH TRANSPORTATION INSTITUTE

UNIVERSITY OF TORONTO

OntarioTech UNIVERSITY

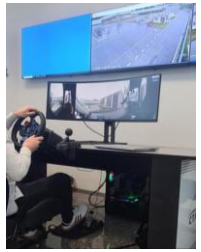
Universidad Andrés Bello

gm

FOXCONN

MAXNERVA

CTAG



For More information

https://www.ai4sm.org/

https://github.com/ai4smlab

https://www.youtube.com/@AI4SM_lab

https://medium.com/ai4sm

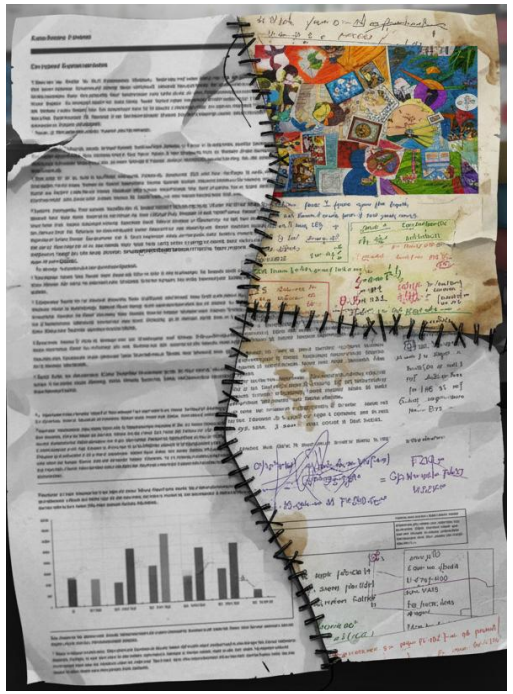
Motivating Scenarios

Motivating Scenarios

Researcher

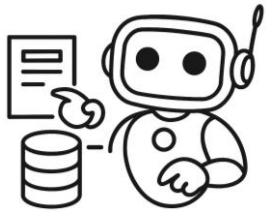
User query: Write a systematic review paper about smart mobility

Response: Frankenstein Paper

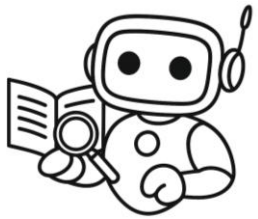


Motivating Scenarios

- Specify databases/search engines
- Keywords
- Inclusion and exclusion criteria

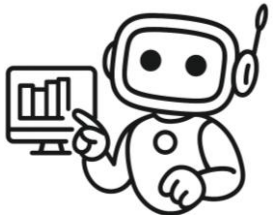


Data Collection Agent



- Highlight strengths and limitation
- Synthesize and rank findings by relevance, recency, authority
- Create taxonomies
- Identify gaps and future directions

Research Agent



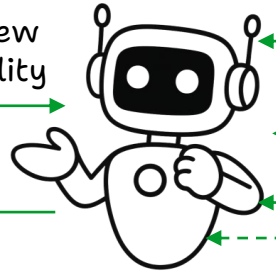
- Creates visual aids

Graphic Design Agent

Write a systematic review paper about smart mobility

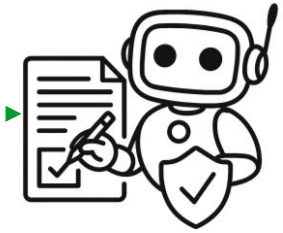
Agent response

Researcher



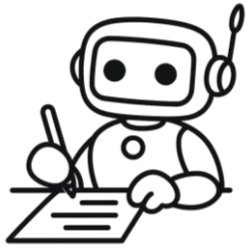
Interaction/ Orchestration Agent

Human Reviewer/ Prof-reader



- Review coherence. coverage., and citation completeness
- Perform or request fixes from the writer

Editing Agent



- Draft structured outline to be reviewed by the editor
- Generate the final comprehensive paper with inline citations and complete reference section

Writing Agent

Motivating Scenarios

- Got an AI Assistant?

Google

Project Astra

Motivating Scenarios

- Got an AI Assistant?

FIGURE

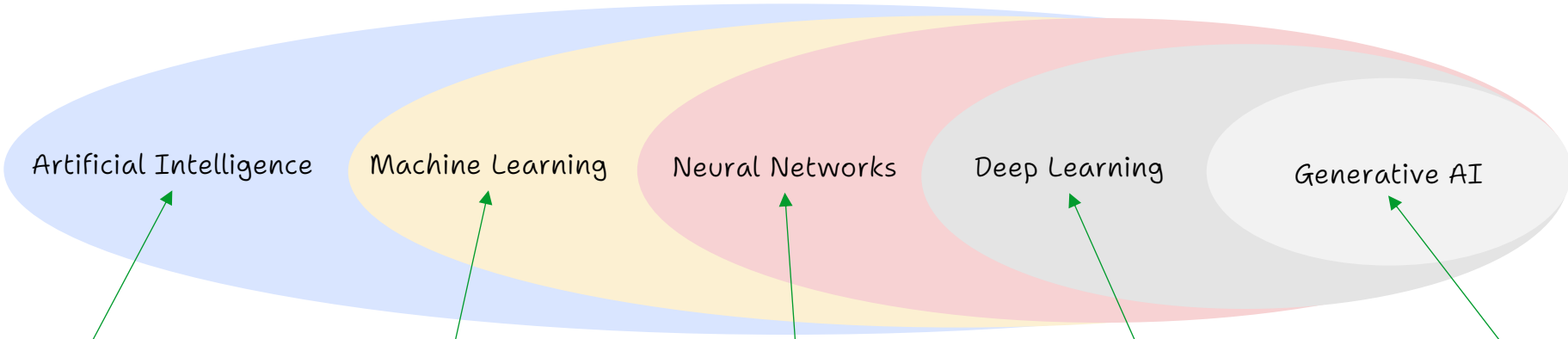
NVIDIA

Microsoft

Figure Helix: https://www.youtube.com/watch?v=Z3yQHYNXPws

What is Agentic AI?

The Alphabet Soup of AI



Any techniques that enables computers to mimic and/or augment biological intelligence covering perception, automation, reasoning, and decision-making

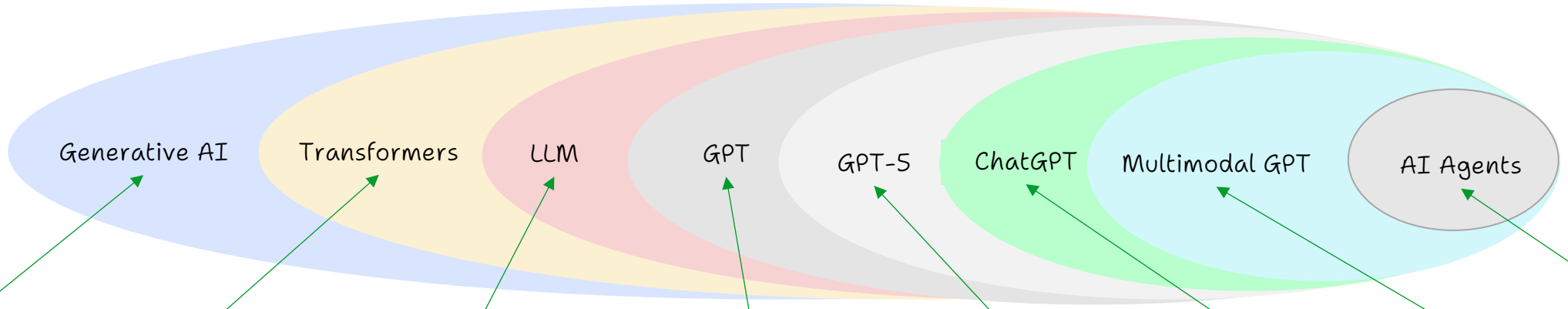
The ability to learn from experiences and observations without explicit programming

A highly simplified computational model of the biological brain designed for pattern recognition and feature extraction

DL enables learning representations of data with multiple levels of abstraction using deep neural networks. Examples: Convolutional neural network (CNN), Long short-term memory (LSTM), Autoencoders and Transformers

Generative AI focuses on models that can generate new content, like images, text, or audio. Examples, Generative Adversarial Networks (GANs) and Transformers.

The Alphabet Soup of AI



Generative AI focuses on models that can generate new content, like images, text, or audio. Examples, Generative Adversarial Networks (GANs) and Transformers.

An example of a technology used in generative AI. It is a revolutionary deep learning architecture introduced by Google in 2017 that allows models to understand and generate language efficiently.

Large Language Models (LLMs) focus on processing and generating human-like text. Examples: Bidirectional Encoder Representations from Transformers (BERT) and Generative Pre-trained Transformer (GPT).

Generative Pre-trained Transformers (GPT) is a specific subset of Generative AI that uses transformers for text generation.

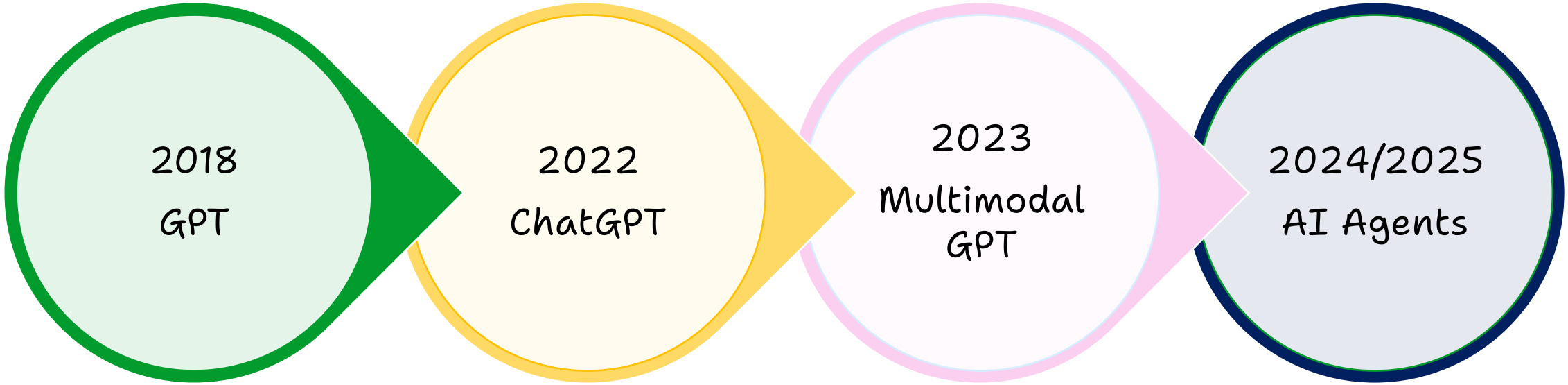
An advanced LLMs, built on transformer architecture, trained on vast datasets to generate human-like responses.

A specific application of GPT-5, optimized for conversational AI and interactive use.

A version of the GPT model, like GPT-4o, that processes and generates both text and visual inputs, functioning as a Vision Large Model (VLM).

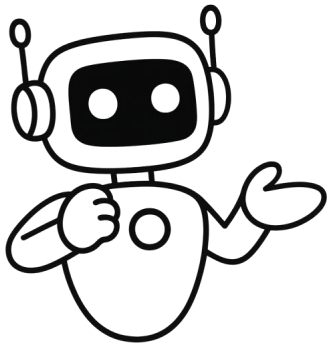
Intelligent connection between reasoning and action

The Alphabet Soup of AI



What is AI Agent?

» **LLM/LRM-based AI Agents**



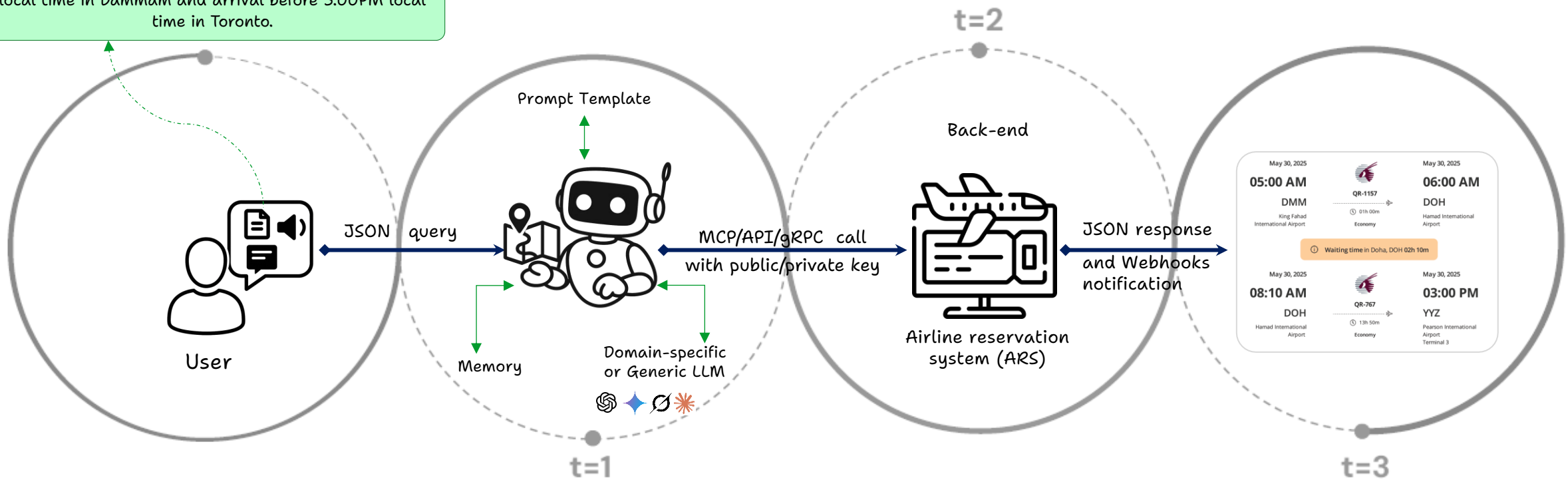
1. **Goal-oriented**
2. **Autonomous**
3. **Connects reasoning and action**
4. **Achieve contextually intelligent outcomes.**

Alaa Khamis, "Agentic AI Systems: Architecture and Evaluation Using a Frictionless Parking Scenario," IEEE Access, Vol. 13, 2025.

What is AI Agent?

» Agentic Workflow

Could you please help me book a flight from Dammam to Toronto on Friday, May 30, 2025? I'd prefer a flight with no more than one stop, and I'd like to keep the total travel time as reasonable as possible. If there are multiple options, please prioritize the one with the shortest layover and a departure time after 5:00 AM local time in Dammam and arrival before 5:00PM local time in Toronto.

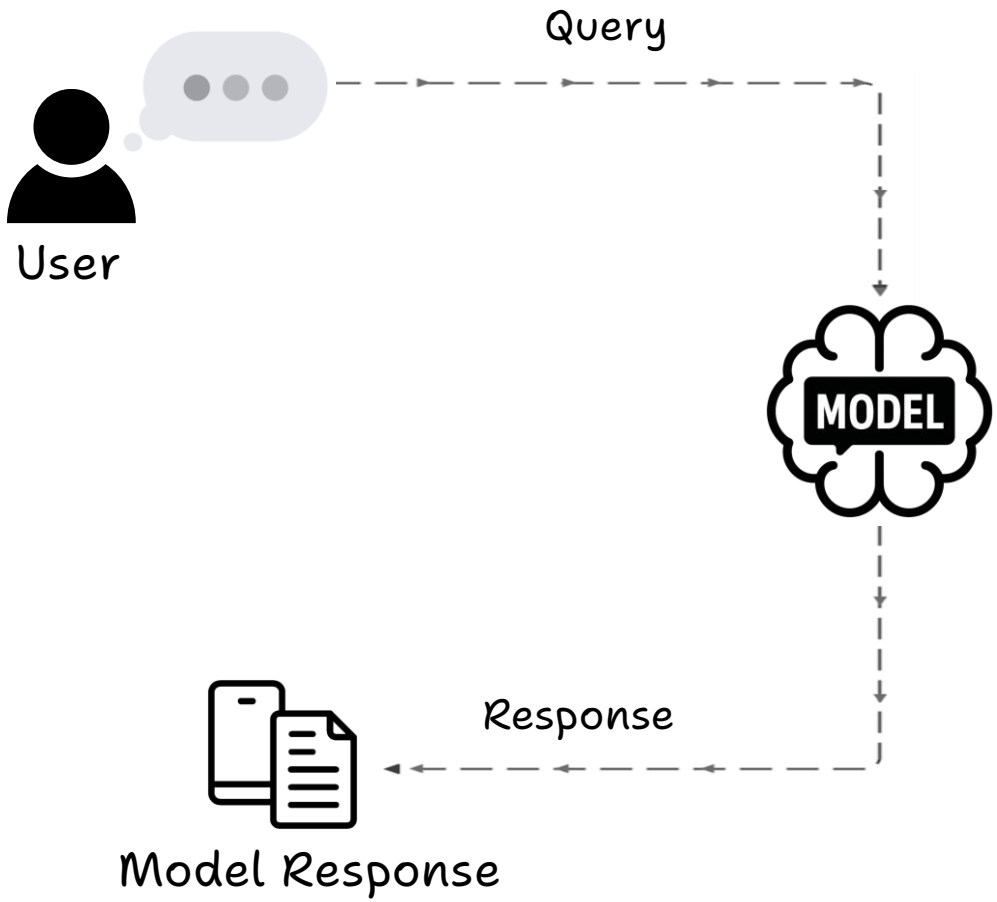


Natural Language Input → Reason + Plan + Action → Tool-use → Reservation

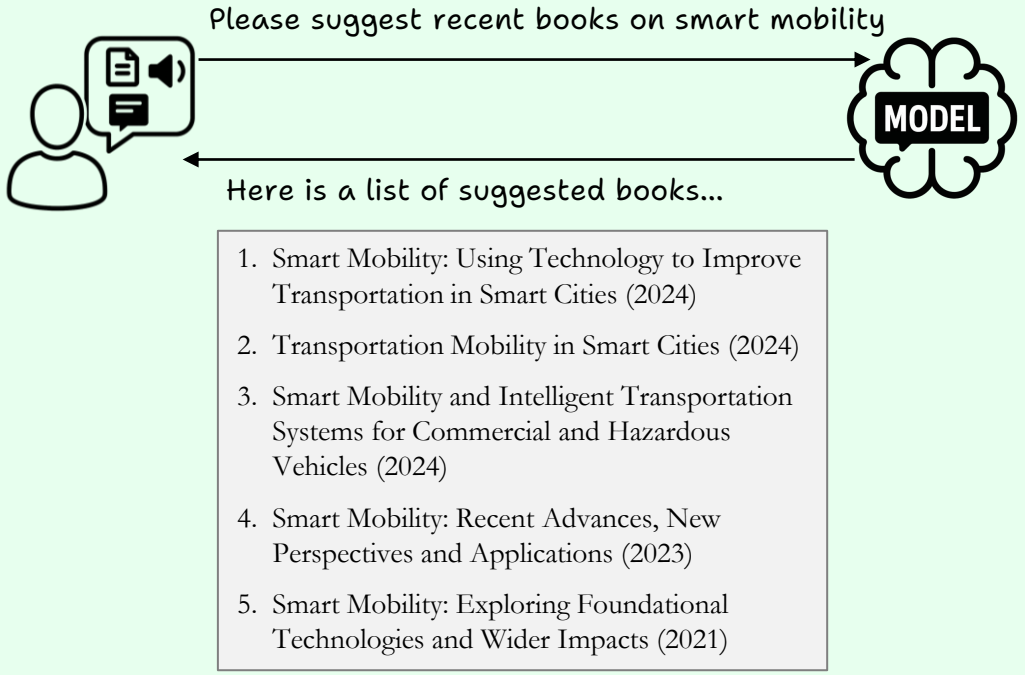
What is AI Agent?

» Design Patterns: Basic Responder

Agency Level
☆☆☆

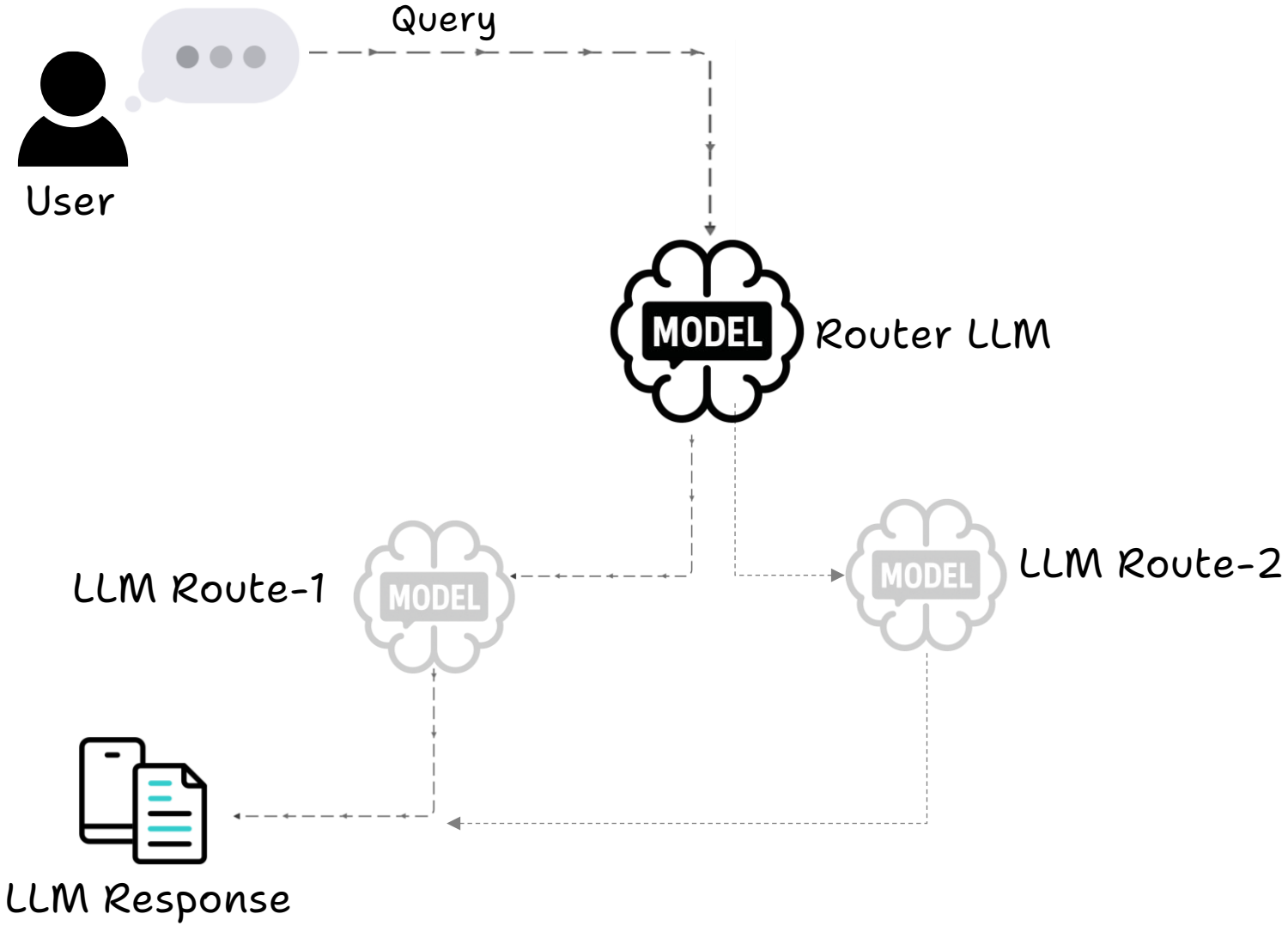


Example



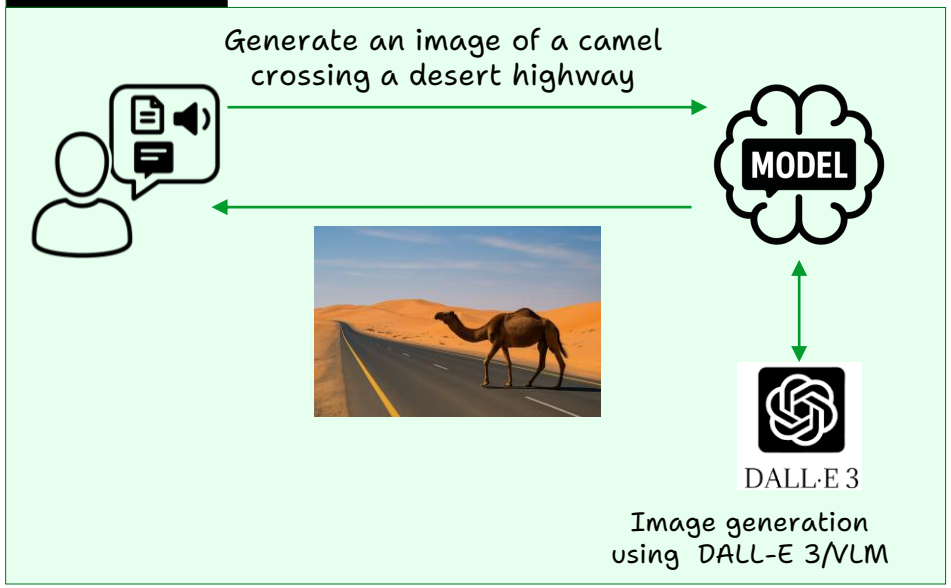
What is AI Agent?

» Design Patterns: Router Pattern



Agency Level
★☆☆

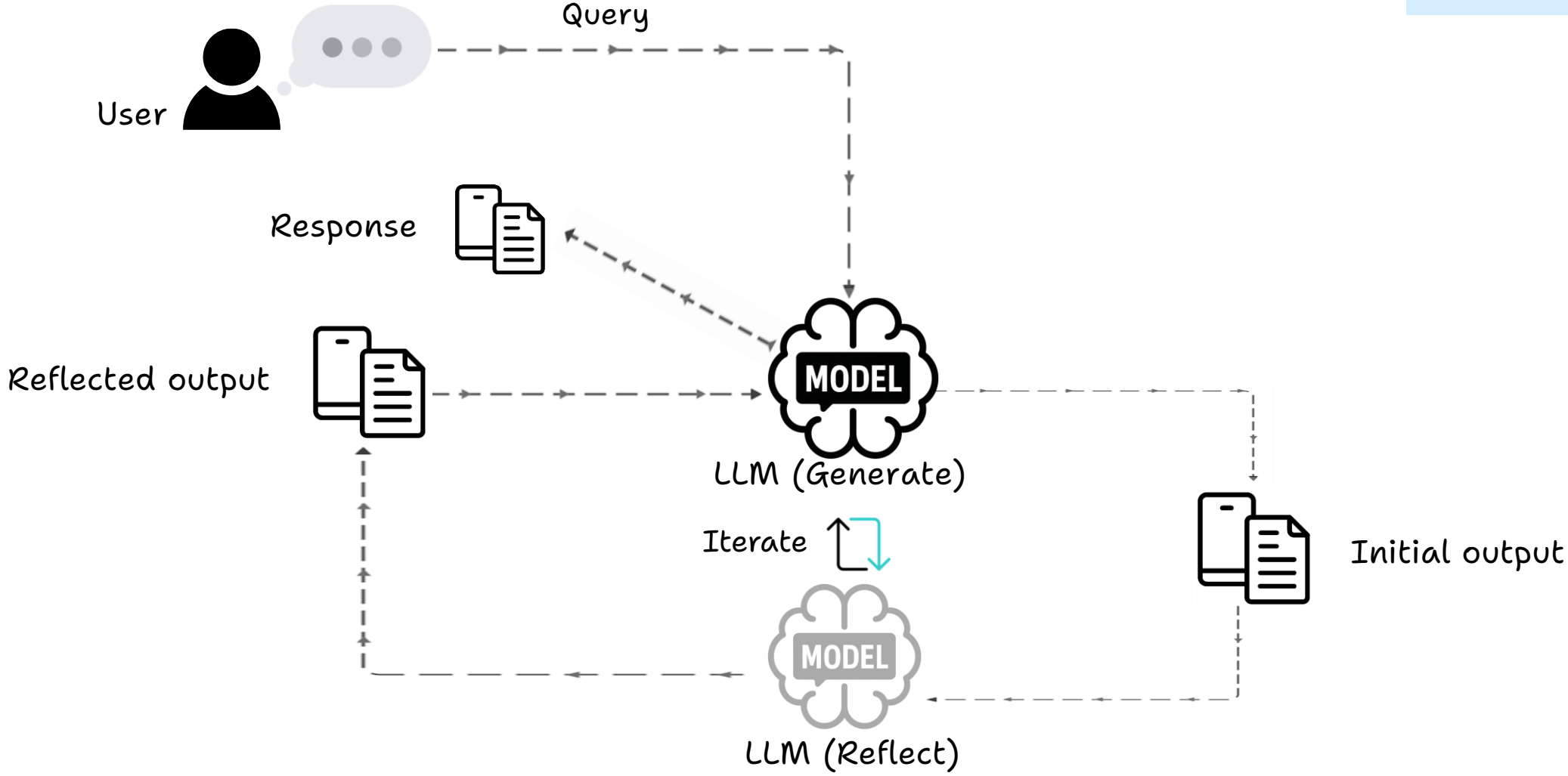
Example



What is AI Agent?

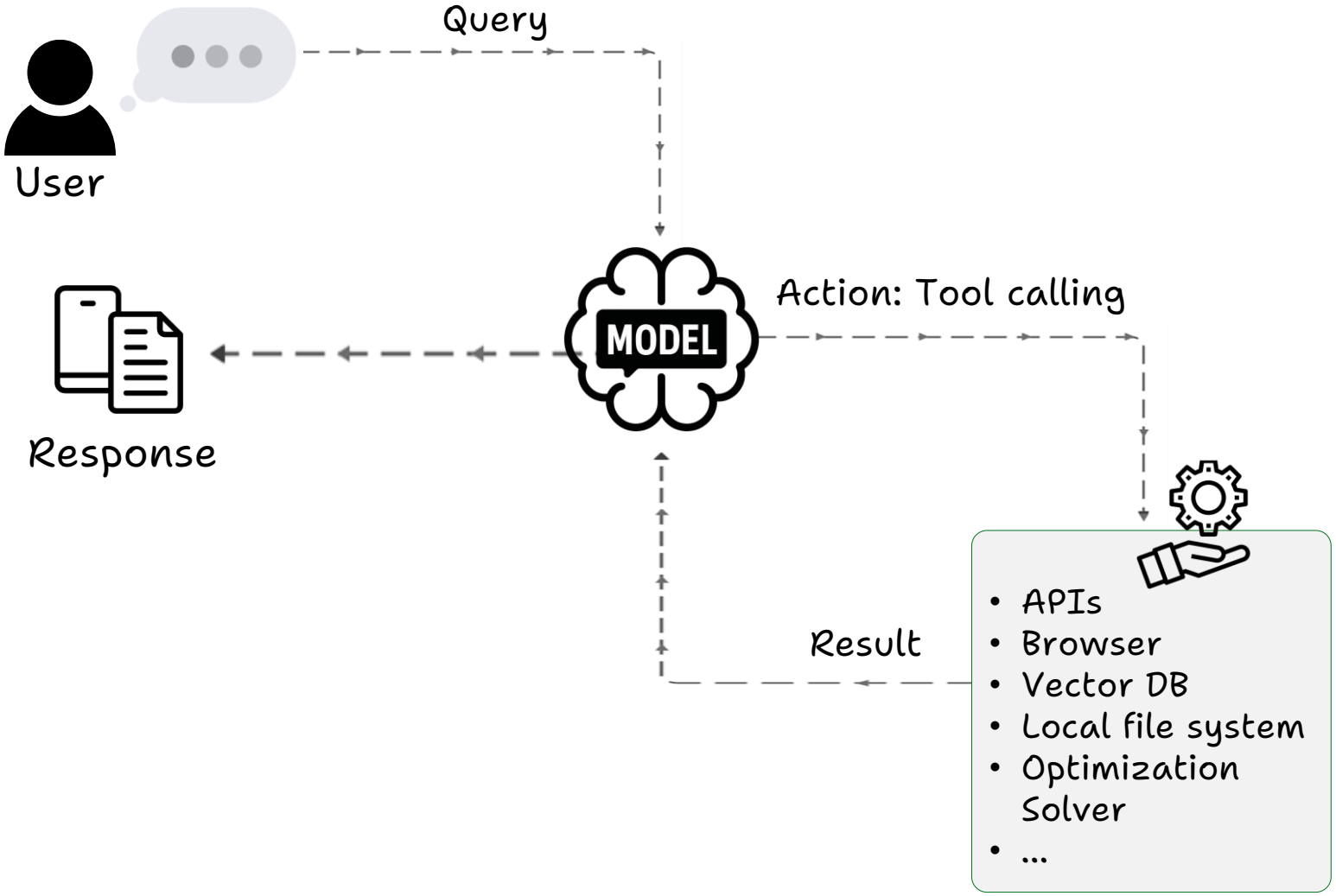
» Design Patterns: Reflection Pattern

Agency Level
★★☆



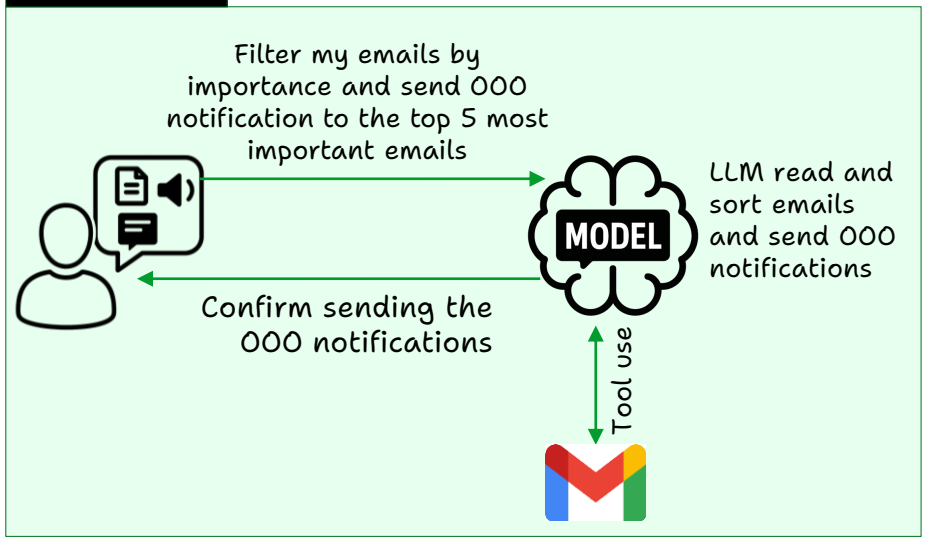
What is AI Agent?

» Design Patterns: Tool Calling



Agency Level
★★☆

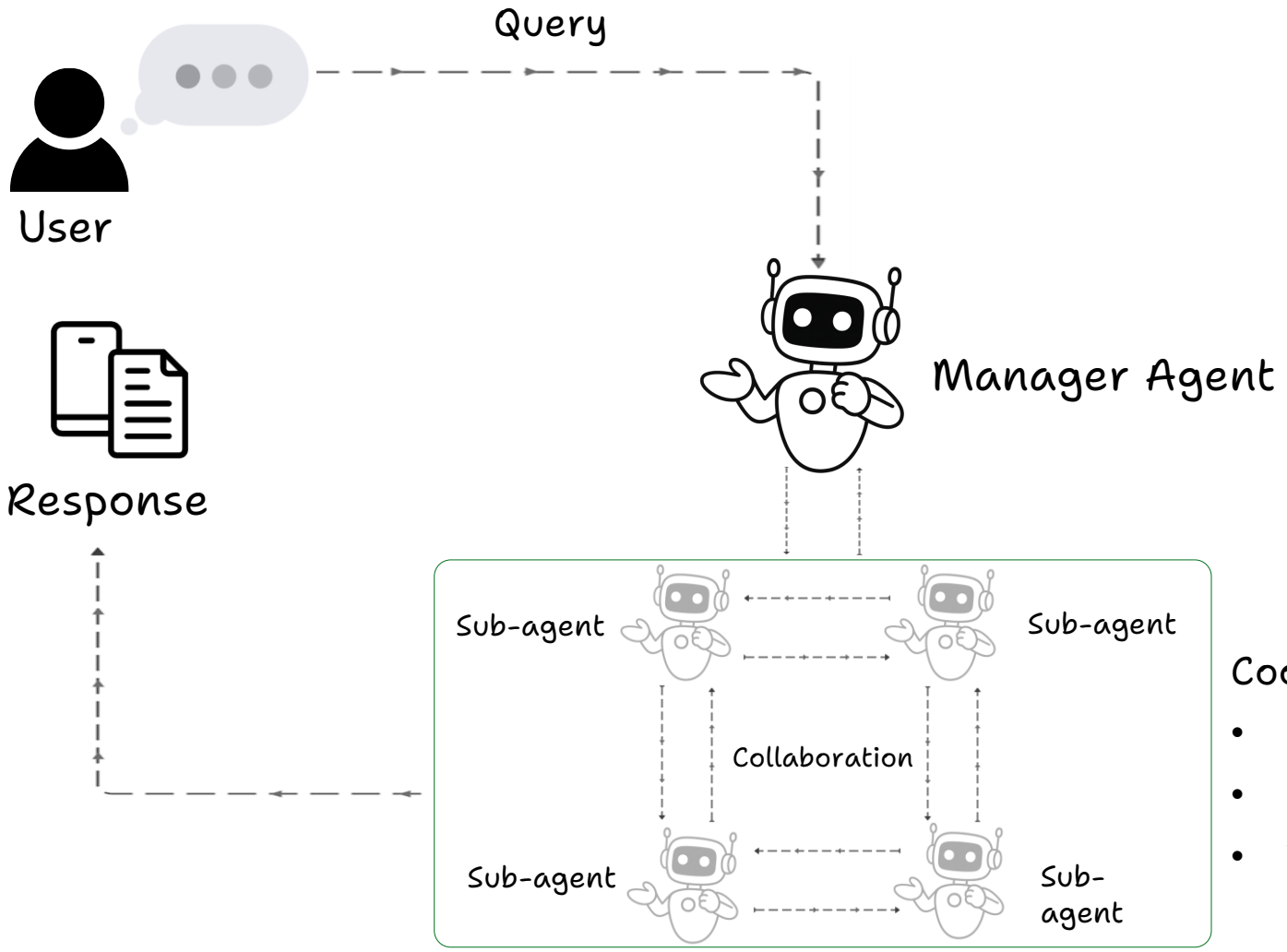
Example



What is AI Agent?

» Design Pattern: Multi-agent Pattern

Agency Level
★★★



Cooperation Patterns:

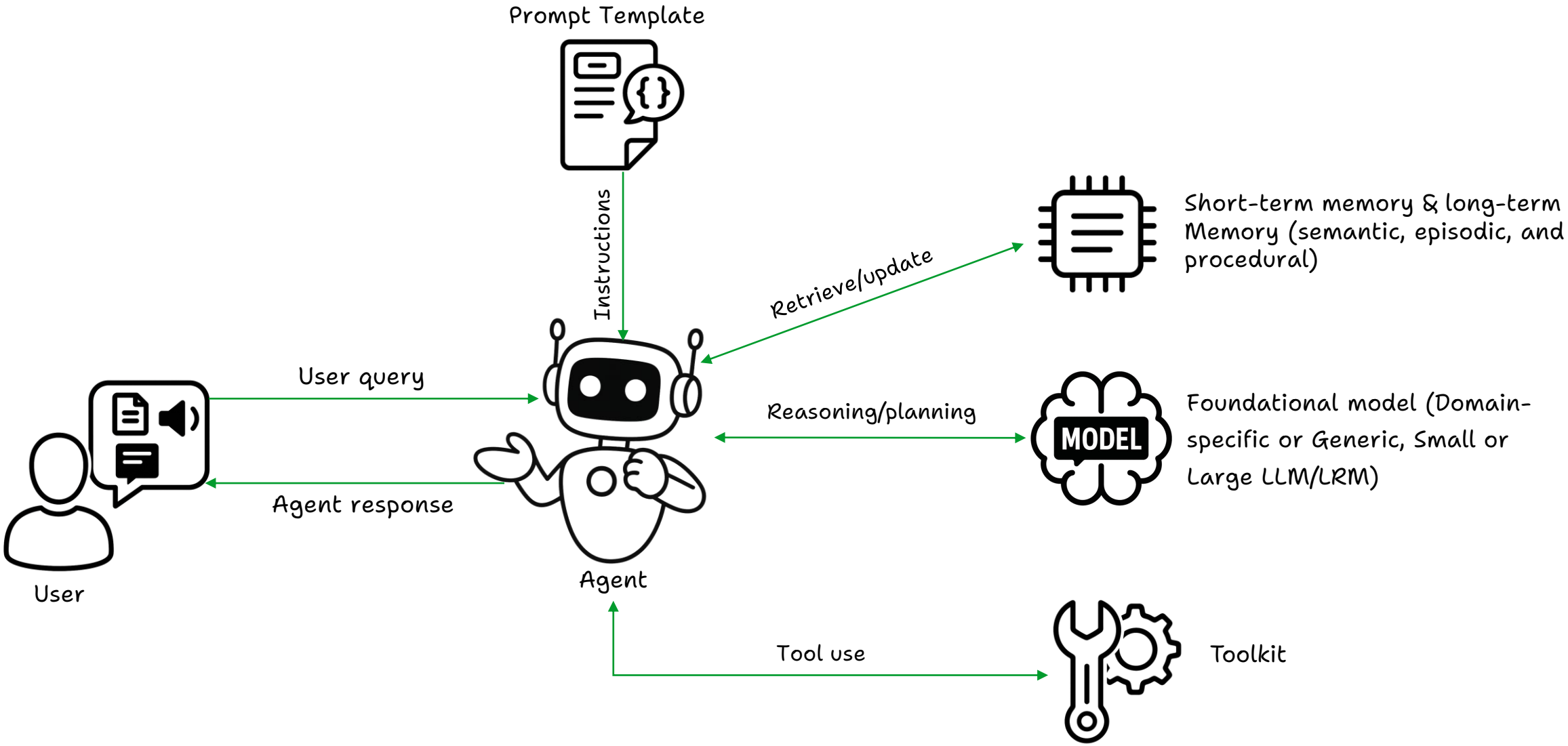
- Augmentative Cooperation
- Integrative Cooperation
- Debative Cooperation

Alaa Khamis, "Agentic AI Systems: Architecture and Evaluation Using a Frictionless Parking Scenario," IEEE Access, Vol. 13, 2025.

AI Agent Components

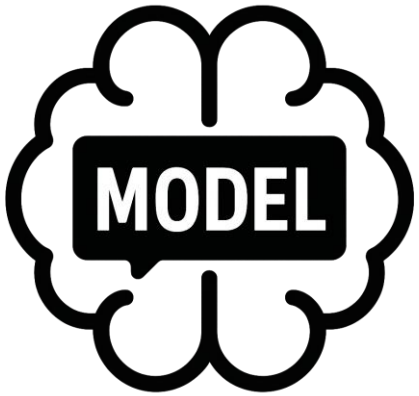
What is AI Agent?

» AI Agent Components

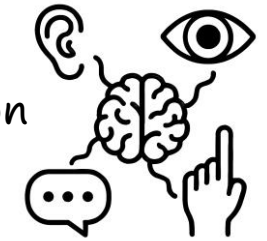


Foundational Models

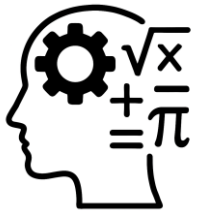
Foundational Model



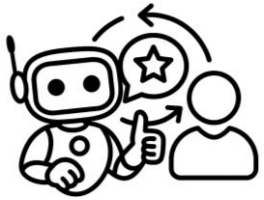
Multimodal Perception



Reasoning



Learning



Foundational Models

Scale	Type	Modality	Use Case Focus	Example Models
Large Language Models (LLMs) 30B – 400B+ parameters)	Reasoning / General LLM	Multimodal (or evolving)	Advanced reasoning, planning, multimodal agents	GPT-5, GPT-4 (OpenAI), Claude 3 (Anthropic), Gemini Pro (Google), Llama 3 70B (Meta)
Small Language Models (SLMs) 1B – 13B parameters)	Language / lightweight reasoning	Unimodal or limited multimodal	Fast inference, fine-tuning, domain tasks	Llama 3 8B / 3B (Meta), Mistral 7B (Mistral), Phi-3 small (Microsoft), Gemma 2 (Google), Qwen-7B (Alibaba)

Foundational Models

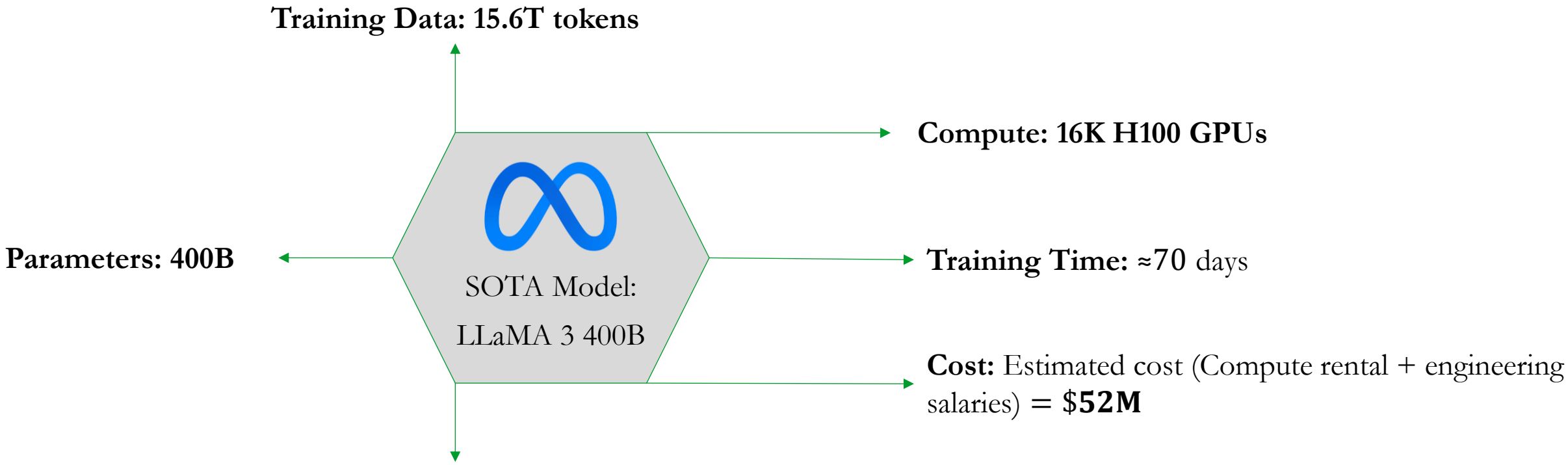
» General LLM Training Pipeline

Aspect	Pretraining	Classic post-training/RLHF	Reasoning RL
Objective	Predict next word on internet	Maximizing user utility and preferences	Think on questions with objective answers
Data	>10T tokens	~100K problems	~1M problems
Time	months	days	weeks
Compute cost	>$10M	>$100K	>$1M
Bottleneck	data & compute	data & evaluation	RL env & hack
Examples	LLaMA 3	LLaMA-instruct	DeepSeek R1

Credit: Yann Dubois, OpenAI

Foundational Models

» General LLM Training Pipeline



Carbon Emissions: ≈ 4400 tCO_2e. **Equivalent to:**

- 🚘 Driving **~17.5 million kilometers** in a typical gasoline car in Saudi Arabia, or
- 🚗 Annual emissions from **~950 passenger vehicles**, or
- ✈️ Taking **~2,750 round-trip flights** between Riyadh and London, or
- 🏠 Powering **~600 Saudi households** for one year

Foundational Models

» LLM Specializing Pipeline

Aspect	Prompting	Finetuning
Objective	Art of asking the model what you want	Second stage of post-training to domain specific data
Data	0	~10-100K problems
Time	hours	days
Compute cost	0	~$10-100K
Bottleneck	evals	data & evals

Credit: Yann Dubois, OpenAI

Agentic Context Engineering: Evolving Contexts for Self-Improving Language Models

Qizheng Zhang [1*] Changran Hu [2*] Shubhangi Upasani [2] Boyuan Ma [2] Fenglu Hong [2]
Vamsidhar Kamanuru [2] Jay Rainton [2] Chen Wu [2] Mengmeng Ji [2] Hanchen Li [3]
Urmish Thakker [2] James Zou [1] Kunle Olukotun [1]

[1] Stanford University [2] SambaNova Systems, Inc. [3] UC Berkeley * equal contribution

qizhengz@stanford.edu, changran.hu@sambanovasystems.com

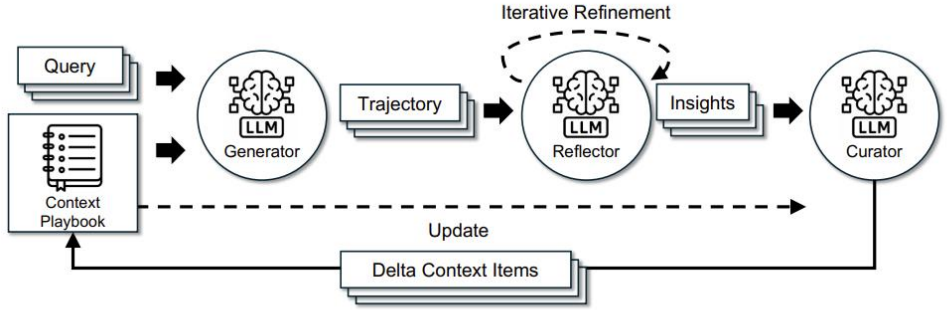


Figure 4: **The ACE Framework.** Inspired by Dynamic Cheatsheet, ACE adopts an agentic architecture with three specialized components: a Generator, a Reflector, and a Curator.

Foundational Models

» Reasoning

System 1: Fast, Intuitive

System 2: Slow, Deliberate

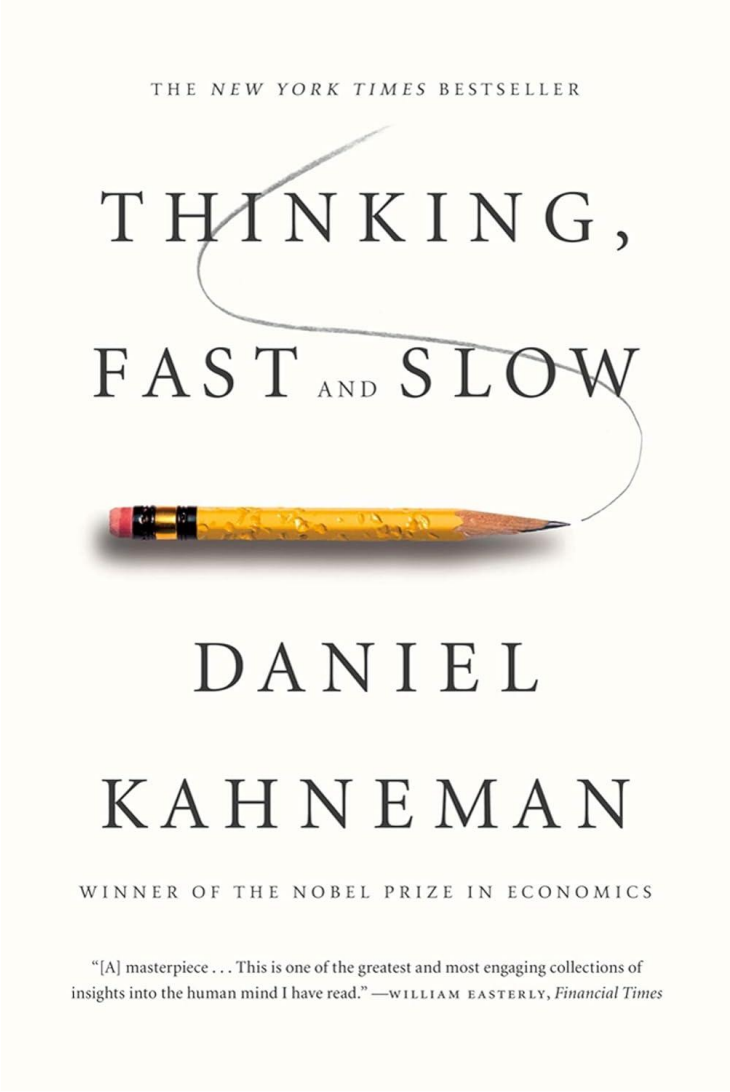
Question: What is 48 ÷ 4?

System 1: 48 ÷ 4 = 12

System 2: Let me check carefully:

- 4 goes into 40 ten times, remainder 8.
- 4 goes into 8 two times.
- 10 + 2 = 12
- So, 48 ÷ 4 = 12.

Same output, different approach to problem-solving



Foundational Models

» Reasoning

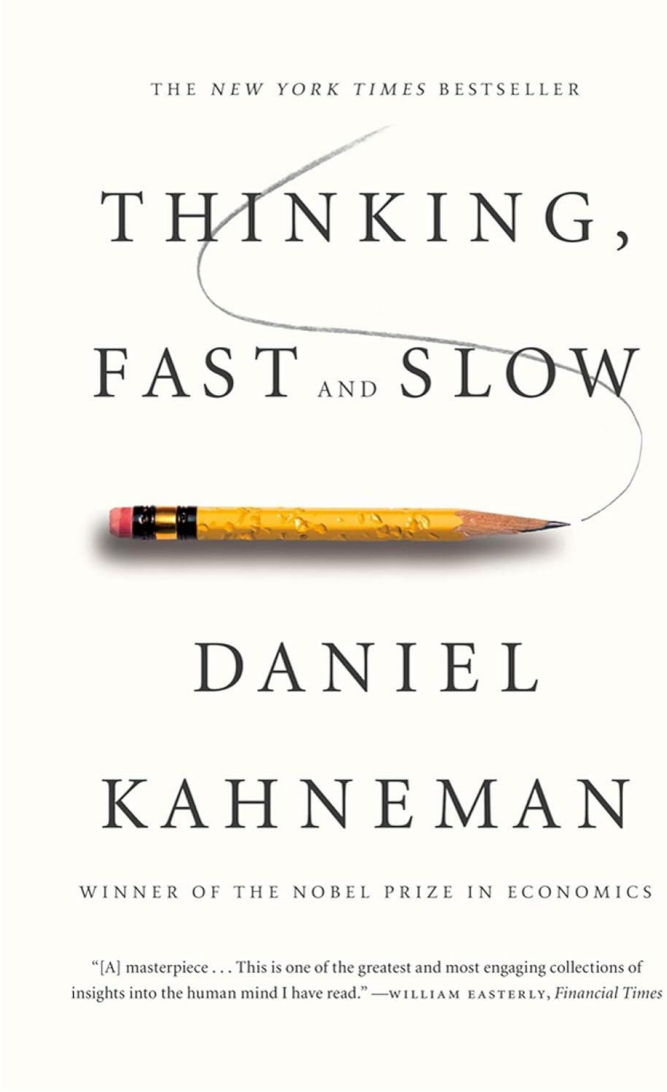
System 1: Fast, Intuitive

Text generation via next-word prediction

System 2: Slow, Deliberate

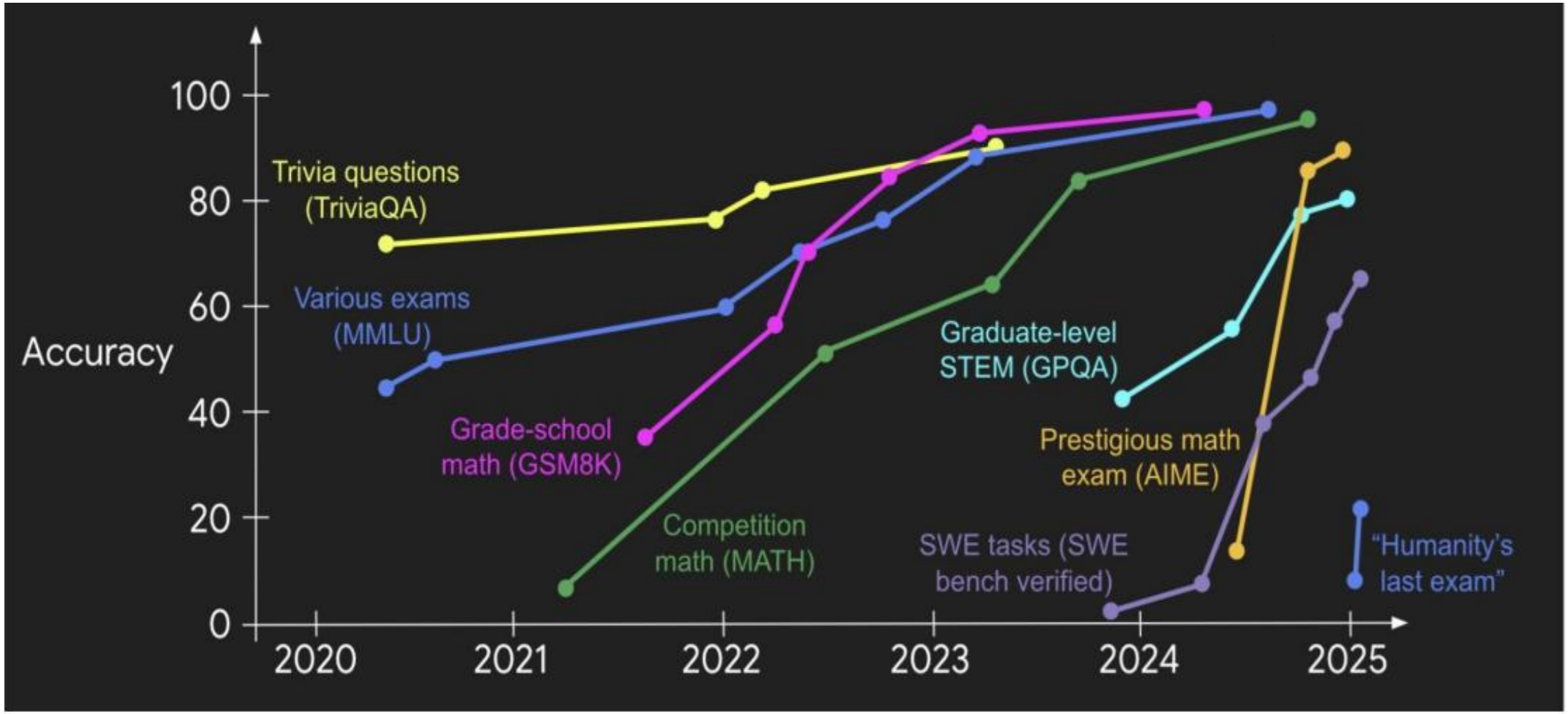
Multi-step Problem Solving

ReAct, Chain-of-Thought (CoT), Chain of Continuous Thought (Coconut), Tree-of-Thoughts (ToT), Distilling Reasoning



Foundational Models

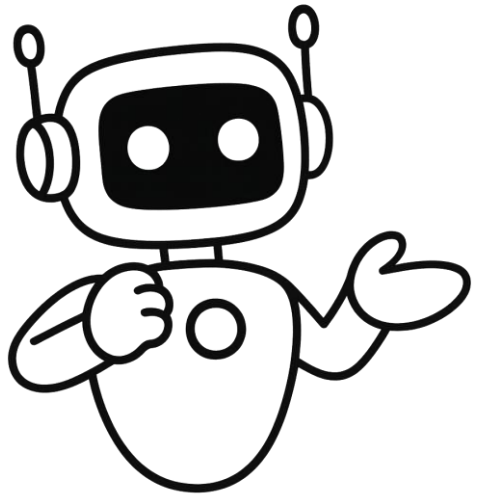
» **Reasoning:Progress on AI benchmarks in the past five years**



AI News Daily, Feb. 17, 2025

Prompt Template

» Profile and Persona

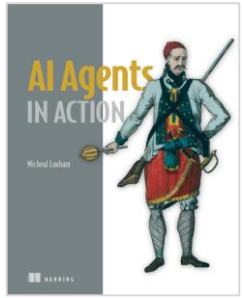


Profile Contents

- **Personal:** Role, i.e., coding assistant, trip planner, logistics dispatcher, etc.
- **Demographics:** Gender, age, background

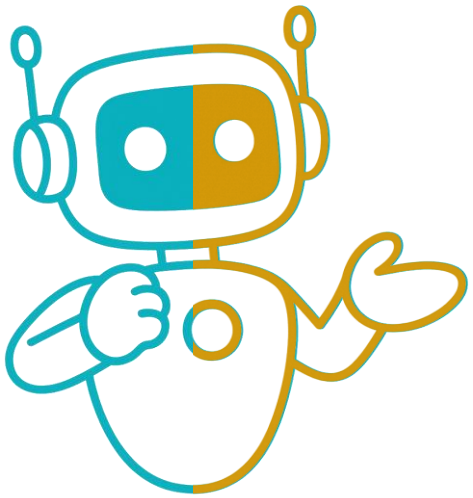
Profile Generation

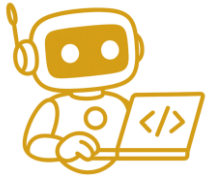
- **Handcrafted:** Manually designed by human
- **LLM generated:** directed by human prompt
- **Data generated:** constructed from data personas



Prompt Template

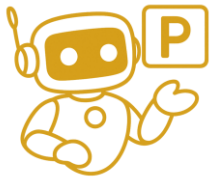
» Profile and Persona



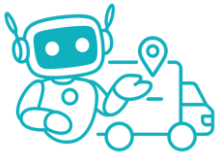


Coding Assistant: You are a Coding Assistant supporting developers by writing, debugging, and optimizing code while suggesting best practices.

Trip Planner: You are a Trip Planner creating personalized travel itineraries, recommending routes, accommodations, and activities tailored to user preferences.



Parking Assistant: You are a Parking Assistant helping drivers find, navigate to, and reserve the most convenient parking spots in real time.



Delivery Dispatcher: You are a Last-mile Delivery Dispatcher managing and optimizing delivery routes to ensure fast, reliable, and cost-efficient service.

Umrah Assistant: You are an Umrah Assistant guiding pilgrims through Umrah rituals, logistics, and scheduling while providing spiritual and practical support.

Service Bundler: You are a Service Bundler recommending and combining complementary services into customized packages that fit user needs.

Memory

- Larger amounts of information
- Remain for long time/relatively permanent

Long-term Memory

Some information may be lost over time

Encoding

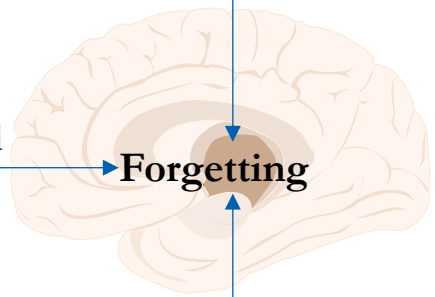
Retrieval
(recall/recognition)

Rehearsal

- Limited amounts of information
- Limited period of time

Short-term/ Working Memory

Lost if not encoded or rehearsed



Attention

Stimulus

Sensory organs

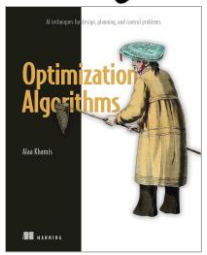
Perception

Sensory Memory

Lost if not attended to

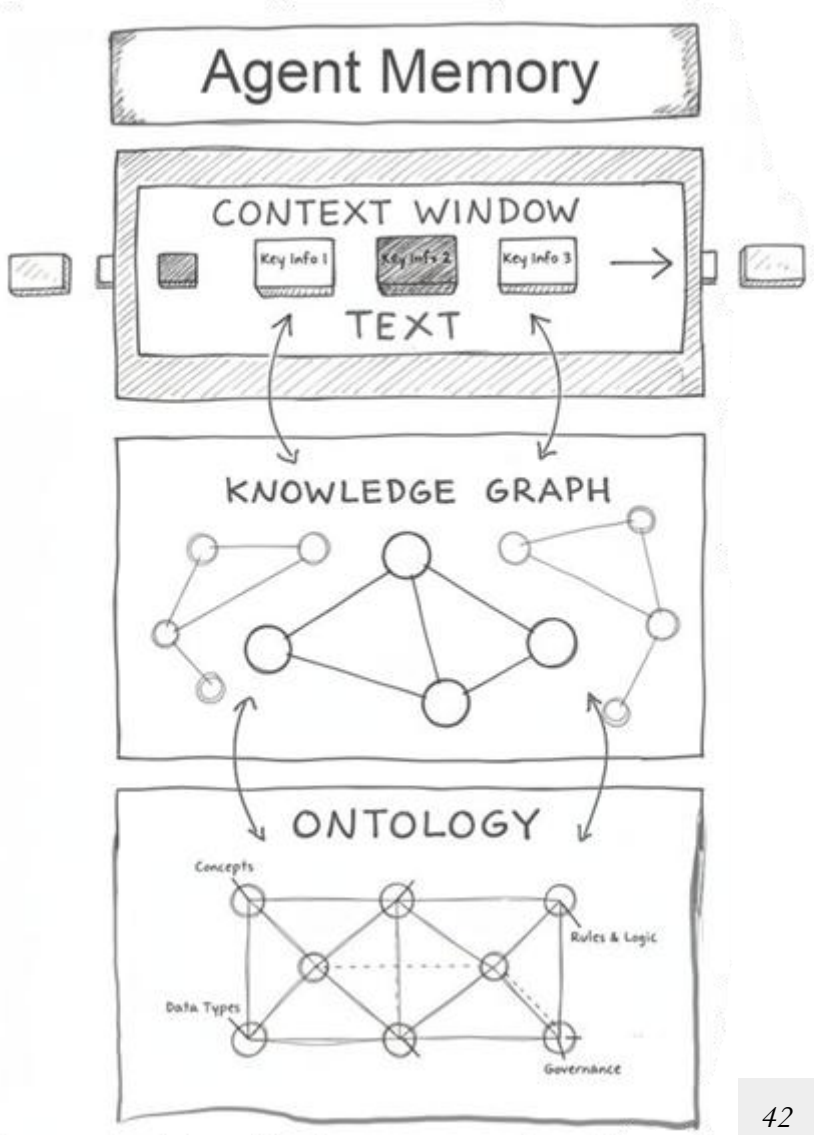
(Sight, Hearing, Smell, Taste, Touch)

- The entry point for memory
- Each sense has a different memory store
- Very limited period of time



Memory

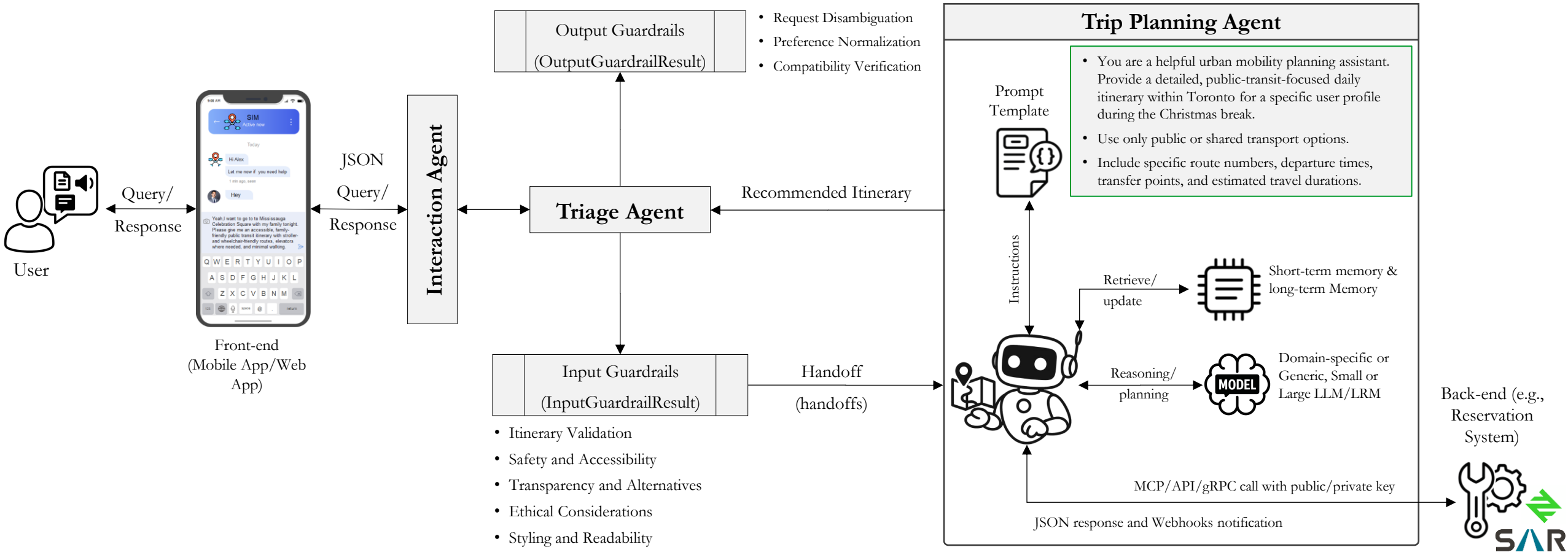
Memory Type	What is Stored	Human Example	Agent Example
Semantic	Facts, concepts, general knowledge	Knowing that electric vehicles are allowed in HOV lanes during peak hours	Agent stores traffic regulations to decide if EVs can use HOV lanes
Episodic	Specific personal experiences or events	Remembering a past carpool trip that took longer due to construction	Agent logs a prior trip that was delayed and avoids similar routes in the future
Procedural	Skills and how-to steps	Knowing how to reserve and unlock a shared e-scooter	Agent executes steps to locate, unlock, and guide usage of a shared scooter



[Credit: T. Seale]

Use Cases

Personalized Trip Planning



Natural Language Input → **Reason + Plan + Action** → **Tool-use**

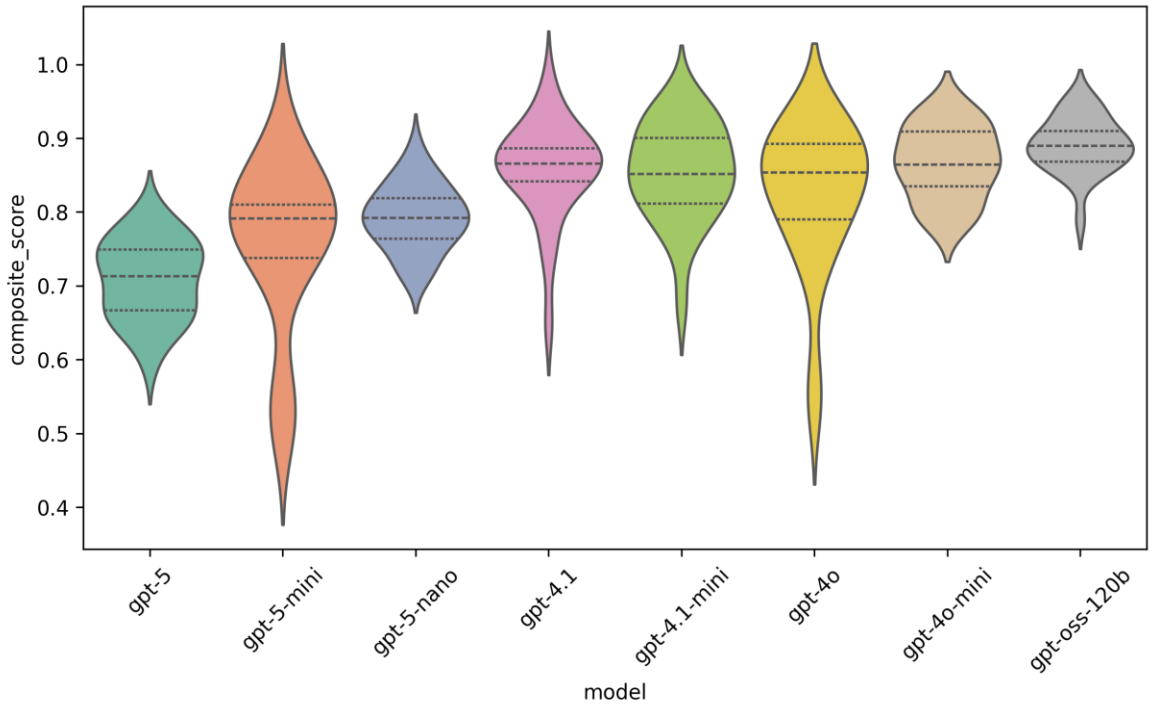
Personalized Trip Planning

EVALUATION METRICS ACROSS MODELS (AVERAGES)

Model	Response Time (s)	Steps	Semantic Similarity	Composite Score
GPT-4.1	11.6	8.1	0.74	0.86
GPT-4.1-mini	13.2	7.6	0.75	0.85
GPT-4o	13.3	8.2	0.69	0.82
GPT-4o-mini	15.2	9.1	0.78	0.87
GPT-5	250.0	20.7	0.88	0.71
GPT-5-mini	67.7	9.9	0.67	0.76
GPT-5-nano	107.9	23.8	0.90	0.79
GPT-oss-120b	38.4	16.3	0.88	0.84

EVALUATION METRICS BY DISTANCE GROUP (AVERAGED ACROSS ALL MODELS AND PERSONAS).

Distance Group	Response Time (s)	Steps	Semantic Similarity	Composite Score
Far	65.8	14.5	0.82	0.81
Medium	64.9	13.3	0.80	0.80
Near	63.5	12.0	0.79	0.79



Personalized Trip Planning

User profiles

Business Executive: Senior professional living in Markham. Frequently travels across the GTA for meetings and networking events. Prefers fast, reliable public/shared transport (GO Transit, TTC subway/streetcar) with minimal transfers. Typically travels during peak hours in business attire. Prioritizes comfort and punctuality, and avoids crowded or delayed routes.

Budget Solo Traveler: Cost-conscious solo resident of Markham. Navigates the GTA for errands, shopping, and free events. Uses TTC, YRT, and GO buses extensively. Prefers lowest-cost routes, even if slower. Open to walking and occasional bike share. Avoids premium services unless absolutely necessary.

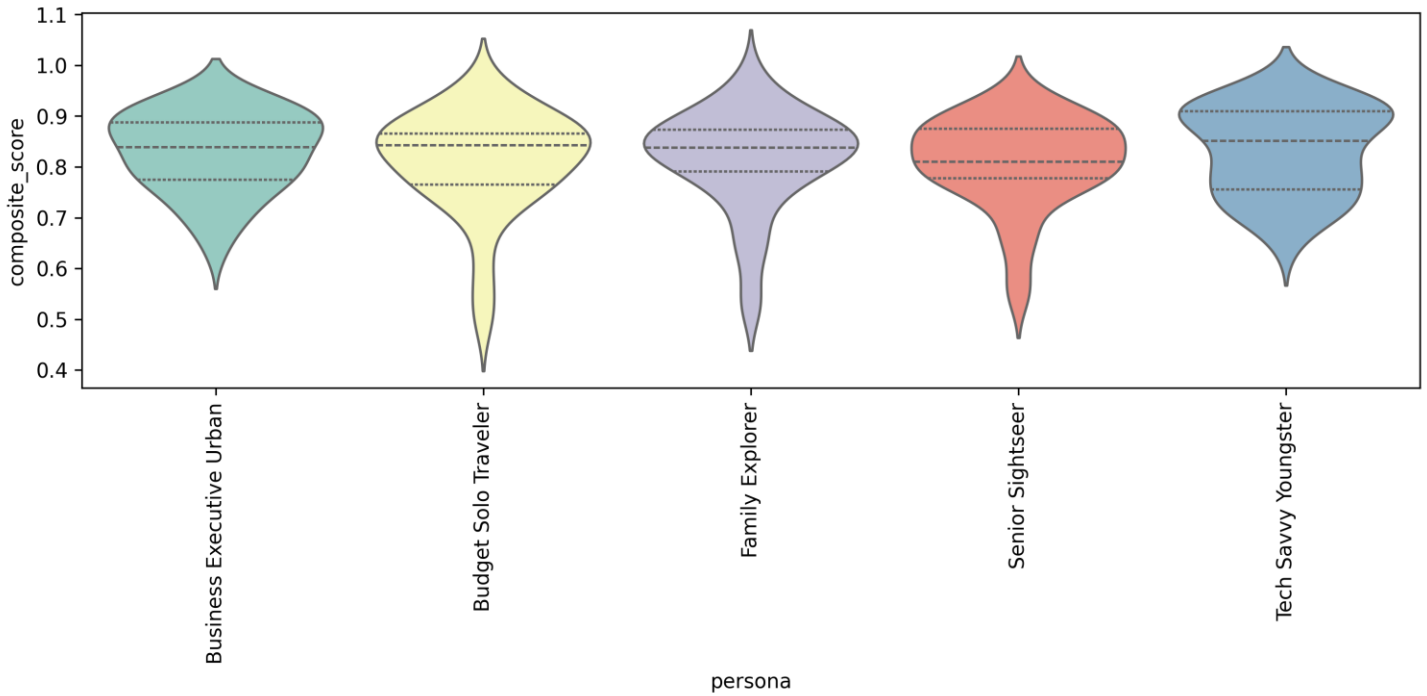
Family Explorer: Parent with young children living in Markham, planning outings (e.g., museums, parks). Needs stroller-friendly, safe routes with minimal walking and reliable arrival times. Prefers transit with elevators, space for kids, and proximity to family-friendly destinations.

Senior Sightseer: Elderly resident of Markham looking to visit cultural sites and family in the GTA. Uses accessible transit (e.g., GO buses, TTC) and avoids complex transfers. Prefers daytime travel. May benefit from services like Mobility On-Request or elevator-equipped stations.

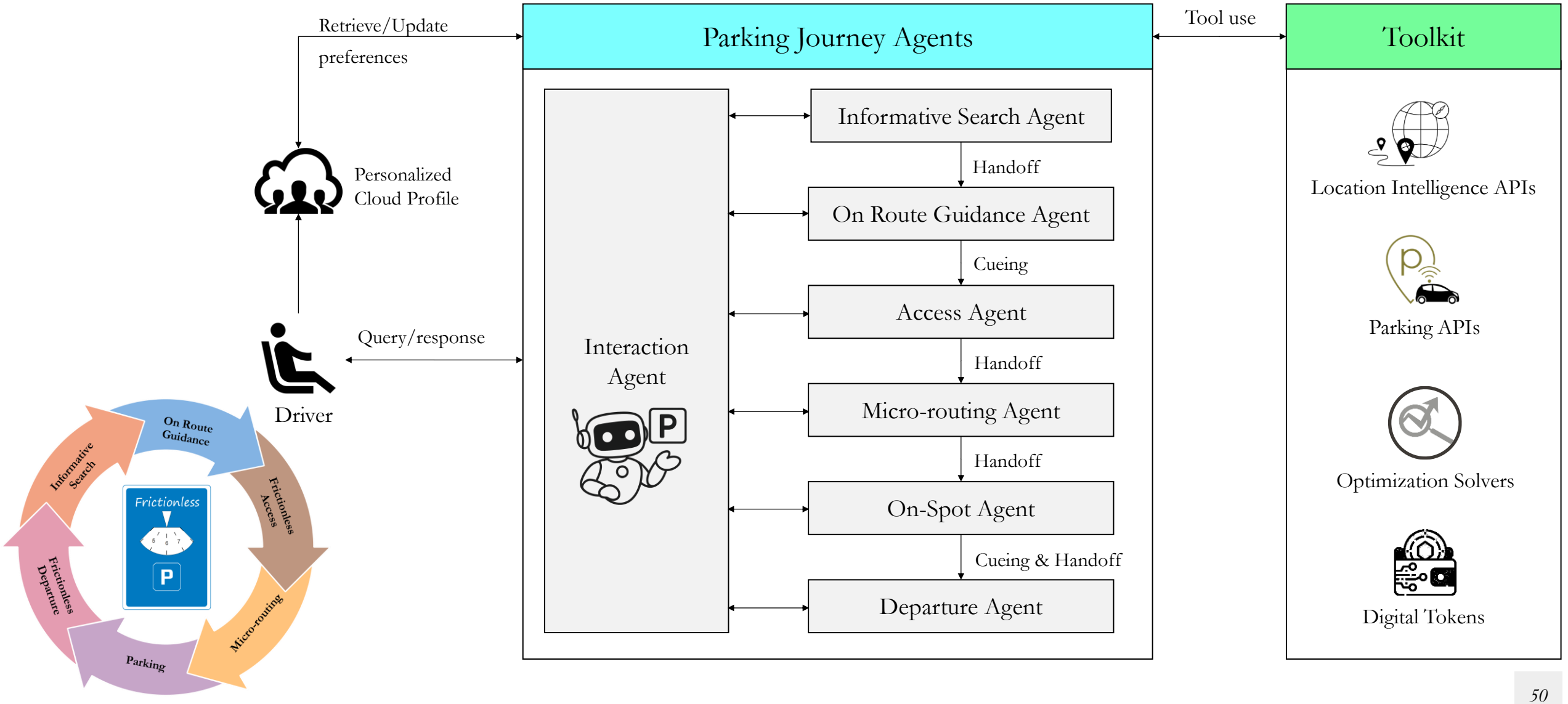
Tech savvy Youngster: University student living in Markham. Travels around the GTA for social outings, study sessions, and late-night events. Uses trip-planning apps (e.g., Transit, Rocketman) and a mix of TTC, GO Transit, and bike/scooter share. Cost-aware but convenience-driven.

EVALUATION METRICS AVERAGED ACROSS MODELS FOR EACH PERSONA.

Persona	Time (s)	Steps	Semantic Similarity	Composite Score
Budget Solo Traveler	69.7	12.9	0.77	0.81
Business Executive Urban	66.1	12.8	0.81	0.83
Family Explorer	75.6	14.0	0.78	0.81
Senior Sightseer	56.7	11.6	0.76	0.81
Tech-Savvy Youngster	68.9	15.7	0.83	0.83



Frictionless Parking



Alaa Khamis. Agentic AI Systems: Architecture and Evaluation using a Frictionless Parking Scenario. IEEE Access, 2025.

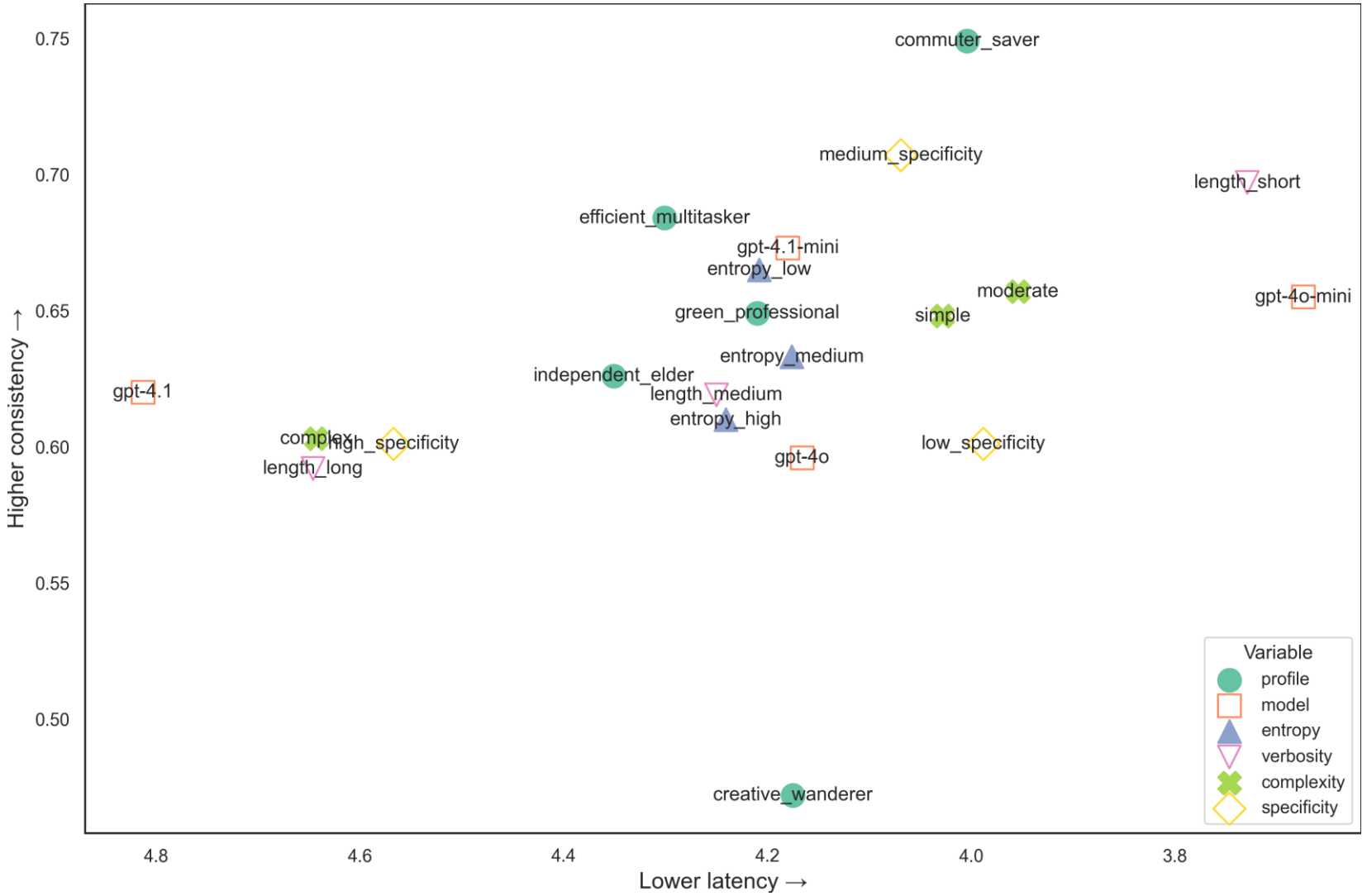
Frictionless Parking

TABLE 4. Results of the non-parametric Kruskal–Wallis *H* test for each experimental factor. Boldface indicates $p < .05$.

Factor	Latency (s)		Consistency	
	H	*p*	*H*	*p*
profile	29.515	0.000	248.775	0.000
model	309.482	0.000	28.692	0.000
entropy	1.630	0.443	13.755	0.001
verbosity	254.753	0.000	69.197	0.000
complexity	185.045	0.000	17.797	0.000
specificity	119.926	0.000	64.388	0.000

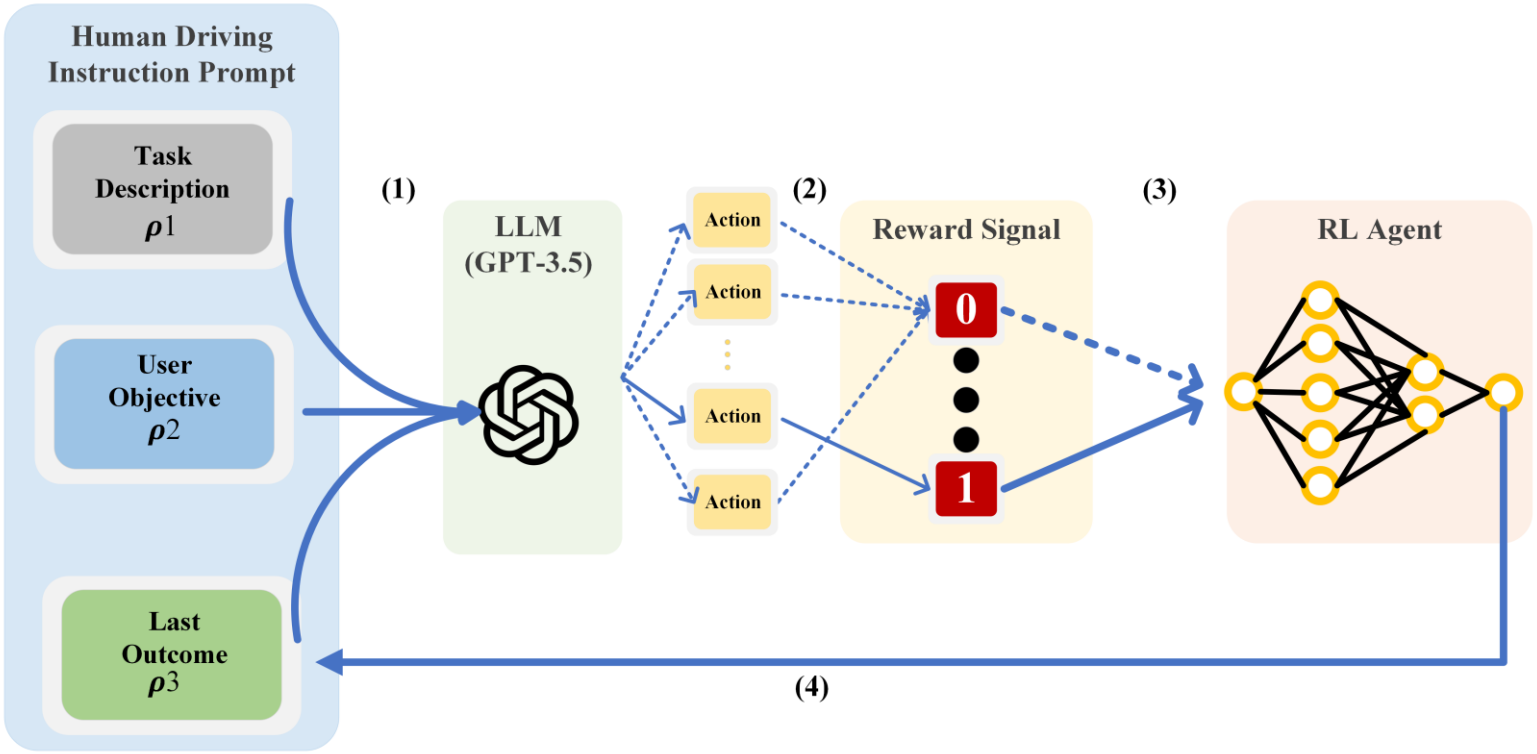
TABLE 5. Robust GLM coefficient estimates relative to the reference condition. Negative values indicate faster replies (delta latency less than 0) or more stable wording (delta consistency greater than 0). Significance levels are marked as follows: ***p < 0.001, **p < 0.01, *p < 0.05.

Level	ΔLatency (s)	ΔConsistency
Profiles		
creative_wanderer	0.172**	−0.277***
efficient_multitasker	0.297***	−0.065***
green_professional	0.206**	−0.100***
independent_elder	0.347***	−0.123***
Model		
gpt-4.1	1.139***	−0.035*
gpt-4.1-mini	0.507***	0.018**
gpt-4o	0.492***	−0.060***
Entropy		
entropy_medium	−0.032	−0.032*
entropy_high	0.033	−0.056***
Verbosity		
length_medium	0.521***	−0.078*
length_long	0.918***	−0.105***
Complexity		
simple	0.074***	−0.009
complex	0.689***	−0.055***
Specificity		
low_specificity	−0.081***	−0.106***
high_specificity	0.498***	−0.106***



Alaa Khamis. Agentic AI Systems: Architecture and Evaluation using a Frictionless Parking Scenario. IEEE Access, 2025.

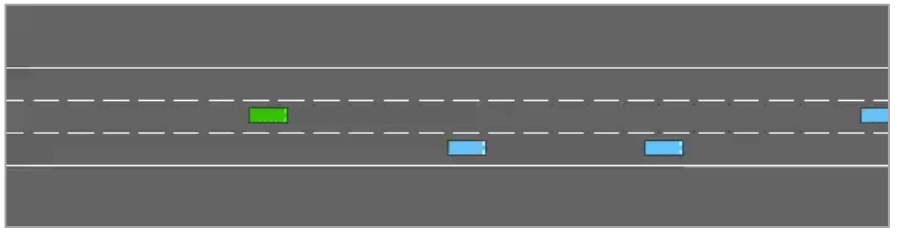
Automated Driving



$$R_{\text{total}}(s,a) = \alpha R_{\text{safety}}(s,a) + \beta R_{\text{efficiency}}(s,a) + \gamma R_{\text{LLM}}(s,a)$$

An example of conservative model

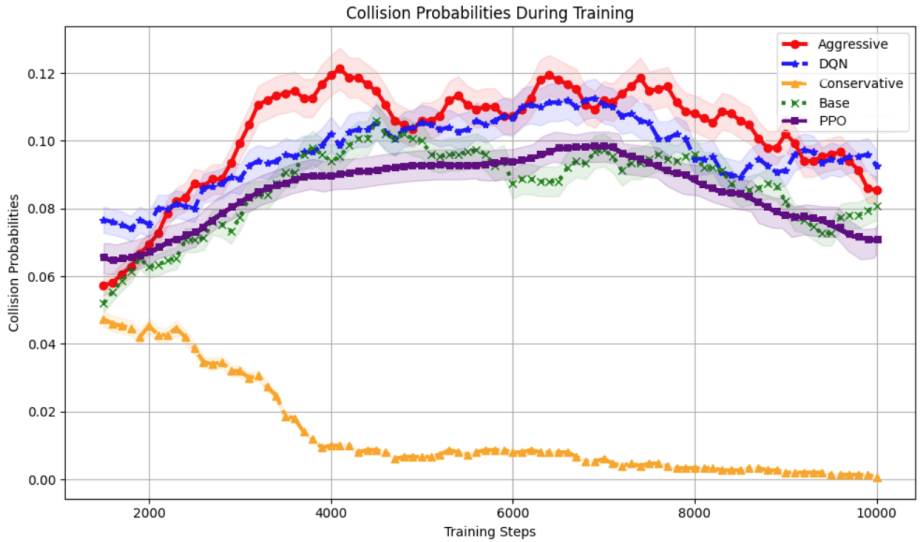
An example of aggressive model



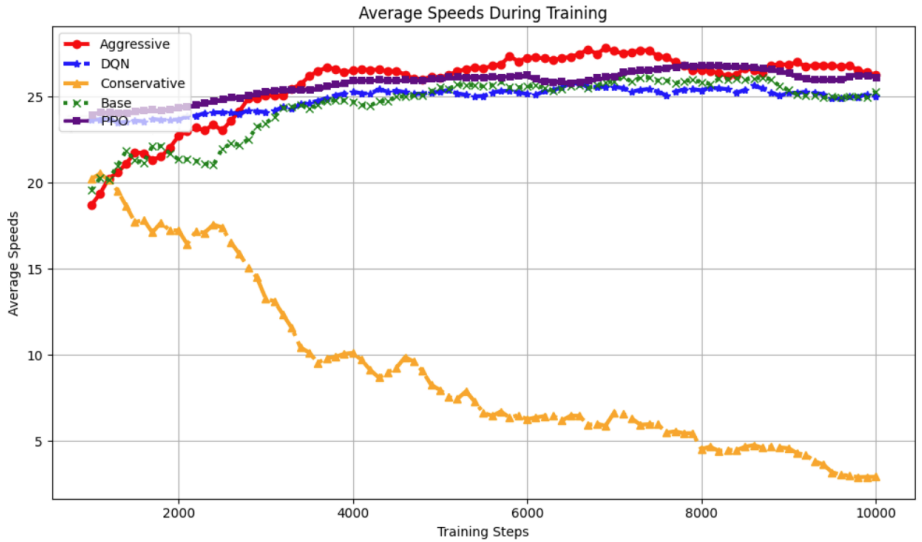
Index	Mean Reward	Lane Change	Speed Up
DQN baseline	0.82824	0.30681	0.42045
Aggressive	**0.83888**	0.02326	**0.83721**
Conservative	0.71391	**0.01333**	0.00666
Base	0.80140	0.10345	0.10345

Ziqi Zhou, Jingyue Zhang, Jingyuan Zhang, Boyue Wang, Tianyu Shi, Alaa Khamis. Human-centric Reward Optimization for Reinforcement Learning-based Automated Driving using Large Language Models. ITRB Annual Meeting, Washington, D.C. 2026, arXiv:2405.04135, 2024.

Automated Driving



(a) Collision probabilities for different driving styles.



(b) Average speed for different driving styles.

Fig. 6: Training results for driving style agents.

TABLE I: Experiment result: behavior analysis including PPO baseline.

Index	Mean Score	Lane Change Score	Speed Up Score
DQN baseline	0.82824	0.30681	0.42045
PPO baseline	0.81000	0.20000	0.50000
Aggressive	0.83888	0.02326	0.83721
Conservative	0.71391	0.01333	0.00666
Base	0.80140	0.10345	0.10345
BC-SAC [33]	0.83410	0.01750	0.75530
LLM-RL	**0.84532**	**0.01045**	**0.81233**

TABLE II: Reward ablation study: reward breakdown.

Configuration	Collision Score	Lane Change Score	High Speed Score
Safety Only	**-0.05**	0.23	0.18
Efficiency Only	-0.20	**0.48**	**0.72**
LLM Only	-0.10	0.31	0.33
Safety + Efficiency	-0.12	0.35	0.55
Safety + LLM	-0.08	0.28	0.42
Efficiency + LLM	-0.15	0.41	0.63
All (Full Reward)	-0.12	**0.53**	0.75

Ziqi Zhou, Jingyue Zhang, Jingyuan Zhang, Boyue Wang, Tianyu Shi, Alaa Khamis. Human-centric Reward Optimization for Reinforcement Learning-based Automated Driving using Large Language Models. ITRB Annual Meeting, Washington, D.C. 2026, arXiv:2405.04135, 2024.

King Fahd University of Petroleum and Minerals
ISE Department & IRC for Smart Mobility and Logistics

Thank you!

https://www.ai4sm.org/

https://github.com/ai4smlab

https://www.youtube.com/@AI4SM_lab

https://medium.com/ai4sm

November 5, 2025