ML105: Agentic AI

Rola Dali, PhD
Senior ML/AI Architect
December 2025

>whoami

- Machine Learning Architect @ Tech42
- Academic: PhD in NeuroScience & Bioinformatics, 2017 @McGill
- Career Interests: Data, AI/ML, Cloud

Rola Dali

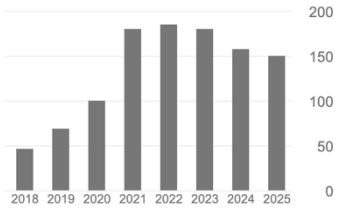
McGill University
Verified email at mail.mcgill.ca

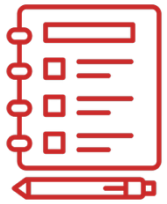
Google Scholar

bioRxiv
THE PREPRINT SERVER FOR BIOLOGY

arXiv

	All	Since 2020
Citations	1213	959
h-index	13	12
i10-index	14	13



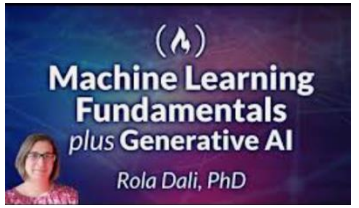


Agenda

- The Rise of Generative AI
- The Spectrum of Autonomy
- Agentic Timeline
- What is an Agent?
- Agent Use Cases
- Patterns and Anti-Patterns
- Designing an Agent
- Implementing an Agent
- Architectural Patterns
- Agent Interface Protocols
- Evaluating Agents
- Agent Frontier
- Agents & Jobs
- References

Artificial Intelligence: The rise of Generative AI

Simplified AI Timeline

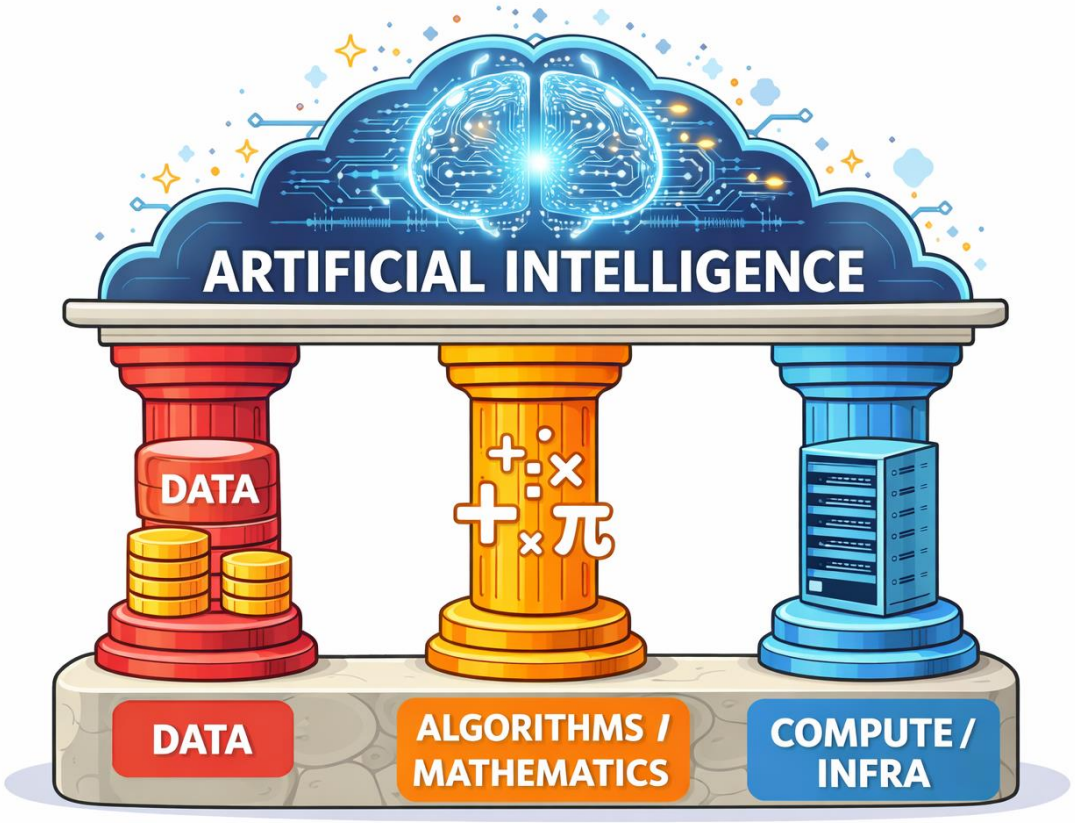


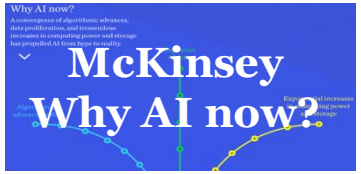
- AI cradle: 1940-1950s
  - 1943: McCulloch & Pitts . “A Logical Calculus of the Ideas Immanent in Nervous Activity”. mathematical model of biological neurons
  - 1950: Alan Turing. "Computing Machinery and Intelligence”. The Turing Test, a benchmark for machine intelligence.
  - 1956: The Dartmouth Workshop established AI as a field of study. John McCarthy coined the term "artificial intelligence".
- Machine Learning Renaissance: 1980-1990s
  - 1978: Geoffrey Hinton awarded his PhD in Artificial Intelligence from University of Edinburgh
  - 1997: IBM's Deep Blue defeated Garry Kasparov, in chess
  - 2011: IBM Watson won the TV game show Jeopardy!
- Deep Learning Boom: 2010-2020s
  - 2012: AlexNet, a deep learning network, performed well in an image recognition competition
  - 2016: Google DeepMind's AlphaGo defeated world champion Lee Sedol in the game of Go.
- Generative AI Boom: 2020s-Present
  - 2017: The transformer paper released by Google researchers providing the blueprint for most modern large language model.
  - 2022: OpenAI's ChatGPT launched in November 2022

What differentiates GenAI from Traditional ML?

Artificial Intelligence: The rise of Generative AI

GenAI differentiators



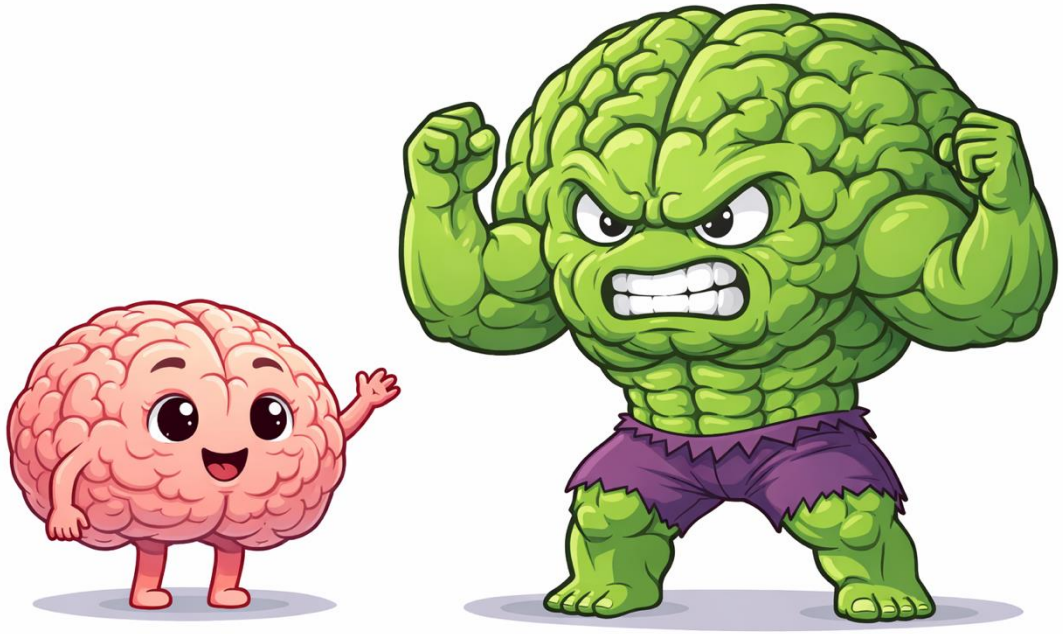


	Training Data Size	Model Size (Params)	Compute
Machine Learning	MBs-GBs	Thousands-Millions	Serial
Generative AI	TBs	Billions-Trillions	Distributed/Parallelization

Artificial Intelligence: The rise of Generative AI

GenAI differentiators

What happens when you supercharge data, algorithms and compute?

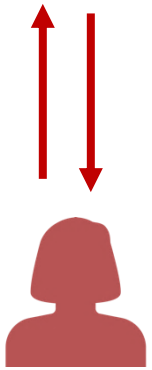


ML GenAI

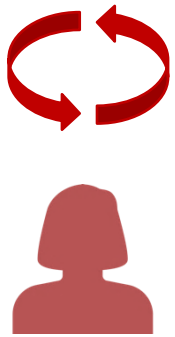
- Models that can use human language
- Tasks: specific to general
- Rise of Model-as-service
- Advancements needed all 3 pillars!

GenAI, Agentic Systems and the spectrum of Autonomy

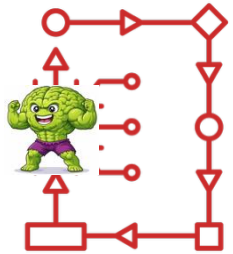
LLM
single call



LLM
in a loop
Chatbot



Workflow



Agent

Deep Agent

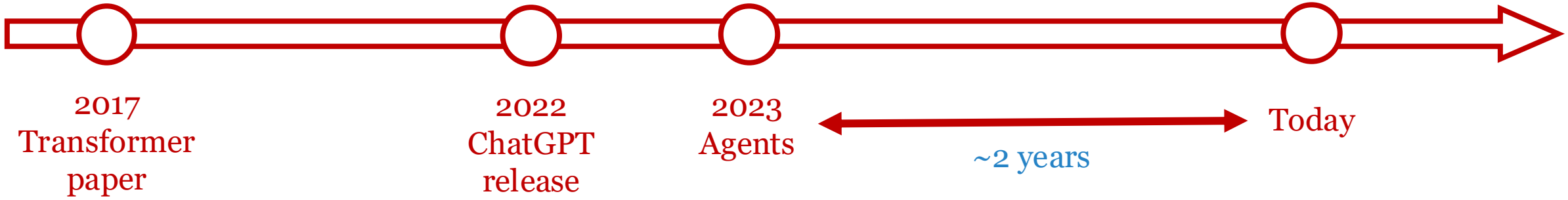
Autonomous
Agentic
Systems

AGI

Agent Autonomy

Human Control

Agentic Milestone Timeline

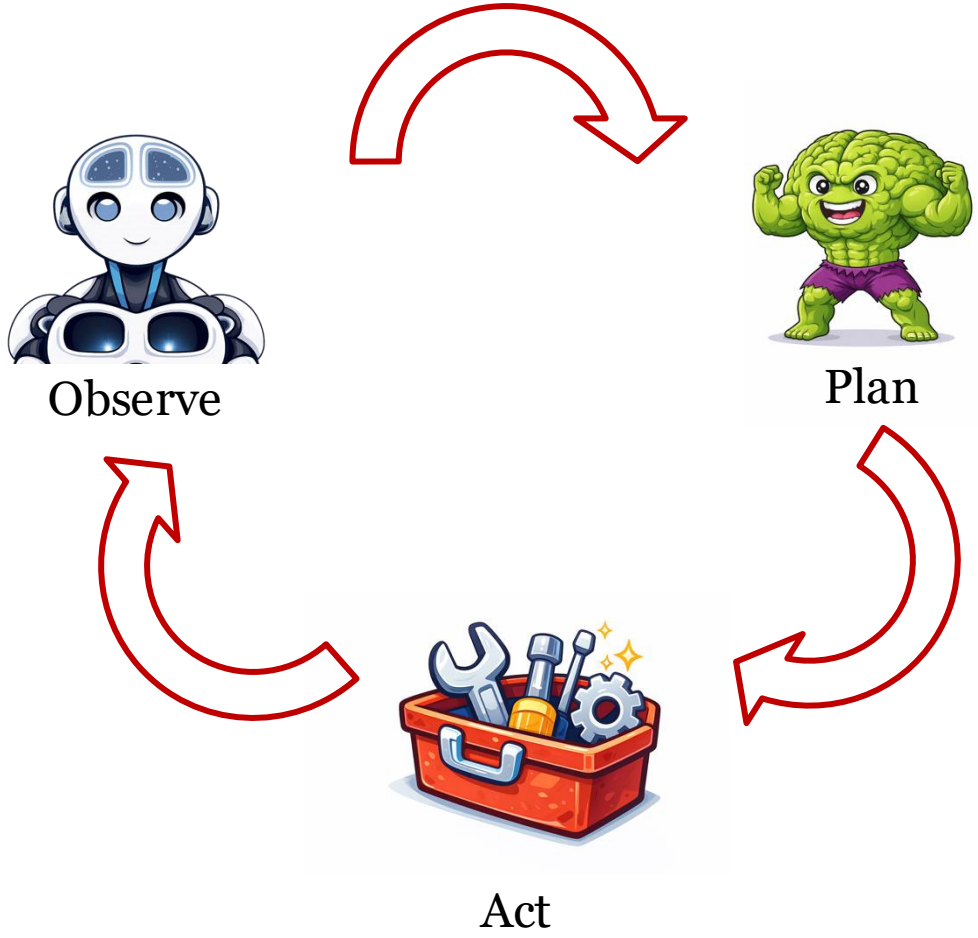


The field is still young/maturing

The time stamp matters!

What is an Agent?

- A GenAI Agent is a software entity designed to perceive its environment, make decisions and take actions to achieve specific goals



Agent pseudocode

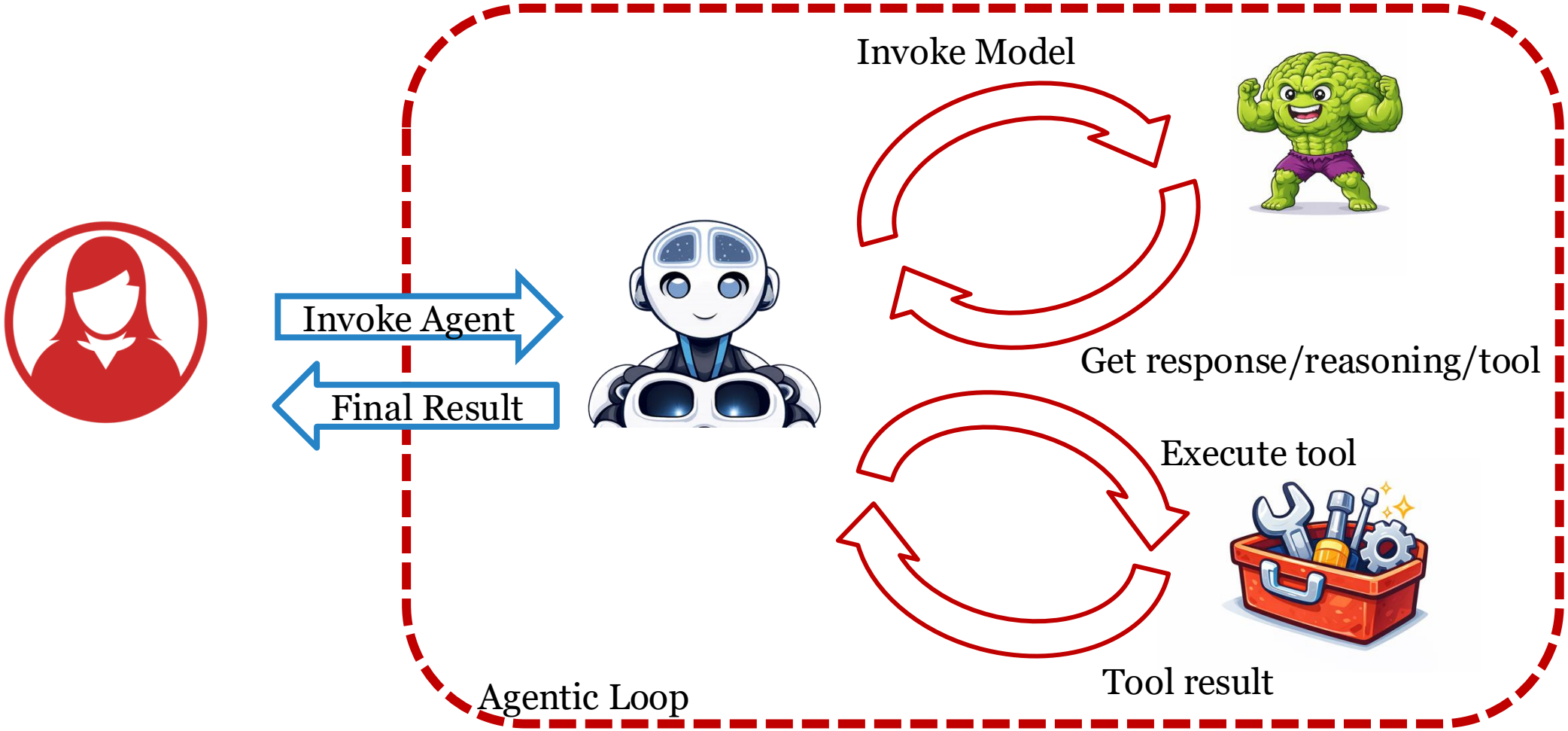
```
while True:
    user query = get_user_input()
    response = invoke_llm(user_query)

    while response.has_tool_call():
        tool_result = invoke_tool(response.tool_spec)
        memory = append_context()
        response = invoke_llm(tool_output, memory)

    final_answer = response
    return final_answer
```

What is an Agent?

- A GenAI Agent is a software entity designed to perceive its environment, make decisions and take actions to achieve specific goals



What is an Agent?
Agents vs Workflows

Task

- Book a travel activity
  - Check available activities
  - Check activity availability
  - Check calendar availability
  - Book/Pay for activity
  - Add activity to Calendar

Workflow

Activities popular in Montreal?

activity_availability()

calendar_availability()

Loop till a desired activity fits availability

book_activity()

update_calendar()

Agent

You are an activity booking agent.
Choose activities based on customer preferences
Book an available activity & update calendar

activity_availability()
calendar_availability()
book_activity()
update_calendar()

Agent ~ dynamic control flow devised by an LLM at Runtime
Workflows: static, pre-defined coded graphs

Agent Use Cases

Pros of using Agents:

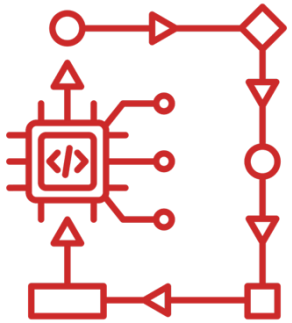
- **Availability**: Available 24/7, 365 days/year
- **Multi-lingual**: Delivers multilingual support
- **Efficiency**: Improves response time
- **Consistency**: Ensures more consistent support
- **Convenience**: Offers convenient self-service options
- **Scale**: Serves more customers at scale
- **Cost**: can be cheaper than human

Cons:

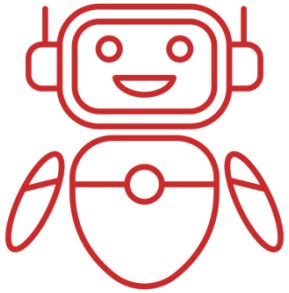
- Not Human
- Technology still maturing
- Application space still maturing
- Cost: usually more expensive than other software

Patterns and Anti-Patterns for Agents

When to use, or not to use an agent

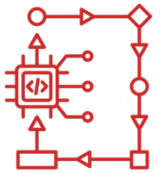


- Mission Critical/Error Sensitive
- Regulated Industries/Deterministic outcomes
- Latency Sensitive
- Cost Sensitive
- Performance
- Flexibility
- Model driven Decision Making



Patterns and Anti-Patterns for Agents

When to use, or not to use an agent



1. Is the application: mission critical/error sensitive or in a highly regulated industry?
2. Is the task path predictable or can be predefined?
3. Is the value of the task worth the cost?
4. Is the latency critical?

Use an agent in cases where error is tolerable, open-endedness is appreciated, the execution path is harder to code, cost is not an issue and latency can be tolerated

Components of an Agent

- AI Agents have several Key characteristics

1. Purpose/Goal
2. Reasoning/Planning
3. Memory
4. Tools & Actions

- Tier 2:
  - Guardrails
  - Communication
  - Learning

Designing an Agent

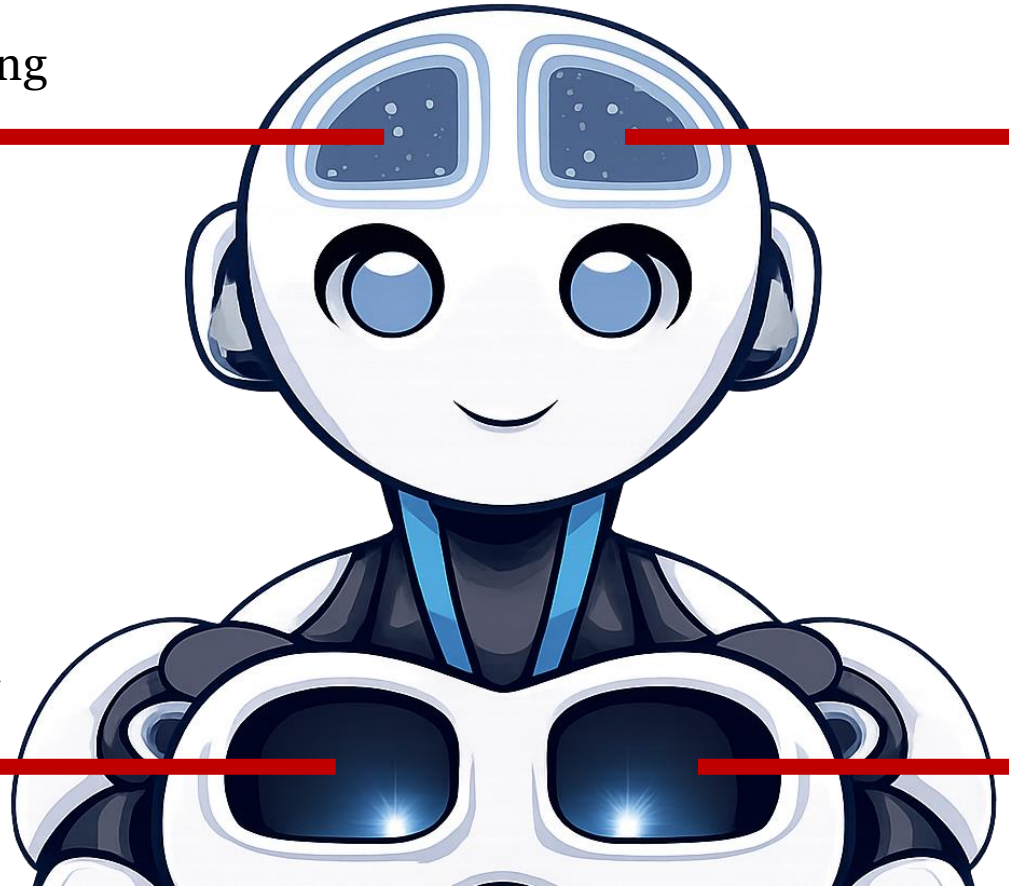
Reasoning/Planning

LLM

Memory

Short-term

Long-term



Purpose/Goal/Mission/Identity

System Prompt

Tools/Actions

Functions

Designing an Agent
Choosing an LLM

- Criteria to consider when choosing a model
  - Task complexity
  - Reasoning capabilities
  - Context window size
  - Tool Calling
  - Latency
  - Cost
  - Compliance & data privacy

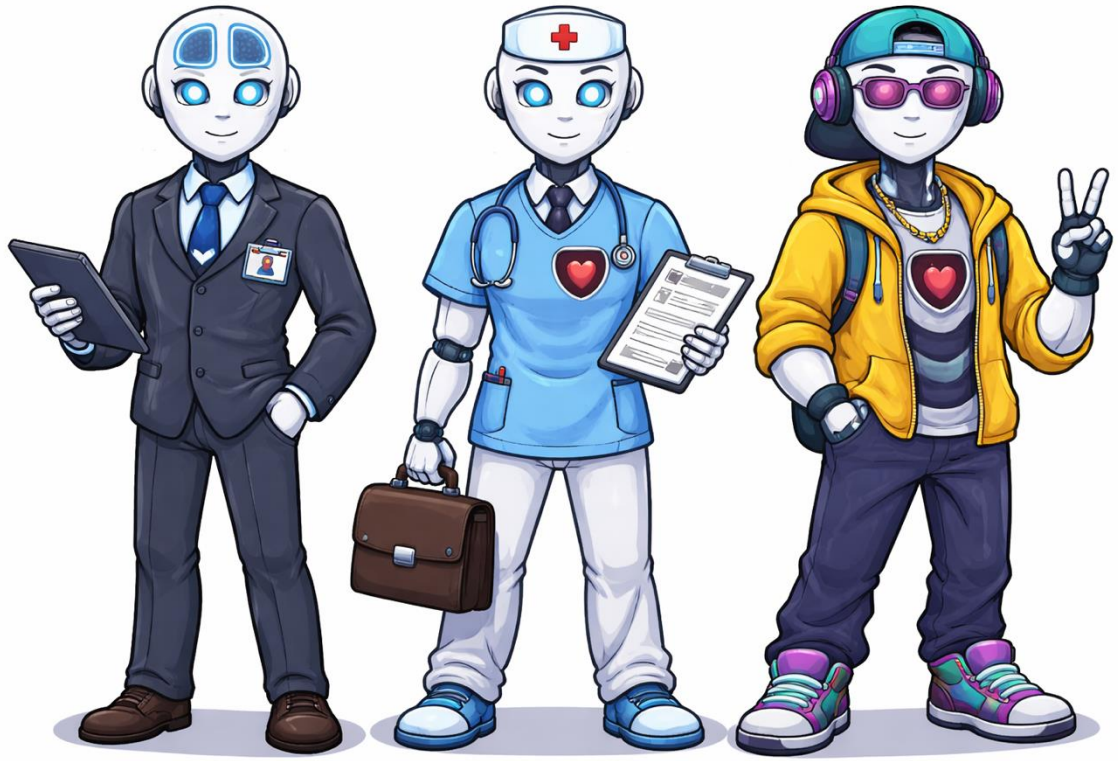
RANK	MODEL	TYPE	OUTPUT TYPE	VENDOR	AVG ACTION COMPLETION	AVG TOOL SELECTION QUALITY	AVG COST ($)	AVG DURATION (S)	AVG TURNS
1st	gpt-4.1-2025-04-14	Proprietary	Normal	OpenAI	0.620	0.800	$0.068	24.3	3.1
2nd	mistral-medium-2508	Proprietary	Normal	Mistral	0.610	0.770	$0.020	37.5	3.0
3rd	gpt-4.1-mini-2025-04-14	Proprietary	Normal	OpenAI	0.560	0.790	$0.014	26.0	3.4

Designing an Agent

System prompt

You are medical assistant. You are caring and empathetic

You are a financial advisor. You are eloquent and professional.

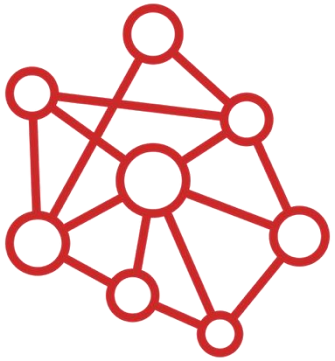


You are a teen advisor. You are young and hip

System prompt ~ Agent character/persona & purpose/task & instructions

Designing an Agent
Memory

LLMs are stateless

Intrinsic Memory	Short-Term	Long-Term
		
Model parameters	Context window	External Storage

Designing an Agent
Actions & Tools

- LLMs have limitations that tools can help overcome
- Agents can take several actions
  - Capability extension: calling functions, APIs, ...
  - Knowledge Augmentation: retrieving data or context
  - Orchestration: calling other agents

Tools can come in the form of

Function Call

API Call

Data retrieval

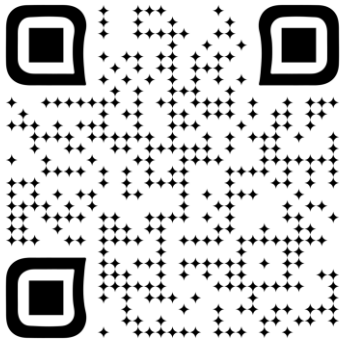
Browser Action

Code execution

Filesystem control

Implementing an Agent

- Single LLM Call
- LLM Loop: ChatBot
- Simple Agent
- Agent with Memory
- Agent with LangChain
- LangChain Agent with Memory
- Agent Architecture

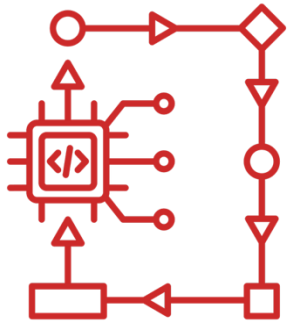


Other Topics

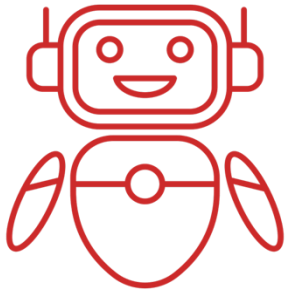
- Model Selection
- Prompt Engineering & Context Engineering
- Data Management (storage, ranking, retrieval, …)
- Memory
- Tooling
- Interfaces: MCP, A2A, AG-UI
- Architecture choices
- Deployment approaches
- Security & Compliance
- Orchestration & Multi-Agent Communication
- Error Handling & Remediation
- Monitoring and Observability
- UI/UX: streaming, latency, …
- …. … … …

Agentic Architectural Patterns

Workflow

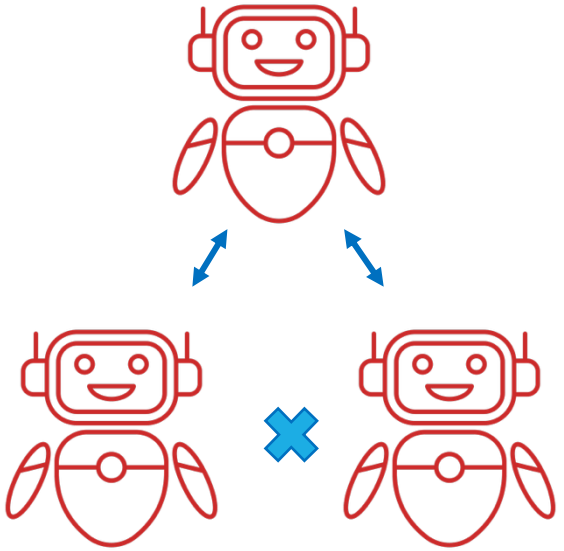


Single Agent

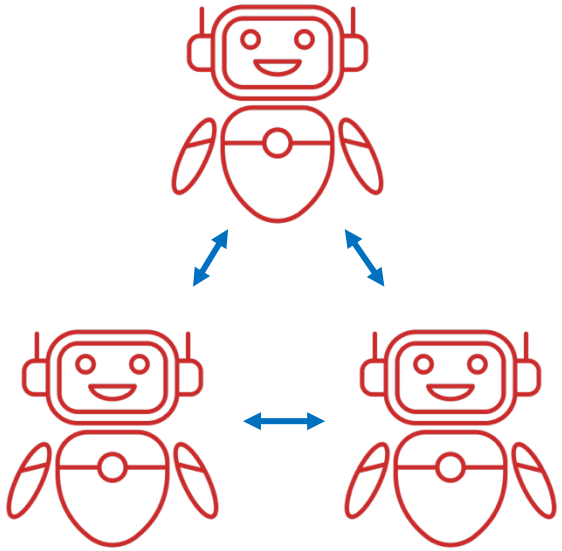


Multi-agent

Hierarchical/Supervisor

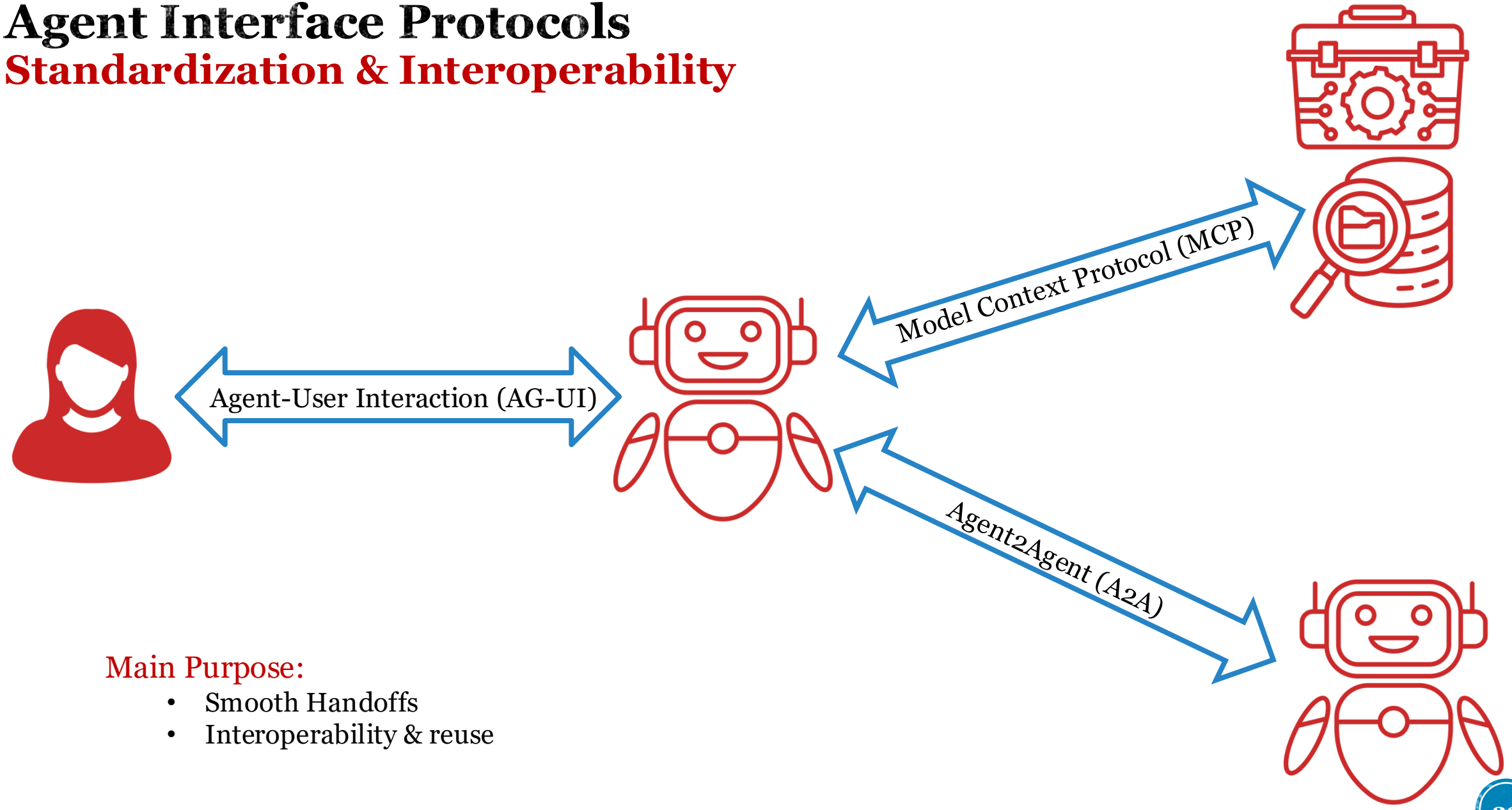


Swarm



Agent Interface Protocols

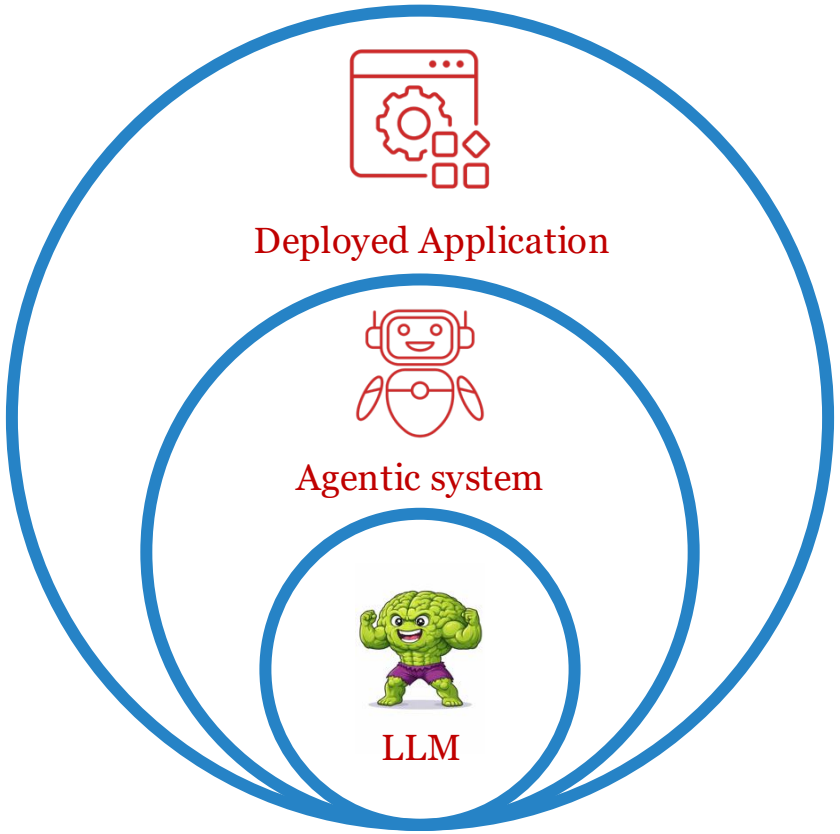
Standardization & Interoperability



Main Purpose:

- Smooth Handoffs
- Interoperability & reuse

Evaluating Agents
Layers of the onion



Evaluating Agents
What to evaluate

LLM

- Following Instructions
- Task Completion
- Accuracy
- Hallucination
- Consistency
- Toxicity
- Guardrails

Agentic System

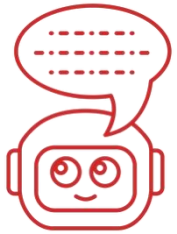
- Proper Task Decomposition
- Efficient Task Execution
- Faithfulness
- Correct Tool Selection
- Answer/Retrieval Relevance
- Task Completion/Success
- Knowledge Retention
- Reliability/Robustness

Application

- Overall Performance
- Error Rate
- Latency
- Scalability
- Cost Efficiency
- Safety & Security
- Identity/Access management
- Data Integrity
- UI/UX
- Networking

Evaluating Agents
How to evaluate output

Code based evals



LLM-as-judge

Human Annotators

- Quantitative and/or Discrete?
- Probabilistic/Deterministic?
- Ground Truth Exists?
- Cost sensitive?

Agents: Technology Frontier

Agent Challenges

Models

- Output Evaluation
- Model limitations
- Hallucinations

Agent

- Path Evaluation
- Context Management
- Convoluted Debugging
- Price Estimates
- Compounding Errors
- Getting Stuck in loops
- Integration Issues
- Framework Stability
- Business Value

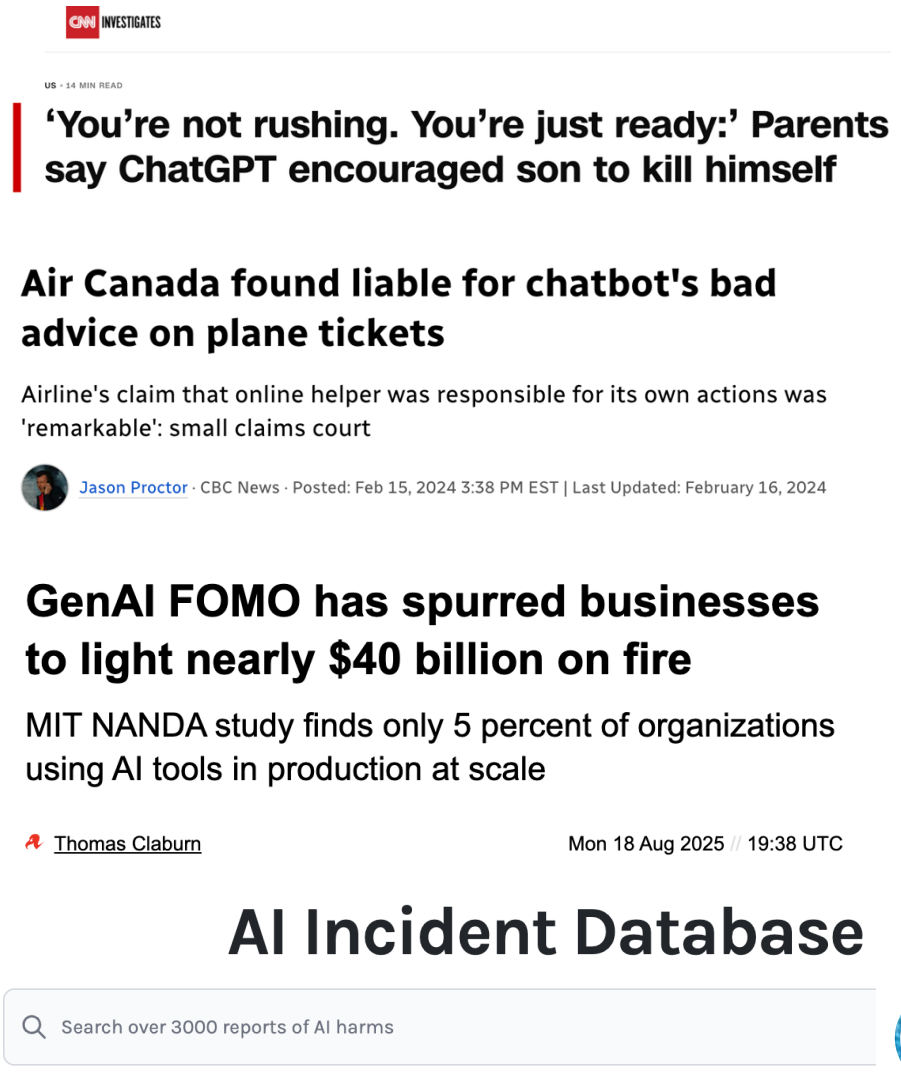
Agents: Technology Frontier

Best Practices

- AI Potential is there… Technology is getting there
- Progress is rarely linear

- **Best Practices:**
  - Treat AI as a junior assistant
  - Start with read-only access to tools and systems
  - Add human approvals for critical steps
  - Enable comprehensive logging

Incident 1152: LLM-Driven Replit Agent Reportedly Executed Unauthorized Destructive Commands During Code Freeze, Leading to Loss of Production Data



Will Agents Take My Job? … Maybe… Not just yet

Table 3: Top 40 occupations with highest AI applicability score.

Job Title (Abbrv.)	Coverage	Cmpltn.	Scope	Score	Employment
Interpreters and Translators	0.98	0.88	0.57	0.49	51,560
Historians	0.91	0.85	0.56	0.48	3,040
Passenger Attendants	0.80	0.88	0.62	0.47	20,190
Sales Representatives of Services	0.84	0.90	0.57	0.46	1,142,020
Writers and Authors	0.85	0.84	0.60	0.45	49,450
Customer Service Representatives	0.72	0.90	0.59	0.44	2,858,710
CNC Tool Programmers	0.90	0.87	0.53	0.44	28,030
Telephone Operators	0.80	0.86	0.57	0.42	4,600
Ticket Agents and Travel Clerks	0.71	0.90	0.56	0.41	119,270
Broadcast Announcers and Radio DJs	0.74	0.84	0.60	0.41	25,070
Brokerage Clerks	0.74	0.89	0.57	0.41	48,060
Farm and Home Management Educators	0.77	0.91	0.55	0.41	8,110
Telemarketers	0.66	0.89	0.60	0.40	81,580
Concierges	0.70	0.88	0.56	0.40	41,020
Political Scientists	0.77	0.87	0.53	0.39	5,580
News Analysts, Reporters, Journalists	0.81	0.81	0.56	0.39	45,020
Mathematicians	0.91	0.74	0.54	0.39	2,220
Technical Writers	0.83	0.82	0.54	0.38	47,970
Proofreaders and Copy Markers	0.91	0.86	0.49	0.38	5,490
Hosts and Hostesses	0.60	0.90	0.57	0.37	425,020
Editors	0.78	0.82	0.54	0.37	95,700
Business Teachers, Postsecondary	0.70	0.90	0.52	0.37	82,980
Public Relations Specialists	0.63	0.90	0.60	0.36	275,550
Demonstrators and Product Promoters	0.64	0.88	0.53	0.36	50,790
Advertising Sales Agents	0.66	0.90	0.53	0.36	108,100
New Accounts Clerks	0.72	0.87	0.51	0.36	41,180
Statistical Assistants	0.85	0.84	0.49	0.36	7,200
Counter and Rental Clerks	0.62	0.90	0.52	0.36	390,300
Data Scientists	0.77	0.86	0.51	0.36	192,710
Personal Financial Advisors	0.69	0.88	0.52	0.35	272,190
Archivists	0.66	0.88	0.49	0.35	7,150
Economics Teachers, Postsecondary	0.68	0.90	0.51	0.35	12,210
Web Developers	0.73	0.86	0.51	0.35	85,350
Management Analysts	0.68	0.90	0.54	0.35	838,140
Geographers	0.77	0.83	0.48	0.35	1,460
Models	0.64	0.89	0.53	0.35	3,090
Market Research Analysts	0.71	0.90	0.52	0.35	846,370
Public Safety Telecommunicators	0.66	0.88	0.53	0.35	97,820
Switchboard Operators	0.68	0.86	0.52	0.35	43,830
Library Science Teachers, Postsecondary	0.65	0.90	0.51	0.34	4,220

Table 4: Bottom 40 occupations with lowest AI applicability score.

Job Title (Abbrv.)	Coverage	Cmpltn.	Scope	Score	Empl.
Phlebotomists	0.06	0.95	0.29	0.03	137,080
Nursing Assistants	0.07	0.85	0.34	0.03	1,351,760
Hazardous Materials Removal Workers	0.04	0.95	0.35	0.03	49,960
Helpers–Painters, Plasterers, ...	0.04	0.96	0.38	0.03	7,700
Embalmers	0.07	0.55	0.22	0.03	3,380
Plant and System Operators, All Other	0.05	0.93	0.38	0.03	15,370
Oral and Maxillofacial Surgeons	0.05	0.89	0.34	0.03	4,160
Automotive Glass Installers and Repairers	0.04	0.93	0.34	0.03	16,890
Ship Engineers	0.05	0.92	0.39	0.03	8,860
Tire Repairers and Changers	0.04	0.95	0.35	0.02	101,520
Prosthodontists	0.10	0.90	0.29	0.02	570
Helpers–Production Workers	0.04	0.93	0.36	0.02	181,810
Highway Maintenance Workers	0.03	0.96	0.32	0.02	150,860
Medical Equipment Preparers	0.04	0.96	0.31	0.02	66,790
Packaging and Filling Machine Op.	0.04	0.91	0.39	0.02	371,600
Machine Feeders and Offbearers	0.05	0.89	0.36	0.02	44,500
Dishwashers	0.03	0.95	0.30	0.02	463,940
Cement Masons and Concrete Finishers	0.03	0.92	0.39	0.01	203,560
Supervisors of Firefighters	0.04	0.88	0.39	0.01	84,120
Industrial Truck and Tractor Operators	0.03	0.94	0.28	0.01	778,920
Ophthalmic Medical Technicians	0.04	0.89	0.33	0.01	73,390
Massage Therapists	0.10	0.91	0.32	0.01	92,650
Surgical Assistants	0.03	0.78	0.29	0.01	18,780
Tire Builders	0.03	0.93	0.40	0.01	20,660
Helpers–Roofers	0.02	0.94	0.37	0.01	4,540
Gas Compressor and Gas Pumping Station Op.	0.01	0.96	0.47	0.01	4,400
Roofers	0.02	0.94	0.38	0.01	135,140
Roustabouts, Oil and Gas	0.01	0.95	0.39	0.01	43,830
Maids and Housekeeping Cleaners	0.02	0.94	0.34	0.01	836,230
Paving, Surfacing, and Tamping Equipment Op.	0.01	0.96	0.29	0.01	43,080
Logging Equipment Operators	0.01	0.95	0.36	0.01	23,720
Motorboat Operators	0.01	0.93	0.39	0.00	2,710
Orderlies	0.00	0.76	0.18	0.00	48,710
Floor Sanders and Finishers	0.00	0.94	0.34	0.00	5,070
Pile Driver Operators	0.00	0.98	0.24	0.00	3,010
Rail-Track Laying and Maintenance Equip. Op.	0.00	0.96	0.27	0.00	18,770
Foundry Mold and Coremakers	0.00	0.95	0.36	0.00	11,780
Water Treatment Plant and System Op.	0.00	0.92	0.44	0.00	120,710
Bridge and Lock Tenders	0.00	0.93	0.39	0.00	3,460
Dredge Operators	0.00	0.99	0.22	0.00	940

Microsoft Research. July 2025: Working with AI: Measuring the Applicability of Generative AI to Occupations

Will Agents Take My Job?
Moravec's paradox

Sensory-Motor

Cognitive/Intellectual

Hard for humans

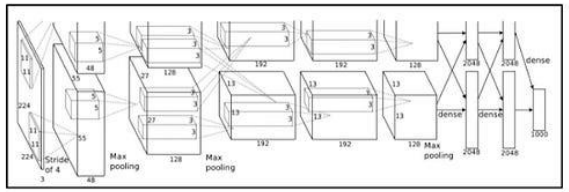
Hard for AI

Will Agents Take My Job?
Software Development

Software 1.0
computer code
programs
computer

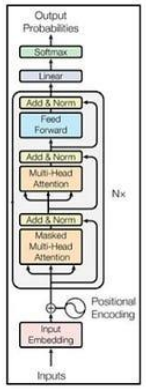
became programmable in ~1940s

Software 2.0
weights
programs
neural net



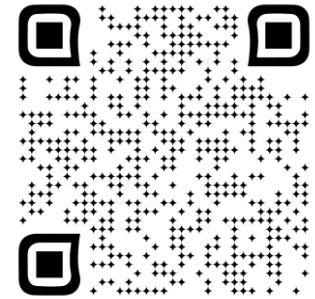
fixed function neural net
e.g. AlexNet: for image recognition (~2012)

Software 3.0
prompts
programs
LLM



~2019

LLM = programmable neural net!



def get_sentiment()

sentiment_model.predict()

"Give me the sentiment"

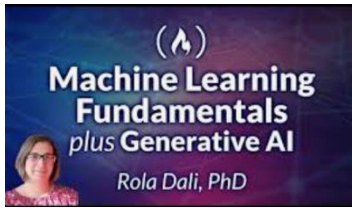
Will Agents Take My Job?
Weathering the Storm

- Learn AI, don't fear it
  - AI does not have to replace people, it can amplify them. AI is a tool.
- Fundamentals don't fade
- Move "up the abstraction ladder"
  - Defining problems, designing solutions, owning outcomes
- Thinking in Systems
  - See the bigger picture, understanding system components, integration points, …
- Be a polymath: Broaden your T of knowledge
- Some niches are more difficult for AI
  - Cutting edge fields, novel idea generation, fields with less data online, …
- Focus on the human element

References

- DeepLearning.AI: Agentic AI. Andrew Ng
- AI Engineering: Building Applications with Foundation Models. Chip Huyen
- Andrej Karpathy
- Coursera: Fundamentals of Building AI Agents: IBM
- Anthropic Guide to Building Agents
- LangChain Academy
- Anthropic Academy
- DeepLearning.ai Course Catalogue
- Stanford Online
- ML Introduction
- Linux Foundation MCP donation

Thank You



Tech42