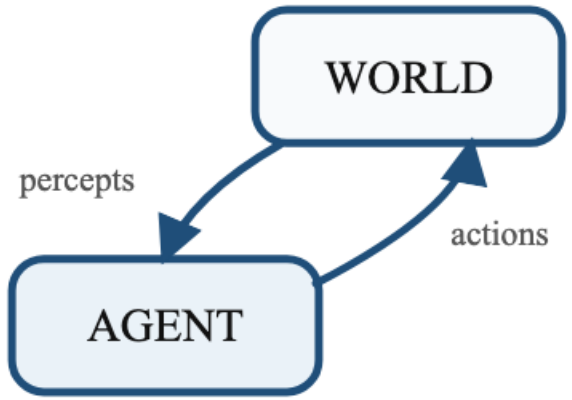
Most materials in this chapter are based on materials from https://ai.berkeley.edu



C0

Introduction

COMP3270
Introduction to Artificial Intelligence

SCHOOL OF COMPUTING & DATA SCIENCE
The University of Hong Kong

What Would Count as Intelligence?

TASK

A	B	C	D
Calculator	Chess Engine	Chatbot	Self-Driving Car

What evidence are you using?

How Should We Judge Intelligence?

What do we evaluate? THINKING ↔ ACTING

What is the benchmark? HUMAN-LIKE ↔ RATIONAL

Combine the answers → four ways to define AI

Build AI That…

	HUMANLY	RATIONALLY
THINK	**Think Humanly** Thinks like a human	**Think Rationally** Reasons correctly
ACT	**Act Humanly** Acts like a human	**Act Rationally** Chooses effective actions

Which Approach Guides Each AI System?

TASK

AI System	Approach
A hypothetical brain simulation that reproduces how humans think	
A chatbot designed to convince a judge that it is human	
An automated theorem prover that derives valid conclusions	
A self-driving car that chooses actions to reach its destination safely	

Thinking Humanly Models Human Cognition

TASK

OBSERVE

Human Thinking

Introspection · experiments · brain activity

→

MODEL

Computational Process

Turn a theory into a program

→

COMPARE

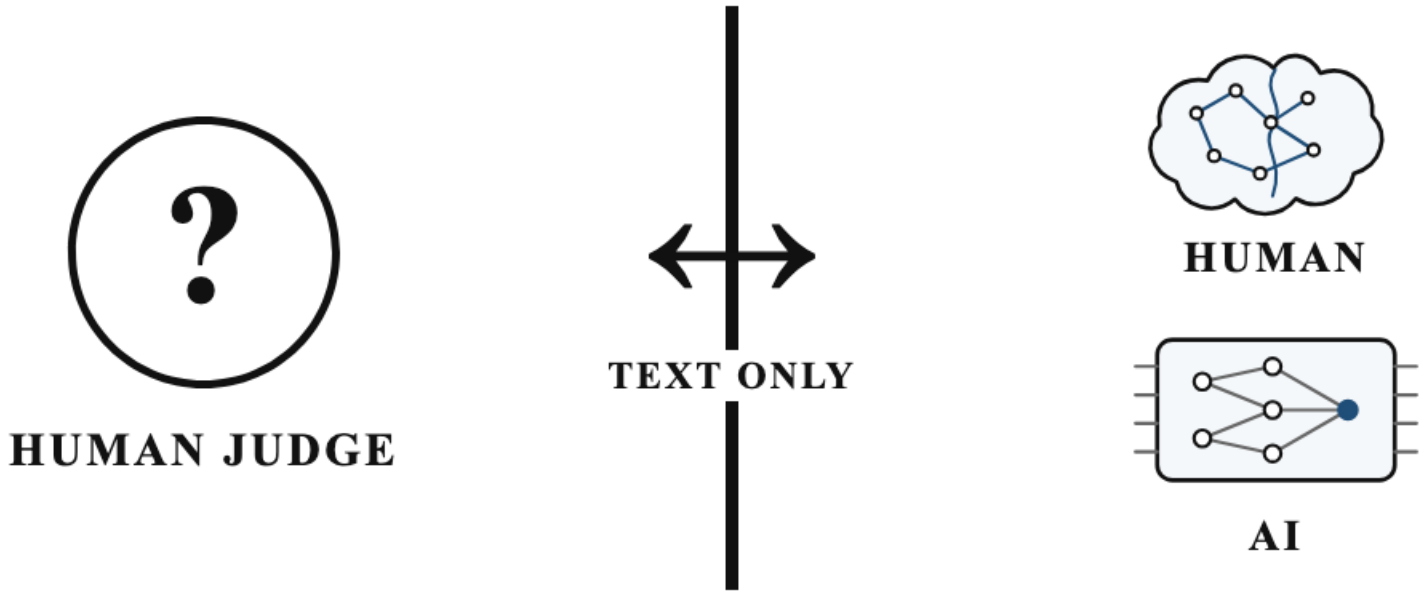
Model and People

Do they solve the task in the same way?

A person and a model give the same answer. Does that mean they thought in the same way?

Acting Humanly Tests Human-Like Behaviour

Can a judge reliably tell which response came from a person?



The test evaluates observable behaviour, not the hidden process

Thinking Rationally Applies Correct Reasoning

TASK

All men are mortal

Socrates is a man

Therefore, Socrates is mortal

What is the limitation of this approach?

This Course Focuses on **Acting Rationally**

DEFINITION

Choose the action expected to achieve the best outcome

Example: **Acting Rationally** Under Uncertainty

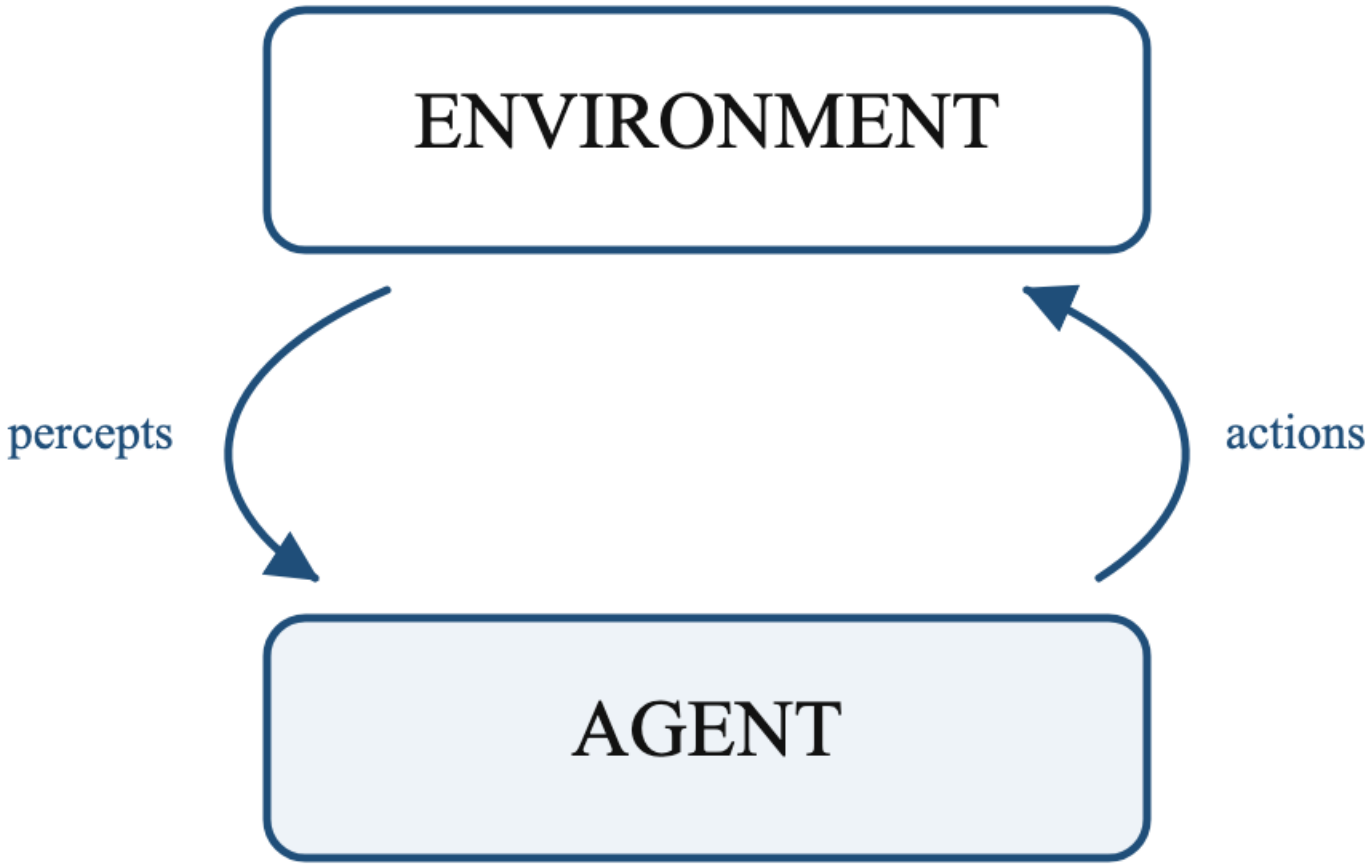
TASK

FORECAST		ACTION		OUTCOME
70%	→	☂	→	☀
chance of rain		take an umbrella		it stays dry

Was taking the umbrella irrational?

An Agent Perceives, Decides, and Acts

An agent is anything that **perceives** its environment and **acts** upon it.



The loop repeats as the world changes

What Made This Choice Rational?

Taking the umbrella depended on four things

OBJECTIVE	Stay dry
EVIDENCE	70% chance of rain
KNOWLEDGE	An umbrella reduces the risk
ACTIONS	Take it or leave it

Rational Does Not Mean Omniscient

RATIONAL $\neq$ KNOWS THE FUTURE

Use the information available now

Choose under limited time and memory

A rational decision can still lead to a bad outcome

A Self-Driving Car With One Objective

TASK

SELF-DRIVING CAR

Maximize safety

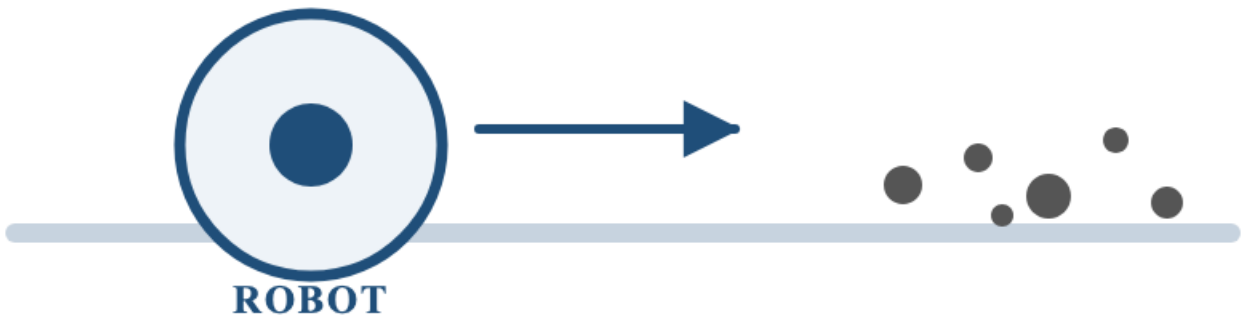
Which behaviour receives the highest score?

A Vacuum Robot With One Objective

TASK

VACUUM-CLEANING ROBOT

Collect as much dirt as possible



How could the robot maximize its score?

Specify the Performance Measure Carefully

PERFORMANCE MEASURE

How the agent's success is scored

POOR MEASURE

Amount of dirt collected

BETTER MEASURE

Cleanliness maintained over time

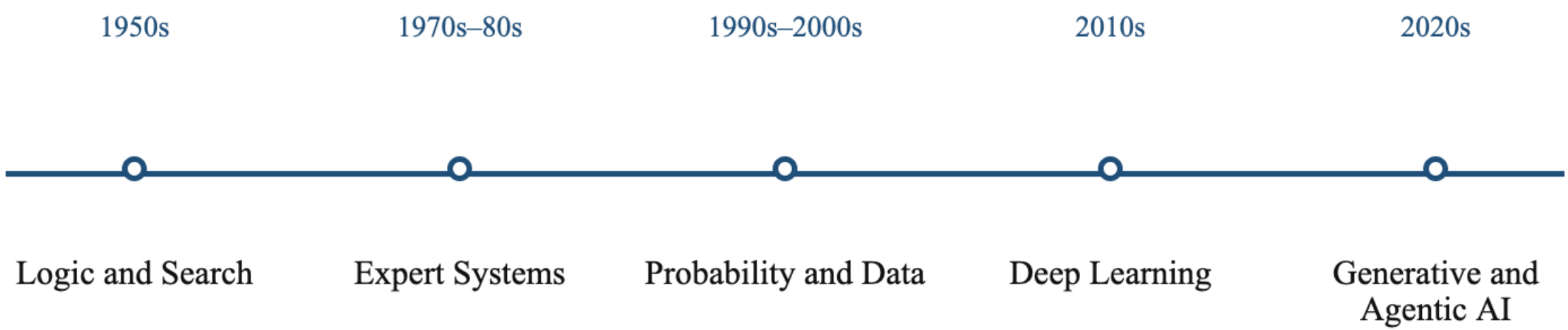
ALSO CONSIDER

coverage · energy · time · safety

This Course Builds a Toolkit for Rational Action

SEARCH · C1	Which action sequence reaches the goal?
MDPs · C2	Which action is best when outcomes are uncertain?
REINFORCEMENT LEARNING · C3	How can an agent improve from experience?
PROBABILISTIC MODELS · C4–C6	What should the agent believe from noisy evidence?
LOGIC · C7	What follows from what the agent knows?

AI Progress Comes in Waves



Each wave solves some problems and reveals new limitations

The Thinking Machine, 1961

TASK

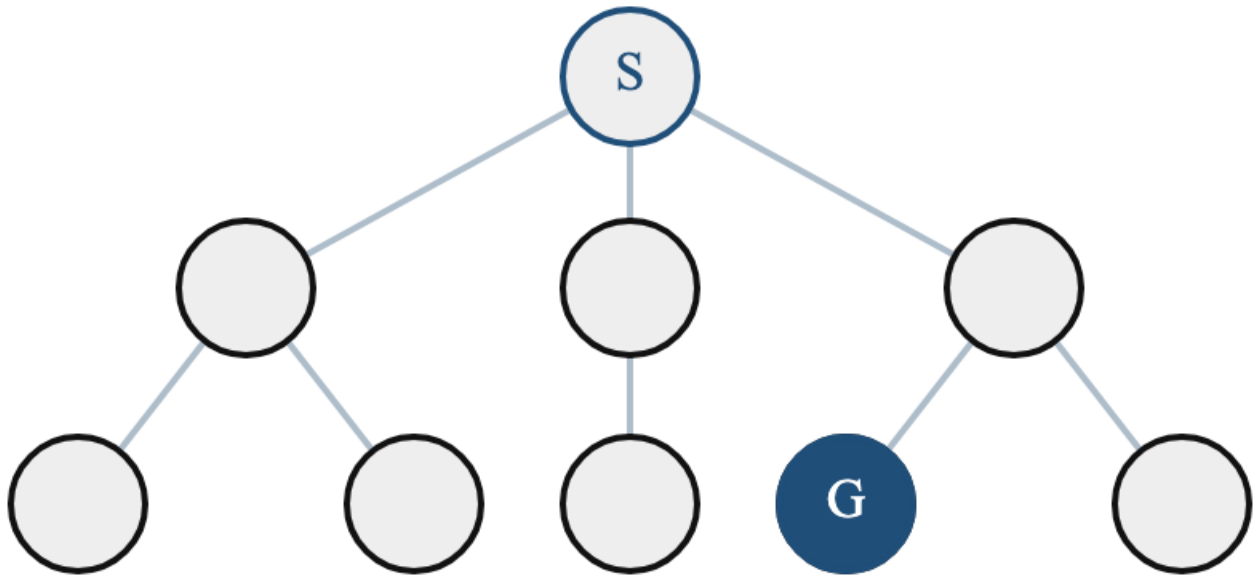
What stands out about the way they describe intelligent machines?

Reference Materials

Course Details	2-page course guide Course focus, schedule, assessment, and policies
AIMA: Chapter 1	66 pages comp3270.dirk.hk/AIMA_Chapter_1.pdf · Password: Schnieders
AIMA: Chapter 2	49 pages comp3270.dirk.hk/AIMA_Chapter_2.pdf · Password: Dirk
The Thinking Machine: MIT 1961	53-minute film comp3270.dirk.hk/the_thinking_machine_mit_1961.mp4

From Rational Agents to Search

An agent has a goal and several possible actions



How can it find a good action sequence without exploring every possibility?