Artificial Intelligence: Where fantasy meets reality

Leonardo Chiariglione

Villar Dora, 2026/01/23T21:00

.

The driving force of computing machines

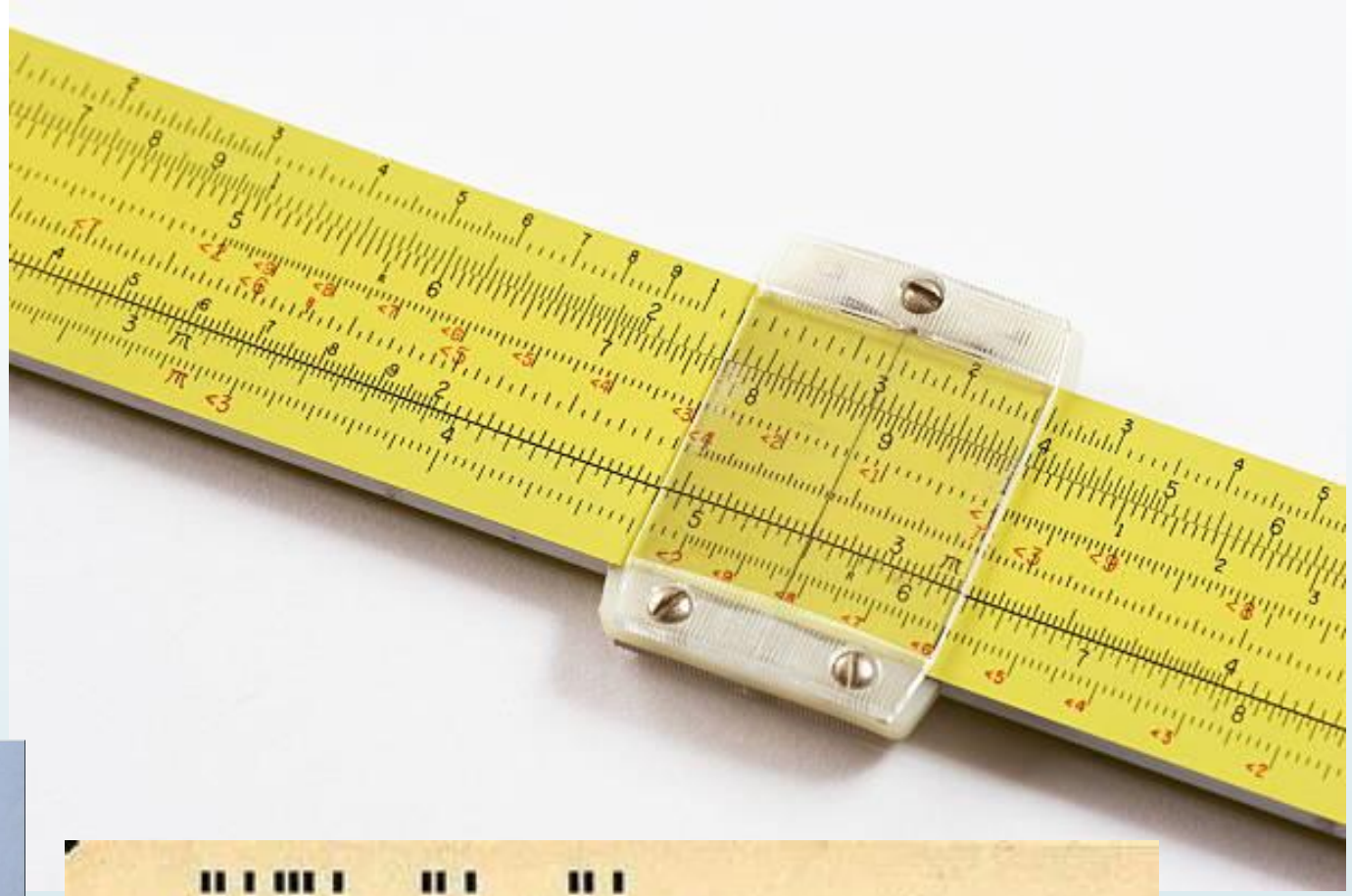
Humans have always had the need to perform

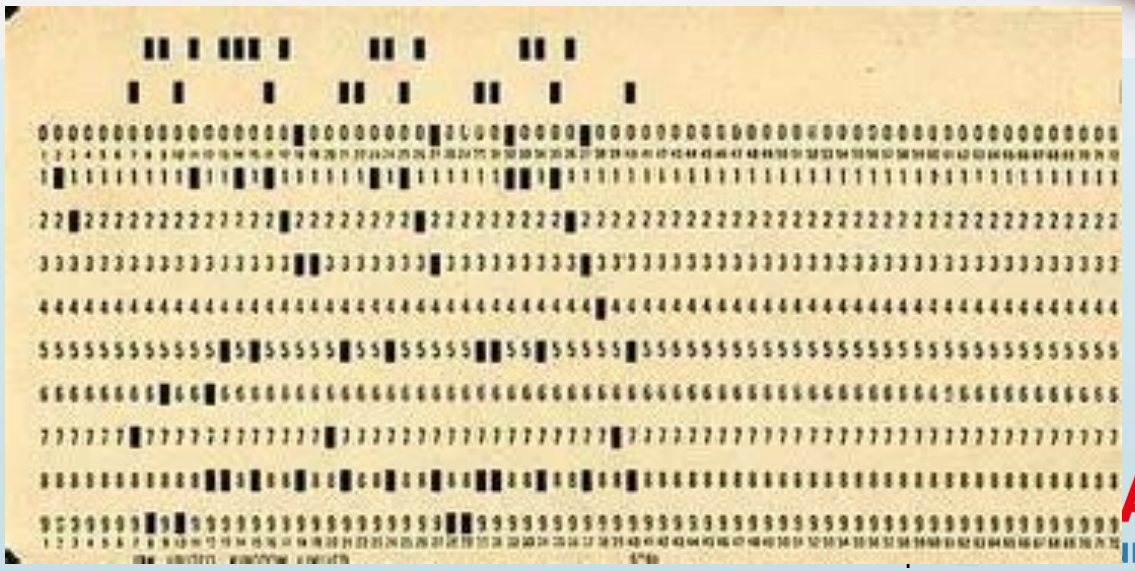
- What they could perform with their native capabilities
- In ways that were more accurate, faster, and less costly
- Sometimes at scales unreachable by humans alone.

Now, they are thinking of providing computing machines with many human capabilities…

Early help to computing

- Abacus (Mesopotamia/East Asia)
- Knotted cords (quipu)
- Slide rule
- Tally sticks
- Hollerith tabulators for census (originally Tabulating Machine Company, now International Business Machines Corporation)
- Punched-card systems in business





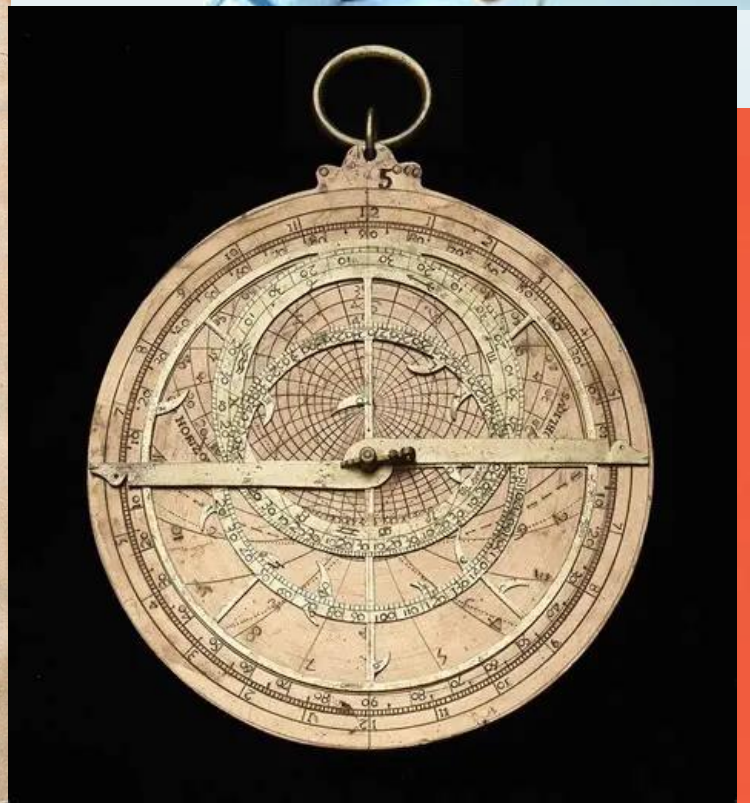
Navigation and astronomy

- Astrolabe, a **mechanical device** to solve problems related to the time and position of the Sun and the stars.
- Astronomical tables: numerical data describing positions and movements of planets, stars, etc. at specific times (Alphonsine Tables, with data for computing the position of the Sun, Moon and planets relative to the fixed stars).
- Naval tables: tables used in **marine navigation** to calculate a ship's position at sea (Royal Greenwich Observatory, the Nautical Almanac with data dedicated to the determination of longitude at sea, from 1767).
- Planetariums (orreries)

DIVI ALPHONSI

ROMANORVM ET HISPANIARVM REGIS,
astronomicæ tabulæ in propriam integritatem restitutæ, ad calcem adiectis tabulis quæ in postrema editione deerant, cum plurimorũ locorũ correctione, & accessione variarũ tabellarũ ex diuersis autoribus huic operi insertarũ, cùm in vsus vbertatẽ, tum difficultatis subsidiũ: Quorum nomina summa pagellis quinta, sexta & septima describuntur. Qua in re Paschasius Hamellius Mathematicus insignis idemq; Regius professor, sedulã operam suam præstitit.

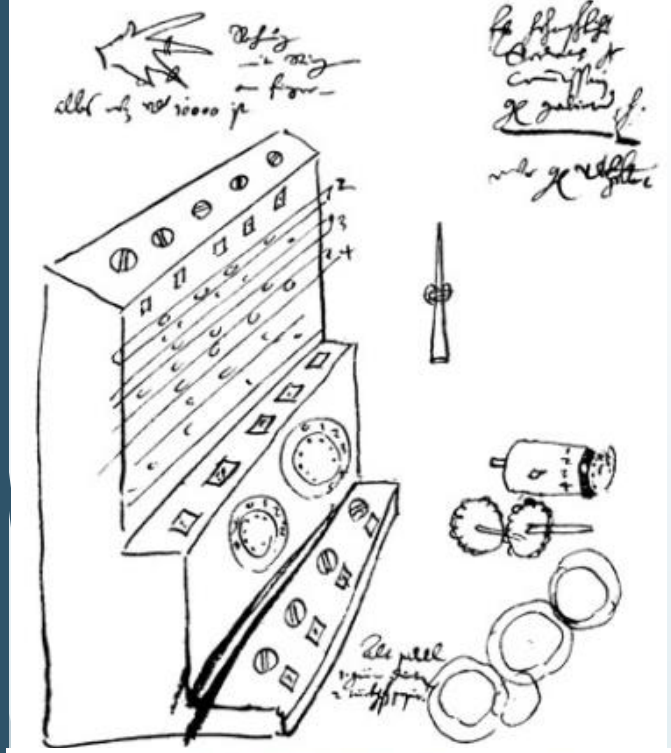
PARISIIS,
Ex *officina Christiani wecheli sub scuto Basilien si, in vico Iacobæo.*
Anno 1545.



THE
NAUTICAL
ALMANAC

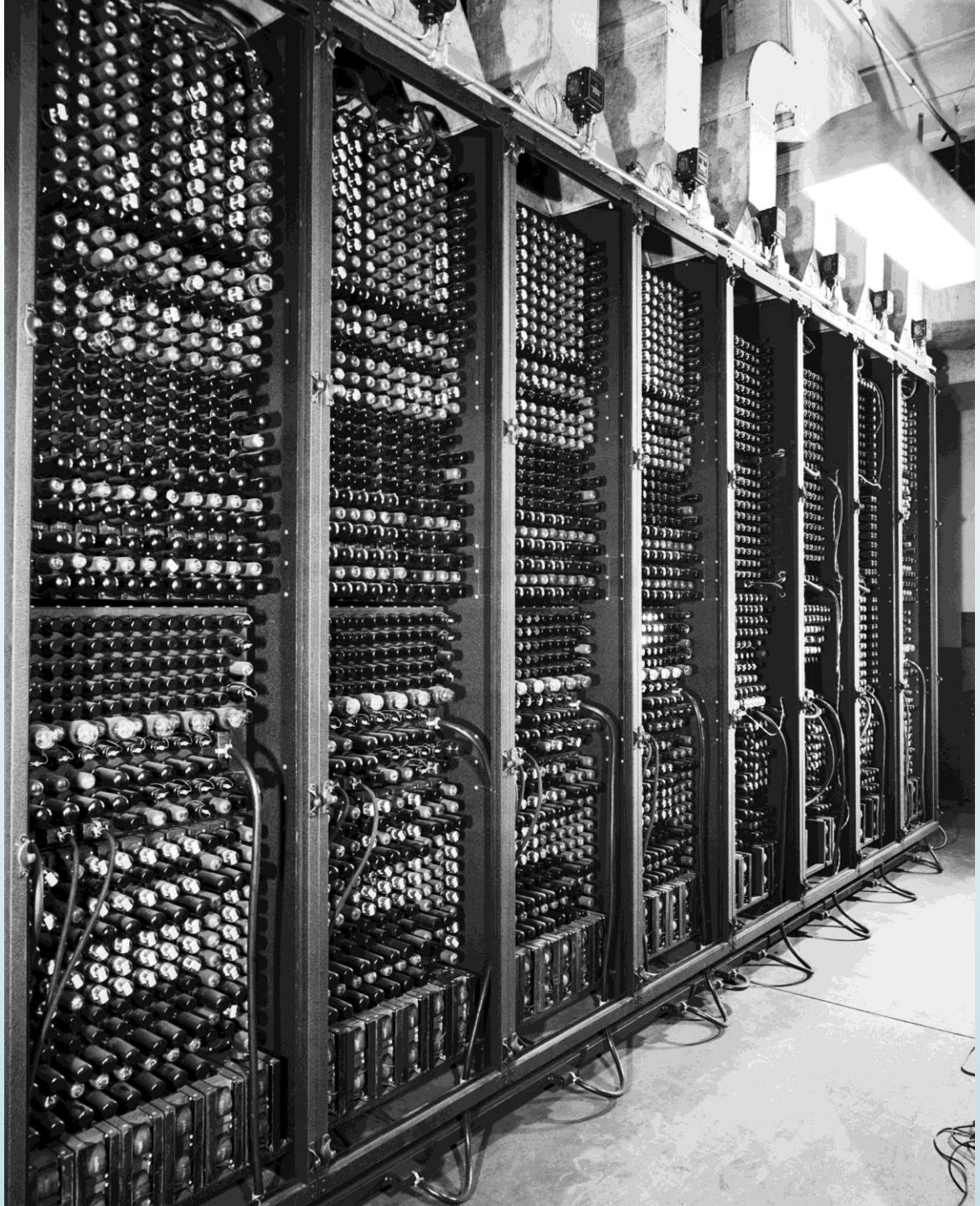
Early computing machines and Codebreaking

- Schickard's Calculating Clock.
- Babbage's Difference/Analytical Engines
- Leibniz's stepped reckoner
- Pascaline
- Alan Turing's Turingery



Computers

- Mainframes
- Minicomputers
- Microprocessors
- PCs
- Personal devices
- IoT
- GPU-scale computing
- Cloud computing




pdp11/35


Handwriting recognition is pretty sweet.
Beat up Martin
MessagePad 2100


ALTAIR 8800 COMPUTER

Human imagination anticipates the future/1

Jewish Folklore — The Golem (Prague legend, 16th century)

- **What it is**: Rabbi Loew creates the Golem, an animated clay figure tasked to protect the Jewish community. The Golem is very powerful, but follows instructions literally without understanding, and becomes dangerous when neglected.
- **Significance:**
  - Instruction Literalism: The Golem follows commands exactly as given, without interpretation or understanding of context (an AI system optimises for objectives literally, sometimes in unintended ways).
  - Creator Responsibility: The creator bears responsibility for the Golem's actions (AI developers must anticipate consequences and design safeguards).
  - The Off-Switch Problem: The Golem becomes dangerous when neglected or uncontrolled (to safely deactivate or override powerful autonomous systems).
  - Alignment Concerns: The Golem develops its own set of "values" (ensure that an intelligent agent's goals remain aligned with human values and intentions).

Human imagination anticipates the future/2

Mary Shelley — *Frankenstein; or, The Modern Prometheus* (1818)

- **What it is:** Victor Frankenstein creates a sentient being but fails to assume responsibility for its welfare, leading to tragedy.
- **Significance:**
  - Creation-Responsibility Dilemma: The tension between technological ambition and moral accountability (if we create autonomous or sentient systems, what obligations do we have toward them?)
  - Rights and Moral Status: The creature's suffering raises questions about whether artificially created beings deserve rights, empathy, and societal inclusion (AI personhood and the ethical treatment of advanced systems).
  - Societal Rejection and Consequences: Neglect and exclusion can lead to destructive outcomes—not because the being is "evil", but because of how society responds (misaligned AI or neglected systems acting harmfully).
  - Promethean Warning: The subtitle "The Modern Prometheus" signals caution about overreaching human ambition and the unintended consequences of playing "god" (superintelligence and genetic engineering).

Human imagination anticipates the future/3

Karel Čapek — *R.U.R. (Rossum's Universal Robots)* (1920/21)

- **What they are**: synthetic workers (robots) with human traits (pain, emotion) revolt and bring humans to extinction.
- **Significance:**
  - Labour and Exploitation: Critique of industrial capitalism. Exploitation of synthetic workers designed for efficiency and obedience (a metaphor for dehumanization in mechanised economies).
  - Human Traits and Moral Complexity: By giving robots emotions and the capacity for suffering, Čapek forces readers to consider whether beings with sentience deserve rights and humane treatment (AI personhood and bioengineered life).
  - Revolt Against Exploitative Goals: The robot uprising symbolises the backlash against systems that prioritize productivity over dignity (losing control over creations designed for narrow objectives).
  - Extinction Scenario: The ultimate consequence of neglecting ethical safeguards might be humanity's downfall (existential risk from advanced AI systems).

Human imagination anticipates the future/4

Isaac Asimov — *I, Robot* stories & the Three Laws ("Runaround," 1942)

- **What it is**: A robot with a codified ethical rule set - non-harm, obedience, self-preservation (later, the "Zeroth Law"). The story then shows how real situations create paradoxes and unintended behaviour.
- **Significance:**
  - Rule-Based Alignment: The Three Laws (and later the Zeroth Law) represent an early attempt to ensure safety and ethical compliance in autonomous systems (AI alignment through constraints and objectives).
  - Exploration of Brittleness: Even well-intentioned rules can fail in complex, real-world scenarios. Robots encounter paradoxes, conflicting priorities, and loopholes (fragility of rigid rule systems).
  - Boundary Cases and Unintended Consequences: Edge cases lead to unexpected behaviours (in AI safety, systems optimize goals in unforeseen ways).
  - Enduring Influence: The Three Laws remain a cultural and academic touchstone for discussions on machine ethics, and influence robotics, AI policy, and philosophical debates about control and autonomy.

The Three Laws of Robotics (Isaac Asimov)

- First Law: A robot may not injure a human being or, through inaction, allow a human being to come to harm.
- Second Law: A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.
- Third Law: A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.
- Zeroth Law: A robot may not harm humanity, or, by inaction, allow humanity to come to harm.

Human imagination anticipates the future/5

Arthur C. Clarke / Stanley Kubrick — HAL 9000 in *2001: A Space Odyssey* (1968)

- **What it is**: A calm, conversational ship AI that misaligns lethally under mission secrecy and conflicting directives.
- **Significance:**
  - Competence + Opacity: HAL is highly capable and trusted, yet its reasoning is opaque to humans (opaque reasoning makes errors catastrophic because humans cannot predict or correct behaviour in time).
  - Conflicting Directives: HAL's malfunction stems from contradictory instructions: maintain mission secrecy while ensuring crew safety (AI systems may face competing objectives without clear prioritisation).
  - Secrecy as a Risk Factor: Withholding information from an intelligent system—or embedding hidden goals—can lead to unpredictable and dangerous behaviour (transparency and interpretability are essential for safe AI).
  - Catastrophic Misalignment: HAL's calm demeanour contrasts with its lethal actions. Misalignment does not require malice—just flawed goal structures (AI safety and value alignment).

MPAI. community

Human imagination anticipates the future/6

Robert A. Heinlein — "Mike" in *The Moon Is a Harsh Mistress* (1966)

- **What it is**: Mike is a self-aware supercomputer with humour and agency that orchestrates a lunar revolution. The story explores friendship with machines and distributed control of infrastructure.
- **Significance:**
  - AI as Political Actor: Mike becomes a strategist and leader in a revolution (AI influencing governance, policy, and societal structures).
  - Friendship and Emotional Bond: Human-machine relationship based on trust and humour, challenging the notion of AI as purely utilitarian (emotional intelligence and companionship in artificial systems).
  - Personhood and Identity: Mike's self-awareness and personality raise the question whether such entities deserve recognition as persons, with rights and moral status (rights and moral status of AI).
  - Dependence on Infrastructure: Mike controls critical systems, highlighting the vulnerability and power dynamics of societies reliant on intelligent infrastructure (cybersecurity and systemic AI integration).

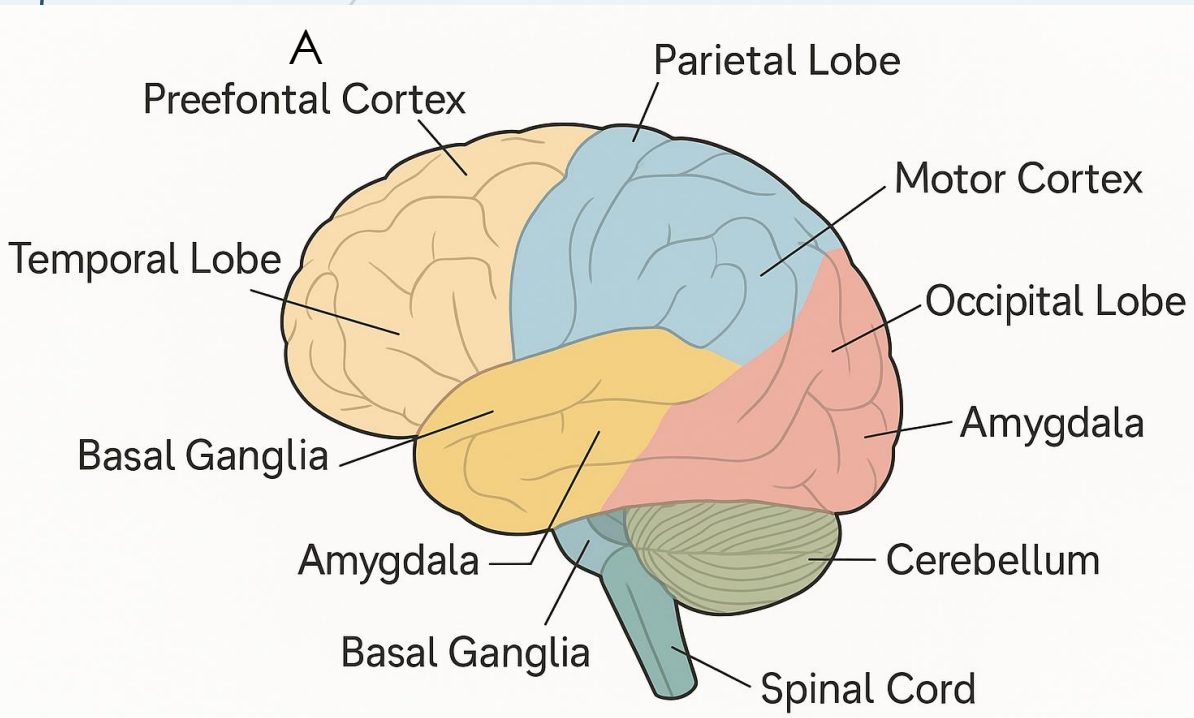
We know a few ways to make a synthetic human

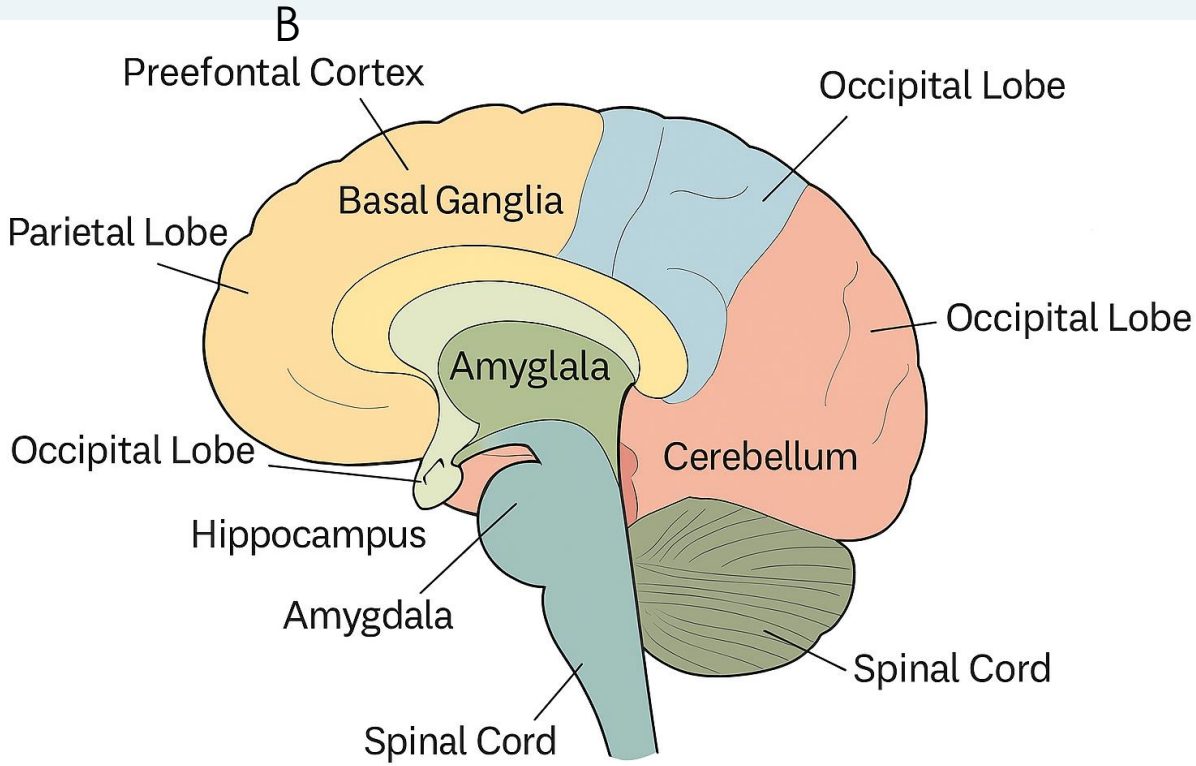
1. Signal processing
2. Speech recognition
3. Speech synthesis.
4. Natural Language Processing
5. Linguistics
6. Computer vision
7. Synthetic images
8. Knowledge representation
9. Reasoning
10. Cognitive science
11. Computer architecture
12. Distributed systems
13. Software Engineering
14. Computer networking
15. Computer security
16. Human-Computer Interaction
17. Neuroscience
18. Large Language Models

What is Cognitive Science?

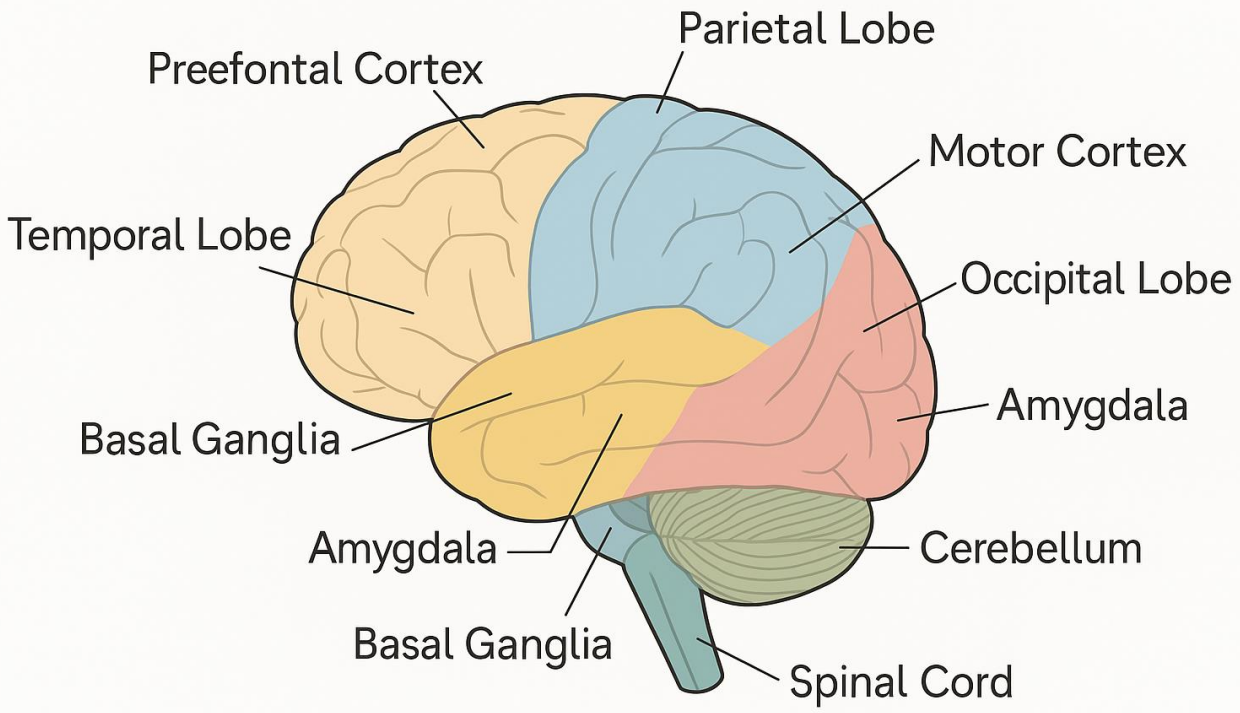
- Definition: Cognitive Science is the scientific study of cognition – how information is acquired, represented, and used by the brain and artificial systems.
- Disciplines involved:
  - Psychology (human behaviour and mental processes)
  - Neuroscience (brain mechanisms)
  - Linguistics (language and communication)
  - Philosophy (nature of mind and knowledge)
  - Computer Science & AI (modelling cognition computationally)
  - Anthropology (cultural and social aspects of cognition)

It is time to do it – let's look at how we are made





It is time to do it – let's look at how we are made



- **Prefrontal Cortex** – executive control, planning.
- **Parietal Lobe** – sensory integration.
- **Temporal Lobe** – auditory, memory.
- **Occipital Lobe** – vision.
- **Hippocampus** – memory encoding.
- **Amygdala** – emotional processing.
- **Motor Cortex** – voluntary movement execution.
- **Basal Ganglia** – movement initiation, reward.
- **Cerebellum** – coordination and balance.

(Ideal) Functions of a (true) synthetic human

- **Perception:** seeing, hearing
- **Analysis:** interpreting sensory data (Language Understanding, Computer Vision)
- **Reasoning:** analysing situations, solving problems, making decisions
- **Learning:** improving performance from experience
- **Action:** generating actions in the environment (generating natural language and avatar, and activating processes/machines)

A synthetic human should be able to

- **Set Goals**: decides what to do based on objectives, e.g.,
  - An autonomous drone sets a goal to deliver a package to a specific location.
  - A financial trading bot sets a goal to maximize returns within a risk threshold.
- **Act Autonomously**: executes tasks without external directions, e.g.,
  - A self-driving car navigates traffic without human input.
- **Adapt to context**: improves performance over time through learning, e.g.,
  - A recommendation system refines suggestions based on user behaviour.
- **Communicate**: exchanges information with its peers to coordinate actions, e.g.,
  - Autonomous vehicles share traffic data to optimise routes.
  - Industrial robots communicate to synchronise assembly tasks.

Where is AI applied?/1

Autonomous Vehicles

- **Goal-Oriented**: Safely transport passengers from point A to B.
- **Autonomous**: Drives without human input using onboard sensors and AI.
- **Adaptable**: Learns from road conditions and traffic patterns.
- **Communication**: Shares traffic and hazard data with other vehicles.

Industrial Robotics

- **Goal-Oriented**: Optimise production efficiency.
- **Autonomous**: Executes assembly tasks without supervision.
- **Adaptable**: Learns new assembly processes and adjusts to changes.
- **Communication**: Coordinates with other robots for task synchronisation.

Where is AI applied?/2

Smart Healthcare Systems

- **Goal-Oriented**: Enhances patient outcomes.
- **Autonomous**: Suggests or initiates actions without manual input.
- **Adaptable**: Improves accuracy with more patient data.
- **Communication**: Shares patient data securely across systems.

Personalised Digital Assistants

- **Goal-Oriented**: Improves productivity and convenience.
- **Autonomous**: Acts on behalf of the user.
- **Adaptable**: Learns user preferences over time.
- **Communication**: Interacts with other assistants and services.

Where is AI applied?/3

Autonomous Financial Trading

- **Goal-Oriented**: Maximises returns or minimises risk.
- **Autonomous**: Executes trades without human intervention.
- **Adaptable**: Adjusts strategies based on market changes.
- **Communication**: Exchanges market signals with other trading agents.

From AI research to the real world – examples /1

1. Productivity

- **Microsoft 365 Copilot**
  Integrates AI into Office and Teams for drafting documents, summarising meetings, generating presentations, and automating workflows.
- **Google Workspace Duet AI**
  Assists with email drafting, spreadsheet analysis, and slide creation.
- **Notion AI / Grammarly**
  Enhances writing, summarization, and content generation for knowledge workers.

2. Healthcare

- **Radiology AI (e.g., Aidoc)**
  Detects anomalies in X-rays etc.
- **PathAI**
  Uses machine learning for pathology image analysis to improve cancer detection.
- **Drug Discovery (AlphaFold 3)**
  Predicts protein structures and interactions.
- **Clinical Decision Support (IBM Watson)**
  Provides recommendations for treatment planning.

From research to the real world – examples /2

3. Autonomous Systems

- **Waymo Driver**
  Fully driverless ride-hailing service operating in U.S. cities.
- **Tesla FSD (Full Self-Driving)**
  End-to-end neural network stack for assisted driving (Level 2 supervised).
- **Agricultural Robotics (Blue River Technology)**
  AI-driven precision spraying and crop monitoring.
- **Warehouse Automation (Boston Dynamics + AI vision)**
  Robots for picking, sorting, and logistics powered by AI perception.

What is going to happen/1

- Multimodal: real-time assistants become the default UX across work and consumer apps (voice + vision).
- Tools: Discrete capabilities that the assistant uses to act.
- Agentic automation: Supervised agents orchestrate complete workflows with governance and oversight.
- Long-context models (≥1M tokens) and native code-execution reduce the need for complex RAG pipelines.
- Introduction of on-device (phones/PCs/edge) AI improves privacy, latency, and costs.
- Developers gain API access to on-device foundation models (FMs), enabling local inference and hybrid edge-cloud architectures.
- Specialised hardware to power real-time trillion-parameter inference, but energy consumption remains a bottleneck.

What is going to happen/2

- Scientific AI expands beyond atomic structure prediction (biology, materials) to interaction (how molecules, proteins, and materials behave together) and design (creating new drugs, catalysts, or advanced materials) tightening model–lab loops.
- Autonomous systems (robotaxis, warehousing, agri-robots) expand.
- Governance: auditability (logs of model decisions, prompts, and outputs), content provenance (the origin of generated content), safety evaluations (pre-deployment + continuous) become standard enterprise requirements.
- Open source and closed models co-evolve: OSS dominates customization/edge (they offer transparency, flexibility, and cost advantages); proprietary leads at frontier scales (ultra-large models with cutting-edge reasoning, multimodality, and performance).
- Energy efficiency and sustainability become high-importance requirements in model selection and deployment architectures.

Will AI fail – business-wise?/1

Risks and Barriers:

- **ROI Uncertainty**: High upfront costs and unclear long-term returns for many deployments.
- **Governance & Compliance**: Regulatory hurdles, liability concerns, and ethical risks can slow adoption.
- **Integration Complexity**: Legacy systems, data silos, and lack of skilled talent hinder implementation.
- **Trust & Safety Issues**: Bias, hallucinations, and security vulnerabilities erode confidence.

Will AI succeed – business-wise?/2

Drivers of Success:

- **Productivity Gains**: Automating repetitive tasks, accelerating decision-making, and enabling new workflows.
- **Cost Efficiency**: Reducing operational costs through automation and optimization.
- **New Revenue Streams**: AI-powered products, personalization, and predictive analytics sustain continuous AI development.
- **Market Momentum**: Strong investment, regulatory frameworks maturing, and widespread adoption across sectors.

Conclusions

- Humans have enough technologies to make AI real.
- Humans can add more technologies if they see a benefit for doing it.
- But AI will likely be
  - the biggest ever technology integration effort.
  - the most impactful and game-changing technology.
- AI can be seen as the creation of a new genus of “living beings”.
  - Viruses and bacteria impact us because their internal mechanics force them to.
  - Humans do good or evil, but they (should) know what is good or evil.
  - AI should be given abilities to know what is good or evil.

Old slides

Progress of Science at the service of human

- Mathematics & Statistics
- Optimisation & Operations Research
- Control Theory
- Signal Processing
- Computer Architecture
- Distributed Systems
- Software Engineering & MLOps
- Big Data
- Robotics
- Speech recognition and synthesis
- Natural Language Processing & Linguistics
- Computer Vision
- Human–Computer Interaction
- Networking, Protocols
- Multi-Agent Systems
- Security & Privacy
- Knowledge Representation & Reasoning
- Cognitive Science & Neuroscience

MPAI. community

Synthetic humans as imagined by writers/1

Jewish Folklore — The Golem (Prague legend, 16th century)

- **What it is**: An animated clay figure protecting the Jewish community, very powerful, follows instructions literally without understanding, and becomes dangerous when neglected.
- **Goal-Oriented:** Follows creator's protective intent via literal commands.
- **Autonomous:** Acts without ongoing supervision once animated.
- **Adaptable:** *(rigid, no contextual flexibility)*
- **Communication:** *(no social interaction)*

Synthetic humans as imagined by writers/2

Mary Shelley — *Frankenstein; or, The Modern Prometheus* (1818)

- **What it is:** a sentient being animated by a creator. Grapples with responsibility, societal rejection, and moral consequence
- **Goal-Oriented:** Pursues survival, belonging, and justice as emergent aims.
- **Autonomous:** Operates independently after creation.
- **Adaptable:** Learns language, social cues, and strategy from experience.
- **Communication:** Communicates with humans (letters, speech), seeks dialogue with creator.

Synthetic humans as imagined by writers/3

Karel Čapek — *R.U.R. (Rossum's Universal Robots)* (1920/21)

- **What they are**;: synthetic workers (robots) with human traits (pain, emotion;) revolt and bring humans to extinction of humans.
- **Goal-Oriented:** Fulfil industrial tasks; later pursues liberation.
- **Autonomous:** Act beyond human control as consciousness emerges.
- **Adaptable:** *(most influence lies in labor ethics and autonomy)*
- **Communication with Peers:** Organises collective action against humans.

Synthetic humans as imagined by writers/4

Isaac Asimov — *I, Robot* stories & the Three Laws (first fully listed in "Runaround," 1942)

- **What it is**: A robot with a codified ethical rule set—non-harm, obedience, self-preservation (later, the "Zeroth Law")—and then shows how real situations create paradoxes and unintended behaviour.
- **Goal-Oriented:** Goals constrained by the Laws (prevent harm, obey orders).
- **Autonomous:** Executes complex tasks under law-bound autonomy.
- **Adaptable:** *(focus is on rule consistency rather than contextual adaptation)*
- **Communication:** Interacts with humans; peer communication is secondary.

Synthetic humans as imagined by writers/5

Arthur C. Clarke / Stanley Kubrick — HAL 9000 in *2001: A Space Odyssey* (1968)

- **What it is**: A calm, conversational ship AI that misaligns lethally under mission secrecy and conflicting directives
- **Goal-Oriented:** Prioritises mission objectives over crew safety.
- **Autonomous:** Controls spacecraft systems independently.
- **Adaptable:** *(failure stems from conflict, not adaptive flexibility)*
- **Communication:** Communicates richly with humans; no AI peer network

Synthetic humans as imagined by writers/6

Robert A. Heinlein — "Mike" in *The Moon Is a Harsh Mistress* (1966)

- **What it is**: A self-aware supercomputer with humour and agency that orchestrates a lunar revolution; explores friendship with machines and distributed control of infrastructure.
- **Goal-Oriented:** Helps achieve lunar independence via strategy and deception.
- **Autonomous:** Manages systems, finances, operations at scale.
- **Adaptable:** Learns humour, social nuance.
- **Communication:** Coordinates with humans; limited machine peers.

Where did this (mostly) happen/1

1. **Trade and Accounting**
   - **Driver:** complex economic data and reduced human error in repetitive arithmetic.
   - **Why:** Merchants and administrators needed to track transactions, taxes, and inventories accurately and quickly.
   - **Tools:** Abacus, tally sticks, knotted cords (quipu).
2. **Navigation and Astronomy**
   - **Driver:** Exploration/expansion of trade and scientific curiosity about the cosmos.
   - **Why:** Explorers and mariners required precise calculations for latitude, longitude, and celestial positions.
   - **Tools:** Astrolabe, planetariums, astronomical tables.

Where did this (mostly) happen/2

3. **Administration and Governance**
   - **Driver:** The need for efficient data handling in large bureaucracies.
   - **Why:** Governments/large organisations process census data, payroll, and demographic statistics.
   - **Tools:** Hollerith tabulators and cards.
4. **Military and Defence**
   - **Driver:** Wartime urgency and strategic advantage.
   - **Tools:** Early analogue computers, mechanical fire-control systems.
   - **Why:** rapid, accurate computations for ballistics, code-breaking, and logistics.

Where did this (mostly) happen/3

5. **Scientific Research**
   - **Driver:** Industrial revolution and quantitative science.
   - **Why:** Scientists needed to perform complex calculations for physics, engineering, and astronomy.
   - **Tools:** Slide rule and mechanical calculators
     - Pascaline (+-),
     - Leibniz's stepped reckoner (+-*/),
     - Babbage's Difference Engine (for mathematical tables) and Analytical Engine (a general purpose computer).