FLASHFEED STUDY GUIDE

The Infinity Machine (book)

Unofficial study guide by FlashFeed

November 2026 edition · pack version 1.0.24

flashfeedlearningapp.com

The Infinity Machine (book): unofficial study guide

Sebastian Mallaby traces Demis Hassabis from chess prodigy to DeepMind cofounder: the Atari-playing DQN, AlphaGo's Move 37 against Lee Sedol, the AlphaFold breakthrough in protein folding, the OpenAI rivalry, and Google's Gemini comeback.

Unofficial study guide for *The Infinity Machine: Demis Hassabis, DeepMind, and the Quest for Superintelligence* by Sebastian Mallaby. FlashFeed is not affiliated with the author or publisher.

Edition	November 2026 (2026-11-21)
Pack version	1.0.24
Contents	12 lessons · 567 cards in the app · 273 self-test questions · 23 pictures
In the app	flashfeedlearningapp.com/packs/infinity-machine

How to use this guide

Take one lesson at a time: read the study guide, then go over any key facts. Where a lesson has a self-test, try it from memory before you check the answer key at the back. Follow the Go deeper links when you want more.

Free, with no ads and no account to make.

Contents

01	The Sweetness — Why this book matters	4
02	Destiny — Hassabis as a child prodigy	11
03	Cambridge and Elixir	18
04	Gang of Three — Legg and Suleyman	26
05	Founding DeepMind	32
06	Atari to AlphaGo	40
07	The OpenAI rivalry	50
08	AlphaFold — Fermat for biology	58
09	RaceGPT and Bletchley	67
10	Step by Step — Gemini and the comeback	76
11	Quick Recall — Key people and places	85
12	Key Concepts	90
	Answer key	98
	Sources	120
	Image credits	121
	Useful links	123
	About FlashFeed	124

LESSON 1 OF 12

The Sweetness — Why this book matters

Opening the book · The book's argument

WHAT YOU'LL BE ABLE TO DO

- Identify the author (Sebastian Mallaby), central subject (Demis Hassabis), and publisher of The Infinity Machine
- Explain 'the sweetness' as Hassabis's name for the instant a hard problem breaks open, and its tie to the AlphaFold Nobel Prize in Chemistry
- Define AGI (artificial general intelligence) and recognise 'the infinity machine' as Mallaby's label for it
- Explain Mallaby's thesis that DeepMind is a research lab pursuing AGI as a primary goal, built by people whose biographies and rivalries determine the outcome

Opening the book

Sebastian Mallaby — previously known for writing about hedge funds and venture capital — opens *The Infinity Machine* (Penguin Press, 2026) on a single charged moment: Demis Hassabis collecting the Nobel Prize in Chemistry. That scene earns its place as the book's frame because it crystallises what Mallaby is actually chasing.

The Sweetness of discovery

The sweetness is Hassabis's private name for the feeling that arrives at the instant a long-resistant problem finally breaks open. It is not public celebration — it is the moment *before* the announcement, alone with the solution. Mallaby uses it as his organising metaphor because it explains why anyone pours a decade into a single research problem: not for the prize or the paper, but for that internal signal. The 2024 Nobel in Chemistry, awarded for **AlphaFold**'s prediction of protein structures from amino-acid sequences, is the book's opening proof that the feeling is not romantic — it maps directly onto measurable, world-altering results.

What is the infinity machine?

The infinity machine is Mallaby's label for **artificial general intelligence**: a system capable, in principle, of solving any problem a human mind can frame — not one narrow task, but the full cognitive range. The phrase is deliberately grandiose, because Mallaby argues the ambition behind DeepMind has always been grandiose in exactly that way. The book's premise is that DeepMind offers the clearest window into how such a machine is actually being constructed — not through a single dramatic leap but through successive, linked breakthroughs in **reinforcement learning**, **deep learning**, and the transformer

architectures that followed. The title frames everything that comes after: this is a story about building something that may have no upper bound.

The book's argument

Mallaby is not writing a corporate history or a technical primer. He is making a specific argument about why DeepMind's particular path — its people, its bets, its rivalries — shaped what **AGI** (artificial general intelligence) looks like in practice. Three interlocking storylines carry that argument.

Three strands of the story

Mallaby braids three threads that cannot be read in isolation. First, Hassabis's personal arc: chess prodigy at thirteen, game designer in his twenties, neuroscience PhD, Nobel laureate at fifty-three. Second, the technical lineage — from the **Atari DQN** paper through **AlphaGo**, **AlphaFold**, and **Gemini** — tracing how each breakthrough made the next one imaginable. Third, geopolitics: the arms-race dynamic between Google DeepMind, OpenAI, Anthropic, and China's state-backed programmes. Mallaby's skill is showing that each strand reshapes the others; the science only makes sense alongside the biography, and both only make sense against the competitive pressure.

Why games matter to the plot

Hassabis built his early career on games — he designed *Theme Park* as a teenager and played chess to master-level. His conviction, which became DeepMind's founding wager, is that a **closed game environment** provides a clean reward signal against which an agent can measure its own intelligence and iterate fast. Winning at **Go**, **StarCraft**, and Atari was never the end goal; it was a proving ground for learning algorithms that could then face open-ended biological and physical problems. Every major DeepMind milestone began in a game.

The book's thesis in one line

Mallaby argues that the most consequential technology of the century is being built by people whose biographies and rivalries are not incidental to the outcome — they are *determinative*. DeepMind, founded out of a London poker game and a pitch at a Singularity Summit, has had influence disproportionate to its size. The lab's defining posture — a **research lab pursuing AGI as a primary goal**, not a product feature — shaped an entire field's self-understanding. Mallaby is explicit that Hassabis does not believe any single actor can slow the race down; he is a participant, not a brake.

Key terms

AGI

Artificial general intelligence — a system that can match or surpass human intelligence across the full range of cognitive tasks, not just one narrow domain. Coined by Shane Legg in 2002 in conversations with Marcus Hutter and adopted as DeepMind's founding mission.

Key concepts

The sweetness

- Hassabis's private name for a particular internal signal, and Mallaby's organising metaphor for the whole book.
- It arrives at the instant a long-resistant problem finally breaks open — the moment before any announcement, alone with the solution.
- Not the prize or the paper but this feeling, Mallaby argues, is why anyone pours a decade into a single research problem.

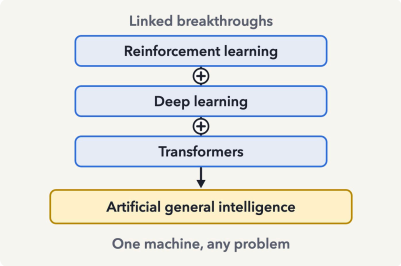
Nobel Prize in Chemistry

- A top international science award, presented in one of its scientific categories.
- The 2024 honour that Mallaby's book opens on, with Hassabis stepping up to collect it.
- It was awarded for AlphaFold's prediction of protein structures from amino-acid sequences.

The infinity machine

- Mallaby's deliberately grandiose name for the kind of system DeepMind has always aimed to build.
- His label for a mind that could, in principle, solve any problem a human can frame — the full cognitive range, not one narrow task.
- The book's central metaphor: a constructed intelligence the author suggests may have no upper bound.

Key facts



FlashFeed original diagram · CC-BY-4.0

What is the infinity machine?

The title is Mallaby's metaphor for artificial general intelligence: a single machine that could, in principle, solve any problem a human mind can frame. The book argues that DeepMind's history is the clearest case study of how such a machine is actually being built — not by a moonshot, but by linked breakthroughs in reinforcement learning, deep learning, and now transformers.

> **Written by Sebastian Mallaby**
>
> Sebastian Mallaby, previously known for writing about hedge funds and venture capital, is the author of The Infinity Machine.



Wikimedia Commons · John Sears · CC BY-SA 4.0

Hassabis, the book's subject

The Infinity Machine centers on Demis Hassabis — chess prodigy, game designer, neuroscience PhD, and Nobel laureate.

Penguin Press, 2026

The Infinity Machine was published by Penguin Press in 2026.

Three strands of the story

Mallaby braids three storylines: Hassabis's personal arc from chess prodigy to Nobel laureate; the technical lineage from Atari DQN through AlphaGo, AlphaFold and Gemini; and the geopolitics of an arms race between Silicon Valley labs, Google DeepMind, Anthropic, and China. None of the three is intelligible without the other two.

Why games matter to the plot

The book keeps returning to games — chess, Diplomacy, Theme Park, Atari, Go, StarCraft. Hassabis's conviction is that games provide a closed environment in which you can measure intelligence against a clear reward, then transfer the learning to the messy world. Every major DeepMind milestone first proved itself in a game.

The book's thesis in one line

Mallaby's argument: the most consequential technology of the century is being built by a small group of people whose biographies and rivalries matter — and DeepMind, born of a London poker game and a Singularity Summit pitch, has had outsized influence on what gets built and how.

DeepMind's AGI-first mission

Mallaby treats DeepMind's founding mission as a research lab pursuing AGI as a primary goal — a framing built into the company's earliest pitches and incorporation documents from 2010.

Hassabis can't slow the race

Mallaby writes that Hassabis doesn't believe any single actor can slow down the AI race — he sees himself as a participant in it, not someone positioned to act as a brake.

Matched pairs

Each item on the left goes with the one on the right.

Author of The Infinity Machine → Sebastian Mallaby

Central subject of the book → Demis Hassabis

Self-test

Answer each question from memory, then check your work against the answer key.

1. Which publishing house released The Infinity Machine in 2026?

- A MIT Press
- B Random House
- C Knopf
- D Penguin Press

2. Who wrote The Infinity Machine?

- A Sebastian Mallaby
- B Marcus Hutter
- C Shane Legg
- D Demis Hassabis

3. Who is the central subject of The Infinity Machine?

- A Shane Legg
- B Sam Altman
- C Demis Hassabis
- D Sebastian Mallaby

4. Before The Infinity Machine, Sebastian Mallaby was best known for writing about which subjects?

- A Neuroscience and cognitive psychology
- B Hedge funds and venture capital
- C Chess and competitive gaming
- D Nuclear weapons policy

5. Which statement about The Infinity Machine is NOT accurate?

- A It was published in 2026
- B Penguin Press published it
- C Demis Hassabis is the book's author
- D Sebastian Mallaby is the book's author

6. In the book, what does Hassabis mean by 'the sweetness'?

- A The moment a research grant is finally approved
- B The instant a long-resistant problem finally breaks open
- C The satisfaction of publishing ahead of a rival lab
- D The public ceremony where a discovery is honoured

7. The scene Mallaby opens the book with is Hassabis collecting which award?

- A The Turing Award
- B The Nobel Prize in Physiology or Medicine
- C The Nobel Prize in Physics
- D The Nobel Prize in Chemistry

8. The Nobel that frames the book's opening was awarded for which achievement?

A. Predicting protein structures from amino-acid sequences
B. Building the Gemini family of models
C. Training an agent to play Atari games from raw pixels
D. Defeating a world champion at the board game Go

9. Why does Mallaby make 'the sweetness' his organising metaphor?

A. It argues that discovery is a purely romantic notion
B. It explains why researchers pour a decade into one problem
C. It proves scientific discovery is mostly luck
D. It shows that prizes are the true motive for research

10. Which moment would Hassabis NOT count as 'the sweetness'?

A. Being alone with a solution before anyone knows
B. The private rush the instant an answer appears
C. Seeing a decade-old problem suddenly crack
D. Standing on stage as the prize is announced

11. 'The infinity machine' is Mallaby's label for what?

A. Artificial general intelligence
B. Reinforcement learning
C. DeepMind's Gemini model family
D. The transformer architecture

12. How does the book define AGI?

A. A machine that predicts protein structures at experimental accuracy
B. A system that beats human experts at one narrow, well-defined task
C. A system matching or surpassing humans across the full range of cognitive tasks
D. Any model trained with reinforcement learning on a closed game

13. Which system would Mallaby count as an infinity machine itself, rather than a step toward one?

A. A system that learns to play many arcade games from raw screen pixels
B. A system that masters a single board game well enough to beat the world's champions
C. A system able, in principle, to solve any problem a human mind can frame
D. A system that predicts a protein's three-dimensional structure from its amino-acid sequence

14. According to the book, how is the infinity machine actually being constructed?

A. By scaling one narrow game-playing agent indefinitely
B. Through successive, linked breakthroughs rather than one dramatic leap
C. Through a single moonshot project inside Google
D. By combining existing narrow systems into a committee

15. Mallaby's 'infinity machine' labels AGI. As the book defines it, such a system is capable in principle of doing what?

- A Mastering one narrow task better than any human
- B Predicting protein structures from amino acids
- C Winning closed games like Go and StarCraft
- D Solving any problem a human mind can frame

16. Mallaby treats DeepMind's mission statement as which kind of project?

- A A research lab pursuing AGI as a primary goal
- B A pure software product company
- C A consultancy advising other AI labs
- D A defense contractor for the UK government

17. What is Mallaby's one-line thesis about the century's most consequential technology?

- A The technology's path is determined by compute budgets alone
- B The builders' biographies and rivalries are determinative of the outcome
- C The outcome was fixed once the transformer was invented
- D Government regulation will decide what ultimately gets built

18. DeepMind was founded out of a London poker game and a pitch delivered at which event?

- A The Turing Institute launch
- B The Royal Society's AI symposium
- C A Google I/O keynote
- D The Singularity Summit

19. What does Mallaby say about Hassabis's view of slowing the AI race?

- A Hassabis has withdrawn from the race on principle
- B Hassabis argues regulation has already halted the race
- C Hassabis does not believe any single actor can slow it; he is a participant
- D Hassabis believes DeepMind alone can pause the race at will

20. Which is NOT part of Mallaby's thesis about DeepMind?

- A Its influence has been proportionate to its modest size
- B Its people's rivalries shaped what got built
- C It framed AGI as a primary goal, not a product feature
- D It shaped an entire field's self-understanding

Go deeper

ONLINE This lesson on the web flashfeedlearningapp.com/study-guides/infinity-machine/the-sweetness-why-this-book-matters

STUDY The Infinity Machine (book) in the FlashFeed app flashfeedlearningapp.com/packs/infinity-machine

LESSON 2 OF 12

Destiny — Hassabis as a child prodigy

Finchley and chess · Bullfrog and the games path

WHAT YOU'LL BE ABLE TO DO

- Identify Hassabis's 1976 birth and his Chinese Singaporean and Greek Cypriot parentage
- Explain Hassabis's chess prodigy years — master strength by 13, the Deep Blue vs Kasparov landmark, and his intuition that cognition has hardware ceilings
- Recognise the ZX Spectrum and Gödel, Escher, Bach as the formative computer and book behind Hassabis's early belief that intelligence could be built
- Trace Hassabis's move into game design, co-designing Theme Park at Peter Molyneux's Bullfrog studio

Mallaby titles his first chapter after a phrase Peter Thiel used in 2010: "Demis's destiny." But the chapter is about how that destiny was actually made — the specific north London childhood, the specific games, the specific decision to leave competitive chess. Mallaby's argument is that Hassabis was not born inevitable. He was, by his own description, born obsessive, and then made the right choices about what to be obsessed by.

Finchley and chess

Demis Hassabis was born in July 1976 in Finchley, a quiet residential corner of north London. His mother was Chinese Singaporean; his father, Greek Cypriot. The family lived above the toy shop his father ran, and from toddlerhood Demis was the kind of child who wanted to disassemble and understand every game in stock. He picked up chess at four, watching his father play. Within months he was beating adults.

By thirteen he had reached master strength — briefly the second-highest-rated player in the world for his age group. Mallaby spends a careful page on what competitive chess gave the young Hassabis: a deep intuition for adversarial systems, the discipline of sustained calculation, and (less obviously) an early intuition that human cognition has ceilings. He could feel his own brain straining at certain positions in a way that suggested the bottleneck was hardware, not effort.

At eight, with his chess winnings, he bought a ZX Spectrum home computer. He learned BASIC and machine code from manuals, and within a year he was writing his own programs — including, by his own later account, a primitive neural network that he didn't yet have the theory to understand. He just wanted the computer to play games against itself. The straight line from this experiment to AlphaZero, thirty years later, is the book's first and clearest example of obsession turning into vocation.

Bullfrog and the games path

The teenage Hassabis kept playing chess but lost interest in tournaments. The thing he wanted to build was bigger than a board. After his A-levels, he took a gap year before Cambridge and walked into Peter Molyneux's Guildford studio, Bullfrog Productions, looking for work. Molyneux — already a celebrity inside game design — gave the seventeen-year-old a job. Hassabis became a co-designer and lead programmer on *Theme Park*, the management sim that would go on to sell millions.

Three influences from this period matter for the rest of the book. First: Molyneux himself, who became one of Hassabis's most important mentors and modelled the importance of audacious goals. Second: Douglas Hofstadter's *Gödel, Escher, Bach*, which Mallaby singles out as Hassabis's most-loved book of his teens. Its braided account of self-reference, formal systems and strange loops of mind gave Hassabis his first articulate sense that intelligence might be something you could build out of layered patterns rather than something mystical. Third: an early brush with PDP-style backpropagation in 1991, which Hassabis could feel was important but underpowered. The technique stayed parked in his mind for two decades, until GPUs made it tractable.

By the end of his gap year, Hassabis had decided what he was. He was a games person; games were where intelligence could be measured cleanly. Mallaby is careful to puncture the standard biography here, which tends to present the games years as a colourful prelude to the real career in research. In Mallaby's reading, games were the real career. The PhD in neuroscience and the founding of DeepMind were what Hassabis did to build better games — first virtual ones, and then, when the technology was ready, games played against nature itself.

Key concepts

ZX Spectrum

- The home computer an eight-year-old Hassabis bought with his chess winnings.
- On it he taught himself BASIC and machine code straight from the manuals.
- Within a year he used it to write a primitive neural network — what the book calls a straight line to AlphaZero.

Theme Park

- The management sim the teenage Hassabis co-designed at Bullfrog.
- He was co-designer and lead programmer on it, working under Peter Molyneux.
- This management simulation went on to sell millions of copies.

Gödel, Escher, Bach

- Mallaby singles it out as Hassabis's most-loved book of his teens.
- Douglas Hofstadter's braided account of self-reference, formal systems and strange loops of mind.
- It gave Hassabis his first articulate sense that intelligence might be built from layered patterns rather than something mystical.

Bullfrog Productions

- The Guildford game studio a seventeen-year-old Hassabis walked into looking for work before Cambridge.
- Peter Molyneux's outfit, where Hassabis became a co-designer and lead programmer.
- The company behind Theme Park — the place Mallaby argues was Hassabis's real early career.

Key facts

Chess master at 13

Hassabis learned chess at four after watching his father play, and within months was beating adults. He reached master strength by age 13, briefly the second-highest-rated child player in the world for his age. The chess world taught him how to think about adversarial systems — but also showed him that humans had cognitive ceilings he could already feel pressing in.

ZX Spectrum and self-taught code

With chess winnings, eight-year-old Demis bought a ZX Spectrum home computer. He devoured manuals and taught himself BASIC and machine code. He coded a neural network long before he understood the theory — he simply wanted to make the computer play games against itself. The mental thread from those early experiments runs directly to AlphaZero.

What chess taught Hassabis

Competitive chess gave young Hassabis an early intuition that human cognition has limits — that the bottleneck is hardware, not effort, hinting intelligence might be something you could build.

His mother's heritage

Demis Hassabis's mother was Chinese Singaporean. His father was Greek Cypriot, giving Hassabis a mixed heritage often noted in profiles of his early life.

Gödel, Escher, Bach

Mallaby singles out a single book — Douglas Hofstadter's Gödel, Escher, Bach — as the most important text of Hassabis's teenage years. Its braided account of self-reference, formal systems and the strange loops of mind gave Hassabis his first sense that intelligence might be something you could build out of layered patterns rather than something mystical.

Quitting tournament chess

Hassabis stopped playing competitive chess in his late teens. The trigger, he later told Mallaby, was watching adults sacrifice everything to be slightly stronger than each other on a sixty-four-square board: "I started to wonder what else I could do with all that thinking." The chess habit never left him — but he wanted a bigger board.

> **A first brush with neural nets**
> In 1991, still a teenager, Hassabis read about the PDP (Parallel Distributed Processing) movement and experimented with backprop on his home machine. He could feel that the technique was important but underpowered. Twenty years later, when graphics cards made it powerful, he was already mentally rehearsed.

> **Co-designing Theme Park**
> As a teenager, Demis Hassabis co-designed the Bullfrog game Theme Park.

> **Mentor: Peter Molyneux**
> Peter Molyneux, already a celebrity in game design, ran the studio that hired the young Hassabis and became one of his most important mentors.

By the numbers

Age at which Hassabis reached chess master strength
13 years

Year Demis Hassabis was born
1976 year

Year Deep Blue defeated Garry Kasparov
1997 year

Matched pairs

Each item on the left goes with the one on the right.

Hassabis's mother's heritage → Chinese Singaporean

Hassabis's father's heritage → Greek Cypriot

Peter Molyneux's game studio → Bullfrog

Self-test

Answer each question from memory, then check your work against the answer key.

1. In which year was Demis Hassabis born?

- A 1980
- B 1984
- C 1976
- D 1972

2. What was the heritage of Hassabis's mother?

- A Chinese Singaporean
- B Singaporean Indian
- C Chinese Malaysian
- D Greek Cypriot

3. What was the heritage of Hassabis's father?

- A Turkish Cypriot
- B Chinese Singaporean
- C Greek mainland
- D Greek Cypriot

4. Which of these statements about Hassabis's origins is NOT accurate?

- A As a toddler he wanted to take apart every game in the shop
- B He was born in Guildford, Surrey
- C The family lived above the shop his father ran
- D He was born in Finchley, north London

5. In which month was Demis Hassabis born?

- A March
- B January
- C July
- D October

6. Beyond calculation skills, what less obvious insight did Mallaby say competitive chess gave the young Hassabis?

- A An intuition that human cognition has ceilings — that the bottleneck was hardware, not effort
- B A conviction that raw memorization beats intuition in complex games
- C A belief that computers could never rival human strategic intuition
- D A preference for teamwork over solitary calculation

7. At what age did Hassabis reach chess master strength?

- A 15
- B 11
- C 17
- D 13

8. Besides an intuition about cognitive limits, what did chess give the young Hassabis, per Mallaby?

- A Certainty that machines could never rival human minds
- B A deep feel for adversarial systems and calculation
- C Formal training in mathematical probability theory
- D A love of team-based collaborative problem-solving

9. At his peak as a junior player, what world ranking did Hassabis briefly reach for his age group?

- A Fifth-highest-rated
- B Second-highest-rated
- C Tenth-highest-rated
- D The top-ranked (1st)

10. Which of these statements about Hassabis's chess years is NOT accurate?

- A He took up chess at eight through a school chess club
- B He picked up chess at four, watching his father play
- C Within months of learning he was beating adults
- D He kept playing chess but lost interest in tournaments as a teenager

11. Which home computer did the eight-year-old Hassabis buy with his chess winnings?

A Amstrad CPC 464
B BBC Micro
C Commodore 64
D ZX Spectrum

12. Which book does Mallaby single out as the most-loved text of Hassabis's teenage years?

A The Emperor's New Mind
B The Selfish Gene
C Gödel, Escher, Bach
D The Society of Mind

13. Within a year of getting his home computer, what did Hassabis write that he did not yet have the theory to understand?

A A chess engine that beat his father
B A text adventure game
C A primitive neural network
D A BASIC compiler

14. What did Hofstadter's braided account of self-reference, formal systems and strange loops give Hassabis?

A A first sense that intelligence could be built from layered patterns rather than being mystical
B A method for training neural networks with backpropagation
C A conviction that intelligence requires a biological brain
D A formal proof that machine intelligence is impossible

15. Which of these was NOT part of Hassabis's self-teaching on his home computer?

A Learning BASIC from manuals
B A structured programming course run by his school
C Learning machine code from manuals
D Writing programs that made the computer play games against itself

16. Which Bullfrog game did the teenage Hassabis co-design?

A Dungeon Keeper
B Populous
C Black & White
D Theme Park

17. What was the teenage Hassabis's role on the Bullfrog management sim he worked on?

A Composer and sound designer
B Co-designer and lead programmer
C Junior QA tester
D Producer and marketing lead

18. Who ran the studio that hired Hassabis and became one of his most important mentors?

A Sid Meier
B David Braben
C Peter Molyneux
D Richard Garriott

19. How old was Hassabis when Molyneux gave him a job at the studio?

A 17
B 19
C 21
D 15

20. How does Mallaby's reading of Hassabis's games years differ from the standard biography?

A The standard version says he left games for good after his gap year; Mallaby says he returned to Bullfrog later

B The standard version dates the Bullfrog job too early; Mallaby moves it after his PhD

C The standard version overstates Molyneux's influence; Mallaby credits Cambridge instead

D The standard version treats games as a colourful prelude; Mallaby argues games were the real career

Go deeper

ONLINE This lesson on the web flashfeedlearningapp.com/study-guides/infinity-machine/destiny-hassabis-as-a-child-prodigy

STUDY The Infinity Machine (book) in the FlashFeed app flashfeedlearningapp.com/packs/infinity-machine

LESSON 3 OF 12

Cambridge and Elixir

Cambridge: deep philosophical questions · Elixir Studios

WHAT YOU'LL BE ABLE TO DO

- Identify the Cambridge influences on Hassabis — his friendship with David Silver, the board game Go, and the book Gödel, Escher, Bach
- Explain Hassabis's UCL neuroscience finding — hippocampal-damage patients couldn't imagine future scenes — leading him to view the brain as a generative model
- Recall that Hassabis founded Elixir Studios in 1998 (aged 22) and define what Elixir Studios was
- Identify Elixir's two best-known games — the political simulation Republic: The Revolution and the spy-villain game Evil Genius

Cambridge: deep philosophical questions

Hassabis arrived at Cambridge in 1994 three years older than most freshers, seasoned by his time at Bullfrog. For the first time he was surrounded by minds at least as fast as his own — and the experience focused rather than deflated him.

Cambridge: Queens' College

Hassabis read computer science at **Queens' College, Cambridge**, graduating in 1997 with a **double first**. The gap between school and university — spent as a professional game designer at Bullfrog — meant he arrived older and more directed than his peers. Cambridge was where, for the first time, his own intellect did not feel exceptional, and that shift in environment hardened his ambition rather than blunting it. He began to think seriously about what he actually wanted to build over the long term, beyond the next game.

Meeting David Silver

In his first weeks at Queens', Hassabis met a fellow fresher named **David Silver** in the Junior Common Room. They bonded over board games — **Go** in particular — and a shared frustration that the computer science syllabus was sidestepping what Hassabis called "the deep philosophical questions": what is intelligence, and what would it mean to construct it? It was the beginning of the longest professional partnership in Hassabis's career. Silver would go on to lead the **AlphaGo** project at DeepMind and later contribute to the reinforcement-learning work inside **AlphaFold**.

The deep philosophical questions

Cambridge crystallised Hassabis's conviction that the most interesting problem in computer science was the nature of mind itself. The tools of the mid-1990s were nowhere near ready — the compute didn't exist, the algorithms were immature — so he formed a long-range plan: spend a few years making money in games, then return to the question properly. Mallaby treats this chapter of the book as the

moment Hassabis stops being a precocious games designer and becomes, in intention at least, an AI researcher with a patient, multi-decade strategy.

PhD in cognitive neuroscience

After Elixir wound down, Hassabis enrolled at **UCL** to do a doctorate in **cognitive neuroscience** under **Eleanor Maguire**. His central finding: patients with damage to the **hippocampus** could not imagine future scenes, establishing that the same brain region responsible for storing memories is also responsible for constructing hypothetical futures. The paper appeared in *PNAS* in 2007 and was named one of the top ten scientific breakthroughs of that year by *Science*. For Hassabis it was proof of concept: the brain is a **generative model**, not a lookup table — and that idea would quietly structure DeepMind's entire research agenda.

Elixir Studios

Between Cambridge and the founding of DeepMind, Hassabis spent seven years running a games studio he built from scratch. It failed commercially and taught him lessons he would not have learned any other way.

Elixir Studios founded, 7 July 1998

Hassabis founded **Elixir Studios** on **7 July 1998**, aged 22, in central London. The studio's premise was the same one **Bullfrog** had taught him: ambitious AI baked into games at a level no other studio would attempt. Elixir's two best-known releases were **Republic: The Revolution** — a political simulation modelling thousands of citizens with individual goals, opinions, and relationships set against a fictional Eastern European revolution — and ***Evil Genius***, a satirical spy-villain strategy game released in 2004. Both were prized for their AI ambition; neither was a commercial smash.

The end of Elixir

Republic: The Revolution shipped in 2003. Critics admired the concept but found the gameplay sluggish; the hardware of the early 2000s could not support what Hassabis was asking of it. *Evil Genius* performed better but could not rescue the studio's finances. By **2005**, Elixir was wound down. Hassabis sold the rights to its completed games and walked away. He enrolled almost immediately at **UCL** for a PhD in cognitive neuroscience — a pivot that most people in his position would not have made, and that Mallaby treats as the most consequential single decision of Hassabis's pre-DeepMind career.

The lesson of Elixir

Mallaby extracts three durable lessons from the Elixir years. First, **ambition without focus is fatal**: trying to build multiple games simultaneously, each requiring a fresh AI architecture, spreads a small studio impossibly thin. Second, **hardware timing matters as much as ideas** — many of Elixir's concepts were correct in principle but premature in compute. Third, **research and product cannot easily share a deadline**: under shipping pressure, research always loses. At DeepMind, Hassabis would enforce the opposite discipline — one well-resourced **strike team** per grand-challenge problem, with the product question deferred until the science was settled.

Key terms

Elixir Studios

Elixir Studios was the games company Hassabis founded after leaving Bullfrog. Elixir Studios built ambitious AI-driven games but struggled commercially and closed in 2005.

Key concepts

Go

- A board game two computer-science freshers bonded over in their college's Junior Common Room in 1994.
- The shared passion of Hassabis and David Silver that sparked the longest partnership of Hassabis's career.
- This board game later became the subject of the celebrated DeepMind project David Silver led.

hippocampus

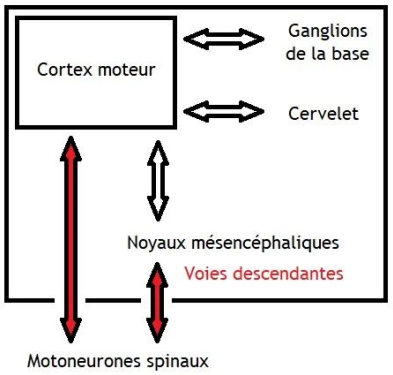
- A brain region Hassabis studied during his cognitive-neuroscience doctorate at UCL under Eleanor Maguire.
- He found that patients with damage to it could not imagine future scenes.
- His 2007 PNAS finding: the same structure that stores memories also constructs hypothetical futures.

generative model

- Hassabis's core conclusion about what the brain fundamentally is.
- Not a lookup table, he argued, but something that actively constructs — an idea that quietly structured DeepMind's research agenda.
- In machine-learning terms, the two-word name for a system that produces new data rather than merely retrieving stored answers.

Key facts

> **Meeting David Silver**
> In his first weeks at Cambridge, Hassabis met another freshman, David Silver, in the JCR. They bonded over Go, board games and a shared sense that the computer science syllabus was not asking the deep philosophical questions. Silver would later lead the AlphaGo project and, decades later, help design AlphaFold's reinforcement learning.



Wikimedia Commons · Mewtow · CC BY-SA 4.0

The deep philosophical questions

Mallaby titles a chapter after Hassabis's phrase. Cambridge convinced Hassabis that the most interesting question in computer science was not the next algorithm but the nature of mind itself. He resolved that, after a few years making money in games, he would return to neuroscience and AI for the long haul.

PhD in cognitive neuroscience

Hassabis's UCL doctorate, with Eleanor Maguire, found that patients with hippocampal damage could not imagine future scenes — the same brain region that stores memories also constructs hypothetical futures. The paper made the cover of Science in 2007 and gave Hassabis a working theory of the brain that would shape DeepMind's design.

Brain as generative model

Hassabis's UCL neuroscience PhD found that patients with hippocampal damage couldn't imagine future scenes, convincing him the brain works as a generative model that builds scenes rather than a lookup table that retrieves them.

Elixir Studios founded, 7 July 1998

On 7 July 1998, aged 22, Hassabis founded Elixir Studios in London. The plan was to build the kind of ambitious AI-driven games he had imagined since Bullfrog. Elixir would produce Republic: The Revolution and Evil Genius — neither a financial smash, but both prized for their AI ambition.

The end of Elixir

Elixir struggled commercially. By 2005, with the studio winding down, Hassabis sold the rights to the games it had completed and shut Elixir's gates. He returned to academia, enrolling at UCL to do a PhD in cognitive neuroscience — looking, specifically, at the brain mechanism behind memory and imagination.

The lesson of Elixir

Hassabis later said Elixir taught him that ambition without focus is fatal. Trying to build six games at once and reinvent AI for each one is how studios die. At DeepMind he would impose ruthless prioritisation: a single, well-resourced strike team on each grand-challenge problem.

Republic: The Revolution

Elixir's flagship political simulation was Republic: The Revolution, a 2003 release modelling thousands of citizens with their own goals — critics admired its ambition even though the AI overshot what the era's hardware could handle.

> **Founded Elixir at 22**
> Demis Hassabis was just 22 when he founded Elixir Studios, a year after graduating from Cambridge in 1997.

By the numbers

Year Hassabis founded Elixir Studios
1998 year

Matched pairs

Each item on the left goes with the one on the right.

Cambridge friend and lifelong DeepMind collaborator → David Silver

Elixir Studios' satirical spy-villain game → Evil Genius

Self-test

Answer each question from memory, then check your work against the answer key.

1. Which fellow fresher did Hassabis meet in the Junior Common Room at Queens', beginning the longest professional partnership of his career?

A Peter Molyneux
B Mustafa Suleyman
C David Silver
D Shane Legg

2. Hassabis and David Silver bonded over board games as Cambridge freshers. Which single game in particular did they bond over?

A Chess
B Go
C Backgammon
D Diplomacy

3. Mallaby singles out which book as the single most important one in Hassabis's intellectual formation?

A Gödel, Escher, Bach
B The Selfish Gene
C The Society of Mind
D A Brief History of Time

4. Which of the following is NOT true of David Silver?

A He later contributed to the reinforcement-learning work inside AlphaFold
B He would go on to lead the AlphaGo project at DeepMind
C He met Hassabis in the Junior Common Room at Queens'
D He was a Bullfrog colleague Hassabis recruited to Cambridge

5. Hassabis and Silver shared a frustration that the CS syllabus sidestepped what Hassabis called what?

A The hard problem of consciousness
B The deep philosophical questions
C The grand-challenge problems
D The limits of symbolic AI

6. Hassabis's UCL neuroscience PhD — patients with hippocampal damage couldn't imagine future scenes — led him to see the brain as what, an idea that shaped DeepMind's agenda?

A A generative model that constructs memories
B A blank slate shaped entirely by experience
C A relational database of stored memories
D A rule-based expert system

7. Which of the following is NOT true of Hassabis's UCL doctoral work?

A It convinced him the brain is a generative model
B It was supervised by Eleanor Maguire
C He carried it out in the years before he founded Elixir Studios
D It studied patients with damage to the hippocampus

8. Damage to which brain region left Hassabis's patients unable to imagine future scenes?

A The amygdala
B The cerebellum
C The prefrontal cortex
D The hippocampus

9. What did Hassabis's finding establish about the brain region that stores memories?

A It stores memories as exact, retrievable copies
B It is separate from the region handling imagination
C It is also responsible for constructing hypothetical futures
D It only becomes active during sleep consolidation

10. A patient with hippocampal damage is asked to describe lying on a beach they have never visited. Per Hassabis's finding, what happens?

A They construct it only after being shown a photograph
B They produce only fragmented, disjointed images and cannot bind them into a coherent scene
C They can construct it, but cannot recall any real past beach
D They construct it vividly but get the details wrong

11. On what date did Hassabis found Elixir Studios?

A 1 January 2005
B 7 July 1994
C 5 May 2003
D 7 July 1998

12. How old was Hassabis when he founded Elixir Studios?

A 22
B 26
C 17
D 30

13. Which best defines Elixir Studios?

A. The Cambridge research group where Hassabis and Silver studied Go
B. The games company Hassabis founded after leaving Bullfrog, building ambitious AI-driven games
C. The UCL neuroscience lab Hassabis joined for his PhD
D. The publisher that funded and distributed Bullfrog's games

14. Where did Hassabis set up Elixir Studios, and what premise did he build it on?

A. Central London, on Bullfrog's premise of ambitious AI baked into games
B. Cambridge, on the premise of commercialising academic neuroscience
C. Central London, on the premise of low-cost casual games at high volume
D. Guildford, on the premise of licensing other studios' AI engines

15. Which of the following is NOT true of Elixir Studios?

A. It built games prized for their AI ambition
B. It was founded in 1998 when Hassabis was 22
C. It was a commercial success that Hassabis sold at a profit
D. It was wound down by 2005

16. What was Elixir's flagship political simulation, modelling thousands of citizens with their own goals?

A. Tropico
B. Civilization V
C. Black & White
D. Republic: The Revolution

17. Which Elixir game was a satirical spy-villain strategy game released in 2004?

A. Theme Park
B. No One Lives Forever
C. Dungeon Keeper
D. Evil Genius

18. Republic: The Revolution was set against what backdrop?

A. A South American drug cartel war
B. A near-future Martian colony uprising
C. A fictional Eastern European revolution
D. A Cold War espionage network in Berlin

19. How did Elixir's two best-known games perform?

A. Both were prized for AI ambition, but neither was a commercial smash
B. Both were cancelled before release when Elixir closed
C. Both were commercial smashes but panned for weak AI
D. Republic sold well while Evil Genius was never finished

20. Critics admired Republic: The Revolution's concept but found the gameplay sluggish. Why?

A. Its publisher forced a rewrite into a different genre
B. Early-2000s hardware could not support what Hassabis asked of it
C. The studio shipped it a year before the AI was written
D. The game modelled only a dozen citizens, too few to feel alive

Go deeper

ONLINE This lesson on the web flashfeedlearningapp.com/study-guides/infinity-machine/cambridge-and-elixir

STUDY The Infinity Machine (book) in the FlashFeed app flashfeedlearningapp.com/packs/infinity-machine

LESSON 4 OF 12

Gang of Three — Legg and Suleyman

Shane Legg and the word AGI · Mustafa Suleyman

WHAT YOU'LL BE ABLE TO DO

- Identify Shane Legg as the coiner of 'AGI' (brainstorming a book title with Marcus Hutter, ~2002) alongside collaborator Jürgen Schmidhuber.
- Recognise IDSIA as the Swiss AI lab where Legg studied with Hutter and Schmidhuber
- Identify Mustafa Suleyman as DeepMind's third cofounder, his nickname 'Moose', and his 1984 birth
- Recognise the personal ties behind the founding — George Hassabis, who introduced Suleyman, and investor Luke Nosek, who heard Thiel's 'Demis's destiny' remark

Hassabis did not found DeepMind alone. Two cofounders matter as much as he does to the early story: Shane Legg, who brought the language of AGI and the safety conscience, and Mustafa Suleyman, who brought the operational ferocity. Mallaby treats the three of them as a study in complementarity — and, eventually, in fracture.

Shane Legg and the word AGI

Shane Legg was born in New Zealand in 1973 and drifted through cognitive science, finance and machine learning before landing at IDSIA, the Swiss AI institute in Lugano. There he worked with Marcus Hutter on theories of universal intelligence and with Jürgen Schmidhuber on recurrent neural networks. Around 2002, while Legg and Hutter were brainstorming a title for a planned book of essays, Legg proposed *Artificial General Intelligence*. The phrase stuck. By the time Hassabis met Legg, the term was the agreed name for the field's most ambitious goal.

What Hassabis found in Legg, however, was not just terminology. Legg was a serious thinker about AI safety years before it was fashionable. On 31 October 2009, Legg gave a Gatsby Unit lecture at UCL titled "The Halloween Scenario." His argument: powerful AI was no longer a thought experiment, and serious safety planning had to begin now. In the audience was Hassabis. They started talking that night and never really stopped.

Mallaby's Legg is the conscience of the early DeepMind triangle. He is the one who keeps asking the harder version of every question, who has, by Mallaby's account, been quietly worried about advanced AI for the entire history of the lab. The "Halloween Scenario" lecture is one of those small events the book pulls out and gives weight to: it was, in retrospect, the conversation that made DeepMind possible.

Mustafa Suleyman

The third corner of the founding triangle was, in many ways, the most unlikely. Mustafa "Moose" Suleyman was born in August 1984 to a Syrian taxi-driver father and an English mother in north London.

He dropped out of Oxford to start a Muslim youth helpline. By his mid-twenties he was a conflict-resolution consultant working in Copenhagen on climate diplomacy. He had no technical training in AI.

What he did have was Hassabis's brother George, who had been at school with him. Through George, Hassabis met Suleyman, and a relationship that began as friendship turned into a business partnership in 2010. Mallaby spends real time on this introduction because it is a reminder that DeepMind's third leg was not a computer scientist but an operator: somebody who could find capital, manage people, and pitch a billionaire at midnight. The lab would not have survived its first three years without him.

The crucial weekend Mallaby reconstructs in detail is mid-August 2010. On 15 July the three were playing poker at the Vic Casino in west London. By the end of the night they had agreed: their first fundraising target would be Peter Thiel. A month later, on 14 August 2010, the trio cornered Thiel at the Singularity Summit in San Francisco. Suleyman engineered the introduction. Hassabis pitched. The next morning at Thiel's Palo Alto house, over breakfast, Founders Fund was effectively committed. As they walked the three founders back to their car, Thiel muttered to his partner Luke Nosek: "Demis's destiny."

The book is careful about the limits of this triangle. By 2019, Suleyman would leave DeepMind under a cloud — a Bloomberg story published on 21 August 2019 detailed bullying allegations. Hassabis remained loyal in private and, Mallaby reports, was instrumental in arranging Suleyman's transition to Mountain View and then, years later, to Microsoft AI. The three north Londoners did not stay in the same building, but they stayed in roughly the same business — each running, by 2024, a different corner of one of the world's most powerful AI configurations.

Key concepts

Marcus Hutter

- An AI theorist based at IDSIA, the Swiss institute in Lugano.
- He and Shane Legg developed theories of universal intelligence together.
- Around 2002 he and Legg were brainstorming a book title when Legg proposed the phrase 'Artificial General Intelligence.'

Jürgen Schmidhuber

- A researcher at the Swiss AI institute IDSIA in Lugano.
- Shane Legg worked with him there on recurrent neural networks.
- A pioneer of recurrent neural networks, he was among Legg's collaborators before Legg helped found DeepMind.

George Hassabis

- The family link through whom Demis first met Mustafa Suleyman.
- Demis's brother, who had been at school with DeepMind's future operations man.
- It was through this sibling that Demis met the operator who became the lab's third founding leg.

Luke Nosek

- A partner at Peter Thiel's venture firm.
- He helped walk the three DeepMind founders back to their car after the Thiel breakfast.
- The Founders Fund partner to whom Thiel muttered 'Demis's destiny.'

Key facts

Shane Legg — coiner of "AGI"

Born in New Zealand in 1973, Shane Legg drifted through cognitive science, finance and machine-learning before landing at IDSIA in Lugano under Marcus Hutter and Jürgen Schmidhuber. While brainstorming a book title with Hutter around 2002, he proposed "Artificial General Intelligence". The phrase stuck. Legg would later cofound DeepMind.

The Halloween Scenario

On 31 October 2009 Legg gave a Gatsby Unit lecture at UCL titled "The Halloween Scenario". His thesis: powerful AI was no longer a thought experiment, and serious safety planning had to begin now. In the audience was Hassabis. They started talking that night and never really stopped.

Three cofounders, three temperaments

Mallaby's central triangle: Hassabis as the obsessive games-and-neuroscience polymath; Legg as the cerebral safety-aware theorist; Suleyman as the relentless operator who could pitch a billionaire at midnight. Each of the three would later run a different kind of AI organisation.

Legg's IDSIA collaborator

At the Swiss AI institute IDSIA, Shane Legg worked with Jürgen Schmidhuber on recurrent neural networks, and separately with Marcus Hutter on theories of universal intelligence.

A book title brainstorm

The term 'Artificial General Intelligence' was born around 2002, when Shane Legg and Marcus Hutter were brainstorming a title for a planned collection of essays.

Poker at the Vic, 15 July 2010

On 15 July 2010, Hassabis, Legg and Suleyman were playing poker at the Vic Casino in west London — they were already incubating DeepMind. By the end of the night they had agreed: the first major fundraising target would be Peter Thiel. Five weeks later they would corner him in California.

Singularity Summit, 14 August 2010

At the 14 August 2010 Singularity Summit in San Francisco, the trio targeted Peter Thiel. Suleyman engineered an introduction. Hassabis pitched a research lab whose explicit mission would be to build AGI safely. Thiel listened, said little, and invited the team to his Palo Alto house for breakfast the next day.

> **Suleyman's birth year**
> Mustafa Suleyman was born in August 1984 in north London — a detail easy to mix up with cofounder Shane Legg's birth year of 1973.

By the numbers

Year Mustafa Suleyman was born
1984 year

Matched pairs

Each item on the left goes with the one on the right.

Coined the term "AGI" → Shane Legg

Swiss AI lab where Legg studied with Hutter → IDSIA

Suleyman's nickname inside DeepMind → Moose

Self-test

Answer each question from memory, then check your work against the answer key.

1. Around 2002, in what context did Shane Legg propose the phrase 'Artificial General Intelligence'?

(A) Titling his doctoral dissertation at IDSIA
(B) Brainstorming a title for a planned book of essays with Marcus Hutter
(C) Naming a new AI research lab he was founding
(D) Christening a recurrent-neural-network project with Jürgen Schmidhuber

2. Who proposed the phrase that became the field's agreed name for its most ambitious goal?

(A) Demis Hassabis
(B) Marcus Hutter
(C) Shane Legg
(D) Jürgen Schmidhuber

3. Around what year did Shane Legg propose the term "Artificial General Intelligence"?

(A) 2002
(B) 2010
(C) 2009
(D) 1997

4. At IDSIA, which collaborator did Legg work with on recurrent neural networks?

(A) Geoffrey Hinton
(B) David Silver
(C) Marcus Hutter
(D) Jürgen Schmidhuber

5. By the time Hassabis met Legg, what standing did the phrase Legg had proposed already have?

A It had been dropped in favour of "universal intelligence"
B It was still a private joke between Legg and Hutter
C It had not yet been proposed
D It was the agreed name for the field's most ambitious goal

6. Which Swiss AI institute did Shane Legg land at, working with Marcus Hutter and Jürgen Schmidhuber?

A Idiap Research Institute
B IDSIA
C ETH Zurich
D EPFL

7. In which city is IDSIA located?

A Lugano
B Lausanne
C Geneva
D Zurich

8. Which of these is NOT connected to IDSIA?

A Jürgen Schmidhuber
B Marcus Hutter
C Demis Hassabis
D Shane Legg

9. What did Legg and Hutter work on together at IDSIA?

A Theories of universal intelligence
B Climate-diplomacy modelling
C Reinforcement learning for Atari games
D Protein-structure prediction

10. Before landing at IDSIA, Shane Legg had drifted through which fields?

A Neuroscience, chess and video-game design
B Cognitive science, finance and machine learning
C Conflict resolution and climate diplomacy
D Law, journalism and public policy

11. Who was the third cofounder of DeepMind alongside Hassabis and Legg?

A Ilya Sutskever
B Geoffrey Hinton
C Mustafa Suleyman
D David Silver

12. What was Mustafa Suleyman's nickname inside DeepMind?

A Sulley
B Mus
C Mo
D Moose

13. In what year was Mustafa Suleyman born?

A 1990
B 1973
C 1980
D 1984

14. Which statement about Suleyman before DeepMind is NOT true?

A He was a conflict-resolution consultant in Copenhagen
B He dropped out of Oxford
C He held a doctorate in machine learning
D He started a Muslim youth helpline

15. In Mallaby's triangle of the three cofounders, how is the third cofounder characterised?

A The relentless operator who could pitch a billionaire at midnight
B The recurrent-neural-network specialist from Lugano
C The obsessive games-and-neuroscience polymath
D The cerebral, safety-aware theorist

16. Through whom did Demis Hassabis first meet Mustafa Suleyman?

A Luke Nosek
B George Hassabis
C Shane Legg
D Peter Thiel

17. What was George Hassabis's relationship to the two men he connected?

A Demis's brother, who had been at school with Suleyman
B Demis's business partner, who met Suleyman at Oxford
C Demis's brother, who met Suleyman at the Vic Casino
D Demis's cousin, who worked with Suleyman in Copenhagen

18. Peter Thiel muttered "Demis's destiny" to which person?

A Shane Legg
B George Hassabis
C Luke Nosek
D Mustafa Suleyman

19. Who was Luke Nosek in this episode?

A An organiser of the Singularity Summit
B DeepMind's first employee
C Suleyman's colleague from his Copenhagen consulting work
D Peter Thiel's partner at Founders Fund

20. In what setting was the "Demis's destiny" remark made?

A On stage at the Singularity Summit in San Francisco
B Walking the founders back to their car after breakfast at Thiel's Palo Alto house
C During Legg's "Halloween Scenario" lecture at UCL
D Over poker at the Vic Casino in west London

Go deeper

ONLINE This lesson on the web flashfeedlearningapp.com/study-guides/infinity-machine/gang-of-three-legg-and-suleyman

STUDY The Infinity Machine (book) in the FlashFeed app flashfeedlearningapp.com/packs/infinity-machine

LESSON 5 OF 12

Founding DeepMind

Thiel and the pitch · Google takes the wheel

WHAT YOU'LL BE ABLE TO DO

- Explain the DeepMind name's origin (first named Solaria, after an Asimov novel) and that the lab was founded and registered in 2010
- Recall Google's 2014 acquisition of DeepMind for roughly £400 million / $600 million and Larry Page's role
- Recognise the Silicon Valley backdrop — Peter Thiel's Founders Fund as first backer and the October 2012 Musk–Page dinner where Musk warned of a 'digital god'
- Define Deep Q-Networks and Q-learning, identify Vlad Mnih as the Atari DQN lead, and explain why experience replay made deep RL stable
- Define DeepMind as the London AI lab Hassabis cofounded in 2010 and Google acquired in 2014

Thiel and the pitch

DeepMind's founding moment was equal parts intellectual ambition and tactical fundraising — a London startup pitching Silicon Valley money on the strength of a vision most investors couldn't evaluate.

The name DeepMind

The company's first working name was **Solaria**, borrowed from an Isaac Asimov novel. Shane **Legg** pushed for something plainer. The founders settled on two words that doubled as a research manifesto: **"Deep"** nodded to **deep learning** — the neural-network approach they were betting on — and **"Mind"** declared the destination. The name was registered in late **2010**, and the London company itself was incorporated that same year, officially launching in November 2010. Choosing the name before they had a product was itself a statement: they were not building another software tool, they were building toward cognition.

Google buys DeepMind, January 2014

Larry Page and Sergey Brin had been watching DeepMind since 2012. In **January 2014** Google closed the acquisition for roughly **£400 million** — then the largest AI acquisition in European history, for a lab with no revenue and fewer than 100 employees. **Peter Thiel's Founders Fund** had led the earliest outside funding, a notable irony given Thiel's later public skepticism about unchecked AI development. The price signaled that Google understood deep learning was not a research curiosity but infrastructure for everything it built.

Musk vs Page on the lawn

On **8 October 2012**, Elon Musk attended a dinner at Larry Page's home in Palo Alto. The evening turned philosophical. Musk argued that a sufficiently advanced AI could become **"a digital god"** — and that humanity had no plan for what happens next. Page disagreed: the species that built such an intelligence

would be its natural heirs. Sebastian Mallaby, whose account this draws from, frames that night as the moment the informal Google-Musk alignment cracked, planting the seed of what would later become **OpenAI**.

The (largely silent) ethics board

Hassabis made one structural demand before signing the Google deal: an independent **Ethics and Safety Review Board**, with external members who had to be informed before any AGI-relevant breakthrough was deployed. In practice the board met rarely and was later absorbed into Google's broader governance structures. The gap between the commitment and the reality previewed a recurring pattern across the industry — **ethics boards** as legitimacy signals rather than operational constraints. The template, flawed as it was, seeded every subsequent attempt at AI governance.

Google takes the wheel

Inside a freshly acquired DeepMind, a small team quietly ran the experiment that would define the company's next decade — teaching a neural network to play video games it had never seen, from nothing but pixels and a score.

The Atari strike team

The project that put DeepMind on the map came from a handful of researchers led by **Vlad Mnih**, trained in reinforcement learning at Toronto and Alberta. The challenge was deliberately blunt: a single neural network, no game-specific code, raw pixels as input, and a score signal as the only feedback — could it learn to play any of the classic **Atari 2600** library? The constraint forced a general solution rather than a tailored one, which was the point.

Why experience replay mattered

Deep Q-Networks (DQN) combined two existing ideas in a way that finally made deep reinforcement learning stable. The first was **Q-learning** — a technique in which an agent estimates the long-term expected reward of every possible action from every state, iteratively refining those estimates through play. The second was **experience replay**: storing past observations in a memory buffer and training the network on randomly sampled batches rather than consecutive moments of gameplay. Neural networks trained on sequential data tend to overfit to recent experience and forget earlier lessons; shuffling the buffer broke that correlation and gave the network something closer to independent, identically distributed training examples. That combination made deep RL reliable enough to be a research tool.

The day Pong clicked

On **7 July 2013** — the day of the Wimbledon men's final — the team watched the DQN agent begin to learn **Pong** from a blank slate. Hours in, the agent's paddle started tracking the ball. By evening it was defeating the built-in opponent consistently. Several people in the room reportedly wept. The moment was not just a demo; every subsequent DeepMind milestone — **AlphaGo**, **AlphaFold**, **Gemini** — traces its intellectual lineage through the architecture and confidence earned that afternoon.

Key terms

DQN (Deep Q-Network)

A reinforcement-learning architecture that pairs Q-learning — predicting the long-term value of each possible action — with a deep convolutional network that reads raw pixels. DQN, with experience replay for stable training, was DeepMind's first headline result and the model for everything that followed.

Q-learning

A reinforcement-learning algorithm in which an agent learns the expected long-term reward of taking each action from each state. The Q stands for the quality of an action — and the trick is iteratively updating the estimate as new rewards arrive.

DeepMind

DeepMind is the London AI lab Demis Hassabis co-founded in 2010 and Google acquired in 2014. DeepMind's stated mission is to solve intelligence and then use it to solve everything else.

Key concepts

Isaac Asimov

- A science-fiction author whose novel supplied DeepMind's first working name.
- The company was briefly called Solaria, borrowed from one of his books.
- This SF writer's work gave the startup its placeholder name before the founders settled on 'Deep' plus 'Mind.'

Larry Page

- A Google co-founder who had been watching DeepMind since 2012.
- He hosted the October 2012 Palo Alto dinner where the future of AI was argued.
- He countered Musk by insisting the species that built a superintelligence would be its natural heir.

Elon Musk

- A guest at Larry Page's 2012 Palo Alto dinner who clashed with the host over AI.
- He warned that a sufficiently advanced AI could become 'a digital god,' with no plan for the aftermath.
- Mallaby frames his split with Page that night as the seed of what became OpenAI.

Key facts

> **The name DeepMind**
>
> The company's first working name was Solaria, after the Asimov novel. Legg argued for something simpler. "Deep" pointed to deep learning, the technique they would bet on; "Mind" stated the goal. The name was registered in London in late 2010, the same year the company itself was founded.

The (largely silent) ethics board

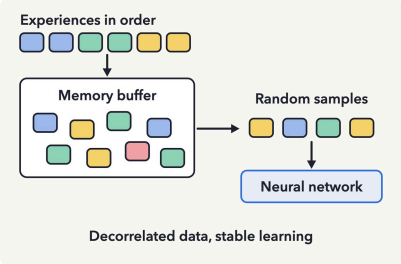
As a condition of the Google deal, DeepMind insisted on an Ethics and Safety Review Board: external members who would have to be informed before any AGI-relevant breakthrough. In practice, Mallaby reports, the board met rarely and was eventually folded into Google's broader processes. It was a victory of optics over governance — but it set a template the industry has been arguing about ever since.

DeepMind's first investor

Peter Thiel's Founders Fund led DeepMind's earliest outside funding round — a notable irony, given Thiel's later public skepticism about unchecked AI development.

The Atari strike team

DeepMind's first publicly recognised milestone came from a tiny team led by Vlad Mnih, a Toronto-Alberta-trained reinforcement-learning specialist. The challenge: build a single neural network that could learn to play any of the classic Atari 2600 games from raw screen pixels and a score signal — no game-specific code.



FlashFeed original diagram · CC-BY-4.0

Why experience replay mattered

Neural networks hate seeing highly correlated training examples in sequence — they overfit to whatever's in front of them. The DQN insight was to store the agent's experiences in a memory buffer and train the network on random samples from that buffer. This decorrelated the data and made deep RL stable enough to actually learn.

The day Pong clicked

On 7 July 2013 — Wimbledon final day — the team watched the DQN agent learn to play Pong from scratch. After a few hours, the bot started to put the paddle in the right place. By that evening it was beating the built-in opponent. Several engineers wept. The intellectual lineage to AlphaGo, AlphaFold and Gemini all run through that afternoon.

Google's $600M deal

Google's purchase of DeepMind is often quoted as roughly $600 million in US dollars — equivalent to about £400 million.



User Balou on en.wikipedia · CC BY 2.0

Atari's one-network rule

The Atari challenge imposed a deliberate constraint: a single network, no game-specific code, learning only from raw pixels and the score — forcing a general solution instead of a tailored one.

By the numbers

Approximate price (£ million) Google paid for DeepMind in 2014

400 £ million

Year the DeepMind name was registered

2010 year

Year Google acquired DeepMind

2014 year

Year DeepMind was founded

2010 year

How many millions of US dollars Google paid for DeepMind

600 million USD

Matched pairs

Each item on the left goes with the one on the right.

DeepMind's first working name → Solaria, taken from the Asimov novel

Where Musk warned AI could become a digital god → At a dinner at Larry Page's home in October 2012

Lead researcher on the Atari DQN agent → Vlad Mnih

Self-test

Answer each question from memory, then check your work against the answer key.

1. What was DeepMind's first working name, before the founders settled on the name they kept?

- A Foundation
- B Aurora
- C Solaria
- D Trantor

2. The company's original working name was borrowed from a novel by which author?

- A William Gibson
- B Arthur C. Clarke
- C Philip K. Dick
- D Isaac Asimov

3. In what year was the DeepMind name registered?

- A 2010
- B 2014
- C 2011
- D 2012

4. Which founder pushed for a plainer name than the original working title?

A. Vlad Mnih
B. Shane Legg
C. Mustafa Suleyman
D. Demis Hassabis

5. In the final name, what does the word "Deep" refer to?

A. Deep Blue, the chess computer that beat Kasparov
B. The depth of the company's funding reserves
C. Deep learning, the neural-network approach they were betting on
D. Deep Q-Networks, the architecture they had already built

6. In which month and year did Google close its acquisition of DeepMind?

A. January 2014
B. July 2013
C. October 2012
D. January 2011

7. Roughly how much did Google pay for DeepMind?

A. £4 billion
B. £40 million
C. £1.2 billion
D. £400 million

8. What record did Google's purchase of DeepMind set at the time?

A. The largest sale of a British company to a US buyer
B. The largest AI acquisition in European history
C. The first AI lab ever bought by a technology company
D. The largest acquisition Google had ever made

9. Which pair of Google leaders had been watching DeepMind since 2012, ahead of the deal?

A. Eric Schmidt and Sundar Pichai
B. Larry Page and Eric Schmidt
C. Sergey Brin and Sundar Pichai
D. Larry Page and Sergey Brin

10. Which statement about the DeepMind acquisition is NOT true?

A. The lab had fewer than 100 employees
B. DeepMind was already profitable on strong revenue at the time of the deal
C. Hassabis demanded an independent ethics board as a condition
D. It was then the largest AI acquisition in European history

11. Which investor's fund led DeepMind's earliest outside funding?

A. Peter Thiel's Founders Fund
B. Marc Andreessen's Andreessen Horowitz
C. Sequoia Capital
D. Elon Musk's personal investment

12. Mallaby treats the October 2012 clash between Musk and Page over advanced AI as planting the seed of which later organisation?

A. xAI
B. DeepMind's Ethics and Safety Review Board
C. OpenAI
D. Anthropic

13. When Musk clashed with Page over advanced AI in October 2012, what phrase did he use for a sufficiently advanced AI?

A "A digital god"
B "A summoned demon"
C "The last invention"
D "An electric brain"

14. Where did the Musk–Page argument over advanced AI take place on 8 October 2012?

A At DeepMind's London office
B At a dinner at Larry Page's home in Palo Alto
C At a tech conference keynote in San Francisco
D At a Google board meeting in Mountain View

15. How did Page respond to Musk's warning that evening?

A Only Google could be trusted to build it safely
B Advanced AI was decades away and not worth worrying about
C The species that built such an intelligence would be its natural heirs
D Regulation should be written before any lab went further

16. Why was experience replay key to making Deep Q-Networks learn stably?

A It slowed the learning rate so the network never overshot
B It let the network memorize game-specific rules for each Atari title
C It ran many agents in parallel and averaged their scores
D It stored past experiences in a buffer and trained on random samples, breaking the correlation between consecutive frames

17. Who led the small team behind DeepMind's Atari DQN agent?

A Vlad Mnih
B Mustafa Suleyman
C Shane Legg
D Demis Hassabis

18. In Q-learning, what does an agent estimate?

A How closely its behaviour matches a human demonstrator's
B The probability that any given pixel changes next frame
C The long-term expected reward of each action from each state
D The shortest path from the current state to the goal state

19. What two components does a Deep Q-Network combine?

A Q-learning with a hand-coded rules engine per Atari game
B A language model with a search tree over future moves
C Supervised learning on human play logs with a scoring heuristic
D Q-learning with a deep convolutional network that reads raw pixels

20. What was the deliberate constraint on the Atari challenge?

- A Access to the games' internal memory state rather than the screen
- B One network, no game-specific code, learning from raw pixels and the score alone
- C Learning only by watching recordings of expert human players
- D A separate network tuned for each of the Atari 2600 titles

21. Which description best defines DeepMind?

- A A Toronto neural-network research group acquired by Microsoft in 2014
- B Google's internal deep-learning team, formed in Mountain View in 2011
- C The London AI lab Demis Hassabis cofounded in 2010 and Google acquired in 2014
- D A Silicon Valley AI lab founded by Elon Musk as a nonprofit in 2015

22. Which city was DeepMind founded and based in?

- A London
- B Mountain View
- C Palo Alto
- D Toronto

23. How is DeepMind best described?

- A A US nonprofit Elon Musk founded to counter Google's AI lead
- B A Google-run research team formed in Mountain View before Hassabis joined
- C An independent ethics board formed to review DeepMind's AGI work
- D The London AI lab Hassabis cofounded in 2010, later bought by Google

24. A profile describes "the London lab behind AlphaGo and AlphaFold, now owned by Google." Which organisation is that?

- A Meta AI (FAIR)
- B DeepMind
- C Google Brain
- D OpenAI

25. Which pairing of years correctly matches DeepMind's founding and its acquisition by Google?

- A Founded 2012, acquired 2014
- B Founded 2010, acquired 2014
- C Founded 2011, acquired 2012
- D Founded 2010, acquired 2011

Go deeper

ONLINE This lesson on the web flashfeedlearningapp.com/study-guides/infinity-machine/founding-deepmind

STUDY The Infinity Machine (book) in the FlashFeed app flashfeedlearningapp.com/packs/infinity-machine

LESSON 6 OF 12

Atari to AlphaGo

Atari and the DQN · AlphaGo and Move 37

WHAT YOU'LL BE ABLE TO DO

- Define reinforcement learning and recall the DQN milestones — published in Nature (2015), human-level or better on 29 of 49 Atari games.
- Define AlphaGo as the DeepMind program that beat Lee Sedol in Seoul in March 2016, after a January 2016 Nature paper reported its methods and its Fan Hui win.
- Identify the people around AlphaGo — engineer Aja Huang, opponents Fan Hui and Lee Sedol, and Sergey Brin who backed the project
- Explain Move 37 — the shoulder-hit AlphaGo's policy network rated a roughly 1-in-10,000 human move against Lee Sedol
- Identify OpenAI's founders — CEO Sam Altman and first chief scientist Ilya Sutskever — formed as a counterweight to Google's DeepMind

Atari and the DQN

DeepMind's Atari breakthrough — a single neural network learning to play dozens of games from raw pixels — was the moment the lab's approach moved from promising to historic. The engine behind it, **reinforcement learning**, would soon be aimed at a far older game.

Forty-nine games, one network

The **Deep Q-Network (DQN)**, published in *Nature* in February 2015, used a single architecture — same code, same hyperparameters — retrained from scratch on 49 different Atari 2600 games. It reached human-level or better performance on 29 of them without any hand-engineered, game-specific features. The input was raw screen pixels and a score signal; nothing more. Before this, game-playing AI was typically hand-tuned per title. DQN showed that **one general algorithm** could master many distinct environments, which was the conceptual leap the field had been waiting for. DeepMind had actually demonstrated early Atari results at **NeurIPS in December 2013**, months before the Google acquisition closed.

Sergey Brin and the Go problem

Not long after the Google acquisition, on a hallway walk in May 2014, **Sergey Brin** asked Hassabis what hard problem DeepMind would tackle next. Hassabis said: "Go." Brin was sceptical — Go has roughly 10^170 possible board positions, a search space that had defeated every serious AI approach for decades — but he provided the resources. The AlphaGo project started that summer. Go mattered because its complexity made brute-force search impossible; cracking it would require the kind of **intuition-like pattern recognition** that neural networks are built for.

Fan Hui, 5–0, October 2015

In October 2015, AlphaGo played **Fan Hui**, the European Go champion, in a closed-door five-game match. AlphaGo won 5–0. The result was kept under embargo until the January 2016 *Nature* cover paper — the same issue that revealed AlphaGo's architecture. Fan Hui was sufficiently moved by the experience that he joined DeepMind as an analyst to help prepare for the next challenge. AlphaGo's first prototype was built by **Aja Huang**, whose work on the tree-search component proved critical to making the system competitive at professional level.

Lee Sedol, Seoul, March 2016

In March 2016, AlphaGo travelled to Seoul to face **Lee Sedol**, widely regarded as the greatest Go player of his generation. Over five games, broadcast live and drawing tens of millions of viewers across Asia, AlphaGo won 4–1. Lee's single victory — game 4 — came from a move so unusual that AlphaGo had not adequately evaluated it. The match made **artificial intelligence a front-page story worldwide**, compressing years of public awareness into one week.

Move 37

In game two, move 37 — a **shoulder-hit on the fifth line** — stopped the broadcast commentators cold. Human grandmasters rated it a mistake; AlphaGo's policy network estimated a human would choose that move roughly **one time in ten thousand**. After twelve minutes away from the board, Lee Sedol returned visibly shaken. AlphaGo's reasoning was correct: the move won territory in a way human intuition would not reach. "Move 37" became shorthand for the idea that trained networks can discover strategic concepts outside the corpus of human play.

AlphaGo and Move 37

AlphaGo's Seoul victory in **2016** didn't just settle a benchmark — it rattled the wider AI world. Within months, a rival lab had materialised in San Francisco, and a talent war was under way that would reshape the industry.

OpenAI is born

In December 2015, **Elon Musk** and **Sam Altman** announced **OpenAI**: a non-profit pledging $1 billion in funding commitments, framed explicitly as a counterweight to Google's ownership of DeepMind. Musk had concluded that concentrating frontier AI capability inside a single commercial giant was dangerous, and DeepMind under Google was his primary concern. OpenAI's founding came about two months after AlphaGo's closed-door victory over Fan Hui, while that result was still confidential — two events close together in time. **Sam Altman** was co-chair alongside Musk; he became OpenAI's CEO in 2019. The January 2016 *Nature* cover that announced AlphaGo's Fan Hui result retroactively validated the very fears about Google's AI lead that had already driven OpenAI's December 2015 founding.

The poaching war begins

OpenAI's first hires targeted the DeepMind and Google Brain alumni networks with deliberate precision. **Ilya Sutskever** — a Geoffrey Hinton student from Toronto who had been working at Google — became **chief scientist**, the role that shaped OpenAI's early research agenda. Pieter Abbeel and others followed.

DeepMind, constrained by Google's standard pay bands, responded with equity-style retention packages unusual for an arms-length subsidiary. The contest for talent between the two labs would intensify over the following decade, driving compensation to levels that reshaped hiring norms across the entire field.

Key terms

Reinforcement learning

Reinforcement learning is a training method in which an agent learns by trial and error, receiving rewards for good actions and penalties for bad ones. DeepMind used reinforcement learning to master Atari games directly from raw pixels.

AlphaGo

AlphaGo is the DeepMind program that defeated world champion Lee Sedol at the board game Go in 2016. AlphaGo combined deep neural networks with Monte Carlo tree search.

Key concepts

Fan Hui

- A reigning European champion of an ancient Asian board game, chosen for AI's first serious professional test.
- In a closed-door five-game match in October 2015, he was beaten without winning a single game.
- Moved by the loss, he joined DeepMind as an analyst; the 5-0 result stayed under embargo until a January 2016 Nature cover.

Lee Sedol

- Widely regarded as the greatest player of his generation at an ancient board game, he was AI's next challenge after a European champion fell.
- In a March 2016 Seoul match watched by tens of millions across Asia, he lost four games to one.
- His lone victory, game four, exploited a move the machine had failed to evaluate; earlier he had left the board shaken by its move 37.

Aja Huang

- A DeepMind engineer who built the very first prototype of the lab's Go-playing system.
- His work on the tree-search component proved critical to making that system competitive at professional level.

Sergey Brin

- A Google co-founder who, soon after the DeepMind acquisition, asked the lab's leader what hard problem it would tackle next.
- On a May 2014 hallway walk he heard the answer 'Go' and was sceptical of a game with about 10^170 positions.
- Despite his doubts he provided the resources, and the project he funded launched that summer.

Key facts



Steve Jurvetson · CC BY 2.0

Sergey Brin and the Go problem

On a hallway walk at Google in May 2014, Sergey Brin asked Hassabis what hard problem DeepMind would tackle next. Hassabis answered: "Go." Brin was sceptical — the game's complexity had defeated AI for decades — but he provided the resources. The AlphaGo project began that summer.



Goban1 · Public domain

Fan Hui, 5–0, October 2015

In October 2015, AlphaGo played the European Go champion Fan Hui in a closed-door match. The result, 5–0 to AlphaGo, was kept under embargo until the January 2016 Nature cover paper. Fan Hui was so impressed by the experience that he joined DeepMind as an analyst to help prepare for Lee Sedol.

> **Move 37**
>
> In game two, on move 37, AlphaGo played a shoulder-hit at the fifth line that human commentators called a mistake. After 12 minutes of human deliberation, Lee Sedol returned to the board visibly shaken. Move 37 became shorthand for AI creativity — a play no human grandmaster would have chosen, and one that the network calculated had a one-in-ten-thousand chance of being played by a human.



Wikimedia Commons · David Shankbone · CC BY 3.0

OpenAI is born

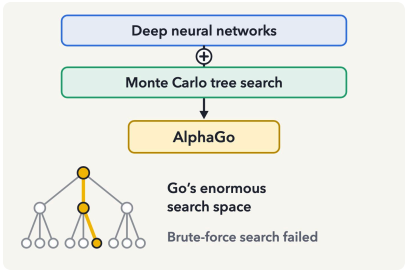
In December 2015, Elon Musk and Sam Altman announced OpenAI: a non-profit AI lab pledging $1 billion in commitments, explicitly framed as a counterweight to Google DeepMind. Mallaby treats the move as a direct reaction to AlphaGo's progress — Musk in particular had decided that DeepMind "under Google" was the most dangerous concentration of AI talent in the world.

The poaching war begins

OpenAI's first hires deliberately targeted the DeepMind alumni network. Ilya Sutskever, a Toronto-trained Hinton student who had been at Google, became chief scientist. Pieter Abbeel and others followed. DeepMind, under Google's pay scales, fought back by offering equity-style retention packages that were unusual for an arms-length subsidiary.

AlphaGo's Nature cover

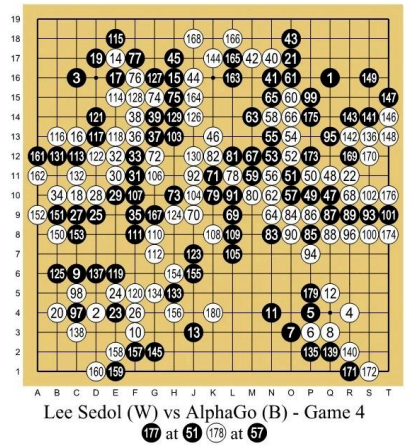
AlphaGo's first major research paper appeared on the cover of the journal Nature in January 2016.



FlashFeed original diagram · CC-BY-4.0

Neural nets plus tree search

AlphaGo paired deep neural networks with Monte Carlo tree search — a combination that succeeded where brute-force search had failed for decades against Go's enormous search space.



Wesalius · CC BY-SA 4.0

Seoul, March 2016: 4-1

AlphaGo defeated Lee Sedol 4–1 in Seoul in March 2016, a very different result from the 5–0 sweep of Fan Hui in the earlier closed-door match.

By the numbers

Year AlexNet won the ImageNet competition
2012 year

Year DQN was published in Nature
2015 year

Number of Atari 2600 games the single DQN architecture was trained and evaluated on
49 games

Number of Atari games on which DQN reached human-level or better
29 games

Year AlphaGo defeated Lee Sedol in Seoul
2016 year

Year AlphaGo beat Lee Sedol in Seoul
2016 year

Matched pairs

Each item on the left goes with the one on the right.

Engineer who built AlphaGo's first prototype → Aja Huang

OpenAI cofounder who became CEO → Sam Altman

OpenAI's first chief scientist → Ilya Sutskever

Self-test

Answer each question from memory, then check your work against the answer key.

1. Which description best defines reinforcement learning as DeepMind applied it to Atari?

(A) Groups are found in unlabelled data by measuring distance between points

(B) An agent learns by trial and error, receiving rewards for good actions and penalties for bad ones

(C) A network compresses inputs and reconstructs them to find hidden structure

(D) A network is shown labelled examples and corrected toward the right label

2. In what year was the Deep Q-Network (DQN) published in Nature?

(A) 2015

(B) 2016

(C) 2013

(D) 2012

3. How many Atari 2600 games was the single DQN architecture retrained on, and on how many did it reach human-level or better?

A 49 games, human-level or better on 49
B 37 games, human-level or better on 10
C 49 games, human-level or better on 29
D 29 games, human-level or better on 49

4. Which of the following was NOT part of what made DQN's Atari result a milestone?

A One general algorithm mastering many distinct environments
B The same code and hyperparameters used across every game
C Raw screen pixels and a score signal as the only input
D Hand-engineered, game-specific features tuned for each title

5. Where and when did DeepMind first demonstrate early Atari results, before the Google acquisition closed?

A The Nature cover, January 2016
B NeurIPS, December 2013
C NeurIPS, December 2015
D ImageNet competition, 2012

6. What journal made AlphaGo's first major paper a cover story in January 2016?

A Science
B Cell
C PNAS
D Nature

7. AlphaGo is best defined as which of the following?

A The DeepMind program that defeated world champion Lee Sedol at Go in 2016
B The DeepMind network that learned 49 Atari games from raw pixels
C The OpenAI system announced as a counterweight to Google DeepMind
D The Toronto network that won the ImageNet competition in 2012

8. What made AlphaGo's March 2016 victory over Lee Sedol significant beyond the 4-1 scoreline?

A It showed that Lee Sedol was not among the strongest players of his generation
B It proved brute-force search could finally solve Go
C It made artificial intelligence a front-page story worldwide
D It was the first time AlphaGo had beaten a professional Go player

9. In which city and year did AlphaGo defeat Lee Sedol, and by what score?

A Seoul, January 2016, 3–2
B London, October 2015, 5–0
C Seoul, March 2015, 5–0
D Seoul, March 2016, 4–1

10. Which statement about AlphaGo is NOT accurate?

A Its architecture was revealed in a January 2016 Nature cover paper
B It combined deep neural networks with Monte Carlo tree search
C It learned Go end to end from raw screen pixels alone
D It was built by DeepMind and defeated Lee Sedol in 2016

11. AlphaGo's 5-0 win over Fan Hui stayed secret until it appeared as a cover story in which publication?

A Science
B Nature
C MIT Technology Review
D Wired

12. In what month and year did Nature run the cover paper announcing AlphaGo?

A January 2016
B October 2015
C February 2015
D March 2016

13. Which engineer built AlphaGo's first prototype, contributing the critical tree-search component?

A Fan Hui
B Pieter Abbeel
C Aja Huang
D Ilya Sutskever

14. Who asked Hassabis on a May 2014 hallway walk what hard problem DeepMind would tackle next, then funded the answer despite being sceptical?

A Sergey Brin
B Sam Altman
C Elon Musk
D Larry Page

15. Fan Hui held which title when he played AlphaGo in October 2015?

A Chinese Go champion
B Korean national Go champion
C World Go champion
D European Go champion

16. Lee Sedol took one game off AlphaGo in Seoul. Which game was it, and what explains his win?

A Game one — AlphaGo's tree search ran out of thinking time
B Game four — he found a move AlphaGo had not adequately evaluated
C Game five — AlphaGo misread the final scoring
D Game two — AlphaGo's shoulder-hit on the fifth line backfired

17. Which pairing of person and role is INCORRECT?

A Aja Huang — engineer whose tree-search work proved critical
B Fan Hui — European champion who later joined DeepMind as an analyst
C Sergey Brin — Google cofounder who funded the Go project despite scepticism
D Lee Sedol — European champion who lost 5–0 in a closed-door match

18. AlphaGo's Move 37 against Lee Sedol was so unusual that its policy network estimated a human would play it about how often?

A Once in 100 times
B Once in 1,000 times
C Once in 10,000 times
D Once in a million times

19. In which game of the Seoul match, and at what board location, did AlphaGo play the move that became known as Move 37?

A Game one, a shoulder-hit on the fourth line
B Game two, a shoulder-hit on the fifth line
C Game four, a shoulder-hit on the fifth line
D Game two, a corner invasion on the third line

20. How did human grandmasters and commentators initially judge Move 37?

A As a mistake
B As a standard professional opening
C As an obvious winning move
D As a deliberate stalling tactic

21. How did Lee Sedol respond immediately after Move 37 was played?

A He replied instantly with a move that won the game
B He resigned the game on the spot
C He requested a review of AlphaGo's legality
D He left the board for about twelve minutes and returned visibly shaken

22. What has "Move 37" come to stand for?

A That AI systems fail in ways humans can easily predict
B That trained networks can discover strategy outside the corpus of human play
C That neural networks reliably reproduce expert human judgement
D That tree search alone can outperform human intuition

23. Which OpenAI cofounder became the organisation's CEO?

A Sam Altman
B Ilya Sutskever
C Pieter Abbeel
D Elon Musk

24. Who became OpenAI's first chief scientist, shaping its early research agenda?

A Sam Altman
B Pieter Abbeel
C Ilya Sutskever
D Geoffrey Hinton

25. When was OpenAI announced, and how much funding was pledged?

A March 2016, $1 billion in commitments
B December 2015, $1 billion in commitments
C December 2015, $100 million in commitments
D May 2014, $10 billion in commitments

26. What was the stated rationale for founding OpenAI as a non-profit?

A To act as a counterweight to Google's ownership of DeepMind
B To commercialise DeepMind's Atari research more quickly
C To build a Go program that could surpass AlphaGo
D To fund academic AI departments left behind by industry

27. Which of the following was NOT part of OpenAI's founding?

A An explicit framing as a counterweight to Google's DeepMind
B A December 2015 announcement by Elon Musk and Sam Altman
C A non-profit structure with $1 billion in pledged commitments
D Ilya Sutskever taking the CEO role while Sam Altman led research

28. How did Mallaby frame the December 2015 founding of OpenAI?

A As a merger with DeepMind
B As a spinoff of DeepMind's safety team
C As a counterweight to Google's DeepMind
D As a government-run AI lab

29. Who concluded that DeepMind 'under Google' was too dangerous a concentration of AI power, driving him to help found OpenAI?

A Elon Musk
B Sam Altman
C Sergey Brin
D Larry Page

Go deeper

ONLINE This lesson on the web flashfeedlearningapp.com/study-guides/infinity-machine/atari-to-alphago

STUDY The Infinity Machine (book) in the FlashFeed app flashfeedlearningapp.com/packs/infinity-machine

LESSON 7 OF 12

The OpenAI rivalry

Musk, Altman and the split · Anthropic and the firing

WHAT YOU'LL BE ABLE TO DO

- Explain the Musk–Altman split — the 2017 email warning of 'absolute control' and Musk's withdrawal after losing operational control.
- Recall OpenAI's 2019 restructuring into a capped-profit company with investor returns capped at 100×
- Identify the April 2023 merger of Google Brain into DeepMind under Hassabis to form Google DeepMind
- Recognize how the founders ended up atop rival giants (Suleyman as Microsoft AI CEO) and the players in OpenAI's 5-day crisis.

Musk, Altman and the split

The tension inside OpenAI between commercial ambition and the original nonprofit mission played out across two years — a private email, a restructuring, and a departure that quietly reshaped the entire industry's power map.

September 2017: a fateful email

On 20 September 2017, **Greg Brockman** and **Ilya Sutskever** sent a memo to Musk and Altman warning that a single person could end up with "absolute control" of AI. The target of concern was Demis Hassabis — but the logic applied equally to anyone at the top of OpenAI. The email is significant because it reveals that even the founders of a lab created to prevent an AI monopoly feared the logic of concentration they were themselves enacting. **Musk** subsequently pushed to run OpenAI himself; when he failed to gain operational control in early 2018, he withdrew his promised funding and resigned from the board.

OpenAI's capped-profit pivot

In **2019**, OpenAI restructured into a nonprofit parent controlling a for-profit subsidiary, with investor returns capped at **100 times** their stake. Altman, now CEO, used the structure to secure a **$1 billion commitment from Microsoft** — the partnership that would eventually make Microsoft the dominant AI-infrastructure player. The move was presented as a safety compromise: profits would be real but bounded. The cap later became a flashpoint: the board cited it as a commitment Altman had implicitly eroded through subsequent fundraising ambitions.

Suleyman exits, August 2019

Mustafa Suleyman co-founded DeepMind alongside Hassabis and Shane Legg in a north-London chess club. By August 2019, Bloomberg's account of internal bullying allegations ended his operational role; he moved to a vice-president position at Google with no direct team. That quiet exit was easy to overlook

at the time. By **2024** he had become CEO of **Microsoft AI**, placing him in charge of the partner company most dependent on OpenAI's models — the third vertex of the original triangle now sitting across the table from his former rivals.

Google DeepMind, April 2023

When **ChatGPT** launched in late 2022 it sent shockwaves through Alphabet. Sundar Pichai's response came in **April 2023**: **Google Brain** — the team that had invented the **transformer architecture** in 2017 — merged with DeepMind to form a single unit called **Google DeepMind**, with **Hassabis** as CEO. The merger ended a decade of internal rivalry between two Google AI factions and handed Hassabis direct control over Google's entire response to OpenAI. The transformer researchers and the reinforcement-learning researchers were finally in the same room.

Three north Londoners, three giants

Sebastian Mallaby's book tracks a striking coincidence: by **2024**, all three **DeepMind co-founders** held commanding AI roles at the world's largest technology companies. Hassabis led **Google DeepMind**; Suleyman led **Microsoft AI**; **Shane Legg** remained DeepMind's chief AGI scientist. Mallaby notes they were "two north Londoners with immigrant parents" heading AI operations at two American tech giants — a detail that underscores how thin the founding cohort of modern AI really is, and how much institutional power flows from a single friendship formed in a London games café.

Anthropic and the firing

Two ruptures defined OpenAI's mid-2020s: a principled walkout that produced its most serious safety-focused rival, and a five-day governance crisis that ended with the acceleration of the AI race rather than its restraint.

Anthropic splits off, 2021

In **2021**, **Dario Amodei** — OpenAI's VP of Research — and his sister **Daniela Amodei** led roughly a dozen safety-focused researchers out of OpenAI to found **Anthropic**. The group included **Paul Christiano**, a prominent figure in AI alignment research, and **Geoffrey Irving**, who later became scientific lead at the UK's AI Safety Institute. Mallaby reads the departure as a recurring structural pattern: the researchers most worried about safety keep leaving the very labs they helped make powerful, unable to slow the pace from inside. Anthropic positioned itself explicitly around **Constitutional AI** and interpretability work, seeking a path to safe systems rather than purely capable ones.

Five days that shook OpenAI

On **17 November 2023**, OpenAI's board fired **Sam Altman** over a video call, citing a loss of confidence in his candor. The decision was made by a small nonprofit board — the governance structure the 2019 restructuring had left formally in charge. Within forty-eight hours, more than **700 of OpenAI's 770 employees** signed a letter threatening to resign. **Microsoft** offered to absorb anyone who left. Five days after the firing, Altman was reinstated and the dissenting board members stepped down. Mallaby's verdict: "The hopes of slowing the AI race had been destroyed." The incident demonstrated that a lab's

nonprofit board held nominal control but zero real power when its workforce and its primary investor were aligned against it.

Nadella's lifeline

Within **36 hours** of Altman's dismissal, **Satya Nadella** publicly announced that Microsoft would create a new AI lab for Altman and offered positions to any OpenAI employee willing to follow. The offer was a calculated application of leverage: Microsoft held the cloud infrastructure and billions in committed capital that OpenAI depended on. The threat of a mass exodus — more than **700 of 770 staff** — made the board's position untenable. They folded within **5 days**. Nadella's move confirmed that **Microsoft** had effectively become the silent guarantor of OpenAI's continuity, and that the nonprofit governance layer was structurally incapable of overriding the economic interests of the workforce and its largest backer. **Mustafa Suleyman** — now running Microsoft AI — was waiting on the other side of that guarantee.

Key concepts

Greg Brockman

- On 20 September 2017 this OpenAI co-founder sent a warning memo to the lab's backers.
- The memo, co-signed with the chief scientist, warned that one person could gain 'absolute control' of AI.
- Its unnamed target of concern was the head of a rival London lab.

Dario Amodei

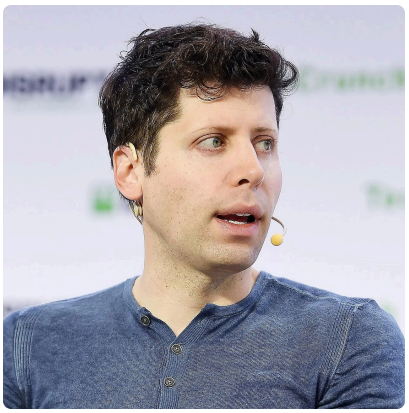
- OpenAI's VP of Research who came to believe safety demanded a clean break.
- In 2021 he and his sister led roughly a dozen researchers out of OpenAI.
- The lab they founded organised itself around Constitutional AI and interpretability rather than raw capability.

Satya Nadella

- The CEO of a cloud giant that had become the silent guarantor of OpenAI's continuity.
- Within 36 hours of a rival CEO's firing, he announced his company would build that ousted leader a new AI lab.
- He offered a job to any of the 700-plus employees threatening to walk, making the board's position untenable.

Key facts

> **September 2017: a fateful email**
>
> On 20 September 2017, Greg Brockman and Ilya Sutskever wrote to Musk and Altman. Their worry: a single individual could end up with "absolute control" of OpenAI. "You are concerned that Demis could create an AGI dictatorship," Sutskever wrote of Musk's view, "so do we." The implication: OpenAI itself needed protection from concentrated power.



TechCrunch · CC BY 2.0

OpenAI's capped-profit pivot

In 2019, OpenAI restructured into a non-profit parent with a for-profit subsidiary capped at 100× return for early investors. Altman, now CEO, used the new structure to raise a $1 billion commitment from Microsoft. The move would later be cited as one reason the board lost confidence in him — they had been told the cap would hold.

Suleyman exits, August 2019

On 21 August 2019, Bloomberg published a long story about bullying allegations against Mustafa Suleyman at DeepMind. By the end of the year he had moved to Google's Mountain View office in a vague vice-president role with no team. Eight years later he would emerge as CEO of Microsoft AI — the third corner of the original triangle, leading the partner of OpenAI.



Photographer: Lukasz Kobus (European Commission) · CC BY 4.0

Google DeepMind, April 2023

After ChatGPT shocked Mountain View, Sundar Pichai consolidated. In April 2023 Google Brain — the team that had invented the transformer in 2017 — was merged into DeepMind to form a single unit called Google DeepMind. Hassabis became its CEO. The merger ended a decade-long internal rivalry and made him directly responsible for Google's response to OpenAI.

Three north Londoners, three giants

Mallaby notes the parallel: by 2024 Hassabis ran Google DeepMind; Suleyman ran Microsoft AI; the third corner of the original DeepMind triangle, Legg, remained the lab's chief AGI scientist. "Two north Londoners with immigrant parents," Mallaby writes, "headed AI operations at two American tech giants."



Wikimedia Commons · TechCrunch · CC BY 2.0

Anthropic splits off, 2021

Dario Amodei and his sister Daniela led a group of safety-focused OpenAI researchers — including Paul Christiano and the future UK AI Safety Institute scientific chief Geoffrey Irving — out of OpenAI in 2021 to found Anthropic. Mallaby treats the spin-off as a recurring pattern: safety-minded technologists keep leaving the labs they helped build.

> **Nadella's lifeline**
>
> Within 36 hours of Altman's firing, Satya Nadella publicly invited him to lead a new Microsoft AI lab with a budget to hire "anyone from OpenAI who wanted to join him." Three days later more than 700 of OpenAI's 770 employees had signed a letter threatening to follow. The board folded; Altman returned.



Jiaqian AirplaneFan · CC BY 3.0

Suleyman leads Microsoft AI

Mustafa Suleyman — one of DeepMind's three cofounders — became CEO of Microsoft AI in 2024.

> **700 staff threatened to quit**
>
> More than 700 of OpenAI's 770 employees signed a letter threatening to resign within 48 hours of Sam Altman's firing — a near-total revolt that made the board's position untenable.

By the numbers

Year OpenAI restructured into a capped-profit company

2019 year

Number of days between Altman's firing and reinstatement at OpenAI

5 days

Matched pairs

Each item on the left goes with the one on the right.

Who wrote the September 2017 email to Musk and Altman → Greg Brockman and Ilya Sutskever

OpenAI's 2019 investor return cap → Profit capped at 100 times for early investors

April 2023 merger that formed Google DeepMind
→ Google Brain merged into DeepMind under Hassabis

Self-test

Answer each question from memory, then check your work against the answer key.

1. According to the section, what did Musk do after failing to gain operational control of OpenAI in early 2018?

- A Founded xAI as a direct competitor
- B Sued Altman for breach of the nonprofit mission
- C Doubled his financial commitment to keep influence
- D Withdrew his promised funding and resigned from the board

2. Who sent the 20 September 2017 memo to Musk and Altman warning that one person could end up with "absolute control" of AI?

- A Shane Legg and Mustafa Suleyman
- B Greg Brockman and Ilya Sutskever
- C Dario Amodei and Paul Christiano
- D Sundar Pichai and Demis Hassabis

3. The September 2017 email named one specific person as the target of its "absolute control" concern. Who was named?

- A Sam Altman
- B Satya Nadella
- C Demis Hassabis
- D Larry Page

4. What did Musk push for at OpenAI before his 2018 departure?

- A To run OpenAI himself
- B To merge OpenAI into DeepMind
- C To convert OpenAI into a capped-profit entity
- D To install Brockman as sole CEO

5. Why is the September 2017 email significant?

- A It proved Musk had already decided to withdraw his funding
- B It showed founders of an anti-monopoly lab feared the very concentration they were enacting
- C It was the first written warning that AGI could arrive before 2030
- D It triggered Google Brain's merger into DeepMind

6. In OpenAI's 2019 restructuring, early investor returns were capped at what multiple of their stake?

- A 100 times
- B 50 times
- C 1,000 times
- D 10 times

7. In what year did OpenAI restructure into a capped-profit company?

A 2021
B 2023
C 2019
D 2017

8. What corporate structure did OpenAI adopt in 2019?

A A public benefit corporation wholly owned by Microsoft
B A for-profit parent with a nonprofit research arm
C A joint venture split evenly with Microsoft
D A nonprofit parent controlling a for-profit subsidiary

9. What did Altman secure using the new 2019 structure?

A A $10 billion commitment from Microsoft
B A $100 million commitment from Musk
C A $1 billion commitment from Microsoft
D A $1 billion commitment from Google

10. Which statement about the 2019 capped-profit pivot is NOT accurate?

A The cap was presented as a way to end Microsoft's influence over OpenAI
B The move was presented as a safety compromise
C The board later cited the cap as a commitment Altman had eroded
D Profits would be real but bounded

11. Which two units merged in April 2023 to form Google DeepMind?

A Google Brain and Google X
B Google Research and DeepMind
C DeepMind and Microsoft AI
D Google Brain and DeepMind

12. Who became CEO of the merged Google DeepMind?

A Mustafa Suleyman
B Demis Hassabis
C Sundar Pichai
D Shane Legg

13. What prompted Sundar Pichai's April 2023 consolidation of Google's AI teams?

A Suleyman's 2019 departure to Google
B Musk's resignation from the OpenAI board
C ChatGPT's launch in late 2022 shocked Alphabet
D Microsoft's $1 billion commitment to OpenAI

14. Which architecture, invented in 2017, came from the team that merged into DeepMind?

A The transformer
B Deep reinforcement learning
C The convolutional neural network
D Constitutional AI

15. What did the April 2023 merger end?

A Alphabet's funding of external AI labs
B DeepMind's independence from Alphabet ownership
C Google's use of the transformer architecture
D A decade of internal rivalry between two Google AI factions

16. Which DeepMind cofounder became CEO of Microsoft AI in 2024?

- A David Silver
- B Mustafa Suleyman
- C Demis Hassabis
- D Shane Legg

17. By 2024, what role did Shane Legg hold?

- A CEO of Google DeepMind
- B DeepMind's chief AGI scientist
- C Head of the UK AI Safety Institute
- D CEO of Microsoft AI

18. How many days passed between Sam Altman's firing and his reinstatement?

- A 10
- B 2
- C 36
- D 5

19. Who publicly offered to create a new Microsoft AI lab for Altman within 36 hours of his dismissal?

- A Greg Brockman
- B Mustafa Suleyman
- C Satya Nadella
- D Sundar Pichai

20. Roughly how many of OpenAI's 770 employees signed the letter threatening to resign?

- A More than 700
- B About 250
- C About 400
- D All 770

Go deeper

ONLINE This lesson on the web flashfeedlearningapp.com/study-guides/infinity-machine/the-openai-rivalry

STUDY The Infinity Machine (book) in the FlashFeed app flashfeedlearningapp.com/packs/infinity-machine

LESSON 8 OF 12

AlphaFold — Fermat for biology

A promise after Seoul · CASP14 and the 200 million

WHAT YOU'LL BE ABLE TO DO

- Define protein folding and recognise Anfinsen's conjecture that a protein's sequence alone determines its folded shape
- Define AlphaFold as the DeepMind system that predicts 3D protein structure from amino-acid sequence, and identify John Jumper as AlphaFold 2's lead
- Explain AlphaFold 1's leap — predicting the continuous distance between amino-acid pairs (a 'distogram') rather than binary contact
- Define the GDT accuracy metric and the biennial CASP contest that scores protein-structure predictions
- Recall AlphaFold 2's landmark results — 92 GDT at CASP14 (2020), the direct-folding pivot, and ~200 million proteins released by July 2022.

A promise after Seoul

The seeds of **AlphaFold** were planted on a South Korean sidewalk in 2016. What followed was a decade-long attempt to crack the central unsolved problem of structural biology: reading a protein's **amino-acid sequence** and predicting the exact three-dimensional shape it folds into.

A promise on a Seoul sidewalk

After AlphaGo defeated Lee Sedol in March 2016, Hassabis walked back to the hotel with David Silver and declared they could solve **protein folding**. Silver was sceptical; Hassabis had been circling the problem since reading the literature as an undergraduate. Within months he had assembled a hackathon team to begin. The moment captures something essential about DeepMind's operating logic: a moonshot target named in the aftermath of one victory, before the next campaign had formally started.

Anfinsen's conjecture

In his 1972 Nobel Prize lecture, Christian Anfinsen conjectured that a protein's sequence alone determined its folded shape — that no additional cellular machinery was required to encode it. Biologists accepted the principle for half a century but could not extract a shape from a sequence computationally. **Protein folding** is the process by which a chain of amino acids twists itself into a specific three-dimensional shape that determines the protein's function. AlphaFold was the algorithmic vindication of Anfinsen's conjecture, arriving roughly fifty years after he stated it.

Jumper arrives, October 2017

Andrew Senior recruited **John Jumper** in October 2017. Jumper, then 32, brought an unusually broad background: mathematics, physics, chemistry, biology, and machine learning, including three years at D. E. Shaw Research modelling proteins with Newtonian physics. He contributed the biological intuition the

team had lacked. Jumper would go on to lead AlphaFold 2 and eventually share the 2024 Nobel Prize in Chemistry with Hassabis for the work.

CASP: the biennial test

Since 1994, Professor John Moult had run **CASP** — the Critical Assessment of Structure Prediction — every two years. Labs received amino-acid sequences for proteins whose true shapes were known only to the experimentalists, then competed to predict them. The headline accuracy metric is **GDT** (**Global Distance Test**), a score measuring how closely predicted atom positions match the experimentally verified ones. CASP gave AlphaFold an external, objective scoreboard at a moment when internal benchmarks could have become circular.

From contact map to distogram

Rival labs trained networks to predict whether two amino acids were in contact — a binary yes or no. AlphaFold's convolutional network instead predicted the **distance** between amino-acid pairs on a continuous scale, producing what researchers called a **distogram**. As one team member described it: like going from black and white to full-colour television. The shift gave the model far richer geometric information to work with and became AlphaFold 1's defining architectural choice.

CASP14 and the 200 million

AlphaFold 1 proved the concept; AlphaFold 2 rewrote the contest's history. This part follows the team through a gutsy architectural pivot, a landmark score of **92 GDT** at **CASP**, and the release of a database that folded roughly **200 million** proteins by July 2022.

AlphaFold wins CASP13, December 2018

At CASP13 in Cancún in December 2018, AlphaFold bested all 97 other entrants — most dramatically in the "free modelling" category where no evolutionary template existed. DeepMind won 25 of 43 hard cases; the nearest rival won three. The result announced the system's existence publicly and set the expectation the team then had to surpass. **CASP** is the biennial Critical Assessment of Structure Prediction contest, where protein-folding methods compete on unsolved structures judged by independent experimentalists.

The direct-folding pivot

After CASP13, Hassabis told the team the job was not finished. The next version would predict the exact **3D coordinates** of every atom directly, rather than assembling them via a post-hoc energy-minimisation search. The **GDT** score initially crashed from 60 to 20. The team kept going. By late 2019 the score had clawed back to 60 and continued rising — a textbook example of accepting short-term regression to unlock a structurally superior approach.

Enter the transformer

In 2019, Jumper's team rebuilt AlphaFold's neural network around a **transformer** architecture — specifically a "tetraformer" combining four variants. They masked amino acids in the UniProt database and trained the model to predict the missing residues, the same self-supervised trick BERT had used on

language. The result: evolutionary co-variation patterns in multiple-sequence alignments turned out to encode the geometry of the folded protein.

CASP14, November 2020

When Professor Moult tallied CASP14 in November 2020, **AlphaFold 2** had scored **92 GDT** — more than 50 percent above any previous contest best. The accuracy reached a threshold where the gold-standard experimental measurement could not detect a meaningful difference from the prediction. "A seismic and unprecedented shift," one observer wrote. Hassabis and Jumper shared the **2024 Nobel Prize in Chemistry** for this work.

200 million proteins

After CASP14, DeepMind partnered with the European Bioinformatics Institute to release the **AlphaFold Structure Database**. Within months it covered all ~20,000 human proteins; by July 2022 it held approximately **200 million** entries — nearly every protein known to science. Over 3 million researchers had accessed it by 2025, enabling drug discovery, disease research, and basic biology at a scale that wet-lab crystallography alone could never have matched.

"Cheese tomorrow"

Late in the book, Hassabis is asked about large language models and is blunt: "Everyone talks about the benefits of language models, but mostly it's just cheese tomorrow. The clearest benefit from AI so far is AlphaFold." He frames the next decade as a race to compress a century of medical discovery into years, with the technology deliberately aimed at diseases of the developing world — the problems that profit motive alone would not have solved.

Key terms

Protein folding

The process by which a chain of amino acids twists itself into a specific three-dimensional shape that determines a protein's function. Predicting the final shape from the sequence alone — biology's Fermat's Last Theorem — was the challenge AlphaFold solved.

GDT (Global Distance Test)

CASP's headline accuracy score. It measures the percentage of a predicted protein's main-chain atoms that lie within a tolerated distance of the ground-truth crystallography structure. Roughly 90 GDT is the threshold at which a prediction is as good as an experimental measurement.

AlphaFold

AlphaFold is the DeepMind system that predicts a protein's three-dimensional structure from its amino-acid sequence. AlphaFold cracked a fifty-year grand challenge in biology and helped win Hassabis a share of the 2024 Nobel Prize in Chemistry.

CASP

CASP is the biennial Critical Assessment of Structure Prediction contest, where protein-folding methods compete on unsolved structures. AlphaFold's runaway victory at CASP in 2020 stunned the field of biology.

Key concepts

Anfinsen's conjecture

- An idea from a 1972 Nobel lecture about what determines a protein's final shape.
- It held that a chain's amino-acid sequence alone dictates the folded form, needing no extra cellular machinery.
- Biologists accepted it for half a century but couldn't compute it, until AlphaFold vindicated it roughly fifty years later.

distogram

- AlphaFold 1's defining output, replacing a simple yes/no prediction about pairs of amino acids.
- Instead of asking whether two residues touch, it predicted the continuous distance between them.
- A team member likened the richer geometric picture to going from black-and-white to full-colour television.

CASP

- A biennial contest, running since 1994, that pits protein-folding methods against one another.
- Run by Professor John Moult, it hands labs amino-acid sequences whose true shapes only the experimentalists know.
- It gave AlphaFold an external scoreboard, from a 2018 Cancun edition to a landmark 2020 edition where the score hit 92 GDT.

GDT (Global Distance Test)

- The headline accuracy score used to judge protein-structure predictions.
- It measures how closely predicted atom positions match the experimentally verified ones.
- During a risky architectural pivot it crashed from 60 to 20; by CASP14 it reached 92, over 50% above any prior best.

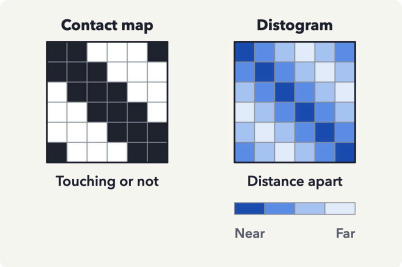
Key facts

> **A promise on a Seoul sidewalk**
> After AlphaGo defeated Lee Sedol in March 2016, Hassabis walked back to the hotel with David Silver. "I'm telling you," he said, "we can solve protein folding." Silver was sceptical. Hassabis had been waiting to tackle the problem since reading the literature as an undergraduate. Within months he had scrambled a hackathon team to try.

> **Anfinsen's conjecture**
> In his 1972 Nobel Prize lecture, Christian Anfinsen conjectured that a protein's sequence alone determined its folded shape — that the information needed was all encoded in the amino acids. For the next half-century, biologists believed him in principle but could not extract the shape from the sequence in practice. AlphaFold was the algorithmic vindication of Anfinsen.

CASP: the biennial test

Every two years since 1994, Professor John Moult had organised CASP — the Critical Assessment of Structure Prediction. Competing labs received amino acid sequences for proteins whose true shapes were known only to the experimentalists. Predictions were scored by Global Distance Test (GDT). For AlphaFold, CASP was the externally verifiable scoreboard.



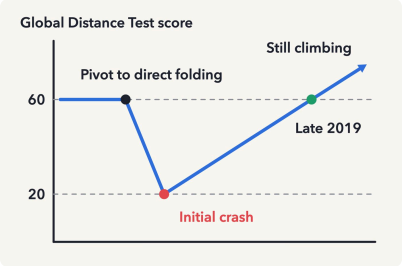
FlashFeed original diagram · CC-BY-4.0

From contact map to distogram

Where rival labs trained networks to predict whether two amino acids touched, AlphaFold trained its convolutional network to predict the distance between them on a continuous scale. "Like going from black and white to a full-colour TV," researcher Marek Barwinski said. The shift to distograms was AlphaFold 1's signature insight.

AlphaFold wins CASP13, December 2018

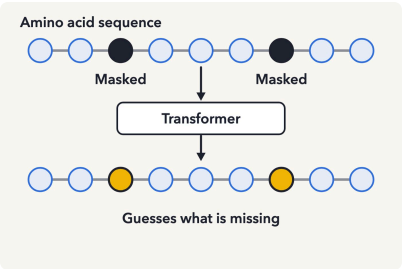
In Cancun in early December 2018, AlphaFold bested the other 97 CASP entrants — particularly in the "free modelling" category where no evolutionary template was known. DeepMind won 25 of 43 hard cases; its nearest rival won three. "What just happened?" became the question of the conference.



FlashFeed original diagram · CC-BY-4.0

The direct-folding pivot

After CASP13, Hassabis told the team they were not done. The next iteration would predict the exact 3D coordinates of every atom directly, instead of relying on a post-hoc search. The GDT score initially crashed from 60 to 20. The team kept going. By late 2019 the new approach was back at 60 and climbing.



FlashFeed original diagram · CC-BY-4.0

Enter the transformer

In 2019 Jumper's team rebuilt AlphaFold's neural network as a transformer — specifically a "tetraformer" combining four variants of the architecture. They masked amino acids in the UniProt database and trained the model to guess what was missing, the same trick BERT had used on language. Evolutionary patterns turned out to encode the geometry of proteins.

"Cheese tomorrow"

Asked late in the book about language models, Hassabis was dismissive: "Everyone talks about the benefits of language models, but mostly it's just cheese tomorrow. The clearest benefit from AI so far is AlphaFold." He wants to spend the next decade compressing a century of medical discovery — and pointing the technology at diseases of the developing world.

By the numbers

AlphaFold 2's GDT score at CASP14

92 GDT

Approximate number of proteins folded by AlphaFold by July 2022 (millions)

200 million

Matched pairs

Each item on the left goes with the one on the right.

Lead scientist on AlphaFold 2 → John Jumper

Self-test

Answer each question from memory, then check your work against the answer key.

1. Protein folding is the process by which a chain of amino acids does what?

- A Splits into shorter peptides that the cell recycles for energy
- B Binds to a crystallography plate so its atoms can be imaged
- C Twists itself into a specific three-dimensional shape that determines the protein's function
- D Copies itself onto a matching strand of messenger RNA

2. What did Christian Anfinsen conjecture about proteins?

- A That most proteins can fold into several equally stable shapes
- B That a protein's sequence alone determines its folded shape
- C That a protein's shape can only be determined by crystallography
- D That extra cellular machinery is required to encode the folded shape

3. Where and when did Anfinsen state his conjecture?

- A In a 1994 keynote at the first CASP contest
- B In his 2004 memoir on structural biology
- C In a 1962 Nobel lecture on the double helix
- D In his 1972 Nobel Prize lecture

4. Which statement about Anfinsen's conjecture is NOT true?

- A Biologists rejected the principle and worked to disprove it for fifty years
- B It concerned the information encoded in the amino acids
- C Biologists accepted the principle but could not compute a shape from a sequence
- D AlphaFold was the algorithmic vindication of the conjecture

5. Roughly how long after Anfinsen stated his conjecture did AlphaFold vindicate it algorithmically?

A About twenty-five years
B About fifty years
C About a century
D About ten years

6. AlphaFold is the DeepMind system that predicts what, from what input?

A A protein's three-dimensional structure, from its amino-acid sequence
B A protein's amino-acid sequence, from a crystallography image
C A gene's mutation rate, from its evolutionary history
D A drug molecule's binding affinity, from a protein's known shape

7. Who was the lead scientist on AlphaFold 2?

A Andrew Senior
B John Moult
C David Silver
D John Jumper

8. A lab has the amino-acid chain of a newly discovered protein but no experimental structure. What would running AlphaFold on it give them?

A The gene sequence that coded for the chain
B A ranked list of drugs that bind the protein
C A predicted three-dimensional structure for that chain
D A crystallography image of the protein's atoms

9. Which prize did John Jumper share with Demis Hassabis for the AlphaFold work?

A The 2020 Turing Award
B The 2024 Nobel Prize in Chemistry
C The 2024 Nobel Prize in Physiology or Medicine
D The 2022 Breakthrough Prize in Mathematics

10. Which of these is NOT an accurate description of AlphaFold or its team?

A AlphaFold predicts 3D structure from an amino-acid sequence
B Jumper was recruited in October 2017 by Andrew Senior, the project's first lead
C Jumper brought biological intuition the team had lacked
D Jumper was recruited in October 2017 by John Moult, who then led AlphaFold 2

11. AlphaFold 1 broke from rival labs by having its network predict what about pairs of amino acids?

A A binary yes-or-no of whether the two were in contact
B The exact 3D coordinates of each atom directly
C The continuous distance between them, producing a "distogram"
D Which amino acid would mutate next over evolution

12. A "distogram" records what information about pairs of amino acids?

A The distance between them on a continuous scale
B Their order along the amino-acid chain
C Whether they touch, as a yes-or-no flag
D The bond angle between their side chains

13. What did rival labs train their networks to predict about two amino acids?

A The evolutionary age of each residue
B A binary label for contact between them
C The exact 3D coordinates of every atom
D The continuous distance separating them

14. A team member compared the shift from contact maps to distograms to which change?

A Going from sheet music to a live orchestra
B Going from a paper map to satellite navigation
C Going from a telescope to a microscope
D Going from black-and-white to full-colour television

15. Which statement about AlphaFold 1's distogram leap is NOT true?

A It was produced by a transformer network built in 2019
B It gave the model far richer geometric information
C It was AlphaFold 1's defining architectural choice
D It was produced by a convolutional network

16. What does the GDT score measure?

A How many labs entered a given CASP contest
B How fast a model produces a structure prediction
C How closely a prediction's atom positions match the experimentally verified structure
D How many amino acids a predicted protein contains

17. What does the acronym CASP stand for?

A Comparative Assay of Structural Proteins
B Critical Assessment of Structure Prediction
C Consortium for Applied Structural Physics
D Computational Analysis of Sequence Properties

18. How often has CASP run, and since when?

A Every two years since 1972
B Every four years since 2004
C Every year since 1994
D Every two years since 1994

19. Roughly what GDT value marks the threshold at which a prediction is as good as an experimental measurement?

A About 50
B About 60
C About 90
D About 20

20. Which statement about CASP is NOT true?

A Competing labs are given the true shapes in advance so they can calibrate
B Predictions are scored by Global Distance Test
C Competing labs receive only amino-acid sequences
D The true shapes are known only to the experimentalists

21. When DeepMind pivoted AlphaFold 2 to predict every atom's 3D coordinates directly, what happened to its GDT score at first?

- A It rose smoothly with no setback
- B It jumped straight from 60 to 92
- C It crashed from about 60 to about 20 before clawing back
- D It held steady near 60 for two years

22. What GDT score did AlphaFold 2 achieve at CASP14?

- A 43
- B 20
- C 60
- D 92

23. How did AlphaFold 2's CASP14 result compare with the contest's previous best?

- A About 10 percent higher than the previous best
- B More than 50 percent higher than any previous best
- C Roughly tied with the previous best
- D Double the previous best, but only in free modelling

24. Approximately how many proteins did the AlphaFold Structure Database hold by July 2022?

- A 200 million
- B 20,000
- C 97
- D 3 million

25. Which statement about AlphaFold 2's landmark results is NOT true?

- A Moult tallied the winning CASP scores in November 2018
- B The database held roughly 200 million proteins by July 2022
- C AlphaFold 2 scored about 92 GDT at CASP14
- D The direct-folding pivot first sank GDT from about 60 to about 20

Go deeper

ONLINE This lesson on the web flashfeedlearningapp.com/study-guides/infinity-machine/alphafold-fermat-for-biology

STUDY The Infinity Machine (book) in the FlashFeed app flashfeedlearningapp.com/packs/infinity-machine

LESSON 9 OF 12

RaceGPT and Bletchley

The ChatGPT shock · Governments arrive

WHAT YOU'LL BE ABLE TO DO

- Recall ChatGPT's 2022 launch and record-fast adoption (100 million users within two months), plus the 2017 Transformer paper behind it
- Define AI alignment, identify Geoffrey Irving, and recognize safety-letter signatory Yoshua Bengio and non-signatory Yann LeCun.
- Trace the US government response — AI czar Ben Buchanan and the July 2023 White House voluntary commitments
- Recall the November 2023 Bletchley Park summit — 28 countries signed the Bletchley Declaration, with the UK and US announcing AI Safety Institutes
- Recognise the 2022 US chip-export ban on China and Ke Jie, the champion whose AlphaGo defeat became China's 'Sputnik moment'

The ChatGPT shock

On 30 November 2022, OpenAI pushed a research preview that most people in the industry expected to cause moderate buzz. What followed was the fastest consumer-technology adoption in recorded history and a forcing event for every government, regulator, and rival lab that had been watching AI from a comfortable distance.

The ChatGPT shock

ChatGPT launched on 30 November 2022 as a polished interface on top of GPT-3.5. A million users arrived within five days; **100 million within two months** — a rate no prior consumer product had matched. For Sebastian Mallaby, the launch is the moment the AI race became public: investors who had filed deep learning under "long horizon" suddenly demanded weekly briefings, and politicians who had never typed the words "large language model" began scheduling hearings. The speed of adoption also accelerated the safety conversation, because a system anyone's grandmother could query was far harder to dismiss than a benchmark score.

Geoffrey Irving joins DeepMind

Geoffrey Irving moved from OpenAI to DeepMind in autumn 2019, bringing with him close professional ties to **Dario Amodei** and **Paul Christiano** — two figures who would later co-found Anthropic. His explicit brief at DeepMind was **alignment**: the technical and conceptual problem of ensuring an AI system reliably pursues human-intended goals even as it becomes more capable. Irving's framing was stark — "If a malevolent actor solves the problem first, civilisation would be in trouble" — and it positioned DeepMind's safety group as an in-house counterweight to the race dynamics driven by the commercial side. Irving would later become a key figure at the **UK AI Safety Institute**.

The one-sentence letter

In May 2023 the **Center for AI Safety** published a single sentence: "Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war." The brevity was tactical — one sentence is harder to nitpick than a white paper. Signatories included **Geoffrey Hinton**, **Yoshua Bengio**, **Demis Hassabis**, **Sam Altman**, and **Dario Amodei**. Mallaby treats the letter as the rhetorical inflection point that pulled governments into the room: once the people who built the systems signed a statement invoking extinction, dismissal became politically untenable. Notably, **Yann LeCun did not sign**, maintaining his long-held scepticism about near-term catastrophic risk.

Governments arrive

The twelve months after ChatGPT's launch produced more AI-specific government action than the previous decade combined. The sequence ran from voluntary pledges to executive orders to an international summit — a compressed arc that revealed both how seriously the administration took the risk and how little legislative appetite existed for binding law.

An AI czar in the White House

In June 2023 the Biden administration created a senior AI coordination role and filled it with **Ben Buchanan**, a Georgetown professor specialising in AI and cybersecurity who had been working on the National Security Council. Buchanan's mandate was to operationalise the concerns Hinton and Bengio had been voicing — without deploying regulatory friction that would hand the lead to China. His position gave the labs a single senior interlocutor and gave the White House a mechanism to move faster than Congress.

Voluntary commitments, 21 July 2023

On 21 July 2023, leaders from **Google, OpenAI, Anthropic, Meta, Amazon, Microsoft**, and Mustafa Suleyman's **Inflection** signed voluntary commitments **at the White House**: pre-release safety testing, cybersecurity investment, and public reporting of capabilities. No major US frontier lab declined. The pledges carried no enforcement mechanism, but Buchanan's reasoning was reputational — any lab seen backsliding publicly would face political consequences. Thin by the standards of financial regulation, the commitments were nonetheless the first coordinated acknowledgment by the industry that external accountability had arrived.

Biden's AI executive order, October 2023

In October 2023 Buchanan's team pushed through an executive order invoking the **Defense Production Act**. Developers training the most powerful AI models would be required to notify the government and share results of **red-team** testing before release. The order also stood up what would become the **US AI Safety Institute**. The administration was writing safety infrastructure into executive authority because Congress had shown no appetite for legislation — a deliberate decision to create facts on the ground before the political window closed.

Bletchley Park summit, November 2023

On **1 November 2023**, delegates from **28 countries** gathered at **Bletchley Park** — Alan Turing's wartime codebreaking home — for the first international AI Safety Summit. The event traced back to a conversation between Hassabis and Prime Minister Rishi Sunak roughly six months earlier. The resulting **Bletchley Declaration** was signed by all 28 delegations. National **AI Safety Institutes** were announced in both the **United Kingdom and the United States**. Yoshua Bengio pressed DeepMind's safety advisers on the pace of deployment. The summit gave governments a multilateral frame that had not existed before.

Chip exports and China

In October **2022** — a month before ChatGPT launched — the Biden administration imposed broad export controls on advanced semiconductors bound for China. Mallaby reads the move as closing the most consequential lane of US-China technology cooperation. The ban meant that by the time Hassabis was pitching a global summit to Sunak, the geopolitical backdrop had already hardened: the AI race and the chip war were now the same story.

China's Sputnik moment

Hassabis has tracked China's competitive intent since **2017**, when AlphaGo's defeat of world champion **Ke Jie** was described inside Beijing as a **Sputnik moment** — a demonstration of Western technological lead that demanded a national response. By 2022, China's State Council had a formal AI strategy with 2030 targets. The framing matters: the West perceived an accelerating race it needed to win; Beijing perceived a race it had been preparing for since AlphaGo played its first move on a Chinese board.

Key terms

Alignment

The technical and conceptual problem of ensuring that an AI system reliably pursues human-intended goals, even as the system becomes more capable. Irving's analogy: rules of safety might be simple (do not harm, do not deceive), but implementing them is as hard as AlphaGo's mastery of Go.

Key concepts

Yoshua Bengio

- A pioneering deep-learning researcher who lent his name to a stark 2023 warning about AI.
- He signed a one-sentence statement ranking AI extinction risk alongside pandemics and nuclear war.
- At the Bletchley Park summit he pressed DeepMind's safety advisers on how fast they were deploying.

Yann LeCun

- A prominent AI pioneer who broke with peers over doomsday framing.
- When a one-sentence statement warned of AI extinction risk, he pointedly declined to sign.
- His refusal reflected a long-held scepticism about near-term catastrophic risk.

Ke Jie

- A reigning world champion of an ancient board game, defeated by AlphaGo in 2017.
- Inside Beijing his loss was called a 'Sputnik moment', proof of a Western technological lead.
- The defeat helped spur China's State Council toward a formal AI strategy with 2030 targets.

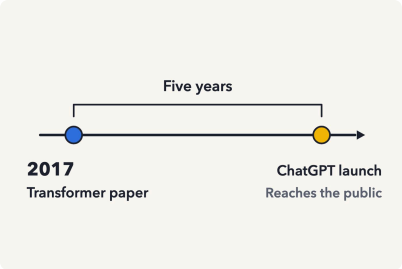
Key facts

Geoffrey Irving joins DeepMind

In autumn 2019 DeepMind hired Geoffrey Irving from OpenAI, where he had been part of a safety circle with Dario Amodei and Paul Christiano. Irving's goal: design alignment into the architecture itself. "If a malevolent actor solves the problem first," he believed, "civilisation would be in trouble."

The one-sentence letter

In May 2023 the Center for AI Safety released a one-sentence letter: "Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war." Hinton, Bengio, Hassabis, Altman, Amodei and hundreds of others signed. Mallaby treats it as the rhetorical inflection point that forced governments into the conversation.



FlashFeed original diagram · CC-BY-4.0

Five years before ChatGPT

The Transformer paper — the architecture ChatGPT ultimately rests on — was published in 2017, five years before ChatGPT's launch put it in front of the general public.

The model behind ChatGPT

ChatGPT launched as a polished interface on top of GPT-3.5, presented at the time as a research preview rather than a finished product.



Wikimedia Commons · Matt H. Wade · CC BY-SA 3.0

Voluntary commitments, 21 July 2023

On 21 July 2023, leaders from Google, OpenAI, Anthropic, Meta, Amazon, Microsoft, and Suleyman's Inflection signed voluntary commitments at the White House: pre-release safety testing, cybersecurity investment, public reporting of capabilities. Not a single major US frontier lab declined. The pledges were thin on enforcement but, Buchanan reasoned, public backsliding would now be visible.



Wikimedia Commons · Adam Schultz · Public domain

Biden's AI executive order, October 2023

In October 2023 Buchanan orchestrated an executive order invoking the Defense Production Act. Developers training the most powerful AI models would have to notify the government and share red-team results. The order also created what would become the US AI Safety Institute. The administration was building the rails before Congress could be convinced to legislate.



Wikimedia Commons · Ordercrazy · CC0

Chip exports and China

In October 2022, a month before ChatGPT, the Biden administration rolled out a wide-ranging ban on the supply of advanced semiconductors to China. Mallaby reads the move as the closing of the most consequential lane of US-China cooperation. By the time Hassabis pitched Sunak on a global AI summit, the geopolitics had already hardened.

> **China's Sputnik moment**
>
> Hassabis has been aware of China's determination to compete since 2017. AlphaGo's defeat of Ke Jie that year was widely described inside Beijing as a Sputnik moment. By 2022 the State Council had a national AI strategy with 2030 ambitions. The race the West thought it was running was, from Beijing, a race China had already been preparing for.

By the numbers

Year the Transformer paper was published

2017 year

Year ChatGPT was released

2022 year

Number of countries that signed the 2023 Bletchley Declaration

28 countries

Year of the first international AI Safety Summit at Bletchley Park
2023 year

Year the Biden administration banned advanced semiconductor supply to China
2022 year

Matched pairs

Each item on the left goes with the one on the right.

DeepMind safety pioneer who joined the UK AI Safety Institute → Geoffrey Irving

Biden's June 2023 AI czar → Ben Buchanan, a Georgetown AI and cybersecurity professor

Where AI labs signed voluntary commitments in July 2023
→ At the White House, with no major US frontier lab declining

Self-test

Answer each question from memory, then check your work against the answer key.

1. How fast did ChatGPT reach 100 million users after its 30 November 2022 launch?

A Within two months
B Within a year
C Within one week
D Within five days

2. For Mallaby, what did ChatGPT's launch mark?

A The first time deep learning beat a human benchmark
B OpenAI's abandonment of its research-preview model
C The moment the AI race became public
D The invention of the Transformer architecture

3. In what year was the Transformer paper — the architecture ChatGPT ultimately rests on — published?

A 2015
B 2012
C 2020
D 2017

4. ChatGPT reached its first million users in how long after launch?

A Five hours
B Five days
C Five months
D Five weeks

5. ChatGPT launched as a polished interface on top of which underlying model?

A Gemini
B GPT-2
C GPT-3.5
D GPT-4

6. In AI, what is 'alignment'?

A. Tuning chip supply to match a lab's compute demand
B. Ensuring an AI reliably pursues human-intended goals as it grows more capable
C. Coordinating release dates across competing AI labs
D. Matching a model's training data to its benchmark test set

7. Which researcher moved from OpenAI to DeepMind in autumn 2019 with an explicit alignment brief, and later became a key figure at the UK AI Safety Institute?

A. Geoffrey Irving
B. Dario Amodei
C. Paul Christiano
D. Demis Hassabis

8. Which prominent AI researcher did NOT sign the Center for AI Safety's May 2023 one-sentence letter?

A. Yoshua Bengio
B. Sam Altman
C. Geoffrey Hinton
D. Yann LeCun

9. The May 2023 one-sentence letter compared AI extinction risk to which other societal-scale risks?

A. Cyberattacks and financial collapse
B. Pandemics and nuclear war
C. Asteroid impacts and supervolcanoes
D. Climate change and famine

10. Which group correctly lists researchers who signed the May 2023 one-sentence letter?

A. Hinton, Bengio, Hassabis, Altman and Amodei
B. Bengio, LeCun, Altman and Hassabis
C. LeCun, Irving, Buchanan and Sunak
D. Hinton, LeCun, Hassabis and Amodei

11. Whom did the Biden administration appoint to its new senior AI coordination role in June 2023?

A. Mustafa Suleyman
B. Geoffrey Irving
C. Ben Buchanan
D. Dario Amodei

12. What was Ben Buchanan's academic and professional background before becoming AI czar?

A. A DeepMind alignment researcher recruited from OpenAI
B. A semiconductor export-control lawyer at Commerce
C. A Center for AI Safety policy director
D. A Georgetown professor of AI and cybersecurity on the National Security Council

13. Where were the July 2023 voluntary AI commitments signed?

A. At the Georgetown School of Foreign Service
B. At OpenAI's headquarters
C. At Bletchley Park
D. At the White House

14. Which was NOT among the voluntary commitments the labs signed in July 2023?

A. Public reporting of capabilities
B. Cybersecurity investment
C. Binding financial penalties for backsliding
D. Pre-release safety testing

15. How did major US frontier labs respond to the July 2023 voluntary commitments?

A All declined pending congressional legislation
B No major US frontier lab declined to sign
C Only OpenAI and Anthropic signed
D Meta and Amazon publicly refused

16. Which two countries announced national AI Safety Institutes around the November 2023 Bletchley summit?

A United Kingdom and United States
B Israel and Singapore
C France and Germany
D Canada and Japan

17. How many countries sent delegations that signed the Bletchley Declaration?

A 18
B 38
C 28
D 50

18. In what month and year was the first international AI Safety Summit held at Bletchley Park?

A November 2023
B July 2023
C October 2023
D November 2022

19. Why was Bletchley Park a pointed venue for the first AI Safety Summit?

A It was DeepMind's original London laboratory
B It was where the Transformer paper was written
C It housed the UK's largest AI supercomputer
D It was Alan Turing's wartime codebreaking home

20. Whose conversation with Prime Minister Rishi Sunak, about six months earlier, led to the Bletchley summit?

A Ben Buchanan
B Demis Hassabis
C Sam Altman
D Geoffrey Hinton

21. In what year did the Biden administration impose broad export controls on advanced semiconductors bound for China?

A 2017
B 2023
C 2020
D 2022

22. What did the Biden administration's October 2022 controls actually restrict?

A Export of trained model weights to China
B Chinese investment in US AI laboratories
C The supply of advanced semiconductors to China
D Chinese researchers' access to US AI conferences

23. Which world champion's defeat by AlphaGo was described inside Beijing as a 'Sputnik moment'?

A Ke Jie
B Lee Sedol
C Garry Kasparov
D Fan Hui

24. From what year has Hassabis tracked China's competitive intent — the year AlphaGo's win over the reigning Go champion was called a 'Sputnik moment'?

A 2015
B 2017
C 2020
D 2022

25. How does Mallaby read the significance of the chip-export ban?

A It was the direct cause of the Bletchley Declaration
B It gave China an unbeatable lead in advanced semiconductors
C It closed the most consequential lane of US-China technology cooperation
D It was a symbolic gesture with no effect on the AI race

Go deeper

ONLINE This lesson on the web flashfeedlearningapp.com/study-guides/infinity-machine/racegpt-and-bletchley

STUDY The Infinity Machine (book) in the FlashFeed app flashfeedlearningapp.com/packs/infinity-machine

LESSON 10 OF 12

Step by Step — Gemini and the comeback

Gemini ships · Mallaby's closing register

WHAT YOU'LL BE ABLE TO DO

- Define Gemini as Google DeepMind's flagship multimodal model family (launched 2023) and identify Gemini Ultra
- Explain the MMLU benchmark (57 subjects) and Gemini Ultra's announced 90% score
- Recall Gemini 1.5 Pro (2024) with its million-token context window and sparse mixture-of-experts design, and why RL couldn't crack post-training
- Identify the two triggers Mallaby gives for the November 2023 Altman firing — undisclosed Gulf fundraising and pushing out critical board members
- Explain Mallaby's closing register — Hassabis as both Nobel laureate and CEO, preferring the label 'realist'

Gemini ships

The merger between Google Brain and DeepMind was sold as a synergy story. Gemini was the first test of whether it was actually true — and getting there required resolving a culture war as much as a research programme.

Gemini, December 2023

On 6 December 2023 **Google DeepMind** launched **Gemini**, its first frontier model built jointly by the merged London and Mountain View teams. Two smaller variants shipped immediately; the headline **Gemini Ultra** was previewed against GPT-4 on **MMLU (Massive Multitask Language Understanding)** — a benchmark of 57 subjects spanning maths, humanities, ethics, and law. Gemini Ultra's announced MMLU score was **90%**, edging past GPT-4's published number. The launch arrived roughly fourteen months after ChatGPT; late by Silicon Valley standards, but the first proof that the merger could ship at frontier scale.

Cultures collide

Building Gemini exposed a deep friction between the two halves of the merged organisation. Mountain View's culture was "crash projects out quickly"; London's was "measure each tweak". Mallaby describes 80-hour weeks, two-time-zone Slack channels, and **Sergey Brin** re-engaging after years away from day-to-day work to personally clear roadblocks. Speed eventually won the internal debates — Mountain View's instinct to ship and iterate beat London's instinct to wait until the science was settled. It was a preview of a tension that would not disappear with the launch.

David Silver and the RL wall

David Silver — the AlphaGo architect and one of DeepMind's most celebrated researchers — agreed after ChatGPT to lead Gemini's post-training. His instinct was to apply sophisticated **reinforcement learning** techniques to language model alignment, the approach that had conquered Go. After roughly six months of frustration he quit the post-training team. The core problem: RL requires a clear reward signal, and language quality resists easy scoring. "It wasn't clear how machine RL could distinguish a good haiku from a great one," Mallaby explains. A Bard veterans team that concentrated on data cleaning and **RLHF** widgets shipped the upgrade instead.

Mechanistic interpretability

Geoffrey Irving's safety group at DeepMind extended a line of research that had earlier taken shape at Anthropic: identifying the specific circuits inside a network responsible for specific behaviours. Mallaby calls **mechanistic interpretability** "chinks of light in the black box" — a rare alignment programme that produced findings with credible commercial uptake, not just academic citations. Irving later left DeepMind to become chief scientist at the UK's **AI Safety Institute**, carrying the field's credibility with him into government-backed safety work.

Mallaby's closing register

Gemini 1.5 Pro shifted the internal power balance at Google DeepMind, the Altman episode rattled Silicon Valley's governance pretensions, and Mallaby closes his book on a paradox that the AI race has yet to resolve.

Gemini 1.5 Pro, February 2024

On 15 February **2024** Google DeepMind released **Gemini 1.5 Pro**, whose headline capability at launch was a **million-token context window** paired with a **sparse mixture-of-experts** architecture — meaning only a fraction of the model's parameters activate for any given input, keeping inference costs manageable. The model was built by a London-led "strike team" of the kind that had produced AlphaFold: small, scientifically minded, given unusual autonomy. Among the team were **Sebastian Borgeaud** and **Jack Rae**, who had returned to DeepMind after a period at OpenAI. It was the first Gemini release where London set the research agenda.

"First innings of the game"

Mallaby catches **Hassabis** on Christmas vacation in late 2023, still calibrating his position relative to OpenAI. "OpenAI had the scaling and engineering and product focus that we didn't have. But now with Gemini we are matching that. And I think we still have the better ideas." His view was that ChatGPT represented "just the start" — the AI race was in its first innings, not its ninth. It is the kind of framing that reads differently depending on whether you find it energising or alarming, and Mallaby declines to adjudicate.

Why the OpenAI board fired Altman

Mallaby identifies two distinct triggers for the November 2023 firing. First, **undisclosed fundraising overtures to investors in the Persian Gulf** — conversations Altman had not disclosed to the board.

Second, a **pattern of attempting to push out independent board members** who criticised him. The board, whose charter charged it with pursuing "safe AI for the benefit of the world", concluded Altman could not be trusted with that mission. What the board underestimated was how thoroughly Silicon Valley and OpenAI's workforce would close ranks behind him within days.

Mallaby's closing register

Mallaby ends the book with Hassabis simultaneously a **Nobel laureate** and the **chief executive of one of the most powerful AI labs in the world** — a dual status that captures the book's central tension. By the epilogue, Hassabis says he prefers the word **"realist"** to either "doomer" or "accelerationist" to describe his own position on AI. Mallaby does not pretend to resolve the paradox he closes on: that the most morally serious major figure in the race is also racing as hard as anybody to build the technology he has spent his career warning the world about.

Key terms

MMLU (Massive Multitask Language Understanding)

A benchmark of 57 subjects spanning maths, humanities, ethics and law. In 2020, GPT-3 scored 44%; in 2023 GPT-4 reached 86%; Gemini Ultra hit 90%, the first AI system to exceed top human experts. Each generational bump on MMLU is roughly a year of student-level cognition gained.

Gemini

Gemini is Google DeepMind's flagship family of multimodal AI models, launched in 2023. Gemini was the lab's answer to OpenAI's GPT-4 and marked its return to the research frontier.

Key concepts

Gemini Ultra

- The headline variant of Google DeepMind's first frontier model built jointly by its London and Mountain View teams.
- Launched 6 December 2023, it was previewed head-to-head against GPT-4.
- Its announced score of 90% on a 57-subject benchmark edged past GPT-4's published number.

MMLU

- A benchmark used to preview Google DeepMind's flagship model against GPT-4.
- It spans 57 subjects, from maths and humanities to ethics and law.
- The London-Mountain View model's announced 90% on it edged past GPT-4's published number.

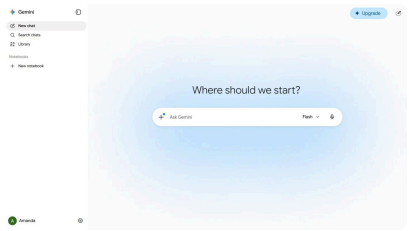
Gemini 1.5 Pro

- A February 2024 model release built by a London-led 'strike team' of the kind that had produced AlphaFold.
- Its headline capability was a million-token context window.
- It was the first release of Google DeepMind's flagship line where London, not Mountain View, set the research agenda.

sparse mixture-of-experts

- An efficiency-minded architecture paired with a million-token context window in a February 2024 model.
- Only a fraction of the model's parameters activate for any given input, keeping inference costs manageable.

Key facts



Wikimedia Commons · Google · Public domain

Gemini, December 2023

On 6 December 2023 Google DeepMind launched Gemini, its first frontier model designed jointly by the merged London and Mountain View teams. Two smaller versions shipped immediately; the headline Gemini Ultra was previewed against GPT-4 on MMLU. The launch was Hassabis's response to ChatGPT — late, but the first proof that the merger could ship at all.

Cultures collide

Building Gemini was as much a management challenge as a research one. Mountain View's culture was "crash projects out quickly"; London's was "measure each tweak". Mallaby describes 80-hour weeks, two-time-zone Slack channels, and Sergey Brin re-engaging "after years of playboy life" to clear roadblocks. Speed eventually won the internal debates.

David Silver and the RL wall

After ChatGPT, David Silver — the AlphaGo veteran — agreed to lead Gemini's post-training. He hoped to apply sophisticated machine RL to language models. After six months of frustration, he quit the post-training team. "It wasn't clear how machine RL could distinguish a good haiku from a great one," Mallaby explains. A Bard veterans team that focused on cleaning data and RLHF widgets shipped the upgrade instead.

Mechanistic interpretability

Geoffrey Irving's safety group at DeepMind extended a line of research started at Anthropic: identifying the specific circuits inside a network responsible for specific behaviours. Mallaby calls the field — mechanistic interpretability — "chinks of light in the black box". By 2024 it was one of the few alignment programmes with credible commercial uptake.

Gemini Ultra's preview-only debut

At Gemini's launch, two smaller variants shipped immediately, but the headline Gemini Ultra was only previewed — benchmarked against GPT-4 rather than released to the public right away.

Gemini 1.5 Pro, February 2024

On 15 February 2024 Google DeepMind released Gemini 1.5 Pro — a million-token context window, sparse mixture-of-experts architecture, and the first model from a London-led, scientifically minded "strike team" of the kind that had built AlphaFold. Among the team were Sebastian Borgeaud and Jack Rae, who had returned from a stint at OpenAI.

"First innings of the game"

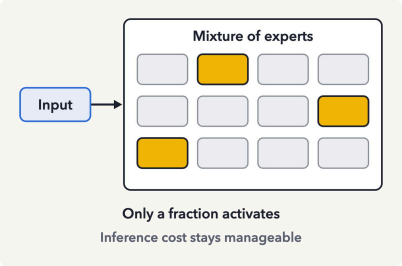
Mallaby's Hassabis, on Christmas vacation in late 2023: "OpenAI had the scaling and engineering and product focus that we didn't have. But now with Gemini we are matching that. And I think we still have the better ideas." He insisted ChatGPT was "just the start" — the AI race was in its first innings, not its ninth.

Why the OpenAI board fired Altman

Mallaby identifies two triggers. First, undisclosed fundraising overtures to investors in the Persian Gulf. Second, a pattern of trying to push out independent board members who criticised him. The board, charged with "safe AI for the benefit of the world", concluded Altman could not be trusted with the mission. They underestimated how thoroughly Silicon Valley would close ranks behind him.

Mallaby's closing register

Mallaby ends the book in 2025–26, with Hassabis simultaneously a Nobel laureate and the chief executive of one of the most powerful AI labs in the world. The book closes on a paradox he does not pretend to resolve: that the most morally serious major figure in AI is also racing as hard as anybody to build the technology he has spent his life warning the world about.



FlashFeed original diagram · CC-BY-4.0

Only a fraction fires

Gemini's sparse mixture-of-experts architecture means only a fraction of the model's parameters activate for any given input, which keeps inference costs manageable despite its overall size.

By the numbers

Gemini Ultra's announced MMLU score (percentage)

90 %

Year Google DeepMind released Gemini 1.5 Pro

2024 year

Matched pairs

Each item on the left goes with the one on the right.

Gemini 1.5 Pro's headline capability at launch
→ A million-token context window with a mixture-of-experts design

Two triggers Mallaby gives for the Altman firing
→ Undisclosed Gulf fundraising and pushing out critical board members

Hassabis's dual status as the book closes in 2025-26
→ A Nobel laureate and chief executive of a leading AI lab

Self-test

Answer each question from memory, then check your work against the answer key.

1. What is Gemini?

- A Meta's open-weights model family released alongside Llama
- B OpenAI's successor to GPT-4, launched in 2023
- C Google DeepMind's flagship family of multimodal AI models, launched in 2023
- D Anthropic's mechanistic-interpretability research programme

2. On what date did Google DeepMind launch Gemini?

- A 6 December 2023
- B 15 February 2024
- C 30 November 2022
- D 6 December 2022

3. At the launch, which Gemini variant was the headline model previewed against GPT-4 rather than shipped immediately?

- A Gemini Flash
- B Gemini Nano
- C Gemini 1.5 Pro
- D Gemini Ultra

4. Which statement about Gemini's December 2023 launch is NOT true?

- A Gemini Ultra was previewed against GPT-4 on the MMLU benchmark
- B Every variant, including the headline model, shipped to users immediately
- C It was the first frontier model built jointly by the merged London and Mountain View teams
- D The launch arrived roughly fourteen months after ChatGPT

5. In Mallaby's account, why did the Gemini launch matter for Google DeepMind beyond the model itself?

- A It was the first model trained entirely in London without Mountain View
- B It was the first time Google published its full training data
- C It was the first proof the Brain–DeepMind merger could ship at frontier scale
- D It was the first commercial product DeepMind had ever sold

6. What does the acronym MMLU stand for?

- A Massive Multitask Language Understanding
- B Multi-Metric Language Uniformity
- C Multimodal Machine Learning Utility
- D Massive Model Learning Universe

7. How many subjects does the MMLU benchmark cover?

A 100
B 57
C 40
D 24

8. What was Gemini Ultra's announced MMLU score?

A 44%
B 86%
C 95%
D 90%

9. Which of these is NOT one of the areas MMLU's 57 subjects are described as spanning?

A Humanities
B Maths
C Computer vision
D Law

10. Which sequence correctly ranks the MMLU scores for GPT-3 (2020), GPT-4 (2023) and Gemini Ultra?

A 44% → 86% → 90%
B 44% → 90% → 86%
C 56% → 80% → 90%
D 60% → 86% → 95%

11. According to Mallaby, why did David Silver's reinforcement-learning approach fail to crack Gemini's post-training?

A Reinforcement learning could not scale to a million-token context
B Human raters generated too much data to process
C The pretrained base model was never shared with his team
D Language quality offers no clear reward signal for RL to optimize

12. What was Gemini 1.5 Pro's headline capability at launch?

A A 32,000-token context window
B A million-token context window
C Native real-time voice conversation
D A 90% score on the MMLU benchmark

13. When did Google DeepMind release Gemini 1.5 Pro?

A 6 December 2023
B 15 February 2024
C 30 November 2023
D 15 February 2025

14. What does a sparse mixture-of-experts architecture mean for how the model runs?

A The model is retrained from scratch for each specialist domain
B Human expert raters score every output before it is returned
C Several separate models vote and the majority answer wins
D Only a fraction of the model's parameters activate for any given input

15. After David Silver quit the post-training team, which approach actually shipped Gemini's post-training upgrade?

A. A London strike team using sparse mixture-of-experts
B. Silver's machine-RL pipeline, scaled up by a larger team
C. A Bard veterans team focused on data cleaning and RLHF widgets
D. Geoffrey Irving's mechanistic-interpretability circuits

16. In what year did Google DeepMind release Gemini 1.5 Pro?

A. 2024
B. 2025
C. 2022
D. 2023

17. On what date did Google DeepMind release Gemini 1.5 Pro?

A. 17 November 2023
B. 15 February 2024
C. 22 January 2024
D. 6 December 2023

18. Which pair does Mallaby identify as the two triggers for the OpenAI board's November 2023 firing of Altman?

A. A failed safety evaluation, and misreporting compute spending
B. The 100x profit cap, and an undisclosed deal with Microsoft
C. Undisclosed Gulf fundraising, and a pattern of pushing out critical board members
D. A leaked model, and refusing to sign the White House safety commitments

19. Undisclosed fundraising overtures to investors in which region formed one of the two triggers?

A. Latin America
B. Scandinavia
C. Southeast Asia
D. The Persian Gulf

20. A news report says Altman was fired over a failed safety evaluation. On Mallaby's account, what is wrong with that claim?

A. Neither trigger concerned model testing; both were about undisclosed conduct and board control
B. Evaluations only began after the firing, so none could have failed
C. The evaluation did fail, but the board judged the lapse too minor to act on
D. Running safety evaluations was the board's own job, not Altman's

21. The OpenAI board's charter charged it with pursuing what mission — the standard against which it judged Altman?

A. "Compute at any cost"
B. "Open research, openly published"
C. "Safe AI for the benefit of the world"
D. "Artificial general intelligence for shareholders"

22. What does Mallaby say the board underestimated when it fired Altman?

A. How much regulators would intervene in the dispute
B. How quickly Microsoft would withdraw its funding
C. How far ahead Google DeepMind's Gemini already was
D. How thoroughly Silicon Valley and OpenAI's workforce would close ranks behind him

23. In the book's epilogue, which label does Hassabis say he prefers for his own stance on AI?

A. Realist
B. Accelerationist
C. Optimist
D. Doomer

24. As Mallaby closes the book, what is Hassabis's dual status?

A. A Nobel laureate and chief scientist of the AI Safety Institute
B. A Nobel laureate and chief executive of one of the most powerful AI labs
C. A chess grandmaster and OpenAI board member
D. A Turing Award winner and a UK government minister

25. What paradox does Mallaby close the book on, and decline to resolve?

A. That the most morally serious figure in the race is also racing hardest to build the technology he warns about
B. That larger models became cheaper to run than smaller ones
C. That AI safety research is funded mostly by the labs it scrutinises
D. That the UK produced the science while the US captured the profits

26. Which of these is NOT part of how Mallaby closes the book on Hassabis?

A. His standing as a Nobel laureate
B. His job running one of the most powerful AI labs in the world
C. His preference for "realist" over "doomer" or "accelerationist"
D. His decision to step back from the lab to work full-time on safety

27. A reader says Mallaby ends by judging Hassabis a hypocrite. Why does that misread the closing register?

A. Mallaby argues the tension dissolved once Gemini 1.5 Pro shipped
B. Mallaby settles the question by crediting the Nobel Prize
C. Mallaby explicitly declines to resolve the paradox he closes on
D. Mallaby concludes Hassabis is at heart a doomer

Go deeper

ONLINE This lesson on the web flashfeedlearningapp.com/study-guides/infinity-machine/step-by-step-gemini-and-the-comeback

STUDY The Infinity Machine (book) in the FlashFeed app flashfeedlearningapp.com/packs/infinity-machine

LESSON 11 OF 12

Quick Recall — Key people and places

People and places

WHAT YOU'LL BE ABLE TO DO

- Recall DeepMind's founding trio — Hassabis (2024 Nobel), Legg (coined AGI), Suleyman (Microsoft AI CEO)
- Identify the scientists behind DeepMind's milestones — David Silver (AlphaGo), John Jumper (AlphaFold 2), Vlad Mnih (Atari DQN)
- Identify DeepMind's first investor Peter Thiel (the ~£400m Google acquisition) and OpenAI's protagonists Sam Altman and Ilya Sutskever
- Identify the AI safety and policy figures and the summit site — Geoffrey Irving (DeepMind→UK AISI), Ben Buchanan (Biden's AI czar), and Bletchley Park

This lesson is intentionally light on prose and heavy on flashcards. *The Infinity Machine* covers a dozen pivotal figures and a handful of places that recur across the book — and once you know who's who, the narrative reads as a much tighter braid. Use these pairs to drill the names until they come without hesitation.

People and places

The DeepMind triangle is the place to start. **Demis Hassabis** is the central subject — chess prodigy, neuroscientist, cofounder and CEO. **Shane Legg** is the cerebral New Zealander who coined "AGI" with Marcus Hutter in 2002 and gave the Halloween Scenario lecture that drew Hassabis in. **Mustafa Suleyman** is the operator, the third cofounder, who ended up running Microsoft AI by 2024. **David Silver** is Hassabis's Cambridge friend, the architect of AlphaGo and a long-running technical second.

Two figures defined the science. **Vlad Mnih** led the Atari DQN work — the foundational reinforcement-learning system on which every later DeepMind result built. **John Jumper** arrived at DeepMind in October 2017 and went on to lead AlphaFold 2, sharing the 2024 Nobel Prize in Chemistry with Hassabis.

Two figures defined the money. **Peter Thiel**, through Founders Fund, anchored DeepMind's first round after the Singularity Summit pitch in 2010 and was the investor who muttered "Demis's destiny" walking to the car. **Larry Page and Sergey Brin** brought DeepMind into Google in 2014 for roughly £400 million.

Two figures defined the rivalry. **Sam Altman**, OpenAI's CEO, was fired on 17 November 2023 and reinstated five days later — the moment Mallaby treats as the death of any plausible institutional brake on the race. **Ilya Sutskever**, OpenAI's chief scientist and the engineer of the board revolt, reversed himself when Microsoft offered to absorb the company.

Two figures defined the response. **Geoffrey Irving** built DeepMind's alignment programme and later joined the UK AI Safety Institute. **Ben Buchanan** was the Biden White House's AI czar, the architect of the July 2023 voluntary commitments and the October 2023 executive order.

One place anchors the story: **Bletchley Park**, Alan Turing's wartime workplace, where twenty-eight countries signed the first AI safety declaration in November 2023. If you know all of these names — and the place — you have the social skeleton of the book in your head, and you can read the rest of *The Infinity Machine* with the right cast of characters already in view.

Key concepts

Peter Thiel

- An early investor who anchored DeepMind's very first funding round.
- He backed the company through Founders Fund after its 2010 Singularity Summit pitch.
- Walking to the car after that pitch, he reportedly muttered "Demis's destiny."

Sam Altman

- The chief executive of a rival lab at the center of the AI race.
- He led OpenAI, and Mallaby treats a crisis around him as the death of any institutional brake on the race.
- Fired on 17 November 2023, he was reinstated just five days later.

Ilya Sutskever

- A leading AI scientist who helped engineer a dramatic corporate revolt.
- As OpenAI's chief scientist, he was central to the board's move against its CEO.
- He reversed himself once Microsoft offered to absorb the company.

Geoffrey Irving

- A researcher focused on making AI systems safe rather than merely capable.
- He built DeepMind's alignment programme.
- He later joined the UK AI Safety Institute.

Ben Buchanan

- A government official who shaped US policy on artificial intelligence.
- He served as the Biden White House's AI czar.
- He was the architect of the July 2023 voluntary commitments and the October 2023 executive order.

Matched pairs

Each item on the left goes with the one on the right.

DeepMind cofounder who won the 2024 Nobel Prize in Chemistry → Demis Hassabis

Cofounder who coined "AGI" → Shane Legg

Cofounder who became Microsoft AI CEO → Mustafa Suleyman

Cambridge friend who led AlphaGo → David Silver

Scientist who led AlphaFold 2 → John Jumper

First investor in DeepMind → Peter Thiel

Researcher who built the Atari DQN → Vlad Mnih

OpenAI CEO fired and rehired in 2023 → Sam Altman

OpenAI cofounder who led the board revolt → Ilya Sutskever

Biden's AI czar from June 2023 → Ben Buchanan

DeepMind safety lead who joined UK AISI → Geoffrey Irving

Site of the first global AI Safety Summit → Bletchley Park

Self-test

Answer each question from memory, then check your work against the answer key.

1. Which DeepMind cofounder coined the term "AGI" with Marcus Hutter in 2002?

A Shane Legg
B Demis Hassabis
C Mustafa Suleyman
D David Silver

2. Which DeepMind cofounder was running Microsoft AI by 2024?

A Shane Legg
B Mustafa Suleyman
C Demis Hassabis
D Ilya Sutskever

3. Which DeepMind cofounder shared the 2024 Nobel Prize in Chemistry?

A Shane Legg
B Mustafa Suleyman
C Demis Hassabis
D David Silver

4. In what year did Shane Legg and Marcus Hutter coin the term "AGI"?

A 1997
B 2010
C 2014
D 2002

5. Which of these people was NOT one of DeepMind's three cofounders?

A Mustafa Suleyman
B David Silver
C Shane Legg
D Demis Hassabis

6. Who led the Atari DQN work — the foundational reinforcement-learning system every later DeepMind result built on?

A Vlad Mnih
B John Jumper
C David Silver
D Geoffrey Irving

7. Which scientist does the book describe as Hassabis's Cambridge friend and a long-running technical second at DeepMind?

A John Jumper
B Geoffrey Irving
C David Silver
D Vlad Mnih

8. Which scientist arrived at DeepMind in October 2017 and went on to lead AlphaFold 2?

A Geoffrey Irving
B David Silver
C Vlad Mnih
D John Jumper

9. John Jumper shared the 2024 Nobel Prize in Chemistry with which DeepMind figure?

A Shane Legg
B Vlad Mnih
C David Silver
D Demis Hassabis

10. Which scientist–project pairing is NOT correct?

A Vlad Mnih — AlphaFold 2
B John Jumper — AlphaFold 2
C David Silver — AlphaGo
D Vlad Mnih — Atari DQN

11. Who was DeepMind's first investor, anchoring the round after the 2010 Singularity Summit pitch and muttering "Demis's destiny" walking to the car?

A Sergey Brin
B Peter Thiel
C Sam Altman
D Larry Page

12. Which pair brought DeepMind into Google in 2014?

A Sam Altman and Ilya Sutskever
B Peter Thiel and Marcus Hutter
C Larry Page and Sergey Brin
D Eric Schmidt and Sundar Pichai

13. Roughly how much did Google pay to bring DeepMind in-house in 2014?

A About £4 billion
B About £1.2 billion
C About £400 million
D About £40 million

14. On what date was OpenAI CEO Sam Altman fired — the firing Mallaby treats as the death of any institutional brake on the race?

A 22 November 2023
B 17 November 2023
C 1 October 2023
D 30 November 2022

15. Ilya Sutskever, OpenAI's chief scientist, engineered the board revolt — then reversed himself when what happened?

A Microsoft offered to absorb the company
B The board reinstated Altman on its own
C Founders Fund threatened to pull its investment
D Google offered to buy OpenAI outright

16. About how much did Google pay to acquire DeepMind in 2014, per Mallaby?

A £800 million
B £4 billion
C £40 million
D £400 million

17. Which figure is closest to the price tag of Google's 2014 DeepMind deal?

A £40 million
B £4 million
C £400 million
D £4 billion

18. Who built DeepMind's alignment programme?

A David Silver
B Geoffrey Irving
C Ben Buchanan
D Shane Legg

19. Who served as the Biden White House's AI czar?

A Ben Buchanan
B Geoffrey Irving
C Peter Thiel
D Mustafa Suleyman

20. How many countries signed the first AI safety declaration at the November 2023 global AI Safety Summit?

A One hundred and ninety-five
B Twelve
C Forty
D Twenty-eight

21. The first global AI Safety Summit, in November 2023, was held at which site?

A The Royal Society, London
B Bletchley Park
C King's College, Cambridge
D Chequers

22. Which figure–role pairing is NOT correct?

A Geoffrey Irving — UK AI Safety Institute
B Ben Buchanan — Biden White House AI czar
C Ben Buchanan — July 2023 voluntary commitments
D Geoffrey Irving — October 2023 executive order

Go deeper

ONLINE This lesson on the web flashfeedlearningapp.com/study-guides/infinity-machine/quick-recall-key-people-and-places

STUDY The Infinity Machine (book) in the FlashFeed app flashfeedlearningapp.com/packs/infinity-machine

LESSON 12 OF 12

Key Concepts

Key People · Key Things · Key Events

WHAT YOU'LL BE ABLE TO DO

- Recognise the key people of the book — Hassabis, Legg, Suleyman, Silver, Mnih, Jumper — including Suleyman's conflict-resolution background
- Recognise the foundational AI techniques — reinforcement learning, deep learning, neural networks, and AGI
- Recognise DeepMind's landmark systems — AlphaGo, DQN, and AlphaFold — including what made DQN notable (learning from raw pixels)
- Recognize the pivotal events: Google's DeepMind acquisition, the day Pong clicked, the Singularity Summit, the Atari strike team, the Musk–Page dinner.

A quick-recognition gallery of the key people, places, things, and events from The Infinity Machine (book).

Key People

Demis Hassabis — He is a co-founder of a leading AI company. He received the 2024 Nobel Prize in Chemistry for his work on AlphaFold. He was a chess prodigy, mastering the game by age 13.

Shane Legg — He is known for coining a term related to advanced AI. He co-founded a prominent AI research lab in London. He proposed the concept of Artificial General Intelligence during discussions in 2002.

Mustafa Suleyman — He is one of the co-founders of a well-known AI company. He has a background in conflict resolution and diplomacy. He was instrumental in pitching to a billionaire investor for initial funding.

David Silver — He is a key figure in the development of AI at a major lab. He led a project that achieved a significant milestone in the game Go. He met the main subject of the book during their university years.

Vlad Mnih — He is recognized for his pioneering work in reinforcement learning. He led a team that created a neural network to play classic video games. His research contributed to a major breakthrough in AI gaming.

John Jumper — Recruited to a London AI lab in 2017, he led its leap into structural biology. He had bounced across physics and chemistry before finding his defining problem. In 2024 he shared the Nobel Prize in Chemistry with the lab's founder for predicting protein shapes.

Key Things

Reinforcement Learning — A type of machine learning where agents learn by interacting with an environment. It focuses on taking suitable actions to maximize cumulative reward. This learning paradigm was crucial for DeepMind's early successes.

Deep Learning — A subset of machine learning inspired by the structure of the human brain. It uses neural networks with many layers to analyze various forms of data. This technique underpins many of DeepMind's advancements, including AlphaGo.

Neural Network — A computational model inspired by the way biological neural networks in the human brain process information. They consist of interconnected nodes or 'neurons' that work together to solve complex problems. DeepMind utilized this structure to develop systems like AlphaFold.

AlphaGo — An AI program developed by DeepMind to play the game of Go. It famously defeated the world champion Lee Sedol in 2016, marking a significant milestone in AI. This program showcased the power of reinforcement learning and deep learning combined.

DQN (Deep Q-Network) — A reinforcement learning algorithm that combines Q-learning with deep learning techniques. It was designed to learn how to play Atari games directly from raw pixel data. This architecture was pivotal in demonstrating the potential of general AI systems.

AlphaFold — An AI system developed by DeepMind to predict protein structures. It utilizes deep learning to solve complex biological problems. This technology has significant implications for drug discovery and biology.

AGI (Artificial General Intelligence) — A theoretical form of AI that can understand, learn, and apply intelligence across a wide range of tasks. It aims to replicate human cognitive abilities in machines. DeepMind's mission is often associated with the pursuit of this concept.

Key Events

Google buys DeepMind — In January 2014, a major tech company made a significant acquisition in the AI field. This acquisition was Europe's largest at the time, valued at around £400 million. The deal included a condition for an independent ethics board to oversee AI applications.

The day Pong clicked — On 7 July 2013, a significant milestone in AI development occurred during a major tennis event. A neural network learned to play a classic arcade game from scratch, impressing its creators. By the evening, the AI was able to beat the built-in opponent, moving towards human-level performance.

Singularity Summit — In August 2010, a significant conference focused on the future of technology took place in San Francisco. This event featured discussions on advanced AI and its implications for society. A trio of AI researchers pitched a groundbreaking lab concept aimed at building AGI safely.

The Halloween Scenario — In October 2009, a lecture at UCL addressed the urgent need for AI safety planning. The presenter argued that advanced AI was no longer just a theoretical concept. This talk sparked ongoing discussions about the implications of powerful AI technologies.

The Atari strike team — In 2013, a small team at a leading AI lab tackled a challenge involving classic video games. They aimed to create a single neural network that could learn to play multiple games. This project marked the lab's first major public success in reinforcement learning.

Musk vs Page on the lawn — In October 2012, a dinner conversation at a tech mogul's home shifted to the topic of AI. One guest warned about the potential dangers of advanced AI, while the host had a contrasting view. This discussion is seen as the moment a key alliance in AI development began to fracture.

Key concepts

Demis Hassabis

- He is a co-founder of a leading AI company.
- He received the 2024 Nobel Prize in Chemistry for his work on AlphaFold.
- He was a chess prodigy, mastering the game by age 13.

Shane Legg

- He is known for coining a term related to advanced AI.
- He co-founded a prominent AI research lab in London.
- He proposed the concept of Artificial General Intelligence during discussions in 2002.

Mustafa Suleyman

- He is one of the co-founders of a well-known AI company.
- He has a background in conflict resolution and diplomacy.
- He was instrumental in pitching to a billionaire investor for initial funding.

David Silver

- He is a key figure in the development of AI at a major lab.
- He led a project that achieved a significant milestone in the game Go.
- He met the main subject of the book during their university years.

Vlad Mnih

- He is recognized for his pioneering work in reinforcement learning.
- He led a team that created a neural network to play classic video games.
- His research contributed to a major breakthrough in AI gaming.

John Jumper

- Recruited to a London AI lab in 2017, he led its leap into structural biology.
- He had bounced across physics and chemistry before finding his defining problem.
- In 2024 he shared the Nobel Prize in Chemistry with the lab's founder for predicting protein shapes.

Reinforcement Learning

- A type of machine learning where agents learn by interacting with an environment.
- It focuses on taking suitable actions to maximize cumulative reward.
- This learning paradigm was crucial for DeepMind's early successes.

Deep Learning

- A subset of machine learning inspired by the structure of the human brain.
- It uses neural networks with many layers to analyze various forms of data.
- This technique underpins many of DeepMind's advancements, including AlphaGo.

Neural Network

- A computational model whose design is inspired by the way the human brain processes information.
- They consist of interconnected nodes or 'neurons' that work together to solve complex problems.
- DeepMind utilized this structure to develop systems like AlphaFold.

AlphaGo

- An AI program developed by DeepMind to play the game of Go.
- It famously defeated the world champion Lee Sedol in 2016, marking a significant milestone in AI.
- This program showcased the power of reinforcement learning and deep learning combined.

DQN (Deep Q-Network)

- A reinforcement learning algorithm that combines Q-learning with deep learning techniques.
- It was designed to learn how to play Atari games directly from raw pixel data.
- This architecture was pivotal in demonstrating the potential of general AI systems.

AlphaFold

- An AI system developed by DeepMind to predict protein structures.
- It utilizes deep learning to solve complex biological problems.
- This technology has significant implications for drug discovery and biology.

AGI (Artificial General Intelligence)

- A theoretical form of AI that can understand, learn, and apply intelligence across a wide range of tasks.
- It aims to replicate human cognitive abilities in machines.
- DeepMind's mission is often associated with the pursuit of this concept.

Google buys DeepMind

- In January 2014, a major tech company made a significant acquisition in the AI field.
- This acquisition was Europe's largest at the time, valued at around £400 million.
- The deal included a condition for an independent ethics board to oversee AI applications.

The day Pong clicked

- On 7 July 2013, a significant milestone in AI development occurred during a major tennis event.
- A neural network learned to play a classic arcade game from scratch, impressing its creators.
- By the evening, the AI was able to beat the built-in opponent, moving towards human-level performance.

Singularity Summit

- In August 2010, a significant conference focused on the future of technology took place in San Francisco.
- This event featured discussions on advanced AI and its implications for society.
- A trio of AI researchers pitched a groundbreaking lab concept aimed at building AGI safely.

The Halloween Scenario

- In October 2009, a lecture at UCL addressed the urgent need for AI safety planning.
- The presenter argued that advanced AI was no longer just a theoretical concept.
- This talk sparked ongoing discussions about the implications of powerful AI technologies.

The Atari strike team

- In 2013, a small team at a leading AI lab tackled a challenge involving classic video games.
- They aimed to create a single neural network that could learn to play multiple games.
- This project marked the lab's first major public success in reinforcement learning.

Musk vs Page on the lawn

- In October 2012, a dinner conversation at a tech mogul's home shifted to the topic of AI.
- One guest warned about the potential dangers of advanced AI, while the host had a contrasting view.
- This discussion is seen as the moment a key alliance in AI development began to fracture.

Self-test

Answer each question from memory, then check your work against the answer key.

1. Which DeepMind cofounder came from a background in conflict resolution and diplomacy?

A Vlad Mnih
B Mustafa Suleyman
C Shane Legg
D David Silver

2. Which figure from the book was a chess prodigy who had mastered the game by age 13?

A Vlad Mnih
B Shane Legg
C Demis Hassabis
D David Silver

3. Who proposed the concept of Artificial General Intelligence during discussions in 2002, coining the term?

A Shane Legg
B John Jumper
C Demis Hassabis
D Mustafa Suleyman

4. Who led the team that created a neural network to play classic Atari video games, contributing a major AI gaming breakthrough?

A Mustafa Suleyman
B John Jumper
C David Silver
D Vlad Mnih

5. Which person-and-contribution pairing is NOT correct?

A Vlad Mnih — pioneering work in reinforcement learning
B David Silver — led a project that hit a major milestone in the game Go
C Shane Legg — pitched a billionaire investor for the lab's initial funding
D John Jumper — led the London lab's leap into structural biology after 2017

6. Which technique has agents learn by interacting with an environment, taking suitable actions to maximize cumulative reward?

A Deep learning
B Reinforcement learning
C Artificial General Intelligence
D Neural networks

7. Deep learning is best described as which of the following?

A A method where agents learn purely from environment rewards
B A theoretical AI that applies intelligence across any task
C A hand-built expert system that encodes human rules explicitly
D A subset of machine learning using many-layered networks inspired by the brain

8. A neural network is built from interconnected units that work together to solve complex problems. What are those units called?

A Nodes, or 'neurons'
B Layers
C Weights
D Connections

9. Which concept names a theoretical AI that can understand, learn, and apply intelligence across a wide range of tasks, replicating human cognitive abilities?

A Deep Q-Network
B Deep learning
C Artificial General Intelligence
D Reinforcement learning

10. A system learns a game by trying actions and improving from a score signal, while a many-layered network reads the screen. Which two techniques does this combine?

A Supervised learning and Monte Carlo tree search
B Reinforcement learning and Monte Carlo tree search
C Supervised learning and deep learning
D Reinforcement learning and deep learning

11. What made DeepMind's DQN (Deep Q-Network) notable in how it learned to play Atari games?

A It learned directly from raw screen pixels
B It read the game's internal RAM state
C It was trained on recordings of expert human players
D It was given the rules of each game in advance

12. Which DeepMind program famously defeated world champion Lee Sedol in 2016?

A AlphaZero
B AlphaGo
C DQN (Deep Q-Network)
D AlphaFold

13. Which DeepMind system predicts protein structures, with significant implications for drug discovery and biology?

A AlphaGo
B AlphaFold
C AlphaZero
D DQN (Deep Q-Network)

14. DeepMind's Atari-playing system paired deep learning with which reinforcement learning algorithm?

A Gradient descent
B Monte Carlo tree search
C Backpropagation
D Q-learning

15. Which system-and-domain pairing is NOT correct?

A AlphaGo — predicting the three-dimensional structures of proteins
B DQN — classic Atari arcade games
C AlphaFold — protein structure prediction for biology and drug discovery
D AlphaGo — the board game Go

16. Google's January 2014 acquisition of DeepMind — Europe's largest at the time (~£400 million) — came with an unusual condition. What was it?

A Relocating DeepMind's headquarters to California
B Keeping all of DeepMind's research permanently closed-source
C An independent ethics board to oversee AI applications
D Barring DeepMind from taking any government contracts

17. On 7 July 2013, during a major tennis event, a neural network learned a classic arcade game from scratch and by evening beat the built-in opponent. Which game?

A Space Invaders
B Pong
C Breakout
D Asteroids

18. In August 2010, a trio of researchers pitched a lab concept aimed at building AGI safely at a conference on the future of technology. In which city was it held?

A San Francisco
B Palo Alto
C New York
D London

19. The 2013 'Atari strike team' at the London lab set out to build what?

A A network trained on recordings of human expert playthroughs
B An emulator for testing classic arcade hardware
C A separate hand-coded program tuned for each individual game
D A single neural network that could learn to play multiple games

20. At an October 2012 dinner at a tech mogul's home, one guest warned about the dangers of advanced AI while the host disagreed. What is this moment seen as?

A. The founding of the London AI lab's research team

B. The origin of the independent ethics board condition

C. The moment a key alliance in AI development began to fracture

D. The point when Google agreed to acquire DeepMind

Go deeper

ONLINE This lesson on the web flashfeedlearningapp.com/study-guides/infinity-machine/key-concepts

STUDY The Infinity Machine (book) in the FlashFeed app flashfeedlearningapp.com/packs/infinity-machine

Answer key

Lesson 1: The Sweetness — Why this book matters

1. D Penguin Press

2. A Sebastian Mallaby

Sebastian Mallaby is the author, previously known for writing about hedge funds and venture capital. Demis Hassabis is the book's central subject, not its author.

3. C Demis Hassabis

The book centres on Demis Hassabis — chess prodigy, game designer, neuroscience PhD, Nobel laureate. Mallaby is the author, not the subject.

4. B Hedge funds and venture capital

Mallaby came to this book from a career writing about hedge funds and venture capital. Neuroscience and chess belong to Hassabis's biography, not Mallaby's.

5. C Demis Hassabis is the book's author

Hassabis is the central subject, not the author — Mallaby wrote the book. The other three statements are all correct: Penguin Press, 2026.

6. B The instant a long-resistant problem finally breaks open

The sweetness is the private rush at the instant a problem cracks — explicitly not the public celebration or the announcement that follows.

7. D The Nobel Prize in Chemistry

The book opens on Hassabis collecting the 2024 Nobel Prize in Chemistry, awarded for AlphaFold's protein-structure work.

8. A Predicting protein structures from amino-acid sequences

The 2024 Chemistry Nobel recognised AlphaFold's prediction of protein structures from amino-acid sequences. AlphaGo, the Atari DQN and Gemini are milestones, but not the Nobel work.

9. B It explains why researchers pour a decade into one problem

The metaphor explains motivation — the internal signal, not the prize or the paper, is what sustains a decade on one problem. Mallaby uses the Nobel precisely to show the feeling is not merely romantic.

10. D Standing on stage as the prize is announced

The sweetness is the moment before the announcement, alone with the solution — a public stage is exactly what it is not.

11. A Artificial general intelligence

The title is Mallaby's metaphor for artificial general intelligence. Transformers, Gemini and reinforcement learning are components or milestones on the path, not the machine itself.

12. C A system matching or surpassing humans across the full range of cognitive tasks

AGI spans the full cognitive range, not one narrow domain. Beating experts at a single task — Go, or protein folding — is exactly the narrow capability AGI is contrasted with.

13. C A system able, in principle, to solve any problem a human mind can frame

The infinity machine is the full cognitive range — any problem a human can frame. AlphaFold, AlphaGo and the Atari DQN are narrow milestones on the path, however impressive.

14. B Through successive, linked breakthroughs rather than one dramatic leap

Mallaby's premise is that it comes from successive linked breakthroughs — reinforcement learning, deep learning, transformers — not a single dramatic leap or moonshot.

15. D Solving any problem a human mind can frame

Mallaby defines the infinity machine (AGI) as a system capable, in principle, of solving any problem a human mind can frame — the full cognitive range, not one narrow task. The distractors are narrow AI, AlphaFold, and game-playing proving grounds.

16. A A research lab pursuing AGI as a primary goal

From its 2010 founding, DeepMind explicitly framed itself as an AGI-first organisation — the word was in pitches and incorporation documents.

17. B The builders' biographies and rivalries are determinative of the outcome

Mallaby argues the people's biographies and rivalries are not incidental but determinative. He never claims compute or regulation settles the outcome.

18. D The Singularity Summit

The book describes DeepMind as born of a London poker game and a pitch at the Singularity Summit.

19. C Hassabis does not believe any single actor can slow it; he is a participant

Mallaby is explicit: Hassabis doubts any single actor can slow the race down — he is a participant, not a brake.

20. A Its influence has been proportionate to its modest size

Mallaby argues DeepMind's influence has been disproportionate to its size — the opposite of proportionate. The other three are core claims of the thesis.

Lesson 2: Destiny — Hassabis as a child prodigy

1. C 1976

Hassabis was born in 1976. 1984 is a tempting misread — that is roughly when, at age eight, he bought his ZX Spectrum with his chess winnings.

2. A Chinese Singaporean

His mother was Chinese Singaporean. Greek Cypriot is the classic swap — that was his father's heritage, not his mother's.

3. D Greek Cypriot

His father was Greek Cypriot — specifically Cypriot, not mainland Greek. Chinese Singaporean was his mother's heritage.

4. B He was born in Guildford, Surrey

He was born in Finchley, north London. Guildford is where Molyneux's Bullfrog studio was based — a place from later in the chapter, not his birthplace.

5. C July

Mallaby dates the birth precisely to July, in Finchley. The other months are plausible but the chapter names only July.

6. A An intuition that human cognition has ceilings — that the bottleneck was hardware, not effort

Playing chess, Hassabis could feel his own brain straining at certain positions in a way that suggested the limit was hardware, not effort — an early hint that intelligence might be something you could build rather than a fixed gift.

7. D 13

By age 13 Hassabis had reached chess master strength — Mallaby's marker of how far the prodigy had already gone before his teens were out.

8. B A deep feel for adversarial systems and calculation

Mallaby credits chess with giving Hassabis a deep intuition for adversarial systems and the discipline of sustained calculation, in addition to his separate intuition that human cognition has ceilings.

9. B Second-highest-rated

He was briefly the second-highest-rated player in the world for his age group — remarkable, but not the outright top spot.

10. A He took up chess at eight through a school chess club

He picked up chess at four by watching his father, not at eight through a club — eight is the age he bought the ZX Spectrum. The other three statements are all in the chapter.

11. D ZX Spectrum

He bought a ZX Spectrum and taught himself BASIC and machine code from the manuals. The other machines were contemporaries of the Spectrum but not the one he owned.

12. C Gödel, Escher, Bach

Mallaby names Douglas Hofstadter's Gödel, Escher, Bach. The other titles are plausible period books about minds and machines, but the chapter singles out only GEB.

13. C A primitive neural network

By his own later account he coded a primitive neural network, simply wanting the computer to play games against itself — a straight line to AlphaZero thirty years on.

14. A A first sense that intelligence could be built from layered patterns rather than being mystical

GEB gave him his first articulate sense that intelligence might be built out of layered patterns rather than something mystical. Backprop came separately, via his 1991 brush with PDP.

15. B A structured programming course run by his school

The chapter stresses he was self-taught from manuals — BASIC, machine code, and his own game-playing programs. No formal course is mentioned.

16. D Theme Park

17. B Co-designer and lead programmer

He was co-designer and lead programmer on the management sim, which went on to sell millions — not a junior hire fetching bug reports.

18. C Peter Molyneux

Peter Molyneux, already a celebrity inside game design, hired him and modelled the importance of audacious goals. The others are famous designers of the era but had no role here.

19. A 17

Molyneux hired the seventeen-year-old during his gap year after A-levels, before Cambridge — so 19 and 21 are too late in his timeline.

20. D The standard version treats games as a colourful prelude; Mallaby argues games were the real career

Mallaby punctures the prelude framing: in his reading the neuroscience PhD and DeepMind were what Hassabis did to build better games — first virtual, then games played against nature.

Lesson 3: Cambridge and Elixir

1. C David Silver

Hassabis met David Silver in his first weeks at Queens'. Legg and Suleyman were later DeepMind co-founders, not Cambridge freshers; Molyneux was his boss at Bullfrog, before university.

2. B Go

They bonded over board games — Go in particular. Go is the specific game named for the Cambridge friendship, and the one Silver would later attack with AlphaGo.

3. A Gödel, Escher, Bach

Douglas Hofstadter's Gödel, Escher, Bach is the book Mallaby singles out. Minsky's The Society of Mind is a tempting AI-adjacent choice but is not the one named.

4. D He was a Bullfrog colleague Hassabis recruited to Cambridge

Silver was a fellow fresher Hassabis met at Queens', not a Bullfrog colleague he brought with him. The other three statements are all accurate.

5. B The deep philosophical questions

Hassabis's phrase was "the deep philosophical questions" — what is intelligence, and what would it mean to construct it. "Grand-challenge problems" is DeepMind-era vocabulary, not the Cambridge phrase.

6. A A generative model that constructs memories

Finding that the hippocampus both stores memories and constructs hypothetical futures convinced Hassabis the brain is a generative model — it builds scenes rather than retrieving them — a conviction that quietly structured DeepMind's research.

7. C He carried it out in the years before he founded Elixir Studios

The PhD came AFTER Elixir wound down in 2005 — he enrolled almost immediately afterwards. The other three statements are accurate.

8. D The hippocampus

The hippocampus — already known for storing memories — turned out to also construct hypothetical futures. The amygdala handles emotional processing and is not the region in the finding.

9. C It is also responsible for constructing hypothetical futures

The point of the finding is that one region does both jobs — storing memories and constructing imagined futures. The idea that memory and imagination live in separate regions is exactly what the result overturned.

10. B They produce only fragmented, disjointed images and cannot bind them into a coherent scene

Hassabis and Maguire's 2007 study found these patients' imagined scenes lacked spatial coherence: fragmented images they could not bind into a whole. The option that they can imagine it but not recall a real beach inverts this — the deficit is in building a new scene.

11. D 7 July 1998

Elixir Studios was founded on 7 July 1998. 1994 is when Hassabis went up to Cambridge, 2003 is Republic's release, and 2005 is when Elixir wound down.

12. A 22

He was 22 when he founded Elixir on 7 July 1998, a year after graduating Cambridge in 1997 — which also rules out the older ages.

13. B The games company Hassabis founded after leaving Bullfrog, building ambitious AI-driven games

Elixir Studios was Hassabis's own games company, founded post-Bullfrog to build ambitious AI-driven games; it struggled commercially and closed in 2005. It was never a lab, a research group, or a publisher.

14. A Central London, on Bullfrog's premise of ambitious AI baked into games

Elixir was founded in central London on the premise Bullfrog had taught him: ambitious AI baked into games at a level no other studio would attempt — the opposite of a low-cost casual strategy.

15. C It was a commercial success that Hassabis sold at a profit

Elixir failed commercially; by 2005 it was wound down and Hassabis merely sold the rights to its completed games and walked away. The other three statements are accurate.

16. D Republic: The Revolution

Republic: The Revolution was a 2003 release with an enormously ambitious AI back-end that overshot the era's hardware. Critics admired the ambition but found play sluggish.

17. D Evil Genius

Evil Genius (2004) was Elixir's satirical spy-villain strategy game. Theme Park and Dungeon Keeper are Bullfrog titles, not Elixir's.

18. C A fictional Eastern European revolution

Republic modelled thousands of citizens with individual goals, opinions, and relationships against a fictional Eastern European revolution. The espionage setting belongs to Evil Genius's genre, not Republic's.

19. A Both were prized for AI ambition, but neither was a commercial smash

Republic and Evil Genius were both prized for AI ambition and neither was a commercial smash — Evil Genius performed better but still could not rescue the studio's finances. Both did ship.

20. B Early-2000s hardware could not support what Hassabis asked of it

The concept overshot the era's compute — hardware timing, one of Mallaby's three Elixir lessons. The AI did exist and modelled thousands of citizens; it was simply too heavy for 2003 machines.

Lesson 4: Gang of Three — Legg and Suleyman

1. B Brainstorming a title for a planned book of essays with Marcus Hutter

While Legg and Marcus Hutter were brainstorming a title for a planned essay collection around 2002, Legg proposed 'Artificial General Intelligence.' The phrase stuck and became the field's name for its most ambitious goal.

2. C Shane Legg

Legg proposed "Artificial General Intelligence" while brainstorming a book title with Hutter. Hutter was in the room but did not propose the phrase; Schmidhuber worked with Legg on recurrent nets, not the term.

3. A 2002

The coinage dates to around 2002, during the book-title brainstorm with Hutter. 2009 is the Halloween Scenario lecture; 2010 is the poker night and the Thiel pitch.

4. D Jürgen Schmidhuber

Legg worked with Schmidhuber on recurrent neural networks and with Hutter on theories of universal intelligence — Hutter is the tempting swap. Hinton and Silver are not part of Legg's IDSIA years.

5. D It was the agreed name for the field's most ambitious goal

The phrase stuck: by the time Hassabis met Legg, "Artificial General Intelligence" was the agreed name for the field's most ambitious goal. Universal intelligence was Legg and Hutter's IDSIA research topic, not a rival name.

6. B IDSIA

IDSIA is the Swiss AI institute where Legg worked with Hutter and Schmidhuber. EPFL, ETH Zurich and Idiap are real Swiss research institutions, but none is the lab in question.

7. A Lugano

IDSIA, the Swiss AI institute, is in Lugano. Zurich, Geneva and Lausanne are other Swiss research cities but not IDSIA's home.

8. C Demis Hassabis

Shane Legg, Jürgen Schmidhuber, and Marcus Hutter were all researchers at IDSIA in Lugano — Schmidhuber as director, Legg and Hutter as students. Demis Hassabis, co-founder of DeepMind alongside Legg, came up through a different path (games and UCL neuroscience), with no IDSIA connection.

9. A Theories of universal intelligence

At IDSIA, Legg worked with Hutter on theories of universal intelligence; the recurrent-neural-network work was with Schmidhuber. Atari and protein folding are later DeepMind work, and climate diplomacy was Suleyman's consulting.

10. B Cognitive science, finance and machine learning

Legg drifted through cognitive science, finance and machine learning before IDSIA. Games and neuroscience describe Hassabis; conflict resolution and climate diplomacy describe Suleyman.

11. C Mustafa Suleyman

12. D Moose

Mallaby records the nickname as "Moose" — Mustafa "Moose" Suleyman. The other options are plausible shortenings of his name but are not the nickname used.

13. D 1984

Suleyman was born in August 1984 in north London. 1973 is Shane Legg's birth year — the tempting mix-up between the two cofounders.

14. C He held a doctorate in machine learning

Suleyman had no technical training in AI at all — he was the operator, not the researcher. The Oxford dropout, the youth helpline and the Copenhagen consulting work are all accurate.

15. A The relentless operator who could pitch a billionaire at midnight

Suleyman is the operator of the triangle. The cerebral safety theorist is Legg; the games-and-neuroscience polymath is Hassabis; the Lugano recurrent-net specialist is Schmidhuber, who was never a cofounder.

16. B George Hassabis

Demis's brother George had been at school with Suleyman and made the introduction. Nosek and Thiel enter the story later, at the 2010 fundraise; Legg met Hassabis separately at UCL.

17. A Demis's brother, who had been at school with Suleyman

George was Demis's brother and had been at school with Suleyman — that school tie is the hinge. The Vic Casino was the July 2010 poker night among the three cofounders, not where George met Suleyman.

18. C Luke Nosek

Thiel made the remark to his Founders Fund partner Luke Nosek as they walked the three founders back to their car. The cofounders were the subject of the remark, not its audience.

19. D Peter Thiel's partner at Founders Fund

Nosek was Thiel's Founders Fund partner, present as the founders were walked back to their car. Founders Fund committed at the Palo Alto breakfast — but Nosek's role was investor-side, not a DeepMind hire.

20. B Walking the founders back to their car after breakfast at Thiel's Palo Alto house

Thiel muttered it to Luke Nosek while walking the trio back to their car after the Palo Alto breakfast. The Summit was where Suleyman engineered the introduction the previous day; the Vic poker night was a month earlier.

Lesson 5: Founding DeepMind

1. C Solaria

The first working name was Solaria, borrowed from an Isaac Asimov novel. Trantor and Foundation also come from Asimov's fiction, which makes them tempting — but Solaria is the name the founders actually used.

2. D Isaac Asimov

Solaria comes from an Isaac Asimov novel. Clarke, Dick and Gibson are all plausible science-fiction sources, but Asimov is the actual source.

3. A 2010

The name was registered in late 2010, the same year the London company was founded — there was no separate 2011 filing. 2012 is when Page and Brin began watching the lab, and 2014 is the Google acquisition.

4. B Shane Legg

Shane Legg argued for something simpler than Solaria. Hassabis and Suleyman are cofounders too, but the credit goes to Legg; Mnih was a researcher who later led the Atari DQN work.

5. C Deep learning, the neural-network approach they were betting on

"Deep" nodded to deep learning; "Mind" declared the destination. DQN is tempting but came years later — the name was registered in 2010, before they had any product.

6. A January 2014

Google closed the deal in January 2014. October 2012 is the Musk–Page dinner and July 2013 is the day the DQN agent learned Pong — both real dates, but neither is the acquisition.

7. D £400 million

The price was roughly £400 million (about $600 million) — then the largest AI acquisition in European history. The other figures are off by an order of magnitude or more.

8. B The largest AI acquisition in European history

At the time, the purchase was the largest AI acquisition in European history. It is not claimed to be any of the other records: not Google's biggest deal ever, not the largest UK sale to a US buyer, not the first AI lab bought by a technology company.

9. D Larry Page and Sergey Brin

Google's cofounders Larry Page and Sergey Brin were the ones circling DeepMind from 2012. Schmidt and Pichai were Google executives but are not credited with that.

10. B DeepMind was already profitable on strong revenue at the time of the deal

DeepMind had no revenue at all when Google bought it — that was what made the price striking. The other three statements are all accurate.

11. A Peter Thiel's Founders Fund

Peter Thiel's Founders Fund led DeepMind's earliest outside funding — a notable irony given Thiel's later public skepticism about unchecked AI. Google acquired the lab in January 2014 for roughly £400 million.

12. C OpenAI

Mallaby frames that night as the moment the informal Google–Musk alignment cracked, seeding what became OpenAI. The ethics board came out of the Google deal instead, and Anthropic and xAI came years later.

13. A "A digital god"

Musk warned that advanced AI could become "a digital god" and that humanity had no plan for what follows. The other phrases are plausible AI-doom rhetoric but not the one he used in that clash.

14. B At a dinner at Larry Page's home in Palo Alto

The exchange happened at a dinner at Larry Page's Palo Alto home on 8 October 2012 — the informal setting is part of why Mallaby treats the night as a turning point.

15. C The species that built such an intelligence would be its natural heirs

Page argued that a machine intelligence built by humanity would be its rightful heir — a succession, not a threat. Mallaby frames this clash as the seed of what became OpenAI.

16. D It stored past experiences in a buffer and trained on random samples, breaking the correlation between consecutive frames

Networks trained on sequential frames overfit to recent play and forget earlier lessons. Storing experiences and sampling random batches decorrelated the data — closer to independent, identically distributed examples — making deep reinforcement learning reliable.

17. A Vlad Mnih

Vlad Mnih, trained in reinforcement learning at Toronto and Alberta, led the Atari strike team. Legg, Hassabis and Suleyman are cofounders of the lab, not the leads on this project.

18. C The long-term expected reward of each action from each state

Q-learning estimates the long-term expected reward (the "quality") of every action from every state, refining those estimates through play. Matching a human demonstrator describes imitation learning instead.

19. D Q-learning with a deep convolutional network that reads raw pixels

DQN pairs Q-learning with a deep convolutional network reading raw pixels — and adds experience replay for stability. Game-specific hand-coded rules are exactly what the Atari challenge forbade.

20. B One network, no game-specific code, learning from raw pixels and the score alone

The constraint — one network, raw pixels, score as the only feedback, no game-specific code — forced a general solution rather than a tailored one, which was the whole point.

21. C The London AI lab Demis Hassabis cofounded in 2010 and Google acquired in 2014

DeepMind is the London lab Hassabis cofounded in 2010 and Google bought in 2014. The Musk nonprofit describes OpenAI — a different organisation that traces back to a 2012 dinner.

22. A London

DeepMind is a London company — the acquisition was the largest AI deal in European history. Palo Alto hosted the Musk–Page dinner, and Toronto is where Vlad Mnih trained.

23. D The London AI lab Hassabis cofounded in 2010, later bought by Google

DeepMind is the London AI lab Hassabis cofounded in 2010 and that Google acquired in January 2014. The nonprofit description matches OpenAI instead, and the ethics board was a separate condition of the Google deal, not the company itself.

24. B DeepMind

AlphaGo and AlphaFold are DeepMind milestones, and the London base plus Google ownership pins it down. Google Brain was a separate, US-based Google team; OpenAI and Meta AI are neither London-based nor Google-owned.

25. B Founded 2010, acquired 2014

Founded in 2010 and acquired by Google in January 2014. 2011 and 2012 are both distractor years — 2012 is when Page and Brin started watching the lab — and match neither the founding nor the acquisition.

Lesson 6: Atari to AlphaGo

1. B An agent learns by trial and error, receiving rewards for good actions and penalties for bad ones

Reinforcement learning is trial-and-error training driven by rewards and penalties — that is how DQN learned Atari from raw pixels. Learning from labelled examples describes supervised learning, not RL.

2. A 2015

The DQN paper appeared in Nature in February 2015. 2013 is tempting — DeepMind showed early Atari results at NeurIPS that December — but the Nature publication came in 2015.

3. C 49 games, human-level or better on 29

One architecture was retrained from scratch on 49 games and hit human-level or better on 29 of them. The 29/49 figures are often flipped — the smaller number is the successes.

4. D Hand-engineered, game-specific features tuned for each title

The whole point was the absence of hand-engineered per-game features — earlier game-playing AI was hand-tuned per title. DQN used only raw pixels and score, with identical code and hyperparameters throughout.

5. B NeurIPS, December 2013

DeepMind showed early Atari results at NeurIPS in December 2013, months ahead of the acquisition. 2012 was AlexNet's ImageNet win, and January 2016 was the AlphaGo Nature cover.

6. D Nature

7. A The DeepMind program that defeated world champion Lee Sedol at Go in 2016

AlphaGo is DeepMind's Go program that beat Lee Sedol in 2016. The Atari-from-pixels system was DQN, a separate DeepMind project.

8. C It made artificial intelligence a front-page story worldwide

The Seoul match's global significance was cultural, not just technical: broadcast live to tens of millions across Asia, it compressed years of public AI awareness into one week and made AI a front-page story worldwide.

9. D Seoul, March 2016, 4–1

AlphaGo beat Lee Sedol 4–1 in Seoul in March 2016. The 5–0 sweep was the earlier closed-door match against Fan Hui in October 2015.

10. C It learned Go end to end from raw screen pixels alone

Learning from raw pixels with no search describes DQN on Atari, not AlphaGo — AlphaGo's networks worked together with Monte Carlo tree search, and the January 2016 Nature cover revealed that architecture.

11. B Nature

The Fan Hui result was embargoed until the January 2016 Nature cover paper that announced AlphaGo.

12. A January 2016

AlphaGo's Fan Hui result was embargoed until Nature's January 2016 cover paper, which also revealed AlphaGo's architecture.

13. C Aja Huang

Aja Huang built AlphaGo's first prototype; his tree-search work made the system competitive at professional level. Fan Hui joined DeepMind later, as an analyst rather than an engineer.

14. A Sergey Brin

Sergey Brin posed the question, heard "Go," doubted it was winnable, and provided the resources anyway. Musk and Altman were on the other side of this story — they founded OpenAI as a counterweight.

15. D European Go champion

Fan Hui was the European Go champion and lost 5–0 in the closed-door match. Lee Sedol, not Fan Hui, was the world-class player widely called the greatest of his generation.

16. B Game four — he found a move AlphaGo had not adequately evaluated

Lee Sedol won game four with a move AlphaGo had not foreseen — his lone victory in a 4–1 defeat. Game two belonged to AlphaGo; that was the game containing Move 37, which worked exactly as intended.

17. D Lee Sedol — European champion who lost 5–0 in a closed-door match

The 5–0 closed-door loss was Fan Hui's, in October 2015. Lee Sedol was the Korean champion who lost 4–1 in Seoul in March 2016 — a public match, not a closed-door one.

18. C Once in 10,000 times

Move 37 — a shoulder-hit on the fifth line — was called a mistake by human commentators, but AlphaGo's policy network put the odds a human would choose it at roughly one in ten thousand. It won territory human intuition would not reach.

19. B Game two, a shoulder-hit on the fifth line

Move 37 was a shoulder-hit on the fifth line in game two. Game four was the one Lee Sedol won, with a move AlphaGo had not adequately evaluated.

20. A As a mistake

Commentators were stopped cold and rated it a mistake. AlphaGo's reasoning turned out to be correct — the move won territory in a way human intuition would not reach.

21. D He left the board for about twelve minutes and returned visibly shaken

Lee Sedol spent roughly twelve minutes away from the board and came back visibly shaken. He did win game four later, but game two was AlphaGo's.

22. B That trained networks can discover strategy outside the corpus of human play

Move 37 is shorthand for networks finding strategic concepts beyond human play. Reproducing expert judgement is the opposite of the point — no human grandmaster would have chosen it.

23. A Sam Altman

Sam Altman became CEO. Musk cofounded and announced OpenAI alongside him, and Sutskever took the chief scientist role rather than the top job.

24. C Ilya Sutskever

Ilya Sutskever became OpenAI's first chief scientist. Hinton was his doctoral adviser in Toronto but did not take the role; Abbeel was among the hires who followed later.

25. B December 2015, $1 billion in commitments

Musk and Altman announced OpenAI in December 2015 with $1 billion in funding commitments — timing that coincided almost exactly with AlphaGo's closed-door win over Fan Hui.

26. A To act as a counterweight to Google's ownership of DeepMind

OpenAI was framed explicitly as a counterweight to Google's ownership of DeepMind — Musk saw frontier capability concentrated in one commercial giant as dangerous.

27. D Ilya Sutskever taking the CEO role while Sam Altman led research

The roles are reversed: Sam Altman was CEO and Ilya Sutskever was chief scientist. The other three describe the founding accurately.

28. C As a counterweight to Google's DeepMind

Mallaby frames OpenAI's founding as a non-profit counterweight to Google's ownership of DeepMind.

29. A Elon Musk

Musk decided a single commercial giant controlling frontier AI was dangerous, with DeepMind under Google his primary concern.

Lesson 7: The OpenAI rivalry

1. D Withdrew his promised funding and resigned from the board

Musk pushed to run OpenAI himself; when he couldn't win operational control in early 2018, he withdrew his promised funding and resigned from the board — the rupture that quietly reshaped the industry's power map.

2. B Greg Brockman and Ilya Sutskever

Brockman and Sutskever — both OpenAI founders — wrote the memo. Amodei and Christiano were also at OpenAI, but their break came later, in the 2021 walkout that founded Anthropic.

3. C Demis Hassabis

Hassabis was the person the memo named — Sutskever wrote of Musk's fear that "Demis could create an AGI dictatorship." Altman ran OpenAI but was a recipient of the warning, not its named target.

4. A To run OpenAI himself

Musk pushed to run OpenAI himself and left when he failed to gain operational control. The capped-profit conversion came in 2019, after he had already resigned.

5. B It showed founders of an anti-monopoly lab feared the very concentration they were enacting

The email's significance is the irony: OpenAI was founded to prevent an AI monopoly, yet its own founders feared the logic of concentration inside their lab. Musk's withdrawal came later, in early 2018.

6. A 100 times

The cap was set at 100× an investor's stake — profits would be real but bounded. The figure later became a flashpoint the board cited against Altman.

7. C 2019

The restructuring happened in 2019. 2017 was the Brockman-Sutskever email, 2021 the Anthropic split, and 2023 the Altman firing.

8. D A nonprofit parent controlling a for-profit subsidiary

The nonprofit parent retained formal control over a for-profit subsidiary — the very structure that left a small nonprofit board able to fire Altman in 2023.

9. C A $1 billion commitment from Microsoft

The capped-profit structure let Altman raise $1 billion from Microsoft — the partnership that would make Microsoft the dominant AI-infrastructure player.

10. A The cap was presented as a way to end Microsoft's influence over OpenAI

The opposite is true: the structure is what enabled Microsoft's $1 billion commitment and its growing influence. The other three all describe the pivot accurately.

11. D Google Brain and DeepMind

Google Brain merged with DeepMind into a single unit. Google Brain was the separate faction whose decade-long rivalry with DeepMind the merger ended.

12. B Demis Hassabis

Hassabis led the merged unit, handing him direct control over Google's entire response to OpenAI. Pichai ordered the merger but ran Alphabet; Legg stayed DeepMind's chief AGI scientist, not its CEO.

13. C ChatGPT's launch in late 2022 shocked Alphabet

ChatGPT's late-2022 launch sent shockwaves through Alphabet, and the merger was Pichai's April 2023 response. The Microsoft commitment dated to 2019 and did not trigger the merger.

14. A The transformer

Google Brain invented the transformer in 2017. Reinforcement learning was DeepMind's own tradition — the merger finally put both sets of researchers in the same room.

15. D A decade of internal rivalry between two Google AI factions

The merger closed a decade-long rivalry between Google Brain and DeepMind. DeepMind was already Alphabet-owned, and the transformer remained central rather than abandoned.

16. B Mustafa Suleyman

17. B DeepMind's chief AGI scientist

Legg remained DeepMind's chief AGI scientist while Hassabis ran Google DeepMind and Suleyman ran Microsoft AI — all three co-founders atop major AI operations.

18. D 5

The board folded within 5 days. The 48-hour mark is when 700+ staff signed the resignation letter, and 36 hours is when Nadella made his public offer.

19. C Satya Nadella

Nadella made the offer — a calculated use of Microsoft's leverage as OpenAI's cloud provider and largest backer. Suleyman ran Microsoft AI but was not the one who extended the lifeline.

20. A More than 700

More than 700 of 770 signed within 48 hours — near-total, but not unanimous, which made the board's position untenable.

Lesson 8: AlphaFold — Fermat for biology

1. C Twists itself into a specific three-dimensional shape that determines the protein's function

Protein folding is a chain of amino acids twisting into a specific 3D shape, and that shape determines the protein's function. Crystallography is how the shape is measured experimentally, not what folding is.

2. B That a protein's sequence alone determines its folded shape

Anfinsen's conjecture: the information needed for the folded shape is all in the amino-acid sequence — no additional cellular machinery required. That is the opposite of saying extra cellular machinery must encode the shape.

3. D In his 1972 Nobel Prize lecture

Anfinsen stated the conjecture in his 1972 Nobel Prize lecture. CASP began in 1994 under John Moult — a different milestone entirely.

4. A Biologists rejected the principle and worked to disprove it for fifty years

Biologists accepted Anfinsen's principle for half a century — they simply could not extract the shape from the sequence in practice. They did not reject it.

5. B About fifty years

Anfinsen conjectured in 1972 and AlphaFold delivered around 2020 — roughly fifty years, matching the "fifty-year grand challenge" framing used throughout the account.

6. A A protein's three-dimensional structure, from its amino-acid sequence

AlphaFold takes an amino-acid sequence and predicts the 3D structure. The reverse direction — sequence from an image — is not what it does.

7. D John Jumper

John Jumper led AlphaFold 2. Andrew Senior was the project's first lead and the person who recruited Jumper; Moult ran CASP; Silver led AlphaGo work.

8. C A predicted three-dimensional structure for that chain

AlphaFold's job is sequence in, predicted 3D structure out. It does not produce experimental images, work backwards to genes, or screen drugs — those are downstream or different tools.

9. B The 2024 Nobel Prize in Chemistry

Jumper and Hassabis shared the 2024 Nobel Prize in Chemistry for AlphaFold. It was Chemistry, not Physiology or Medicine.

10. D Jumper was recruited in October 2017 by John Moult, who then led AlphaFold 2

Andrew Senior — not John Moult — recruited Jumper in October 2017. Moult was the professor who ran the independent CASP contest.

11. C The continuous distance between them, producing a "distogram"

Rivals predicted a binary contact map (in contact or not). AlphaFold instead predicted the distance between amino-acid pairs on a continuous scale — a distogram — described as going from black-and-white to full-colour TV. It was AlphaFold 1's defining architectural choice.

12. A The distance between them on a continuous scale

A distogram records the continuous distance between amino-acid pairs. The binary touch/no-touch flag is the older contact map that rival labs predicted.

13. B A binary label for contact between them

Rivals predicted a binary contact map. Predicting continuous distance was AlphaFold's departure; predicting exact 3D coordinates directly came later, with AlphaFold 2's pivot.

14. D Going from black-and-white to full-colour television

Researcher Marek Barwinski described it as "like going from black and white to a full-colour TV" — the distogram carried far richer geometric information than a binary contact map.

15. A It was produced by a transformer network built in 2019

AlphaFold 1's distogram came from a convolutional network. The transformer (the "tetraformer") arrived in 2019 with AlphaFold 2 — a later, separate rebuild.

16. C How closely a prediction's atom positions match the experimentally verified structure

GDT measures the percentage of a predicted protein's main-chain atoms lying within a tolerated distance of the ground-truth structure — an accuracy score, not a speed or size measure.

17. B Critical Assessment of Structure Prediction

CASP is the Critical Assessment of Structure Prediction — the contest where protein-folding methods compete on structures known only to the experimentalists.

18. D Every two years since 1994

Professor John Moult has run CASP every two years since 1994 — it is biennial. 1972 is the year of Anfinsen's Nobel lecture, a different milestone.

19. C About 90

Roughly 90 GDT is the threshold where a prediction is as accurate as experiment. 60 and 20 were mid-pivot AlphaFold 2 scores, not the experimental-quality bar.

20. A Competing labs are given the true shapes in advance so they can calibrate

Labs receive only the sequences; the true shapes are withheld and known only to the experimentalists. That blinding is what made CASP an external, objective scoreboard.

21. C It crashed from about 60 to about 20 before clawing back

The direct-folding pivot initially tanked the GDT score from ~60 to ~20. The team pressed on; by late 2019 it had recovered to 60 and kept climbing — a textbook case of accepting short-term regression to unlock a better approach.

22. D 92

AlphaFold 2 scored roughly 92 GDT (92.4) at CASP14 in November 2020. 60 and 20 were scores during the direct-folding pivot; 43 was the count of hard cases at CASP13.

23. B More than 50 percent higher than any previous best

AlphaFold 2's 92 GDT was more than 50 percent above any previous best in CASP's history — accurate enough that experimentalists' gold-standard test could not detect a meaningful difference.

24. A 200 million

By July 2022 the database held roughly 200 million entries. 20,000 is the human proteome it covered within months; 3 million is the number of researchers who had used it by 2025.

25. A Moult tallied the winning CASP scores in November 2018

Moult tallied CASP14 in November 2020, not 2018. December 2018 was CASP13 in Cancún — the earlier AlphaFold 1 win, a different contest.

Lesson 9: RaceGPT and Bletchley

1. A Within two months

ChatGPT hit 1 million users in five days and 100 million within two months — the fastest consumer-technology adoption on record, which is what pulled investors and governments off the sidelines and into the AI race.

2. C The moment the AI race became public

Mallaby treats the launch as the moment the AI race went public: investors demanded weekly briefings and politicians began scheduling hearings. The Transformer had been invented five years earlier, in 2017.

3. D 2017

The Transformer paper was published in 2017, five years before ChatGPT's 2022 launch put the architecture in front of the general public.

4. B Five days

A million users arrived within five days; 100 million took two months. The five-day figure is the first-million milestone, not the 100-million one.

5. C GPT-3.5

ChatGPT was a research preview wrapping GPT-3.5. GPT-2 was an earlier OpenAI model, GPT-4 arrived after the launch, and Gemini was Google's — none underpinned the November 2022 release.

6. B Ensuring an AI reliably pursues human-intended goals as it grows more capable

Alignment is the technical and conceptual problem of an AI reliably pursuing human-intended goals even as capability grows. The other options describe engineering or business coordination, not goal fidelity.

7. A Geoffrey Irving

Geoffrey Irving made that move. Amodei and Christiano were his OpenAI safety colleagues who went on to co-found Anthropic; Hassabis co-founded DeepMind rather than joining it from OpenAI.

8. D Yann LeCun

Yann LeCun declined, consistent with his long-held scepticism about near-term catastrophic risk. Bengio, Hinton and Altman all signed, alongside Hassabis and Amodei.

9. B Pandemics and nuclear war

The single sentence named pandemics and nuclear war as the comparison class. The other pairings are plausible societal risks but were not the ones the letter invoked.

10. A Hinton, Bengio, Hassabis, Altman and Amodei

Hinton, Bengio, Hassabis, Altman and Amodei are among the signatories. Every other option includes Yann LeCun, who pointedly declined to sign.

11. C Ben Buchanan

Ben Buchanan, a Georgetown professor of AI and cybersecurity working at the National Security Council, took the AI czar role. Suleyman ran Inflection and Amodei Anthropic — labs that signed the commitments.

12. D A Georgetown professor of AI and cybersecurity on the National Security Council

Buchanan came from Georgetown, specialising in AI and cybersecurity, via the National Security Council. The DeepMind alignment description fits Geoffrey Irving instead.

13. D At the White House

The 21 July 2023 commitments were signed at the White House. Bletchley Park hosted the separate international summit that November; Georgetown was Buchanan's academic home, not the venue.

14. C Binding financial penalties for backsliding

The pledges carried no enforcement mechanism — Buchanan's leverage was reputational. Safety testing, cybersecurity investment and capability reporting were the three actual commitments.

15. B No major US frontier lab declined to sign

Google, OpenAI, Anthropic, Meta, Amazon, Microsoft and Inflection all signed — not a single major US frontier lab declined, which is what made the pledges a coordinated first.

16. A United Kingdom and United States

17. C 28

Delegates from 28 countries gathered at Bletchley Park on 1 November 2023, and all 28 delegations signed the Declaration.

18. A November 2023

The summit opened 1 November 2023. October 2023 was Biden's executive order and July 2023 the voluntary commitments — the summit came after both.

19. D It was Alan Turing's wartime codebreaking home

Bletchley Park was Alan Turing's wartime codebreaking workplace, which gave the summit its symbolism. DeepMind was founded in London but not at Bletchley.

20. B Demis Hassabis

Hassabis pitched Sunak on a global summit roughly six months before Bletchley. Buchanan was driving the parallel US executive-order track, not the UK summit.

21. D 2022

The controls came in October 2022. 2023 was the year of the voluntary commitments, the executive order and the Bletchley summit; 2017 was the Ke Jie match.

22. C The supply of advanced semiconductors to China

The controls were a wide-ranging ban on supplying advanced semiconductors to China. They targeted hardware, not capital flows, model weights or researcher access.

23. A Ke Jie

AlphaGo's defeat of Ke Jie, played on a Chinese board, prompted the Sputnik framing in Beijing. Lee Sedol's and Fan Hui's earlier losses, and Kasparov's chess defeat, were not the trigger described here.

24. B 2017

The match was in 2017. By 2022 China's State Council already had a formal AI strategy with 2030 targets — the strategy followed the Sputnik moment, it did not precede it.

25. C It closed the most consequential lane of US-China technology cooperation

Mallaby reads the ban as closing the most consequential lane of US-China cooperation, merging the AI race and the chip war into one story. Bletchley traced to Hassabis and Sunak, not to the ban.

Lesson 10: Step by Step — Gemini and the comeback

1. C Google DeepMind's flagship family of multimodal AI models, launched in 2023

Gemini is Google DeepMind's flagship multimodal model family, launched in 2023 as the lab's answer to OpenAI's GPT-4 and its return to the research frontier. It is a Google DeepMind product, not an OpenAI one.

2. A 6 December 2023

Gemini launched on 6 December 2023, roughly fourteen months after ChatGPT (30 November 2022). 15 February 2024 is the later Gemini 1.5 Pro release.

3. D Gemini Ultra

Two smaller variants shipped immediately; the headline Gemini Ultra was only previewed, benchmarked against GPT-4 on MMLU. Gemini 1.5 Pro was a separate release in February 2024.

4. B Every variant, including the headline model, shipped to users immediately

Only two smaller variants shipped immediately — Gemini Ultra was previewed, not released. The other three statements match Mallaby's account of the launch.

5. C It was the first proof the Brain–DeepMind merger could ship at frontier scale

Gemini was the first test of whether the merger's synergy story was real — late by Silicon Valley standards, but proof the merged organisation could ship at frontier scale. London only set the research agenda later, with Gemini 1.5 Pro.

6. A Massive Multitask Language Understanding

MMLU is Massive Multitask Language Understanding — the benchmark on which Gemini Ultra was previewed against GPT-4.

7. B 57

MMLU spans 57 subjects, covering maths, humanities, ethics and law — its breadth is the point, testing general knowledge rather than one narrow skill.

8. D 90%

Gemini Ultra's announced MMLU score was 90%, edging past GPT-4's published 86%. 44% was GPT-3's 2020 score.

9. C Computer vision

MMLU spans maths, humanities, ethics and law. Computer vision is not part of the benchmark — MMLU tests language-based multitask knowledge.

10. A 44% → 86% → 90%

GPT-3 scored 44% in 2020, GPT-4 reached 86% in 2023, and Gemini Ultra hit 90% — the first system announced above top human experts. Each bump is roughly a year of student-level progress.

11. D Language quality offers no clear reward signal for RL to optimize

RL needs a clear reward signal, but language quality resists scoring — as Mallaby puts it, RL couldn't tell a good haiku from a great one. Silver quit after ~6 months; a Bard team using data cleaning and RLHF shipped the upgrade instead.

12. B A million-token context window

Gemini 1.5 Pro's headline capability was a million-token context window, paired with a sparse mixture-of-experts architecture. The 90% MMLU score belonged to Gemini Ultra at the earlier December 2023 launch.

13. B 15 February 2024

Gemini 1.5 Pro was released on 15 February 2024. 6 December 2023 was the original Gemini launch.

14. D Only a fraction of the model's parameters activate for any given input

In a sparse mixture-of-experts design only a fraction of parameters activate per input, which keeps inference costs manageable despite the model's size. It is not an ensemble vote or a human-review step.

15. C A Bard veterans team focused on data cleaning and RLHF widgets

Silver quit after ~6 months because language quality gives RL no clear reward signal. A Bard veterans team concentrating on data cleaning and RLHF widgets shipped the upgrade instead — not sophisticated machine RL.

16. A 2024

Google DeepMind released Gemini 1.5 Pro on 15 February 2024, about two months after the original Gemini launch.

17. B 15 February 2024

Gemini 1.5 Pro shipped on 15 February 2024, built by a London-led strike team of the kind that produced AlphaFold.

18. C Undisclosed Gulf fundraising, and a pattern of pushing out critical board members

Mallaby gives exactly two triggers: fundraising overtures to Gulf investors that Altman had not disclosed to the board, and a pattern of attempting to push out independent board members who criticised him.

19. D The Persian Gulf

Mallaby cites undisclosed fundraising overtures to investors in the Persian Gulf — conversations Altman had not disclosed to the board.

20. A Neither trigger concerned model testing; both were about undisclosed conduct and board control

Mallaby's two triggers were the undisclosed Gulf fundraising and the pattern of pushing out critical board members — questions of trust and governance, not of any safety evaluation result.

21. C "Safe AI for the benefit of the world"

The board's charter charged it with pursuing "safe AI for the benefit of the world"; measured against that mission, it concluded Altman could not be trusted.

22. D How thoroughly Silicon Valley and OpenAI's workforce would close ranks behind him

The board underestimated how completely Silicon Valley and OpenAI's own workforce would rally behind Altman within days — the reaction that undid the firing.

23. A Realist

Mallaby closes with Hassabis rejecting both "doomer" and "accelerationist," preferring "realist" — fitting for a man who is at once a Nobel laureate warning about AI and a CEO racing to build it.

24. B A Nobel laureate and chief executive of one of the most powerful AI labs

Mallaby ends with Hassabis as simultaneously a Nobel laureate and the CEO of one of the most powerful AI labs in the world. It was Geoffrey Irving who became chief scientist at the UK's AI Safety Institute.

25. A That the most morally serious figure in the race is also racing hardest to build the technology he warns about

Mallaby's closing paradox is that the most morally serious major figure in the race is also racing as hard as anybody to build the technology he has spent his career warning the world about — and he does not pretend to resolve it.

26. D His decision to step back from the lab to work full-time on safety

Mallaby's closing register holds the two roles together — Nobel laureate and sitting CEO — alongside the "realist" label. Hassabis never steps back from the lab; that tension is the point of the ending.

27. C Mallaby explicitly declines to resolve the paradox he closes on

Mallaby does not pretend to resolve the paradox — he leaves the two Hassabises standing side by side. He offers no verdict, and the "realist" label is Hassabis's own, not Mallaby's judgement.

Lesson 11: Quick Recall — Key people and places

1. A Shane Legg

Shane Legg, the cerebral New Zealander of the three cofounders, coined "AGI" with Marcus Hutter in 2002 and later gave the Halloween Scenario lecture that drew Hassabis in.

2. B Mustafa Suleyman

Suleyman was the operator of the three cofounders and ended up running Microsoft AI by 2024. Hassabis stayed on as DeepMind's CEO; Sutskever was OpenAI's chief scientist, not a DeepMind cofounder.

3. C Demis Hassabis

Hassabis — the chess prodigy, neuroscientist, cofounder and CEO — shared the 2024 Nobel Prize in Chemistry with John Jumper for the AlphaFold work.

4. D 2002

Legg and Hutter coined "AGI" in 2002. 2010 was the Singularity Summit pitch to Thiel and 2014 was the Google acquisition — later milestones, not the coinage.

5. B David Silver

The founding trio was Hassabis, Legg, and Suleyman. Silver was Hassabis's Cambridge friend and the architect of AlphaGo — a long-running technical second, but not a cofounder.

6. A Vlad Mnih

Vlad Mnih led the Atari DQN work, the reinforcement-learning foundation the book says every later DeepMind result stood on. Irving's DeepMind work was the alignment programme, not a game-playing system.

7. C David Silver

David Silver was Hassabis's friend from Cambridge and DeepMind's long-running technical second. Jumper arrived only in October 2017, and Mnih is never described as a Cambridge connection.

8. D John Jumper

John Jumper joined DeepMind in October 2017 and led AlphaFold 2. Irving's DeepMind work was the alignment programme, not a protein-folding system.

9. D Demis Hassabis

Jumper shared the 2024 Nobel Prize in Chemistry with Demis Hassabis for the AlphaFold work. Silver, Legg, and Mnih were central to DeepMind but were not laureates.

10. A Vlad Mnih — AlphaFold 2

AlphaFold 2 was led by John Jumper; Mnih's contribution was the Atari DQN. The other three pairings are exactly as the book gives them.

11. B Peter Thiel

Peter Thiel, through Founders Fund, anchored DeepMind's first round after the 2010 Singularity Summit pitch and was the source of the "Demis's destiny" line. Altman's role in the book is at OpenAI.

12. C Larry Page and Sergey Brin

Google cofounders Larry Page and Sergey Brin acquired DeepMind in 2014. Thiel was the first investor, not the acquirer; Altman and Sutskever were OpenAI's protagonists.

13. C About £400 million

Larry Page and Sergey Brin acquired DeepMind for roughly £400 million in 2014 — Europe's largest AI acquisition at the time, made on the condition of an independent ethics board.

14. B 17 November 2023

Altman was fired on 17 November 2023 and reinstated five days later, around 22 November — the reinstatement date, not the firing date, is the tempting distractor.

15. A Microsoft offered to absorb the company

Sutskever reversed himself once Microsoft offered to absorb OpenAI. Altman's reinstatement followed that reversal rather than causing it, and no Founders Fund or Google threat is part of the account.

16. D £400 million

Larry Page and Sergey Brin brought DeepMind into Google in 2014 for roughly £400 million.

17. C £400 million

The 2014 Google-DeepMind deal, led by Larry Page and Sergey Brin, was worth roughly £400 million.

18. B Geoffrey Irving

Geoffrey Irving built DeepMind's alignment programme. Silver's work was AlphaGo and Legg's was the AGI coinage — neither ran DeepMind's safety research.

19. A Ben Buchanan

Ben Buchanan was the Biden White House's AI czar from June 2023. Irving's safety work was at a lab and then a national institute, never inside the White House.

20. D Twenty-eight

Twenty-eight countries signed the first AI safety declaration in November 2023 — a broad but far-from-universal group, so the UN-wide figure of 195 is wrong.

21. B Bletchley Park

Bletchley Park — Alan Turing's wartime workplace — hosted the November 2023 summit and gave the declaration signed there its name.

22. D Geoffrey Irving — October 2023 executive order

The October 2023 executive order was Buchanan's work, alongside the July 2023 voluntary commitments. Irving's path ran from DeepMind's alignment programme to the UK AI Safety Institute, not into US policy drafting.

Lesson 12: Key Concepts

1. B Mustafa Suleyman

Mustafa Suleyman, the third cofounder and the lab's operator, had a background in conflict resolution and diplomacy before helping start DeepMind; he was instrumental in pitching the early billionaire investor and later ran Microsoft AI.

2. C Demis Hassabis

Demis Hassabis was a chess prodigy who mastered the game by 13, and later received the 2024 Nobel Prize in Chemistry for his work on AlphaFold. Silver's game milestone was Go, achieved by a project he led — not personal play.

3. A Shane Legg

Shane Legg, who co-founded the London lab, proposed the concept of Artificial General Intelligence in 2002 discussions. Hassabis co-founded the lab too, but the term itself traces to Legg.

4. D Vlad Mnih

Vlad Mnih is recognised for pioneering reinforcement-learning work and led the team behind the network that played classic Atari games. David Silver led the Go milestone instead.

5. C Shane Legg — pitched a billionaire investor for the lab's initial funding

Pitching the billionaire investor for initial funding was Mustafa Suleyman, not Shane Legg. Silver (Go), Jumper (structural biology, recruited 2017) and Mnih (reinforcement learning) are all correctly paired.

6. B Reinforcement learning

Reinforcement learning is defined by agents interacting with an environment and choosing actions that maximize cumulative reward — the paradigm crucial to DeepMind's early successes. Deep learning is about many-layered networks, not reward-driven action.

7. D A subset of machine learning using many-layered networks inspired by the brain

Deep learning is a subset of machine learning inspired by the brain's structure, using neural networks with many layers to analyze data. The reward-driven description belongs to reinforcement learning; the cross-task description is AGI.

8. A Nodes, or 'neurons'

A neural network is a computational model inspired by biological brains, consisting of interconnected nodes or 'neurons'. Agents, rewards and environments are the vocabulary of reinforcement learning, not the network's building blocks.

9. C Artificial General Intelligence

AGI (Artificial General Intelligence) is the theoretical form of AI that generalises across a wide range of tasks and aims to replicate human cognitive abilities; DeepMind's mission is associated with pursuing it. DQN is a specific algorithm, not a general capability.

10. D Reinforcement learning and deep learning

Learning from a reward signal is reinforcement learning; the many-layered network reading raw input is deep learning. AGI is a theoretical goal, not a technique that could be applied here.

11. A It learned directly from raw screen pixels

DQN combined Q-learning with deep learning to learn Atari games directly from raw pixel data — no hand-coded rules or human demos — demonstrating the potential of a general AI system.

12. B AlphaGo

AlphaGo, built to play Go, beat world champion Lee Sedol in 2016 — a landmark showing reinforcement learning and deep learning combined. AlphaFold predicts protein structures and DQN played Atari games.

13. B AlphaFold

AlphaFold uses deep learning to predict protein structures, a breakthrough for drug discovery and biology; it earned Hassabis and Jumper a share of the 2024 Nobel Prize in Chemistry. AlphaGo and DQN play games.

14. D Q-learning

The Atari system combined Q-learning with deep learning techniques to learn from raw pixels. Backpropagation and gradient descent are training mechanics, not reinforcement learning algorithms, and tree search was not what it relied on.

15. A AlphaGo — predicting the three-dimensional structures of proteins

AlphaGo played Go; protein structure prediction was AlphaFold's domain. The other three pairings match the systems as described.

16. C An independent ethics board to oversee AI applications

The deal, Europe's largest acquisition at the time at around £400 million, included a condition that an independent ethics board oversee how the AI was applied.

17. B Pong

The 7 July 2013 milestone was the day Pong clicked: the network learned the game from scratch and by that evening was beating the built-in opponent, moving toward human-level play.

18. A San Francisco

The Singularity Summit took place in San Francisco in August 2010, where the trio pitched their AGI-safety lab. London is where the lab itself was later founded — a tempting mix-up.

19. D A single neural network that could learn to play multiple games

The small 2013 team aimed at one neural network that could learn many classic games — the lab's first major public success in reinforcement learning. Hand-coding per game or copying human demos was exactly what they avoided.

20. C The moment a key alliance in AI development began to fracture

The Musk–Page exchange on the lawn is seen as the moment a key AI alliance began to fracture. The ethics board came with Google's January 2014 acquisition, a separate event.

Sources

Image credits

Each picture in this guide is credited below, with its creator, source and license where there is one.

- **Lesson 1, What is the infinity machine?:** FlashFeed original diagram · CC-BY-4.0
- **Lesson 1, Hassabis, the book's subject:** Wikimedia Commons · John Sears · CC BY-SA 4.0 — https://commons.wikimedia.org/wiki/File:Demis_Hassabis,_2024_Nobel_Prize_Laureate_in_Chemistry_7_(cropped).jpg
- **Lesson 3, The deep philosophical questions:** Wikimedia Commons · Mewtow · CC BY-SA 4.0 — https://commons.wikimedia.org/wiki/File:Aires_c%C3%A9r%C3%A9brales_de_la_motricit%C3%A9_volontaire.jpg
- **Lesson 5, Why experience replay mattered:** FlashFeed original diagram · CC-BY-4.0
- **Lesson 5, Atari's one-network rule:** User Balou on en.wikipedia · CC BY 2.0 — https://commons.wikimedia.org/wiki/File:Atari2600jr.jpg
- **Lesson 6, Sergey Brin and the Go problem:** Steve Jurvetson · CC BY 2.0 — https://commons.wikimedia.org/wiki/File:Sergey_Brin_Ted_2010_(cropped).jpg
- **Lesson 6, Fan Hui, 5–0, October 2015:** Goban1 · Public domain — https://en.wikipedia.org/wiki/Go_(game)
- **Lesson 6, OpenAI is born:** Wikimedia Commons · David Shankbone · CC BY 3.0 — https://commons.wikimedia.org/wiki/File:Elon_Musk_2011_Shankbone.JPG
- **Lesson 6, Neural nets plus tree search:** FlashFeed original diagram · CC-BY-4.0
- **Lesson 6, Seoul, March 2016: 4-1:** Wesalius · CC BY-SA 4.0 — https://commons.wikimedia.org/wiki/File:Lee_Sedol_(B)_vs_AlphaGo_(W)_-_Game_4.jpg
- **Lesson 7, OpenAI's capped-profit pivot:** TechCrunch · CC BY 2.0 — https://commons.wikimedia.org/wiki/File:Sam_Altman_CropEdit_James_Tamim.jpg
- **Lesson 7, Google DeepMind, April 2023:** Photographer: Lukasz Kobus (European Commission) · CC BY 4.0 — https://commons.wikimedia.org/wiki/File:Sundar_Pichai_-_2023_(cropped).jpg
- **Lesson 7, Anthropic splits off, 2021:** Wikimedia Commons · TechCrunch · CC BY 2.0 — https://commons.wikimedia.org/wiki/File:Dario_Amodei_at_TechCrunch_Disrupt_2023_01_(cropped).jpg
- **Lesson 7, Suleyman leads Microsoft AI:** Jiaqian AirplaneFan · CC BY 3.0 — https://commons.wikimedia.org/wiki/File:Building_92_of_Microsoft_Redmond_Campus_-_panoramio.jpg
- **Lesson 8, From contact map to distogram:** FlashFeed original diagram · CC-BY-4.0
- **Lesson 8, The direct-folding pivot:** FlashFeed original diagram · CC-BY-4.0
- **Lesson 8, Enter the transformer:** FlashFeed original diagram · CC-BY-4.0
- **Lesson 9, Five years before ChatGPT:** FlashFeed original diagram · CC-BY-4.0
- **Lesson 9, Voluntary commitments, 21 July 2023:** Wikimedia Commons · Matt H. Wade · CC BY-SA 3.0 — https://commons.wikimedia.org/wiki/File:WhiteHouseSouthFacade.JPG
- **Lesson 9, Biden's AI executive order, October 2023:** Wikimedia Commons · Adam Schultz · Public domain — https://commons.wikimedia.org/wiki/File:Joe_Biden_presidential_portrait.jpg
- **Lesson 9, Chip exports and China:** Wikimedia Commons · Ordercrazy · CC0 — https://commons.wikimedia.org/wiki/File:Embedded_World_KNX_TP_Demo_Board.jpg
- **Lesson 10, Gemini, December 2023:** Wikimedia Commons · Google · Public domain — https://commons.wikimedia.org/wiki/File:Google_Gemini_Screenshot_(2026).png

- **Lesson 10, Only a fraction fires:** FlashFeed original diagram · CC-BY-4.0

Useful links

The book

- *The Infinity Machine: Demis Hassabis, DeepMind, and the Quest for Superintelligence* by Sebastian Mallaby — flashfeedlearningapp.com/packs/infinity-machine

FlashFeed

- This pack in the app: flashfeedlearningapp.com/packs/infinity-machine
- All packs: flashfeedlearningapp.com/packs
- How FlashFeed works: flashfeedlearningapp.com/how-it-works
- Free and open: flashfeedlearningapp.com/free-and-open

About FlashFeed

Learn anything, a little at a time.

FlashFeed turns big subjects into Knowledge Packs of short cards, quick games and study guides. Scroll when you have a minute. Go deeper when you have more. Free, with no ads and no account to make.

flashfeedlearningapp.com

Download on the App Store

Get it on Google Play

Follow

- Instagram — @flashfeedlearningapp · @flashfeed_history (history) · @flashfeed_spanish (Spanish)
- YouTube — @FlashFeedLearningApp
- TikTok — @flashfeed.learningapp
- FlashFeed Study Notes podcast — Spotify · Apple Podcasts · RSS

FlashFeed, by Recurxive LLC. November 2026 edition, built from The Infinity Machine (book) pack version 1.0.24. Typeset in Big Shoulders Display and Figtree (SIL Open Font License). Pictures are credited on the image credits page.