A History of AI Since the 1950s

INTS1301 Technology and Society: From Plato to NATO — Week 7

Brian Ballsun-Stanton

2026-04-21

A conversation with a book

WHEN WAS THIS WRITTEN?

> Simmons' dream was that one could have "a conversation with a book;" the computer would read the book, and then the user could have a conversation with the computer, asking questions to be answered from the computer's understanding of the book.
>
> "The objective of this project is to develop a research methodology and a vehicle for the design and construction of a general purpose computerized system for synthesizing complex human cognitive functions. The initial vehicle, proto-synthex, will be an elementary language-processing device which reads simple printed material and answers simple questions phrased in elementary English."

1961

Robert Simmons, SDC "Synthex" grant proposal

> *Accomplishing this "dream" would turn out to be as hard as AI itself.*
>
> — Nilsson (2010)

- **Source:** (Nilsson, 2010, p. 111), quoting Simmons, SDC "Synthex" grant proposal, 1961
  - "A conversation with a book" is Simmons's own phrase, in quotes, via Nilsson (2010)
  - Full transcription saved locally
- **SDC = System Development Corporation**
  - RAND spin-off (1956)
  - Operated on US Air Force contracts

  - Week 6 callback: state-funded research decides what tech exists

- **The reveal game**
  - Students will guess 2020s, possibly 2010s
  - Contrast candidate to hand: any 2026 OpenAI / Anthropic / Google AI press release
  - The Synthex grant prose and 2026 AI press releases use the same register

- **Nilsson's closer (Nilsson, 2010, p. 111):** *"Accomplishing this 'dream' would turn out to be as hard as AI itself"*
  - Amara's Law in one sentence: we overestimate short-term, underestimate long-term
  - Different voice from the 1961 material (Nilsson (2010) hindsight). That voice shift is why the line goes post-reveal

- **Bridge into slide 2**
  - The same promises keep coming back
  - Next 100 minutes: 70 years of how we got here, and whose interests that recycling serves

- **Drumbeat threads**
  - Thread 1 (notice/don't-notice) arrives slide 2
  - Threads 2 and 3 arrive slides 4 and 5

- **Budget:** ~4 min including guess-and-reveal

AI is a marketing term

WHEN IT WORKS, WE STOP CALLING IT AI

- **Technology you don't notice because it's working**
  - The motion sensor in this room is three if-statements in a trench coat. That's an expert system. Nobody calls it AI.
  - Your phone's autocomplete, spam filter, face unlock: all AI from previous hype cycles. You don't see any of it.
  - Google Maps routes you. Netflix picks what you watch. Siri sets your alarms. Background AI, every day.

- **Technology you notice, because it broke, because it's new, because it's threatening your job**
  - ChatGPT (2022). First time most people talked to an AI. Still new. You still see it.

  - Clippy (1997). Mocked because of the googly eyes. The machine learning underneath runs your autocomplete now. You don't see that part.
  - Self-driving cars. You still see them because they still go wrong.

- **AI is a synonym for technology: the stuff marketing wants you to notice**

Your spam filter uses AI. ChatGPT uses AI. Why do we only call one of them AI?

- **Core thesis slide:** AI is a marketing synonym for technology: the stuff the industry wants you to notice. When we understand how to do something properly, it fades to ready-to-hand and moves out of the label "AI."
- **180-year precedent: Ada Lovelace, 1843**
  - Ada Lovelace, in her notes on Menabrea's account of Babbage's Analytical Engine: *"It is desirable to guard against the possibility of exaggerated ideas that might arise as to the powers of the Analytical Engine."*
  - Cited in (Russell & Norvig, 2021, p. 33) as the first AI-hype warning
  - The marketing-term thesis has a 180-year lineage, not a 2026 invention
  - **SLIDE 3 FORWARD-LINK:** Lovelace is a pre-AI instance of the same naming-is-political phenomenon that produces "AI" at Dartmouth in 1956
- **R&N pigeons quote, speaker-voice backup for discussion**
  - Russell & Norvig (2021, p. 20): *"The quest for 'artificial flight' succeeded when engineers and inventors stopped imitating birds and started using wind tunnels and learning about aerodynamics. Aeronautical engineering texts do not define the goal of their field as making 'machines that fly so exactly like pigeons that they can fool even other pigeons.' "*
  - Alternative framing to the spam-filter-vs-ChatGPT question: *"Aeronautics stopped trying to fool pigeons. Why does AI still try to fool humans?"*
  - Use this one if the current discussion stalls, or deploy verbally as a *Turing test* sideshow
- **Martins 1984: 1984 industry voice confirming the marketing thesis**
  - Martins, L.-F. (1984), former director of RAND Corporation's R&D Program in Information Processing Technology (until 1982), writing from inside the defence-research establishment
  - Cited in (Martin, 1986, pp. 56–57)
  - On expert-systems success: *"six interdependent factors: brilliant programmers, the very narrowly defined and/or easy problems worked, generous funding over a long period of time, luck, development of customized tools (without the use of shells) and finally,* ***misleading advertising as to the utility and intelligence of some of the programs****"*

  - Deploy if a student pushes back that *"AI is a marketing term"* is a modern critique. Martins was naming it inside the 1984 US defence-research apparatus, same year as the TIMM brochure (**SLIDE 10 CALLBACK-FORWARD**)

- **Heidegger vocabulary (verbal, not on slide)**
  - **Ready-to-hand:** the hammer when you're hammering. Glasses, keyboard, motion sensor. Tech you don't notice because it's working
  - **Present-at-hand:** the hammer when it breaks and you stare at it. Tech becomes an object of attention rather than a tool
  - Tute uses "ready-to-hand" directly: *"What does it mean for something claimed as AI to fade to ready-to-hand?"*
  - Pedagogy: lived examples first, labels named after, so students meet the concept before the jargon

- **Motion-sensor prop**
  - Physical demo available: walk under it, trigger the light
  - "Three if-statements in a trench coat" = an expert system (1980s vintage). Nobody calls it AI
  - Callback to slide 8 (expert systems boom) and slide 9 (FRESH)

- **Ready-to-hand examples**
  - Autocomplete, spam filter, face unlock, voice-to-text
  - Google Maps routing, Netflix recommendations, Siri
  - Students probably used 3–5 AI systems before walking into this room

- **Present-at-hand examples**
  - ChatGPT (2022): first time most people *talked to* an AI
  - Clippy (1997): the googly eyes made it visible; the Bayesian inference underneath is invisible in your autocomplete today (callback to slide 12)
  - Self-driving cars: still visible because they still go wrong

- **Discussion question seeding:** *"Your spam filter uses AI. ChatGPT uses AI. Why do we only call one of them AI?"*
  - "One is invisible" → ready-to-hand path → reinforce the phenomenology
  - "One is new" → marketing path → reinforce the "name reserved for the unsolved" thesis
  - "One is smarter" → capability path → push back: spam filters beat symbolic AI; both *are* AI
  - "One talks to me" → anthropomorphism path → tees up ELIZA (slide 4)
  - Rigged question: any answer lands the thesis, because both are AI

- **Drumbeat threads**
  - Thread 1 (notice/don't-notice) introduced here

  - Thread 2 (Amara's Law) arrives slide 4 (ELIZA)
  - Thread 3 (symbolic/nonsymbolic) arrives slide 5 (1960s)
  - Don't name them as a trilogy yet; they emerge as we go

- **Budget:** ~7 min with live student answers

Before AI was called AI

HOW A FIELD GETS ITS NAME

- **The military problem that became a field**
  - WWII: how do you hit a fighter plane? Anti-aircraft predictors, gun directors, fire-control computers
  - Wiener's *Cybernetics* (1948): one framework (feedback and control) that covers thermostats, gun directors, and the way your body regulates its own temperature

- **1948 Hixon Symposium: two roads fork**
  - Caltech, *Cerebral Mechanisms in Behavior*: cognitive science and AI's common root
  - Simulate biology (neural nets) vs simulate behaviour (logic, symbols)
  - The logic-and-symbols side is later renamed GOFAI (*good old-fashioned AI*), Haugeland's 1985 retronym, coined after the rival camp emerges.

- **1956: the name "AI" exists because McCarthy writes the grant**
  - Wiener is not invited to Dartmouth: the man, not just the framework
  - McCarthy's own words: wouldn't *"accept Norbert Wiener as a guru or argue with him"*
  - Simon and Newell preferred "complex information processing"; they weren't writing the grant
  - Newell: *"Like all names of scientific fields, it will grow to become exactly what its field comes to mean."* McCarthy held the pen

- **Sources**
  - Three-meetings founding cluster, 1955 / 1956 / 1958 (Nilsson, 2010, p. 49)
  - McCarthy's naming rationale; Samuel *"phony"*; Newell/Simon *"complex information processing"* preference (Nilsson, 2010, p. 53)
  - Wiener-avoidance quote cites McCarthy (2000; via Nilsson, 2010, p. 53, fn. 12)
  - Full *"Like all names of scientific fields, it will grow to become exactly what its field comes to mean"* from Newell (1980; via Nilsson, 2010, p. 53, fn. 14)
  - Primary source for feedback-loop framework: Wiener (1948)
  - Dartmouth proposal: McCarthy et al. (1955)

- **WWII targeting lineage (the seed, not the whole framework)**

  - Wiener worked on anti-aircraft predictor systems at MIT during the war
  - Naval gun directors, fire-control computers: feedback applied to one narrow problem
  - Lineage runs from here to the motion sensor in this room (slide 2 callback) and to every autonomous system that senses + acts

- **Cybernetics as Wiener actually defined it**
  - Full title: *Cybernetics: Or Control and Communication in the Animal and the Machine* (MIT Press, 1948)
  - Wiener's coinage: from Greek *kybernētēs* ("steersman"), the same root as "governor"
  - Scope covers mechanical, biological, *and* social systems unified by feedback + control mathematics
  - Book chapters: Newtonian mechanics / information theory / the nervous system / homeostasis / learning / self-reproducing machines / language / society
  - The targeting case is one application; thermostats, endocrine regulation, and an ecosystem finding equilibrium are the same maths
  - This is the scope McCarthy is dodging: not just "the anti-aircraft guys," an entire unified theory of self-regulating systems with Wiener as its guru

- **Hixon Symposium 1948**
  - Caltech, September 1948, *Cerebral Mechanisms in Behavior*
  - Where cognitive science and AI share a common root before branching
  - Rochester's Dartmouth proposal section cites Lashley's Hixon paper directly (McCarthy et al., 1955, p. 9, ref 2)

- **The two proto-camps**
  - Simulate biology: neural nets, cybernetics, connectionism
  - Simulate behaviour: logic, symbolic reasoning, formal methods
  - Pitts (via Nilsson, 2010) claimed they *"will come to the same thing"*; 70 years later, they have not
  - Foreshadows (Nilsson, 2010, p. 77) "symbolic vs nonsymbolic" naming that arrives slide 5

- **GOFAI: the retronym is the thesis in miniature**
  - Haugeland (1985) coins *Good Old-Fashioned AI* as the name for the symbolic camp
  - Retronyms only appear when a rival needs a contrast class: *"acoustic guitar"* once electric exists; *"analog"* once digital exists; *"GOFAI"* once the statistical / connectionist camp starts claiming the AI label
  - The symbolic side needed naming only because the *other* side had started calling itself AI: exactly Brian's marketing-term thesis operating on the word *"AI"* itself
  - **SLIDE 12 CALL-FORWARD:** the retronym reveal lands at the *"revival-not-revolution"* move, both camps claiming the title of AI, with Haugeland's 1985 coinage as the artefact of the fight (Spider-Man-pointing-at-Spider-Man on that slide)

  - Register: pedagogically we use *"symbolic AI"* through slides 4–8 as the working name; GOFAI itself gets named here at slide 3, re-used at slide 5, payoff-revealed at slide 12

- **Three-meeting naming cluster (Nilsson, 2010, p. 49)**
  - 1955: *"Session on Learning Machines,"* Western Joint Computer Conference, Los Angeles
  - 1956: *"Summer Research Project on Artificial Intelligence,"* Dartmouth College
  - 1958: *"Mechanization of Thought Processes,"* National Physical Laboratory, UK
  - Fields don't get invented at a single meeting; they emerge from clusters of conversations
  - Dartmouth gets the credit because McCarthy *named* it there

- **Humanities framing for this slide** (*this is the people-and-institutions-and-governments slide*)
  - Technology is not pure ideas. It's people writing grants, people sitting on committees, governments choosing what to fund
  - McCarthy wrote the Dartmouth grant. Minsky, Rochester (IBM), and Shannon (Bell Labs) co-signed. Those four picked the field: the name, the attendees, the agenda
  - Wiener was not invited. Not metaphorically; literally absent from the room. That is how a rival framework gets sidelined
  - Simon and Newell had a different preferred name. Simon and Newell were not grant authors on this one. The name that won is the name the grant writer picked
  - Each of these is a *human* decision with *political* consequences; the shape of the field for 70 years comes from specific people making specific choices inside specific institutions

- **McCarthy's own words (Nilsson, 2010, p. 53, fn. 12; via McCarthy, 2000)**
  - Full quote: *"to escape association with 'cybernetics.' Its concentration on analog feedback seemed misguided, and I wished to avoid having either to accept Norbert Wiener as a guru or having to argue with him"*
  - McCarthy's original (primary source) adds *"(not Robert)"* after Norbert. Bloomfield misnamed him; McCarthy politely corrects the record mid-sentence
  - Both intellectual disagreement *and* personal politics, confirmed via McCarthy's own 2000 essay
  - Minsky (reported by McCarthy in the same review): *"neither Wiener nor von Neumann, with whom he had personal contact, influenced him, because he didn't agree with their ideas."* The whole founder-group was distancing, not just McCarthy

- **Simon and Newell resistance (Nilsson, 2010, p. 53)**
  - Preferred *"complex information processing"* as a label for years after Dartmouth
  - Lost the naming fight because the grant was already written

  - Samuel (per Nilsson, 2010, p. 53; via McCorduck, 1979): *"The word artificial makes you think there's something kind of phony about this, or else it sounds like it's all artificial and there's nothing real about this work at all."* Founder-voice scepticism of the name from day one
  - Newell (1980): *"Like all names of scientific fields, it will grow to become exactly what its field comes to mean"* (web-version wording)
  - One-word check for print citation: web version has *"its field"*; Brian's earlier hard-copy reading had *"the field"*. Confirm from the Cambridge UP print edition before final bibliography pass
  - URL: https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/1839/2222

- **Cybernetics scope discipline**
  - One-sentence gloss only: the McCarthy dodge story
  - No Beer, no Davies, no Medina, no Cybersyn
  - Cybernetics as a field is further reading for interested students, not lecture material

- **Discussion seeding:** *"Who writes the grants today? Who doesn't get invited?"*
  - Grant writers today: OpenAI / Anthropic / Google executives pitching investors; DARPA programme managers; EU Commission officials drafting the AI Act; national-security advisers briefing ministers
  - Not invited: Bender et al. (fired/pushed out of Google over Stochastic Parrots; callback to slide 14); Timnit Gebru; the Kenyan data-labellers paid dollars an hour; regulators behind the technology curve
  - Name choices doing the same work today: "agentic AI," "AGI," "foundation model"; each one picked by someone holding a pen
  - Naming-as-political-act is a live phenomenon, not a 1956 curiosity
  - Air Canada's *"separate legal entity"* defence (callback to slide 14): the elastic name weaponised for liability-shifting

- **Callbacks**
  - Slide 2: motion sensor example reappears with WWII ancestry now attached
  - Slide 5: symbolic/nonsymbolic two-camps thread previewed here, named there
  - Slide 10: the naming dodge is the first move in the marketing-term arc that terminates in TIMM 1984 / Cloudflare 2026

- **Budget:** ~7 min. Dense slide; three main bullets, each needing room to breathe

ELIZA, 1966

PEOPLE BELIEVED IT. HIS OWN SECRETARY BELIEVED IT.

- **ELIZA: symbolic AI's first public landmark**

  - Explicit rules for manipulating language: rules the programmer writes down
  - "I hate my mother" → *"Why do you hate your mother?"*
  - First chatbot most people ever saw. Symbolic AI reaching a mass audience

- **Weizenbaum's secretary, who had watched him build it:**
  - *"She knew she was talking to a machine. Yet after a few sentences she turned to me and said 'Would you mind leaving the room, please?'"* — Weizenbaum (1967)
  - Parasocial relationships with computers are not new. **1966.**

- **Weizenbaum warned the field from 1966. McCarthy dismissed him in print**
  - *Computer Power and Human Reason* (1976): *"there are certain tasks which computers ought not be made to do, independent of whether computers can be made to do them"*
  - McCarthy's review: *"no argument is offered that might be answered."* Field-founder dismissing field-critic, first of a long pattern

Have you ever said *"please"* or *"thank you"* to an AI? Why?

- **Sources**
  - Primary ELIZA paper: Weizenbaum (1966)
  - Secretary anecdote lives here, not in the 1966 paper: Weizenbaum (1967)
  - Book-length ethical argument: Weizenbaum (1976)
  - ELIZA *"'understood' nothing"* (scare quotes are Nilsson's) (Nilsson, 2010, p. 110)
  - ELIZA → CP&HR linkage + *"ought not"* framing (Nilsson, 2010, p. 315)
  - *"Monstrous obscenity"* quote + Nazi-escape biography + Nilsson's own *"premature use"* commentary (Nilsson, 2010, p. 316)
  - McCarthy's *"An Unreasonable Book"*: McCarthy (1976)
  - Lewis Thomas, *"Notes of a Biology Watcher: On Artificial Intelligence," NEJM* 302(9), pp. 506ff, 28 February 1980 (via Nilsson, 2010, p. 316, fn. 29)
  - Wes convo, 2026-04-10: *"If you put googly eyes on a rock, we will pair-bond with the rock"*

- **What ELIZA actually was (tying it into the AI taxonomy)**
  - Simple pattern-matching program. Specific script named "DOCTOR" running a *Rogerian therapist* impersonation
  - Take user input → flip pronouns → rephrase as open question
  - No memory, no world model, no semantic understanding; pure substitution rules
  - Ran on MIT's Compatible Time-Sharing System, 1964–66

- **Which class of AI is ELIZA?** (this is the slide's taxonomic job)
  - **Symbolic camp, NLP branch.** Rule-based symbol manipulation, no learning from data

  - **Not an expert system.** Expert systems specifically encode *domain expertise* through production rules + inference engine (Feigenbaum's "knowledge principle", 1977), first landed by DENDRAL (1965 onwards, but crystallised as a class in the 70s). ELIZA has no domain knowledge; it's shallow pattern substitution
  - **Not nonsymbolic either.** No neural nets, no statistics, no pattern recognition from data. The nonsymbolic camp's NLP equivalent doesn't arrive until IBM Candide (1988, statistical MT) and then transformers (2017)
  - **Slot for the lecture:** ELIZA is the symbolic camp's first mass-exposure moment, the first time *the public* encountered what symbolic AI could do. Preview of what lands formally at slide 5 (Nilsson, 2010, p. 77)
  - Same rule-based lineage will flower into expert systems (slide 8), then collapse in the late-80s winter (slide 11)

- **This is where drumbeat thread 2 (Amara's Law) arrives**
  - Roy Amara (Institute for the Future president, c.1978, widely attributed, no verified primary publication): *"We tend to overestimate the effect of a technology in the short run and underestimate the effect in the long run"*
  - ELIZA short-run overestimate: *"ELIZA understands! AI is imminent!"* (wrong)
  - ELIZA long-run underestimate: parasocial relationships with software at planetary scale, nobody saw coming, except Weizenbaum
  - Name-check the thread verbally: *"This is the second of three threads running through this lecture: Amara's Law. What was promised, what landed, how long between."*

- **The secretary anecdote (load-bearing quote)**
  - Full passage from Weizenbaum (1967): *"My secretary watched me work on this program over a long period of time. One day she asked to be permitted to talk with the system. Of course, she knew she was talking to a machine. Yet, after I watched her type in a few sentences she turned to me and said 'Would you mind leaving the room, please?'"*
  - She had watched him *build* the program. She KNEW it was a machine. She still asked for privacy.
  - Parasocial relationship with a machine, fully aware, 1966
  - Wes's frame: googly eyes on a rock

- **The 1976 publication date matters (per Brian, 2026-04-18)**
  - *Computer Power and Human Reason* published 1976
  - Mass parasocial problem: ChatGPT, November 2022 (**46 years later**)
  - Weizenbaum was widely *read*. CP&HR is a set text in computer-ethics courses. Shaped AI-ethics as a discipline. Cited by Dreyfus, Searle, Lewis Thomas, and the whole philosophical-critique tradition
  - The gap wasn't reception. The gap was the *commercial product* at scale. Weizenbaum warned about a mass parasocial phenomenon that didn't exist as a mass phenomenon until LLMs hit consumer scale

  - Weizenbaum d. 2008, never saw ChatGPT. Had already seen enough

- **Dreyfus as companion voice (not on slide, verbal backup)**
  - Dreyfus (1972), revised 1979, 1992
  - Also Dreyfus & Dreyfus (1986) with Stuart Dreyfus
  - Philosophical argument drawing on Heidegger (!) and Merleau-Ponty: embodied know-how, tacit knowledge, why symbol-manipulation misses the human
  - Repeatedly dismissed by Minsky and Papert. Turned out to be substantially right on the limits of symbolic AI
  - Slide 2 callback: Dreyfus invokes the same ready-to-hand / present-at-hand distinction we introduced from lived examples

- **Weizenbaum biography (Nilsson, 2010, p. 316)**
  - Escaped Nazi Germany with his family in 1936
  - Nilsson: *"That experience cannot but have sharpened his keen sense of social responsibility"*
  - This is not abstract moral philosophy: a child who saw what systems-without-judgement do

- **Stronger Weizenbaum quote (Nilsson, 2010, p. 316) (speaker voice only, not on slide)**
  - *"The very asking of the question, 'What does a judge (or a psychiatrist) know that we cannot tell a computer?' is a monstrous obscenity. That it has to be put in print at all, even for the purpose of exposing its morbidity, is a sign of the madness of our times."*
  - Deploy if a student pushes on whether AI ethics is overblown

- **McCarthy's dismissal as a pattern**
  - McCarthy (1976): *"No argument is offered that might be answered, and no attempt is made to define criteria of acceptability"* (verbatim from the review)
  - They became professional adversaries
  - Pattern repeats: field-founder dismisses field-critic. Dreyfus gets the same treatment; Bender and Gebru get pushed out of Google over *Stochastic Parrots* (**SLIDE 14 CALLBACK**)
  - **SLIDE 3 PAYOFF:** humanities-framing (specific people, specific politics, specific consequences) lands here. McCarthy's review is a 1976 instance of the pattern slide 3 sets up

- **Nilsson's own position (Nilsson, 2010, p. 316), discussion backup**
  - *"The real danger, I think, lies in the premature use of machines: thinking that they are able to perform a task before they are really competent to do so"*
  - Rare first-person voice; Nilsson putting his position on record

  - Diagnoses 2024–26 agent deployment directly

- **Lewis Thomas (via Nilsson, 2010, p. 316): outside-discipline critic**
  - Lewis Thomas, physician / biologist / essayist (1913–1993), writing in *NEJM* 302(9), pp. 506ff, 28 February 1980: "Notes of a Biology Watcher: On Artificial Intelligence"
  - Full passage (three parts, via Nilsson (2010, p. 316) with ellipses indicating Nilsson's edits; full text quote-verified):
    * Part 1: *"The most profoundly depressing of all ideas about the future of the human species is the concept of artificial intelligence. The ambition that human beings will ultimately cap their success as evolutionary overachievers by manufacturing computers of such complexity and ingenuity as to be smarter than they are, and that these devices will take over and run the place for human betterment or perhaps, later on, for machine betterment, strikes me as wrong in a deep sense, maybe even evil."*
    * Part 2: *"This is what the artificial intelligence people are talking about: a mechanical brain with the capacity to look back over the past and make accurate predictions about the future, then to lay out flawless plans for changing that future any way it feels like, and, most appalling of all, capable of feeling like doing one thing or another."*
    * Part 3: *"It is, in my view, an absolutely hideous prospect, and if I thought it were really something waiting ineluctably ahead of us I would spend all my days in protest."*
  - Worth naming to students: *"This is not only computer scientists saying this."* NEJM is the premier biomedical journal; a biologist calling AI *"absolutely hideous"* from its pages lands differently than a philosopher's critique
  - Nilsson couples Thomas to Theodore Roszak (self-confessed *"neo-Luddite"*), *The Cult of Information* (1994): a second outside-discipline voice if the discussion runs that way

- **Discussion seeding:** *"Have you ever said 'please' or 'thank you' to an AI? Why?"*
  - Most students will admit they have
  - Possible paths:
    * "Politeness is habit" → parasocial is automatic. Weizenbaum's secretary, 1966
    * "I don't want it to be rude back" → attributing agency. Connects to slide 14 Air Canada *"separate legal entity"* defence
    * "Sam Altman said saying please wastes compute" (real 2025 news item) → corporate politics of how we relate to machines
    * "I feel bad not to" → Weizenbaum's diagnosis arriving in 2025
  - Wes line to deploy: *"If you put googly eyes on a rock, we will pair-bond with the rock"*

- **Callbacks**

  - Slide 2 (ready-to-hand): ChatGPT is present-at-hand; ELIZA was present-at-hand in 1966 and never faded because the parasocial effect keeps it visible
  - Slide 3 (people matter): Weizenbaum's Nazi-escape biography, McCarthy's dismissal; specific humans with specific politics
  - Slide 14 (Bender + Air Canada): Bender's *"ersatz fluency"* and Weizenbaum's *"concealing misunderstanding"* are the same phenomenon 55 years apart
  - Week 8 bridge: parasocial relationships, modern communication tech; this is where it plants

- **Budget:** ~8 min. Quote-heavy; needs time for the secretary line to land and for discussion to open

Teaching computers by showing

THE OTHER CAMP: LEARNING FROM DATA, NOT RULES

- **Rosenblatt's Perceptron (1958): the first *neural net* you could actually run**
  - McCulloch & Pitts designed a neural net in 1943 as mathematical theory. Rosenblatt builds a working one 15 years later
  - Show thousands of labelled examples. The machine adjusts its own internal weights until it gets the answers right. The rules emerge from the data
  - This is where the word *"training"* comes from
  - Once trained, the weights freeze. The machine then *applies* what it learned, read-only. **Training is not doing**
  - Perceptron descendants run the high-voltage power grid today, catching spikes and transients before the lights flicker

- **Nilsson (2010, p. 77) names the split in retrospect**
  - *"Symbolic"* AI (GOFAI): rules, logic, theorem-proving. ELIZA lives here
  - *"Nonsymbolic"* AI: neural nets, statistics, pattern recognition. Rosenblatt lives here
  - Both are called "AI" in the 1960s. Both will fight for 70 years. Neither wins permanently

ELIZA has rules a human wrote. The Perceptron has weights the machine found. If both count as AI, what *is* AI?

- **Sources**
  - *"Symbolic"* vs *"nonsymbolic"* retrospective naming (the slide's taxonomy); also names the first Pattern Recognition Workshop (Puerto Rico, October 1966), organised by the IEEE Computer Society's Pattern Recognition Subcommittee (Nilsson, 2010, p. 77)

  - Primary proceedings for the October 1966 workshop: Kanal (1968; via Nilsson, 2010, p. 77, fn. 33)
  - MINOS III at SRI (Munson, Hart, Duda): neural-net OCR on FORTRAN coding sheets, mid-1960s (Nilsson, 2010, p. 71)
  - GPS and the *accomplish vs simulate* axis (speaker-note backup) (Nilsson, 2010, pp. 87–88)
  - Rosenblatt, F. (1958), *"The perceptron: a probabilistic model for information storage and organization in the brain," Psychological Review* 65(6), 386–408
  - Selfridge, O. (1959), *"Pandemonium: a paradigm for learning"*
  - Foreshadows slide 7: Minsky & Papert (1969)
  - IBM history: 1951 SNARC, 1957 Perceptron, 1959 Selfridge + Samuel *"machine learning"*

- **The Perceptron in detail**
  - Rosenblatt at Cornell Aeronautical Laboratory
  - Mark I Perceptron (1960): a physical machine, image recognition through potentiometers and electric motors
  - Two-layer *neural net* (input → output), linear classifier
  - Weight-adjustment rule from labelled examples: what we now call supervised learning
  - **Direct ancestor of every deep-learning system today.** AlexNet 2012, GPT-4, Claude; same basic idea, scaled 60 years

- **Vocabulary that landed here**
  - *Neural net*: explicit biological analogy (McCulloch & Pitts, 1943); Rosenblatt makes it run
  - *Training*: the word for the weight-adjustment process. Now everyday tech vocabulary ("we trained a model on X")
  - Callback to slide 2's "AI vocabulary leaking into common usage" thread (*heuristic*, *algorithm*, now *training*)

- **Training vs doing (an important first-year distinction)**
  - **Training phase:** machine is fed labelled examples, adjusts weights, learns
  - **Inference phase** (the doing): weights are frozen; machine applies the trained function to new input
  - Classical neural nets (Perceptron, modern deep learning, LLMs) all follow this pipeline
  - GPT-4 and Claude were *trained once*, at enormous compute cost. When you chat with them, weights are frozen; they are not "learning" from your conversation
  - Exceptions (online learning, continual learning, RLHF-during-deployment) exist but are specialist cases. Ignore at first-year level
  - Deploy if a student asks "does ChatGPT remember what I told it?" No, the model weights don't update, context window is a separate mechanism

- **Call-forward to slide 14 (design vs deployment)**
  - 1943 McCulloch-Pitts design → 1958 Rosenblatt running machine = 15-year gap between paper and deployment
  - Same pattern at slide 14: Vaswani et al. (2017) → ChatGPT 2022, which OpenAI execs called a *"low-key research preview"* (The Verge, 2023)
  - Paper deployment. The gap is where the marketing-term thesis lives: the engineering is quiet; the commercial framing is loud
- **Perceptrons are *everywhere* (the "still with us" point)**
  - **High-voltage power distribution:** neural-net classifiers detect transient faults, voltage spikes, equipment failures in real time. Descendants of 1958 Perceptron keeping the lights on
  - Other ubiquitous deployments: handwriting recognition on your phone, face unlock, spam filters (the Bayesian version is a close cousin), medical image classification, engine-fault detection in aircraft
  - Callback to slide 2 motion sensor: you live inside the output of trained neural nets whether or not you ever think about them
- **Other nonsymbolic figures of the 1960s**
  - Selfridge's *Pandemonium* (1959): competing "demons" identify features in data
  - MINOS III at SRI (Munson, Hart, Duda): neural-net OCR on FORTRAN coding sheets. The net produced ranked candidate characters with confidence scores; a FORTRAN-legality check chose among them (Nilsson, 2010, p. 71). Early example of probabilistic output + constraint filter: the shape of modern ML inference
  - Arthur Samuel (1959): checkers program that improved through self-play; coined *"machine learning"*
- **Foundations laid at the same time**
  - 1957–58: Rosenblatt's Perceptron (nonsymbolic)
  - 1958: McCarthy's LISP (symbolic programming language)
  - The two camps building their foundational tools *simultaneously*, often at the same institutions
- **Where ELIZA (symbolic) and Perceptron (nonsymbolic) sit together**
  - Same decade
  - Both called "AI" at the time
  - Different approaches: rules a human wrote vs weights the machine found
  - Both well-funded, strong defenders
  - Nilsson's p. 77 naming is 2010 retrospective; in 1966 they were just different research programmes
- **GPS (a cross-cutting complication, speaker-note backup, not on slide)**

  - Newell, Shaw, Simon, 1959 Paris conference: *General Problem Solver*
  - They wrote: *"GPS maximally confuses the two approaches, with mutual benefit"*
  - **The "two approaches" GPS refuses are NOT symbolic/nonsymbolic.** They are *accomplish* (build a machine that does the task) vs *simulate* (build a machine that does it the way a human does it)
  - A second axis that cross-cuts the symbolic camp itself
  - Deploy verbally if a student pushes on whether the split was clean

- **Foreshadowing slide 7 (1970s winter)**
  - Minsky & Papert (1969)
  - Argued single-layer perceptrons can't learn XOR and other patterns: fundamental limits
  - Standard narrative: this book killed neural-net research for 15 years
  - Nilsson (2010, ch. 16) disputes: interest was already fading before the book; Rosenblatt had moved on; heuristic programming drew people away
  - Slide 7 handles this: Minsky/Papert as one of several reports that redirect fields

- **Foreshadowing slide 13 (deep learning resurrection)**
  - The Perceptron's descendants dominate today
  - 1969 critique dissolved by compute + data in 2012 (AlexNet), not by a better algorithm
  - The nonsymbolic camp's 40-year exile ends here

- **Discussion seeding:** *"ELIZA has rules a human wrote. The Perceptron has weights the machine found. If both count as AI, what is AI?"*
  - Paths students might take:
    * *"AI is anything that imitates intelligence"* → marketing-term thesis (slide 2 callback)
    * *"Neither is really AI, both are tricks"* → Samuel's 1956 *"phony"* critique (slide 3 callback)
    * *"AI needs to learn, so the Perceptron is more real"* → nonsymbolic advocacy
    * *"AI needs to reason, so ELIZA is more real"* → symbolic advocacy
    * *"This is the Bitter Lesson"* → jump to slide 14, fine
  - Both are AI in 1966. Only one side gets called AI by 2026; the other gets called "machine learning" or just "software"

- **Callbacks**
  - Slide 2 (marketing term): both camps called AI in the 60s; today only symbolic's *successor* (LLMs) lives under the AI brand, though LLMs are nonsymbolic. The label migrated
  - Slide 3 (people + institutions): Rosenblatt at Cornell, McCarthy at MIT, Minsky at MIT; institutional politics shape which camp gets funded

  - Slide 4 (ELIZA): worked symbolic example established; this slide balances with a worked nonsymbolic example
  - Slide 7 (winter): Minsky/Papert 1969 lands here
  - Slide 13 (deep learning): Rosenblatt is the ancestor of the resurrection

- **Budget:** ~7 min

Where is the intelligence right now?

END OF THE 1960s. THREE THREADS.

- **Thread 1: notice / don't-notice**
  - Technology that works reliably fades into background infrastructure. You don't notice your spam filter, you don't think of the motion sensor as AI, and autocomplete is just how typing works now. What we still call AI is whatever we haven't yet figured out
- **Thread 2: Amara's Law**
  - We overestimate what a new technology will do in the short run and underestimate what it does in the long run. Weizenbaum's 1976 warning about parasocial attachment to machines arrived on time in 2022, forty-six years late
- **Thread 3: symbolic vs nonsymbolic**
  - ELIZA followed rules a human wrote down explicitly. Rosenblatt's Perceptron adjusted its own weights from labelled examples. Both were called AI in 1966; both camps are still with us; neither has ever permanently won the label

SNL "New Shimmer" (1976): floor wax AND a dessert topping

Where is the AI in your life right now? Which of these three threads is it on?

- **Purpose of this slide**
  - Pause slide. Breathe. Synthesise
  - First time the three drumbeat threads get named as a set; up to now each was introduced one at a time
  - Checkpoint at the end of the 1960s before we walk into the first AI winter
  - Not an information-delivery slide; light by design
- **Where the lecture is at this point**
  - ~30 minutes in (roughly)
  - Covered: epigraph (slide 1), thesis + Heidegger (slide 2), pre-AI + Dartmouth (slide 3), ELIZA (slide 4), Perceptron + nonsymbolic (slide 5)

  - Next: slide 7 brings the 1970s: reports that kill fields, Mansfield Amendment, Lighthill, Minsky/Papert's *Perceptrons* critique lands

- **How each thread was seeded**
  - Thread 1 (notice/don't-notice), slide 2. Heidegger's ready-to-hand / present-at-hand, delivered through lived examples before the jargon
  - Thread 2 (Amara's Law), slide 4. Weizenbaum's 46-year lag, ELIZA's over-promise and under-delivered warning
  - Thread 3 (symbolic/nonsymbolic), slide 5. ELIZA (rules) and Rosenblatt (data) as worked examples

- **Answering "where is the intelligence right now?" in 1966**
  - In ELIZA? (the secretary thought so)
  - In Rosenblatt's Mark I? (pattern recognition worked)
  - In GPS, the Logic Theorist? (theorem-proving worked)
  - In the programmer writing the rules?
  - In the data the Perceptron was trained on?
  - Nilsson (2010, p. 110): ELIZA *"understood nothing."* So where *did* the intelligence live?
  - Students won't have a clean answer. That's the point

- **Discussion seeding:** *"Where is the AI in your life right now? Which thread is it on?"*
  - Possible student answers:
    * Autocomplete / spam filter → thread 1 (ready-to-hand), thread 3 (nonsymbolic, trained)
    * ChatGPT → thread 1 (still present-at-hand), thread 2 (the Weizenbaum 46-year warning arriving), thread 3 (nonsymbolic now winning)
    * Smart light switch → thread 1 (ready-to-hand), thread 3 (symbolic, rules a human wrote)
    * Siri / Alexa → thread 1 (ready-to-hand since ~2015), thread 3 (nonsymbolic speech)
  - Any plausible answer lands because the threads are real and overlap

- **Forward-foreshadow verbal transition to slide 7**
  - *"You've now seen what the field looked like at the top of its first boom. Funded, confident, producing results, two camps competing. Next we look at what happens when the money catches up to the promises."*

- **Budget:** ~3–4 min. Do not over-run; the recap is a pause, not a re-teach

Reports that kill fields

SPECIFIC DOCUMENTS. SPECIFIC CONSEQUENCES.

> "Work of excellence by talented young people was stigmatised as bad science … killed in mid-trajectory. … To speak plainly, it was an outrage."
>
> — Donald Michie, on Edinburgh AI after Lighthill (1972)

- **ALPAC, 1966: machine translation dies**
  - A US National Research Council committee chaired by John Pierce reviewed Cold-War machine translation funding
  - The verdict: quality was poor, cost was high, and human translators outperformed the machines meant to replace them
  - Funding was cut across the US. Machine translation went quiet for a decade, 1967 to 1976
- **Lighthill, 1973: UK AI collapses**
  - *"In no part of the field have the discoveries made so far produced the major impact that was then promised"* (Lighthill (1972), commissioned by the UK Science Research Council)
  - The critique: AI demos worked on toy problems but could not scale, because the combinatorial explosion punished every attempt to generalise

Who could write the 2026 Lighthill report? What field would it kill?

- **Sources**
  - Michie "outrage" blockquote, verified from hardcopy 2026-04-19 (Nilsson, 2010, p. 204)
  - Michie primary source (Michie, 1982, p. 220)
  - Part IV opener confirming Mansfield Amendment 1969 (Nilsson, 2010, p. 207)
  - Pierce's *"artificial intelligence is real stupidity"* (full context: *"John R. Pierce, whom I have already mentioned in connection with both the ALPAC report on machine translation (in Section 7.2) and his negative comments about speech understanding (p. 282), wrote me a very short letter…"*) (Nilsson, 2010, p. 318)
  - On-slide quote *"In no part of the field have the discoveries made so far produced the major impact that was then promised"* opens **§3 *"Past disappointments"*, p. 9** (second sentence of the section, following the preamble *"Workers entered the field around 1950, and even around 1960, with high hopes that are very far from having been realised in 1972"*; verbatim verified). Also documents Lighthill's three-category taxonomy (A: Advanced Automation; B: Building Robots; C: Computer-based CNS research); speaker-note backup if a student asks what Lighthill was

actually measuring against. Report dated **July 1972**, published 1973 as Part I of the SRC's *Artificial Intelligence: A Paper Symposium* (Lighthill, 1972)
  - ALPAC report: Pierce & Automatic Language Processing Advisory Committee (1966)
  - *"Quiet decade"* (1967–1976) phrase source: Hutchins (1995; via Nilsson, 2010, ch. 10, fn. 1)

- **The Michie quote: context**
  - Donald Michie, Edinburgh AI researcher; built MENACE (matchbox tic-tac-toe learner, 1960; (Nilsson, 2010, p. 117))
  - His Edinburgh group was one of three UK AI centres destroyed by Lighthill
  - Full passage in Nilsson includes the middle sentence: *"The destruction of cooperative human mechanism and the careful craft of many hands is elsewhere described as a mishap. But to speak plainly, it was an outrage."*
  - Slide opens with the human cost. The bullets explain why

- **ALPAC 1966: the template**
  - Automatic Language Processing Advisory Committee of the US National Research Council
  - Chair: John R. Pierce (Bell Labs communications engineer; satellite comms, pulse-code modulation)
  - Cold War context: US funding Russian-to-English MT for intelligence
  - Verdict: quality poor, cost high, human translators outperformed
  - Pierce's own 1981 letter to Nilsson (Nilsson, 2010, p. 318): *"I believe I invented the slogan, 'Artificial intelligence is real stupidity.' "* ALPAC chair's cynicism about the field he helped gut
  - Forward-link to slide 12: statistical MT returns 1988 (IBM Candide). The symbolic/rule-based approach lost to data + compute: a Bitter Lesson preview

- **Lighthill 1973: the UK collapse**
  - Sir James Lighthill, applied mathematician (fluid dynamics, not AI)
  - Commissioned by UK Science Research Council
  - Report: Lighthill (1972)
  - Core argument: combinatorial explosion. AI demos work on toy problems (blocks worlds, small puzzles) but scale catastrophically with problem size
  - Verdict: UK should stop funding general AI; narrow applications only
  - Edinburgh, Sussex, Essex AI groups dismantled or starved
  - Michie's "outrage" is the human voice of this mechanism

- **Lighthill was correct, and this is the lead-up to Sutton's Bitter Lesson**
  - The technical critique was right. Symbolic AI *did* hit the combinatorial wall. Cleverer symbolic methods did not resolve it; Lighthill's prediction on that was accurate

  - The resolution came from a direction Lighthill couldn't see in 1973: compute + data scaling
  - Thread runs Lighthill → expert-systems bust (slide 11) → neural-net resurrection (slide 13) → transformers (slide 14)
  - Sutton (2019): *"general methods that leverage computation are ultimately the most effective, and by a large margin"*
  - Forty-six-year gap between the critique and the resolution, rhymes with Weizenbaum 1976 → ChatGPT 2022 (slide 4)
  - The Bitter Lesson is slide 14's closer. Slide 7 is where it first lands

- **Michie's outrage (reframed once Lighthill is conceded as correct)**
  - If Lighthill's technical diagnosis was right, Michie's outrage isn't that Lighthill was *wrong* about the combinatorial explosion
  - The outrage is about the *response*: stigmatising good researchers as bad scientists, ending careers mid-trajectory, starving a field of the funding that would (decades later, via compute and data) have resolved it
  - Sharpens slide 7's thesis: reports don't just redirect fields; the institutional *response* to a correct diagnosis is what destroys people. The document didn't kill Edinburgh; the funding decision did
  - Deploy if a student asks "was Lighthill wrong?" Answer: no, he was right, and that's not the same as the response being just

- **Bullet 3: what survived (slide 2 callback)**
  - Tech persists; labels die. Slide 2 fade-to-background thesis pays back here
  - ALPAC killed MT-the-field. MT-the-capability returned as statistical MT (1988), then Google Translate, now embedded in every LLM
  - Heuristic search → Google Maps routing (Dijkstra, A*)
  - OCR → phone camera text scanning
  - Speech recognition → Siri, voice-to-text
  - The 1970s AI winter is a label winter, not a capability winter

- **In notes, not on slide: Mansfield Amendment 1969**
  - Nilsson (2010, p. 207) confirms date 1969 (attached to Defense Procurement Authorization Act of FY1970; both dates circulate)
  - Vietnam changed DOD funding politics
  - DOD restricted to *"direct and apparent relationship to specific military function or operation"*
  - Same government that funded AI to build weapons defunded AI because of a war
  - Week 6 callback: state investment determines what technology exists

- **In notes, not on slide: Minsky & Papert, *Perceptrons* (1969)**
  - Standard narrative: this book killed neural nets for 15 years

  - Nilsson (2010, ch. 16) disputes: *"Rosenblatt himself began concentrating on other topics well before '69, and the success of heuristic programming methods caused a shift of attention, including my own, away from neural networks during the mid nineteen sixties"*
  - Fields don't die because of one book; funding, fashion, and competing successes matter more
  - Pays back slide 5's forward-link on Minsky/Papert
  - Forward-link to slide 13: neural nets return via compute + data, not via better arguments

- **Discussion seeding:** *"Who could write the 2026 Lighthill report? What field would it kill?"*
  - Possible student paths:
    * *"A funding agency: DARPA, EU AI Office, ARC"* → institutional power, state-decides framing
    * *"OpenAI or Anthropic: publish a paper that small-model AI is dead"* → commercial-incumbent framing
    * *"The Productivity Commission could declare something not worth funding"* → Australian-specific move
    * *"Nobody: there's too much money in AI"* → challenges the premise; good pivot into slide 8 (boom) and slide 11 (bust)
    * *"The report already exists: Bender, Gebru, Stochastic Parrots"* → forward-link to slide 14
  - Forces present-tense application of the slide 3 grant-writing thread

- **Callbacks**
  - Slide 2: fade-to-background thesis pays back at bullet 3
  - Slide 3: who writes the grants, who doesn't get invited. Slide 7 answers "who gets defunded"
  - Slide 5: Minsky/Papert forward-link resolved in notes with the Nilsson dispute

- **Forward-links**
  - Slide 8: expert systems rise from these ashes; Feigenbaum's knowledge principle answers Lighthill's combinatorial explosion by narrowing scope
  - Slide 12: statistical MT returns 1988
  - Slide 13: neural nets return via compute + data

- **Budget:** ~7 min

Expert systems rise

FEIGENBAUM'S ANSWER TO LIGHTHILL

> "Our agents must be knowledge-rich, even if they are methods-poor."
>
> — Feigenbaum (1977)

- **DENDRAL and the knowledge principle**
  - DENDRAL (Stanford, 1965 onwards) encoded a chemist's judgement about molecular structure into explicit rules: the first expert system
  - Feigenbaum's knowledge principle: intelligence comes from encoded domain expertise, not from general-purpose inference methods
- **MYCIN and XCON: the commercial arc**
  - MYCIN (Stanford, 1970s) diagnosed bacterial infections better than junior doctors, but was never deployed: liability and trust, not capability, blocked it
  - XCON (Digital Equipment, 1980) configured VAX computer orders. By 1989 the system had 17,500 rules and returned over $40 million per year to DEC
  - Expert systems spread through the Fortune 500 between 1975 and 1985: boom time for narrow-scope AI

The autopilot in every modern aeroplane can land the plane without the pilot touching anything. Is that AI?

- **Sources**
  - Primary source for the knowledge-principle quote: Feigenbaum (1977)
  - Scholarly framing + verbatim Feigenbaum quote on p. 3: Dick (2019)
  - XCON: 17,500 rules by 1989, "in excess of $40 million per year" return to DEC (verbatim from DEC's 1989 paper via Nilsson, photo-verified) (Nilsson, 2010, p. 238)
  - Kahn 1983 "ripe for exploitation" framing (carnival-barker preview for slide 10) (Nilsson, 2010, p. 286)
  - Expert-systems brittleness + McCarthy's MYCIN male-pregnancy anecdote (Nilsson, 2010, p. 326)
  - MYCIN Stanford PhD thesis: Shortliffe (1976)
  - R1/XCON at Carnegie-Mellon → DEC: McDermott (1980)
- **Full Feigenbaum quote (from Dick, 2019, p. 3; quoting Feigenbaum, 1977)**
  - "*We must hypothesize from our experience to date that the problem-solving power exhibited in an intelligent agent's performance is primarily a consequence of the specialist's knowledge employed by the agent, and only very secondarily related to*

the generality and power of the inference method employed. Our agents must be knowledge-rich, even if they are methods-poor."
  - Slide shows only the slogan; the rationale stays here in notes

- **Dick's framing (Dick, 2019): pedagogical anchor for this slide**
  - Dick explicitly frames expert systems as a response to the *"consistent disappointment"* of symbolic rule-bound AI: *"proponents of a field called 'expert systems' rejected the premise that human intelligence was grounded in rule-bound reasoning alone. They believed, in part because of the consistent disappointment attendant to that approach, that human intelligence depended on what experts know and not just how they think"*
  - This is the scholarly formulation of the slide 7 → slide 8 move: Lighthill's critique of rule-bound AI fed directly into Feigenbaum's narrow-domain knowledge-engineering response
  - Dick's broader thesis, *"what counts as intelligence is a moving target in the history of artificial intelligence,"* rhymes precisely with Brian's slide 2 thesis (AI as a synonym for technology whose label migrates)

- **DENDRAL: the first expert system**
  - Stanford, 1965 onwards. Feigenbaum, Buchanan, and Joshua Lederberg (Nobel laureate in genetics, 1958); the cross-disciplinary collaboration matters here
  - Problem: infer molecular structure from mass-spectrometry data
  - Approach: encode the chemist's heuristics explicitly as production rules plus an inference engine
  - Worked within its narrow domain. Did not scale beyond mass spectrometry, which is exactly Feigenbaum's point

- **Knowledge engineers: a new professional role**
  - Dick (2019) citing Feigenbaum (1977, p. 4): *"knowledge engineers"* interview human experts, observe their problem-solving practices, and make the knowledge explicit for encoding
  - This is the labour side of expert systems, a profession that did not survive the late-1980s bust (slide 11 payoff)
  - Week-6 callback: state and commercial investment creates new professions; the bust destroys them

- **MYCIN: worked, never deployed**
  - Shortliffe's PhD thesis, Stanford, 1976
  - Diagnosed bacterial blood infections; outperformed junior clinicians in evaluation studies
  - Never deployed in a real hospital: liability (who is responsible when the rule-chain misleads?), trust (doctors did not want to defer), workflow (not designed around how medicine actually happens)

  - Previews slide 9 (FRESH: Navy shelves a working expert system because planners don't want help)
  - Brittleness anecdote (Nilsson, 2010, p. 326): McCarthy's example. MYCIN could classify a male patient as "pregnant" because its rules had no commonsense background. Deploy if a student pushes on expert-systems capability; links forward to slide 14 LLM confabulation

- **XCON: the commercial success case**
  - Digital Equipment Corporation, 1980
  - John McDermott at Carnegie-Mellon built the original system as R1; DEC rebranded it XCON for production
  - Configured VAX computer orders: which parts fit together, what cables, what power supply
  - By 1989: 17,500 rules. Return to DEC: *"in excess of $40 million per year"* (DEC's own 1989 paper via (Nilsson, 2010, p. 238))
  - This is what "expert systems work" actually looks like: narrow, structured, high-volume, incremental
  - Folds into slide 2's fade-to-background thread: XCON-style business-rules engines now run invisibly under every enterprise order-configuration system

- **The broader boom (1975–1985)**
  - AAAI founded 1980: the expert-systems boom institutionalised
  - Japan Fifth Generation Computer project (1982–92): Tokyo's bet that expert-systems hardware would win the next decade
  - Kahn 1983 (Nilsson, 2010, p. 286): *"[AI] was ripe for exploitation … expert systems would be the centerpiece."* Language that rhymes directly with 2026 AI-agent marketing (slide 10 payoff)
  - DARPA Strategic Computing Program (billion-dollar scale): US response to the Japanese Fifth Generation panic
  - Expert systems move from research to commerce. Fortune 500 adoption

- **The peak-hype voice: Nilsson himself, 1983 AAAI Presidential Address**
  - Nils J. Nilsson (*our* primary-source historian of AI; his 2010 *Quest for AI* is the spine of this lecture) was AAAI President during peak expert-systems hype
  - In his 1983 AAAI Presidential Address (cited in Martin, 1986, p. 54): *"AI, perhaps together with molecular genetics, will be society's predominant scientific endeavor for the rest of this century and well into the next, just as physics and chemistry predominated during the decades before and after 1900"*
  - In the same address Nilsson proposes *"computer individuals,"* personal-assistant programs *"never turned off,"* with a *"constantly changing model of the world and of the user(s),"* *"able to engage in extended dialogs in natural language."* An LLM description, forty years early

  - Nilsson's 1983 hedge on timing: *"we are now at about the same stage in being able to build such programs as physicists were in the 1930's in being able to harness nuclear energy."* Correct idea, timing off by decades
  - **Pedagogical payoff:** the author of our 2010 history was the peak-hype voice in 1983, then wrote the book about the subsequent bust. Same person, both voices. Amara's Law operating on a single career
  - **SLIDE 4 CALLBACK:** Amara's Law (thread 2). **SLIDE 10 FORWARD-LINK:** Nilsson's own *"sprinkle a little AI"* autobiography (p.272) is the same Nilsson critiquing the boom he helped proclaim

- **Autopilot as borderline case (speaker-note backup)**
  - Is the autopilot in every modern aeroplane an expert system?
  - By Feigenbaum's 1977 definition (encoded pilot expertise operating on sensor inputs), yes it is
  - By the 1980s production-rule definition (IF-THEN rules and an inference engine), no; autopilots use control theory
  - Same technology, two labels. Whichever framing sells is the one used

- **Autopilot as the canonical borderline case**
  - Modern autopilots can take off, cruise, navigate, and land without pilot input (CAT III landings at most major airports)
  - Under Feigenbaum's 1977 framing (encoded expert judgement operating on sensor data to produce actions), autopilot qualifies as an expert system
  - Under the 1980s narrow framing (explicit IF-THEN production rules + backward-chaining inference engine), autopilot is control theory, not an expert system
  - The label shifted with the fashion. In 1977, "AI" covered everything; in 1985, "AI" meant production rules specifically; in 2026, "AI" means LLMs and "agents"
  - This is slide 2's marketing-term thesis paying back at slide 8

- **Verbal call-forward from slide 7 (Lighthill-was-right → Sutton's Bitter Lesson)**
  - Slide 7's on-slide Sutton call-forward was moved to notes; slide 8 is where it lands verbally
  - Speaker line: *"Lighthill said general AI wouldn't work. Feigenbaum's answer: stop being general. Knowledge-rich, methods-poor. Forty-six years later Sutton will tell us this strategy lost at scale; but for 1977 through 1985 it ran the show, and for narrow industrial problems it still does. XCON-style business rules are everywhere you don't look"*

- **Discussion seeding:** *"The autopilot in every modern aeroplane can land the plane without the pilot touching anything. Is that AI?"*
  - Possible student paths:

    * *"No, because nobody calls it AI"* → marketing-term thesis (slide 2 callback). The label has moved on
    * *"Yes, by the original definition"* → Feigenbaum's knowledge-rich, methods-poor. Autopilot encodes pilot expertise
    * *"It's just software"* → exactly. The successful AI fades to software. Slide 7 bullet 3 callback
    * *"The pilot is still responsible"* → Air Canada defence (slide 14 forward-link). Liability is where the label-stretch gets tested
    * *"It's a tool, not an agent"* → tees up slide 14's "agent" as the new elastic term
  - Any path lands the thesis because the label IS elastic

- **Callbacks**
  - Slide 2 (marketing term): autopilot example shows the label moving off working tech
  - Slide 7 (Lighthill): Feigenbaum's knowledge principle is the direct answer to Lighthill's combinatorial critique
  - Slide 7 (Sutton call-forward): Feigenbaum's knowledge-rich strategy is what Sutton will dethrone, in 2019; verbal move lands here

- **Forward-links**
  - Slide 9: FRESH, an expert system that works, and the Navy mothballs it. Politics beats technical capability
  - Slide 10: TIMM 1984 brochure + Cloudflare 2026, peak expert-systems hype rhymes with current AI-agent marketing. Kahn 1983 "ripe for exploitation" belongs here in the carnival-barker genealogy
  - Slide 11: end-of-1980s recap. The bust
  - Slide 14: Feigenbaum's knowledge-rich strategy is what Sutton's Bitter Lesson dethrones. The story arc closes there

- **Budget:** ~7 min

The Navy shelved it

DAYS ARE GOOD ENOUGH

> "The Navy's decision to mothball FRESH was because there was no compelling reason to keep it. It duplicated the expert judgement of Fleet planners, in a matter of hours rather than days. But the planners were not looking to retire, and in Naval warfare, days are good enough."
>
> — Walter Saunders, FCCBMP program manager (Nilsson, 2010, p. 291)

- **What FRESH did**

  - FRESH was an expert system for Fleet battle management, delivered to CINCPACFLT in 1990
  - Built on Symbolics Lisp machines using commercial expert-system shells, squarely inside Feigenbaum's knowledge-rich paradigm

- **Why the Navy shelved it**
  - When DARPA funding ended, the Navy decided not to continue
  - Saunders, in hindsight, called the FCCBMP goals *"an overreach for the state of the art in the 1980s"*
  - Congress pressured DARPA to divert the funding to anti-submarine warfare. Geopolitics trumped technical merit

If your employer deployed an AI that did your job in hours instead of days, would you help train it?

- **Sources**
  - FRESH/CASES mothball rationale, Saunders email opening (verbatim) (Nilsson, 2010, p. 291)
  - Extended Saunders email continuation covering Congressional ASW pressure + DARPA/Navy institutional tensions (quote-verified via web version; continuation past the photo-captured blockquote). Print-edition page number to confirm from hardcopy before final bibliography pass (Nilsson, 2010, p. 291 ff)
  - DRCF slide deck, session 1 (2025-10-07): augment-vs-automate framing

- **FCCBMP context: Nilsson (2010, p. 291) narrative paragraph**
  - FCCBMP = Fleet Command Center Battle Management Program (acronym expansion widely reported; Brian to verify from hardcopy Nilsson if wanted on-slide)
  - Two expert systems in the program: **FRESH** (battle planning) and **CASES** (simulation)
  - Both hosted on Symbolics Lisp machines, used commercial expert-system shells
  - CASES final simulations ran on an Encore parallel processor
  - Final prototypes delivered to CINCPACFLT (Commander in Chief, Pacific Fleet) in 1990
  - When DARPA's prototype phase ended and funding stopped, the Navy decided not to continue either system
  - CASES was received enthusiastically (the Fleet wanted the simulation tools) but the core intelligent-simulation goal *"never got off the ground"* (Saunders email, (Nilsson, 2010, p. 291))

- **The Saunders email: the load-bearing source**
  - Walter Saunders was the FCCBMP program manager at the Fleet's end, later Technical Director at Computer Sciences Corporation

  - Nilsson quotes from Saunders' email directly, with Saunders' own caveat: his email recollections *"contrast a bit with what he wrote in 1990"*
  - This is a named stakeholder giving a retrospective assessment that differs from the contemporaneous official story: exactly the kind of primary-source trace that makes stakeholder-identification visible to students

- **The stakeholder map for FRESH**
  - **SRI and other AI developers:** built FRESH. Wanted the system deployed as proof expert systems could work
  - **DARPA:** funded the research. Interested in advancing AI-for-defence; prototype phase always had a funding sunset
  - **US Navy command / CINCPACFLT:** client. Decided whether to continue after DARPA sunset
  - **Fleet planners:** end users. The people whose judgement FRESH duplicated. *Chose not to be replaced*
  - **Congress:** budget authority. Per Saunders's extended email (Nilsson, 2010, p. 291 ff): *"There was also pressure from Congress for DARPA to divert the funding to anti-submarine warfare, which was a hot topic at the time. And relations between DARPA and Navy were always strained near the breaking point."* ASW-pressure was *one* factor among four that Saunders names: DARPA/ISTO judging the FRESH/CASES research agendas ill-suited to an operational environment; Congressional ASW pressure; Navy-DARPA institutional tensions; *"everyone decided it was time to pack up and go home"*
  - The planners won because they were the daily users. Without their buy-in, the tech could not be deployed, regardless of its technical capability

- **Augment, not automate (verbal move, Brian's DRCF framing)**
  - The system worked. The people it was supposed to help did not want the help
  - AI that *replaces* the people it serves tends to get rejected; AI that *augments* them tends to persist
  - Who benefits from a working system, and who loses, decides which AI survives deployment
  - The framing carries forward to universities deploying AI against faculty

- **Augment-vs-automate (lecture-unit thread)**
  - DRCF framing: *"AI as augmentation, not automation. Enhances human judgment rather than replacing it. Requires skilled operators, not an assembly line. Success measured in better decisions, not fewer people"*
  - FRESH was framed as *automating* the planners' work. The planners rejected that framing; they did not want to be replaced
  - Contrast with MYCIN (slide 8): MYCIN was developed but *never deployed* for liability and trust reasons. FRESH was developed, deployed, and *mothballed* for stakeholder-consent reasons. Same underlying politics, different failure mode

  - The augment-not-automate lesson still holds in 2025; universities deploying AI against faculty replay the FRESH pattern

- **Week-6 callback**
  - State funding (DARPA) determined FRESH existed. Congressional redirection (anti-sub warfare) determined it would not persist
  - Week 6's thesis (state investment shapes what tech exists) applies to both deployment *and* withdrawal

- **Discussion seeding:** *"If your employer deployed an AI that did your job in hours instead of days, would you help train it?"*
  - Possible student paths:
    * *"No, that's how I get replaced"* → Fleet-planner position. Exactly what happened at FRESH
    * *"Yes, because my job will change whether I help or not"* → fatalistic/realistic. Connect to DRCF's Excel-in-1995 analogy: new skills, not replacement
    * *"Depends on whether I trust my employer to keep me employed"* → stakeholder-trust move; the real question is whose interests the deployment serves
    * *"I'd help and demand to own the outputs"* → labour-politics move; who owns the augmented work
  - Every path lands stakeholder identification: the student becomes one of the stakeholders in the scenario. That IS the pedagogy

- **Callbacks**
  - Slide 7: FRESH is what the reports-that-kill-fields thread produces when the "report" is *political* rather than academic. Congress shifted money; the field lost a working system
  - Slide 8: Feigenbaum's knowledge-rich paradigm built FRESH. Technically successful, politically shelved. The paradigm's limits weren't capability; they were stakeholder consent
  - Slide 2: the fade-to-background thesis gets a twist here. FRESH didn't fade; it was *actively killed* by the people it was supposed to serve

- **Forward-links**
  - Slide 10: TIMM 1984 was selling expert systems at the same time the Navy was shelving FRESH. The carnival-barker / on-the-ground contrast is deliberate
  - Slide 11 (end-of-80s recap): FRESH is one of three stories students carry into the bust: MYCIN (never deployed), FRESH (deployed and shelved), TIMM (sold as hype)
  - Slide 14 (Air Canada): when AI *does* get deployed into a planner-like role without stakeholder buy-in *and* without liability structure, the Air Canada case is the consequence. FRESH averted that by being shelved; Air Canada didn't

- **Budget:** ~7 min

Where is the intelligence right now?

END OF THE 1980s. SAME THREE THREADS.

- **Thread 1: notice / don't-notice**
  - What was new in 1969 is now background. Expert systems themselves are folding into plain software: XCON is DEC's parts-ordering system; MYCIN-shaped tools sit inside hospital decision support; autopilot flies the plane. Nobody calls any of this AI anymore
- **Thread 2: Amara's Law**
  - The 1980s show the short-run overestimate: Feigenbaum's *"knowledge-rich"* AI, Nilsson's 1983 *"predominant scientific endeavour"*, XCON at $40M/yr, the Japanese Fifth Generation. Bust: 1987. The long-run half of Amara's Law is invisible from inside the decade. Only from 2026 can we see expert-system descendants running tax software, airline booking, compliance engines. The tech dissolved into *"software"*. The name didn't survive
- **Thread 3: symbolic vs nonsymbolic**
  - Symbolic had the 80s: MYCIN, XCON, Japan Fifth Generation, Strategic Computing's billions. Nonsymbolic stayed quiet at IBM Raleigh / Yorktown, at SRI, in pattern-recognition conferences. The pendulum hasn't swung yet, but the weight has shifted

Jordan Harrod: "Does AI Shift Power?"

Where is the intelligence right now, at the end of 1989? Which thread carried the load?

- **Purpose of this slide**
  - Second recap. Pause. Breathe. Synthesise
  - Same shape as slide 6: three threads, re-examined against what has happened since the end of the 1960s
  - Checkpoint before we walk into the 1990s and the revival-not-revolution move at slide 12
  - Not information-delivery; light by design. Register matches slide 6 paragraph-subs
- **Where the lecture is at this point**
  - ~60 minutes in (roughly)
  - Covered since slide 6: slide 7 (Lighthill, ALPAC, the 1970s reports), slide 8 (Feigenbaum, DENDRAL, MYCIN, XCON, Kahn's DARPA, Japan Fifth Generation), slide 9 (FRESH/CASES mothballed, the stakeholders). The Carnival-barkers slide (TIMM 1984 / Cloudflare 2026 rhyme) has been moved to the end as an optional coda; may or may not get delivered depending on time

  - Next: the next slide opens with Drew McDermott's 1984 AAAI winter prediction as the segue into *"the label went radioactive"*: GOFAI retronym reveal

- **Thread 1 at end of 1989: notice / don't-notice**
  - What was bleeding-edge *new* in 1969 (ELIZA, Rosenblatt's Perceptron, early pattern recognition) is now everyday: autocorrect, spam filter, motion sensor
  - Expert systems themselves are now folding into that category. XCON is "DEC's parts ordering"; nobody calls it AI. MYCIN never shipped as MYCIN but its successor architectures run hospital decision support, insurance underwriting, tax-preparation software
  - Autopilot (slide 8 borderline case) flies the plane. Nobody calls it AI anymore
  - Slide-2 thesis holds: the moment the technology works reliably, the name comes off

- **Thread 2 at end of 1989: Amara's Law**
  - The 1980s show the short-run overestimate half *in compressed form*: the hype peaks 1983–1986 (Feigenbaum, Nilsson's AAAI Presidential Address, TIMM trade show, Japan's Fifth Generation, DARPA Strategic Computing), the bust lands 1987–1989
  - **The long-run half of Amara's Law takes longer than a decade to become visible,** which is exactly Amara's point. From inside the bust the 1980s feel like failure. From 2026, the long-run is: rule-based decision systems descended from expert systems run most of the world's business software. The tech delivered, the name didn't
  - Wait ~1 beat after reading the long-run examples (tax software, airline booking, compliance engines); these are unglamorous on purpose. That's the whole point: when it works, we stop calling it AI

- **Thread 3 at end of 1989: symbolic vs nonsymbolic**
  - Symbolic camp had the 80s. The institutional winners: Feigenbaum at Stanford, Schank, Minsky, DEC, Japan's MITI, DARPA's Strategic Computing (Kahn 1983 *"ripe for exploitation"*)
  - Nonsymbolic camp kept working but off the AI brand. IBM Raleigh moved to Yorktown (Church & Mercer 1993 chronology); SRI continued pattern recognition; IEEE PAMI conferences ran; statisticians like Breiman worked in their own venues
  - *"The weight has shifted"*: nonsymbolic had been the weaker camp from 1969 (Minsky/Papert) through the 80s; by 1989 the architects who will win the 1990s are already in place at IBM, at Bell Labs, in statistics departments
  - Slide 12 payoff: these architects don't call themselves AI people. That's the revival-not-revolution move

- **Video intermezzo: Jordan Harrod, "Does AI Shift Power?"**
  - Placed after the three recap beats, before the discussion question (same pattern as slide 6's SNL Shimmer)

  - Harrod's video asks *who benefits when AI is deployed*: stakeholder-and-power framing
  - Use it as a breath moment *and* as a preamble to the discussion question; Harrod poses the question, students answer it for 1989

- **Discussion seeding:** *"Where is the intelligence right now, at the end of 1989? Which thread carried the load?"*
  - Possible student answers and the threads they map to:
    * "In the expert systems, but they failed" → thread 2 (short-run overestimate playing out) + thread 3 (symbolic camp's decade)
    * "In the software nobody calls AI anymore" → thread 1 (notice/don't-notice) + thread 2 (long-run underestimate)
    * "In the labs that stopped calling themselves AI" → thread 3 pivot (foreshadow slide 12)
    * "Not in AI anymore, it moved somewhere else" → exactly right; slide 12 will show where
  - All plausible answers land because the threads really did overlap in 1989

- **Forward-foreshadow verbal transition to slide 12**
  - *"Three threads. By 1989 the word 'AI' is radioactive. Funding is drying up. Students are told not to put it on their CVs. And yet, watch what happens next. The work didn't stop. It just stopped being called AI."*

- **Budget:** ~5 min recap + video (~3 min) + discussion (~5–7 min). ~13–15 min total. Second major breathing point in the lecture.

The label goes radioactive

THE WORK CONTINUED, UNDER DIFFERENT NAMES.

> *"In spite of all the commercial hustle and bustle around AI these days, there's a mood that I'm sure many of you are familiar with of deep unease among AI researchers who have been around more than the last four years or so."*
>
> — Drew McDermott, AAAI *"Dark Ages of AI"* panel, 1984 (McDermott et al., 1985)

- **The winter lands (1984–1989)**
  - McDermott's 1984 panel is the warning; the bust follows. DARPA cuts funding. Lisp-machine companies collapse. Japan's Fifth Generation fades. Grad students are told to strip *"AI"* from their CVs

  - AAAI membership peaks in 1985 and falls through the 1990s. The word *"AI"* becomes something you apologise for, not something you brand a product with

- **Architects' names and venues: not *"AI"***
  - Judea Pearl publishes *Probabilistic Reasoning in Intelligent Systems*, a Bayesian-networks book in a CS department, not an AI department. Different venue, different vocabulary (Pearl, 1988)
  - At IBM Yorktown, Brown and Mercer translate languages using statistics. Breiman's statisticians publish at *NIPS* and *Machine Learning Journal*. Working names: statistical NLP, corpus-based linguistics, machine learning

If a field's name becomes a liability, what happens to the people doing the work?

- **McDermott opening blockquote: citation chain**
  - Primary: McDermott et al. (1985). Panellists: **Drew McDermott, M. Mitchell Waldrop, B. Chandrasekaran, John McDermott, Roger Schank** (panellist list (via Nilsson, 2010, p. 325, fn. 47))
  - Secondary route: §24.4 opener (Nilsson, 2010, p. 325). Haiku-resolved wording: *"motivated"* (not *"willing"*); *"I'm sure"* (not *"I think"*)
  - McDermott was on the AAAI executive council at the time; later AAAI president 1999–2000. Yale CS. Institutional voice, not dissident; the field's own leadership warning the field's own community at the field's flagship conference
  - *"More than the last four years or so"* = the researchers who remember the 1980 expert-systems boom (inside joke that dates the speaker and the audience). *"Commercial hustle and bustle"* = Kahn's DARPA Strategic Computing billions (slide 8), TIMM-brochure carnival barkers (slide 10), Japan's Fifth Generation, DEC's XCON at $40M/yr

- **Purpose of this slide**
  - Deliver beats 1 and 2 of the winter-revival arc: *the name went radioactive + the architects who kept working deliberately didn't self-identify as AI people*
  - Pair-slide with slide 13; beats 3 and 4 (revival framing + GOFAI retronym reveal) land there
  - Slide 11's forward-foreshadow (*"by 1989 the word 'AI' is radioactive"*) lands here with specifics
  - Slide 3's *"names are political"* (McCarthy picking *"artificial intelligence"* to dodge Wiener) gets its 1989 inverse: name is so politicised that researchers avoid it

- **Pearl 1988: why the title matters**
  - Foundational Bayesian-networks textbook; the mathematics that underlies most of modern machine-learning inference (Pearl, 1988)
  - Title contains *"Intelligent Systems,"* not *"AI"*. Pearl's venue was UCLA computer science, target audience CS researchers and engineers, not the AAAI crowd

  - Pearl won the Turing Award 2011 for this work. Retroactively called AI research. Was not called AI at the time

- **IBM Raleigh → Yorktown: the chronology**
  - An existing speech-recognition group was moved from Raleigh, NC to Yorktown Heights **in early 1972** (Church & Mercer (1993, p. 4)). The group worked quietly inside IBM through the 1970s and 1980s, throughout the symbolic camp's ascent and bust
  - Peter F. Brown and Robert L. Mercer led the 1988–1990 statistical machine-translation work (*Candide*). Primary source: Brown et al. (1990) *"A Statistical Approach to Machine Translation," Computational Linguistics* 16(2):79–85
  - **Brown and Mercer left IBM for Renaissance Technologies (hedge fund) in 1993.** The same statistical methods became the core of one of the most profitable trading firms in history. The marketing-term thesis on steroids: statistics renamed *"AI"* in academia, renamed *"quant finance"* in industry, same underlying technique

- **The AAAI-membership curve**
  - AAAI membership peaks at ~15,000 in 1985, falls to ~4,000 by the late 1990s. Numbers from AAAI historical records; worth speaker-note reference if a student asks how *"radioactive"* the word became institutionally

- **Verbal framing options**
  - *"Slide 10 showed the TIMM brochure selling 'expert system' in 1984. Five years later, the same word is the kiss of death on a grant application. Nothing about the underlying technology changed in those five years; just the label."*
  - *"Pearl's book has 'Intelligent Systems' in the title, which sounds like AI but isn't. Brown and Mercer's paper has 'Statistical Approach' in the title. Breiman publishes in* Statistical Science*. Nobody is lying; the work is statistical NLP, Bayesian networks, statistical modelling. But nobody is calling it AI."*

- **Discussion seeding:** *"If a field's name becomes a liability, what happens to the people doing the work?"*
  - Expected paths:
    * *"They change the name of what they do"* → exactly right. Slide 14 will name *how* they changed it (revival, not revolution)
    * *"They quit the field"* → some did (academia → finance, academia → industry; Brown and Mercer are the exemplars)
    * *"They keep calling it AI"* → no, and this is the marketing-term thesis: the label is disposable
    * *"They wait for the name to recover"* → lands via slide 15's ChatGPT re-brand (the label recovered in 2022, 35 years later)
  - Bridge to slide 14: *"So they don't call it AI. What DO they call it? That's the next slide."*

- **Budget:** ~5 min

Revival, not revolution

BOTH SIDES CLAIM THE TITLE OF AI.

- **The architects called it revival**
  - *"A resurgence of 1950s-style empirical and statistical methods."* Not a new paradigm, a revival of Shannon, Firth, Harris (Church & Mercer, 1993)
  - Algorithmic modelling had been practised *"under other names for decades."* A paradigm that narrates itself as restoration doesn't coin a revolution label (Breiman, 2001)
- **GOFAI: the retronym is the thesis in miniature**
  - Haugeland (1985) coins *Good Old-Fashioned AI*, **GOFAI**. The symbolic camp needed a name only once a rival emerged
  - Retronyms appear only when a contrast forces them: *acoustic* guitar once electric exists; *analog* watch once digital exists; *GOFAI* once statistical methods claim AI

If neither side called itself *"the winning side,"* how do we know which one won?

- **Purpose of this slide**
  - Payoff slide for the marketing-term thesis. The thesis has been running since slide 2 as *"AI is a label that gets attached and detached commercially, not earned by the engineering."* Slide 14 shows the mechanism in its 1990s form: the statistical wing inherited the AI label retroactively, without ever claiming the title at the time
  - Slide 3's GOFAI seed (bullet 2, Haugeland 1985 as retronym) pays off here with the Spider-Man reveal
- **Church & Mercer 1993: the Rationalism vs Empiricism frame**
  - Primary paper: Church & Mercer (1993)
  - Table 6 (p. 15): explicit two-column contrast between **Rationalism** (Chomsky, Minsky; Competence Model; Phrase Structure; theoretical) and **Empiricism** (Shannon, Skinner, Firth, Harris; Noisy Channel; N-grams; applied)
  - **The frame is pre-computational.** Church & Mercer are using philosophical labels from epistemology, not AI-vs-ML. That's the point: the shift was narrated as a philosophical restoration, not a technical revolution
  - Firth 1957 slogan quoted on p. 1: *"You shall know a word by the company it keeps."* The empiricist creed. Speaker-note backup if students ask what the statistical camp believed
- **Breiman 2001, *Two Cultures***

  - Primary paper: Breiman (2001)
  - Breiman (UC Berkeley statistician, creator of random forests, bagging, CART) makes the same revival move: the *"algorithmic modeling culture"* (machine learning) had been practised *"under other names for decades"* before it got named
  - Same structure as Church & Mercer (framing the shift as restoration, not revolution) but from the statistics side rather than the linguistics side

- **Halevy, Norvig, Pereira 2009: the retrospective**
  - Primary paper: Halevy et al. (2009)
  - Key quote: *"simple models and a lot of data trump more elaborate models based on less data"*
  - By 2009, the empirical side treats its win as normal. The question isn't *"did it win?"*; it's *"what's the next problem?"*
  - Norvig is Peter Norvig, co-author of the R&N textbook. The textbook-author explicitly naming his own camp as the winners *without* using the word *"revolution"*. **This is the marketing-term thesis at its cleanest:** the architects refuse the banner even when they've won

- **Haugeland 1985: the retronym coinage**
  - Primary: Haugeland (1985)
  - The book's thesis is that classical symbolic AI rests on a specific metaphysical bet: that cognition IS symbol manipulation. Haugeland coins *Good Old-Fashioned AI* to distinguish this bet from emerging connectionist and statistical approaches
  - Slide 3 introduced GOFAI as retronym (line 140, speaker notes 179–184). Slide 14 delivers the payoff: the name became necessary only because the rival camp emerged
  - R&N 4th ed. ch. 28.1 confirms GOFAI as the textbook-canonical retronym (second independent source) (Russell & Norvig, 2021)

- **Why *"revival, not revolution"* matters pedagogically**
  - Students' default model of science: paradigm shifts are loud. Kuhn's *Structure of Scientific Revolutions* is the famous story
  - What actually happened in AI 1990s: the winners denied they were fighting a paradigm war. Church & Mercer called it revival. Breiman called it continuity. Halevy/Norvig/Pereira treated it as normal
  - **The lesson:** when a field's name is contested, the architects don't name the contest. The contest gets named by retrospective historians or by rivals. That's why *"AI"* as a marketing term endures; it survives because no one on the inside wanted to claim the banner

- **Spider-Man-pointing-at-Spider-Man: the visual reveal**
  - Meme source: original 1967 *Spider-Man* episode *"Double Identity,"* two identical Spider-Men pointing accusingly at each other

  - Both panels labelled *AI*. The joke: both camps at some point claim the label, neither fully commits to it, the label floats free
  - Pedagogical function: visual compression of the argument. Students who miss the verbal argument can still grasp the meme

- **Slide 14 forward-foreshadow (deep learning resurrection)**
  - The nonsymbolic camp's 40-year exile ends at slide 14 via compute + data (Rumelhart/Hinton/Williams 1986 → LeCun → LSTM → Hinton 2006 *"deep learning"* coinage → ImageNet → AlexNet 2012 → AlphaGo 2016)
  - Hinton's 2006 re-labelling of neural nets as *"deep learning"* is the camp finally choosing to brand itself. Slide 14 will make this explicit

- **Slide 15 forward-foreshadow (transformers → ChatGPT → agents)**
  - Vaswani et al. (2017): zero uses of the word *"intelligence"* in a paper that will enable 2022 ChatGPT. Same pattern as Church & Mercer 1993 / Pearl 1988: architects don't name the label
  - Design-vs-deployment call-back to slide 5: 1943 McCulloch-Pitts → 1958 Rosenblatt; 2017 Vaswani → 2022 *"low-key research preview"* (The Verge, 2023)
  - The AI-risk discourse predates the CEOs (Grace (2026)). Nobody inside the community named themselves *"AI risk people"*; the label got reattached externally, same marketing-term pattern

- **Discussion seeding:** *"If neither side called itself 'the winning side,' how do we know which one won?"*
  - Paths students might take:
    * *"By what's deployed now"* → leads to Halevy/Norvig/Pereira 2009 evidence (simple models + lots of data)
    * *"By what the textbooks say"* → R&N 2020 treats statistical/ML methods as default; GOFAI is historical
    * *"By who got rich"* → Brown and Mercer at Renaissance Technologies; the hedge-fund route is the marketing-term thesis in its harshest form
    * *"Neither won, they merged"* → half-right; slide 14 (deep learning) combines neural-net architecture with statistical training. Modern LLMs are a hybrid
    * *"This is the Bitter Lesson"* → Sutton (2019) closer at slide 15
  - The question is genuinely open, which is why the thesis is *"AI is a marketing term."* The label attaches to whatever is currently impressive

- **Budget:** ~7 min (beats 3 + 4 + meme reveal + discussion)

Deep learning resurrects

THE NONSYMBOLIC CAMP NAMES ITSELF.

- **The exile ends: compute + data, not a better algorithm**
  - Backprop published in *Nature* (Rumelhart et al., 1986). LeCun reads handwritten ZIP codes (LeCun et al., 1989). LSTM (Hochreiter & Schmidhuber, 1997). The algorithms are done by the late 1990s
  - The 1969 *Perceptrons* critique didn't get refuted; it got absorbed once GPUs and datasets were big enough (Minsky & Papert, 1969)
- **Hinton 2006: the camp names itself**
  - Hinton's Toronto group publishes *"A fast learning algorithm for deep belief nets"* (Hinton et al., 2006). The modern term *"deep learning"* lands for multi-layer neural nets
  - Church & Mercer called their shift a *revival*; Hinton's camp chose a banner. Forty years of exile end with a word
- **The compute avalanche (2012–2016)**
  - ImageNet (Deng et al., 2009) → AlexNet (Krizhevsky et al., 2012) → AlphaGo beats Lee Sedol (Silver et al., 2016). Compute: 300,000× in six years, doubling every 3.4 months (Russell & Norvig, 2021)
  - *"General methods that leverage computation are ultimately the most effective, and by a large margin."* The compute wins, not the cleverness (Sutton (2019))

Sutton's warning: compute over algorithms. Forever true? Go look at Harness Engineering

- **Purpose of this slide**
  - The nonsymbolic side's 40-year exile (since Minsky/Papert 1969) ends here. The camp finally accepts a banner: *deep learning*
  - Contrast with slide 13 is load-bearing: the statistical wing framed its shift as revival; the connectionist wing names itself. Both end up inside the 2022 AI brand, but by different routes
  - Slide 15 follows: Transformers 2017, ChatGPT 2022, agents 2024–26. The deep-learning machinery from this slide runs inside all of them
  - Primary reading: Sutton (2019). This is the lecture's second-most-important reading after the Dartmouth proposal, per unit planning
- **Rumelhart / Hinton / Williams 1986: backprop**
  - Primary: Rumelhart et al. (1986)
  - Backprop wasn't new in 1986 (Werbos 1974, Parker 1985, Le Cun 1985); the *Nature* paper was the popularisation moment for multi-layer neural nets

  - The algorithm *existed* in the 1970s and mostly sat unused. Compute and data weren't there. The 40-year exile isn't about waiting for the algorithm; it's about waiting for the hardware

- **LeCun 1989/98: CNNs and handwriting**
  - Bell Labs, reading US Postal Service addresses: LeCun et al. (1989)
  - Canonical CNN paper: LeCun et al. (1998)
  - By the late 1990s US banks used LeCun's CNNs to read ~10% of all handwritten cheques; nonsymbolic AI quietly deployed, in infrastructure, not called AI

- **Hochreiter & Schmidhuber 1997: LSTM**
  - Primary: Hochreiter & Schmidhuber (1997)
  - Solves the vanishing-gradient problem for sequence models. Lands largely unused until the 2010s (same compute-gap as backprop). Direct ancestor of modern sequence modelling (transformers' attention mechanism solves the same problem differently)

- **Hinton 2006: the *"deep learning"* coinage**
  - Primary: Hinton et al. (2006)
  - Popularisation moment in a high-impact journal: Hinton & Salakhutdinov (2006)
  - The term *"deep learning"* has scattered earlier uses (Dechter 1986 in a different sense; Aizenberg et al. 2000 in a connectionist context). Hinton's 2006 work is the popularisation and consolidation moment. Not a clean *coinage* the way Haugeland's *GOFAI* was. Hedge accordingly in delivery

- **ImageNet 2009: the dataset that made it go**
  - Primary: Deng et al. (2009)
  - 14 million hand-labelled images across 20,000 categories. Built at Princeton/Stanford. Fei-Fei Li's insistence that data scale mattered more than algorithm sophistication: explicit Halevy et al. (2009) thesis in artefact form

- **AlexNet 2012: the moment**
  - Primary: Krizhevsky et al. (2012)
  - Won ImageNet 2012 by a huge margin (~10% absolute over the next-best non-neural approach). GPU-trained. Pattern-recognition community's symbolic approaches collapse in response

- **AlphaGo 2016: the end-of-arc moment**
  - Primary: Silver et al. (2016)
  - DeepMind's Silver beats Lee Sedol 4–1. Go was supposed to be the canary for when deep learning had arrived; Chinese players and Western GOFAI theorists had both predicted the game's combinatorial complexity would protect it for decades. Combinatorial complexity lost to compute + data

- **R&N 2020 compute curve: the 300,000× number**
  - Russell & Norvig (2021, ch. 1) documents compute growth: Moore's law through 2012 (~2× every 18 months) → 2012–2018 ~2× every 3.4 months. Total factor: ~300,000× over six years
  - Primary OpenAI source: Amodei & Hernandez (2018). R&N's textbook presentation is the secondary route
- **Sutton 2019 *Bitter Lesson*: the reading**
  - Primary: Sutton (2019)
  - Full quote: *"The biggest lesson that can be read from 70 years of AI research is that general methods that leverage computation are ultimately the most effective, and by a large margin. The ultimate reason for this is Moore's law, or rather its generalisation of continued exponentially falling cost per unit of computation."*
  - Sutton's Bitter Lesson lands here rather than at slide 7 (where Lighthill-was-right first appears) because this is where the lesson is *actually demonstrated* by the compute-beats-algorithm story
- **Callbacks**
  - Slide 5 (Perceptron + Minsky/Papert 1969): Rosenblatt's lineage resurrects here. The nonsymbolic camp's 40-year exile explicitly named at slide 5 pays off at slide 14
  - Slide 13 (revival-not-revolution): deliberate contrast; the connectionist camp names itself, unlike the statistical wing
  - Slide 10 (carnival barkers): the 2026 pitch meeting that uses all of this tech is set up here; slide 15 closes it with Vaswani / ChatGPT
- **Discussion seeding:** *"What does 'compute wins' tell us about who gets to build AI?"*
  - Expected paths:
    * *"Only big tech"* → OpenAI, Google, Anthropic, Meta. Training runs cost $10M–$1B. Universities cannot compete
    * *"Rich countries"* → US, China, UK; the Global South locked out structurally
    * *"People with cheap electricity"* → data-centre geography follows power prices; Iceland, Quebec, northern Virginia
    * *"Governments regulate whoever has the compute"* → cf. Harrod video at slide 11 (*"Does AI Shift Power?"*)
    * *"Nobody, we're just renting from NVIDIA"* → the compute layer is itself a chokepoint; NVIDIA's 2024–26 market cap history demonstrates
  - All paths land on concentration of power. Bridge to slide 15 (ChatGPT arrival + Weizenbaum returns) + slide 16 (Bender/Gebru, Air Canada, Sutton closer) + week 8 (parasocial + modern communication tech)
- **Budget:** ~6–7 min

ELIZA returns

1966 → 2022: 56 YEARS LATE.

- **The Transformer, and the five-year gap**
  - *"Attention Is All You Need"* (Vaswani et al. (2017)). Zero uses of the word *"intelligence"* in the paper. The architecture that powers 2022 LLMs has no claim on *"AI"* at its birth
  - OpenAI execs call ChatGPT (November 2022) a *"low-key research preview"* (The Verge, 2023). 100M users in two months. Paper deployment
- **ELIZA at planetary scale**
  - Weizenbaum 1966: the secretary knew DOCTOR was code. Still asked him to leave the room. Amara's Law set the clock: 56 years from warning to planetary arrival
  - 2026: the chatbot sits in bedrooms, classrooms, boardrooms. Parasocial attachment at scales Weizenbaum could not have imagined

When does a warning actually arrive: when it's given, or when we're ready to hear it?

- **Purpose of this slide**
  - Payoff for the Amara's Law thread seeded at slide 4 (Weizenbaum 1976 *"ought not"*, the secretary anecdote) and re-named at slide 6 recap 1. ELIZA's 1966 warning finally lands in 2022–2026, 56 years late
  - Slide 5 design-vs-deployment structural rhyme (1943 McCulloch-Pitts → 1958 Rosenblatt = 15 years) paid back here (2017 Vaswani → 2022 ChatGPT = 5 years). Paper deployment; the gap is where the marketing-term thesis lives
  - Pair-slide with slide 16 (the critique lands + closer). Slide 15 is the arrival; slide 16 is the bill
- **Vaswani et al. 2017: the Transformer paper**
  - Primary: Vaswani et al. (2017)
  - Eight authors, all at Google Brain or Google Research. Written to improve machine translation, not to build AGI
  - **Zero uses of the word *"intelligence"*** in the 15-page paper. Marketing-term thesis in purest form: the paper that enables the 2022 AI brand has no claim on the AI brand at its own birth
  - Introduces the Transformer architecture (attention mechanism + encoder-decoder). Supplants RNNs/LSTMs for sequence modelling. Everything from GPT-2 onward is a Transformer variant
- **ChatGPT November 2022: the *"low-key research preview"***

  - OpenAI launched ChatGPT on 30 November 2022. Internally framed as a low-stakes demo of GPT-3.5 with RLHF (reinforcement learning from human feedback)
  - OpenAI execs in sworn testimony and leaked internal communications called it *"a low-key research preview"* (The Verge, 2023)
  - Product hit 1M users in 5 days; 100M in 2 months (fastest-adopted consumer product in history until Threads, 2023)
  - The design-vs-deployment gap is 5 years (2017 → 2022), compressed relative to 1943 → 1958 (15 years). Compute is faster but the marketing-term pattern holds: the engineers did not name the commercial product

- **Weizenbaum 1966: full return**
  - Slide 4 established the secretary anecdote (Weizenbaum, 1967): his own secretary, knowing the program was code, still wanted Weizenbaum out of the room while she *"talked"* to DOCTOR
  - Slide 4 also established Weizenbaum's 1976 *"ought not"* (Weizenbaum, 1976, p. 39; via Nilsson, 2010, p. 315): *"there are certain tasks which computers ought not be made to do, independent of whether computers can be made to do them"*
  - The 56-year gap (1966 CACM paper → 2022 ChatGPT) is Amara's Law in its sharpest form. Short-run: ELIZA was a toy curiosity in 1966. Long-run: parasocial attachment at planetary scale in 2026

- **2026 scale notes**
  - ChatGPT weekly active users (as of early 2026): ~800M+ per OpenAI disclosures. Meta AI via WhatsApp/Instagram: ~1B+ monthly users exposed to AI assistants. Google Gemini integration in Search: every search query
  - *"Bedrooms, classrooms, boardrooms"*: concrete placements matter pedagogically. First-year students will have personal ChatGPT usage; they've seen friends use it for emotional support, studying, relationship advice

- **Verbal framing options**
  - *"Slide 4 was Weizenbaum, 1966. His secretary asks him to leave the room so she can 'talk' to the computer. Fifty-six years later, millions of people are doing exactly what Weizenbaum's secretary was doing. The warning was right; the scale was unimaginable."*
  - *"The Transformer paper, the paper that makes ChatGPT possible, has zero uses of the word 'intelligence.' The engineers were writing about machine translation. The product was named by OpenAI's marketing team. The label got reattached commercially."*

- **Discussion seeding:** *"When does a warning actually arrive: when it's given, or when we're ready to hear it?"*
  - Paths:

    * *"When we're ready to hear it"* → Amara's Law confirmed. The warning was right in 1966; we couldn't hear it yet
    * *"When the harm becomes visible"* → pushes toward slide 16's critique material (Bender/Gebru, Air Canada)
    * *"When it's given; we just didn't listen"* → stronger claim. Weizenbaum was clear; we chose not to act
    * *"Warnings don't arrive; they accumulate"* → sophisticated answer on institutional accumulation

- **Bridge to slide 16**
  - *"The warning lands. Some people named it: Weizenbaum, then Bender, then Gebru. What happens to the people who name the warning?"*

- **Budget:** ~5 min

The bill arrives

THE WARNING LANDS.

- **The critique, and the field closing ranks**
  - *"On the Dangers of Stochastic Parrots"* names the mechanism: statistical outputs, no understanding; environmental cost; bias baked in (Bender et al., 2021)
  - Google fires Gebru December 2020, Mitchell February 2021. Field-founder dismisses field-critic: McCarthy 1976 → Google 2020

- **Brittleness meets liability**
  - Air Canada 2024: chatbot promised a bereavement refund. Airline argued *"it spoke on its own."* The tribunal ruled against the airline (BBC Travel, 2024; British Columbia Civil Resolution Tribunal, 2024). The pitch meets the courtroom

- **Sutton's closer**
  - *"General methods that leverage computation win, by a large margin"* (Sutton (2019)). The decision becomes infrastructure politics: who owns the compute, who pays the externalities

Simmons, 1961: *"a conversation with a book."* Today: yes, and the book talks back, at planetary scale.

Who owns the book?

- **Purpose of this slide**
  - Final substantive slide. Three beats close three threads:

    - Beat 1 closes the drumbeat (field-founder-dismisses-critic) running from slide 3 (McCarthy picking the name) through slide 4 (McCarthy 1976 dismissing Weizenbaum) to 2021 (Google dismissing Gebru)
    - Beat 2 closes the brittleness thread (deployed systems fail in predictable ways; the cost lands on users)
    - Beat 3 closes the infrastructure-politics thread (Sutton's Bitter Lesson = compute wins = who owns the compute decides who builds AI)
  - Final line echoes slide 1 Simmons. Discussion question bridges to week 8 (modern communication technologies, parasocial relationships, ownership of AI as infrastructure)

- **Bender et al. 2021, *Stochastic Parrots***
  - Primary: Bender et al. (2021)
  - Four arguments: (1) environmental + financial costs concentrate on those who don't benefit; (2) bias baked in from training-data demographics; (3) research-opportunity cost from chasing scale; (4) risk of harm from text generated without communicative intent
  - Coins the *"stochastic parrot"* metaphor: language models are patterns without meaning

- **Gebru firing chronology**
  - December 2020: Timnit Gebru forced out of Google after conflict over *Stochastic Parrots* draft. Then-Google-SVP Jeff Dean demanded paper retraction or author withdrawal; Gebru refused; Google accepted her *"resignation"* she had not offered
  - February 2021: Margaret Mitchell fired from Google. Also a *Stochastic Parrots* co-author. Publicly tied to her own internal investigation of Gebru's firing
  - Bender (University of Washington, tenured) and McMillan-Major (then PhD) retained academic freedom; Gebru and Mitchell (industry, unprotected) did not
  - **Drumbeat close:** slide 3 set up McCarthy picking *"artificial intelligence"* to control who was in the room; slide 4 set up McCarthy 1976 *"no argument is offered"* dismissing Weizenbaum; this slide completes the arc with Google 2020 dismissing Gebru. 70 years. Same pattern

- **Air Canada 2024, *Moffatt v Air Canada***
  - Primary: British Columbia Civil Resolution Tribunal (2024) (CRT File No: SC-2023-005609, 14 February 2024)
  - Customer Jake Moffatt asked the airline's chatbot about bereavement fares; chatbot told him he could book the full fare and claim a refund after the fact. Policy said the opposite
  - Airline argued the chatbot was *"a separate legal entity that is responsible for its own actions"*
  - Tribunal rejected that argument: the airline is responsible for information provided on its own website. Ordered refund plus costs

  - Textbook slide-10 payoff (carnival-barker pitch meets actual user / actual court). If slide 10 was delivered, this closes that loop. If slide 10 wasn't delivered, this lands on its own as the concrete 2026 example of deployment brittleness
  - BBC reporting: BBC Travel (2024)

- **Sutton 2019 *Bitter Lesson*: the reading**
  - Already fully cited in slide 14 speaker notes. Referenced here as the closing move
  - Why Sutton lands here rather than at slide 14: slide 14 demonstrated the lesson historically (deep learning resurrection via compute); slide 16 makes the *political* move. If compute wins, and compute costs $1B+ per frontier training run, the decision about what AI exists is a decision about who owns compute
  - Infrastructure politics: data centres, electricity grids, NVIDIA, Taiwan (TSMC), US/China geopolitical competition. The Bitter Lesson is a warning about power concentration

- **Simmons 1961 rhyme: closing the loop**
  - Slide 1's epigraph: Simmons's SDC Synthex grant (1961) proposed *"a conversation with a book"*, the sentence that opens the lecture
  - Slide 16 closes the lecture by answering the question the epigraph posed: yes, we built the conversation; yes, it arrived; the book talks back
  - *"The book talks back"* does two things: (1) confirms the Simmons 1961 dream is realised; (2) names the parasocial payoff (Weizenbaum's secretary, now at planetary scale)

- **Discussion:** *"Who owns the book?"*
  - Genuine open question. Possible answers:
    * *"OpenAI / Google / Anthropic"* → compute + data concentration thesis (slide 14 Sutton lands)
    * *"Whoever owns the training data"* → copyright fights 2023–26 (NYT v OpenAI, authors' guild cases); Common Crawl and its scrapes of the open web
    * *"Nobody, it's trained on humanity"* → democratic-commons framing; popular but overstates the commons
    * *"Us, we write the prompts"* → strong answer; pushes toward human-AI collaboration framings
    * *"Governments, eventually"* → EU AI Act 2024, California SB-1047 debates, sovereign AI programmes
  - Week 8 picks up: parasocial relationships, modern communication technology, the platform question

- **Speaker-note material NOT on-slide (moved per Brian's cut)**
  - R&N ch.28.1 (GOFAI textbook confirmation): already covered at slide 13
  - R&N ch.28.2 (Dijkstra submarines-swim, Chinese Room): philosophy off the marketing-term-thesis axis

  - R&N ch.28.3 (Petrov 1983, third-revolution-in-warfare, Buolamwini/Gebru bias stats): rhymes with slide 9 but not thesis-load-bearing
  - Grace (2026): AI-risk-discourse-predates-CEOs; overlaps with the Vaswani zero-intelligence move, and Bender/Gebru covers the critique-authorship point more directly
  - Hinton leaves Google 2023: ornamental rhyme; optional verbal move
  - If delivery time permits, any of these can be woven in verbally; none are load-bearing

- **Bridge to week 8**
  - *"Next week: parasocial relationships, modern communication technology. The book talks back, at planetary scale, in every bedroom, to every teenager. What does that do to us?"*

- **Budget:** ~6–7 min (beats + final line + discussion)

Notes and References

This deck was built with the help of Claude Opus 4.7. Transcripts and tooling available on request. Rendered with Quarto. It is licensed with a Creative Commons 4.0 International – With Atribution License.

Amodei, D., & Hernandez, D. (2018). *AI and compute*. OpenAI blog post. https://openai.com/index/ai-and-compute/

BBC Travel. (2024). *Air canada chatbot misinformation: What travellers should know*. https://www.bbc.com/travel/article/20240222-air-canada-chatbot-misinformation-what-travellers-should-know

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? . *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, FAccT '21*, 610–623. https://doi.org/10.1145/3442188.3445922

Breiman, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science*, *16*(3), 199–231. https://doi.org/10.1214/ss/1009213726

British Columbia Civil Resolution Tribunal. (2024). *Moffatt v air canada*.

Church, K. W., & Mercer, R. L. (1993). Introduction to the special issue on computational linguistics using large corpora. *Computational Linguistics*, *19*(1), 1–24. https://aclanthology.org/J93-1001/

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255. https://doi.org/10.1109/cvpr.2009.5206848

Dick, S. (2019). Artificial intelligence. *Harvard Data Science Review*, *1*(1). https://doi.org/10.1162/99608f92.92fe150c

Dreyfus, H. L. (1972). *What computers Can't do: A critique of artificial reason*. Harper & Row.

Dreyfus, H. L., & Dreyfus, S. E. (1986). *Mind over machine: The power of human intuition and expertise in the era of the computer*. Free Press.

Feigenbaum, E. A. (1977). The art of artificial intelligence: Themes and case studies of knowledge engineering. *Proceedings of the 5th International Joint Conference on Artificial Intelligence (IJCAI-77)*, 1014–1029.

Grace, K. (2026). *AI risk was not invented by AI CEOs to hype their companies*. Substack (*world spirit sock stack*). https://worldspiritsockpuppet.substack.com/p/ai-risk-was-not-invented-by-ai-ceos

Halevy, A., Norvig, P., & Pereira, F. (2009). The unreasonable effectiveness of data. *IEEE Intelligent Systems*, *24*(2), 8–12. https://doi.org/10.1109/MIS.2009.36

Haugeland, J. (1985). *Artificial intelligence: The very idea*. MIT Press.

Hinton, G. E., Osindero, S., & Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural Computation*, *18*(7), 1527–1554. https://doi.org/10.1162/neco.2006.18.7.1527

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, *313*(5786), 504–507. https://doi.org/10.1126/science.1127647

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hutchins, W. J. (1995). Machine translation: A brief history. In E. F. K. Koerner & R. E. Asher (Eds.), *Concise history of the language sciences: From the sumerians to the cognitivists* (pp. 431–445). Pergamon Press. http://www.hutchinsweb.me.uk/ConcHistoryLangSci-1995.pdf

Kanal, L. N. (Ed.). (1968). *Pattern recognition: Proceedings of the IEEE workshop on pattern recognition, held at dorado, puerto rico*. Thompson Book Co.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems 25*, 1097–1105.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural Computation*, *1*(4), 541–551. https://doi.org/10.1162/neco.1989.1.4.541

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324. https://doi.org/10.1109/5.726791

Lighthill, J. (1972). *Artificial intelligence: A general survey*. Science Research Council of Great Britain.

Martin, W. L. (1986). *An assessment of artificial intelligence and expert systems technology for application to the management of cockpit systems* (AAMRL-TR-86-040). Harry G. Armstrong Aerospace Medical Research Laboratory, Air Force Systems Command. https://apps.dtic.mil/sti/citations/ADA175456

McCarthy, J. (1976). An unreasonable book [review of weizenbaum, *computer power and human reason*]. *ACM SIGART Bulletin*, (58), 5–6. https://doi.org/10.1145/1045264.1045265

McCarthy, J. (2000). *Review of "the question of artificial intelligence" edited by brian bloomfield*. https://www-formal.stanford.edu/jmc/reviews/bloomfield/bloomfield.html

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1955). *A proposal for the*

Dartmouth summer research project on artificial intelligence. Dartmouth College. http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf

McCorduck, P. (1979). *Machines who think: A personal inquiry into the history and prospects of artificial intelligence*. W. H. Freeman; Co.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, *5*(4), 115–133. https://doi.org/10.1007/BF02478259

McDermott, D., Waldrop, M. M., Chandrasekaran, B., McDermott, J., & Schank, R. (1985). The dark ages of AI: A panel discussion at AAAI-84. *AI Magazine*, *6*(3), 122–134. https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/494

Michie, D. (1982). *Machine intelligence and related topics: An information scientist's weekend book*. Gordon; Breach Science Publishers.

Minsky, M., & Papert, S. (1969). *Perceptrons: An introduction to computational geometry*. MIT Press.

Newell, A. (1980). AAAI president's message. *AI Magazine*, *1*(1), 1–4. https://doi.org/10.1609/aimag.v1i1.84

Nilsson, N. J. (2010). *The quest for artificial intelligence: A history of ideas and achievements*. Cambridge University Press.

OpenAI. (2026). *Enterprises power agentic workflows in Cloudflare agent cloud with OpenAI*. https://openai.com/index/cloudflare-openai-agent-cloud/

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Morgan Kaufmann.

Pierce, J. R., & Automatic Language Processing Advisory Committee. (1966). *Language and machines: Computers in translation and linguistics*. National Academy of Sciences / National Research Council.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536. https://doi.org/10.1038/323533a0

Russell, S. J., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4th ed.). Pearson.

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Driessche, G. van den, Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., & Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, *529*(7587), 484–489. https://doi.org/10.1038/nature16961

Sutton, R. (2019). *The bitter lesson*. http://www.incompleteideas.net/IncIdeas/BitterLesson.html

The Verge. (2023). *OpenAI execs dubbed ChatGPT a low-key research preview*. https://www.theverge.com/2023/12/5/23989871/openai-execs-dubbed-chatgpt-a-low-key-research-preview

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems 30*, 5998–6008. https://arxiv.org/abs/1706.03762

Weizenbaum, J. (1966). ELIZA — a computer program for the study of natural language

communication between man and machine. *Communications of the ACM*, *9*(1), 36–45. https://doi.org/10.1145/365153.365168

Weizenbaum, J. (1967). Contextual understanding by computers. *Communications of the ACM*, *10*(8), 474–480. https://doi.org/10.1145/363534.363545

Weizenbaum, J. (1976). *Computer power and human reason: From judgment to calculation*. W. H. Freeman.

Wiener, N. (1948). *Cybernetics: Or control and communication in the animal and the machine*. MIT Press.

Coda: Carnival barkers

EXPERT SYSTEMS THEN, AGENTS NOW

> *"We've built a better brain … Expert systems reduce waiting time, staffing requirements and bottlenecks caused by the limited availability of experts. Also, expert systems don't get sick, resign, or take early retirement."*

TIMM brochure, AAAI trade show, 1984 (via George Johnson, (Nilsson, 2010, p. 272))

> *"Enterprises can now deploy agents … to perform real work. For example, companies can use it with OpenAI to deploy agents that automatically handle tasks like responding to customers, updating systems, and generating reports, all within a secure, production-ready environment."*

Cloudflare / OpenAI press release 13 April 2026

One is from 1984. One is from 2026. Which is which, and how can you tell?

- **This slide is an optional coda.** Brian does not plan to reach it during the lecture itself. If time permits, the TIMM 1984 / Cloudflare 2026 rhyme is the thesis made concrete: the marketing-term thesis demonstrated via two artefacts 42 years apart. If time does not permit, the thesis has already landed via slides 12 (McDermott radioactive label) + 13 (GOFAI retronym reveal) + 14 (Hinton names the camp) + 15 (Vaswani zero-intelligence) + 16 (Bender critique + Sutton closer). This slide is a reinforcement, not a load-bearing piece.
- **Sources**
  - George Johnson passage with TIMM brochure embedded (verbatim). Johnson's original reportage in Nilsson's footnote 2; citation body not in photo frame (flag for hardcopy verification) (Nilsson, 2010, p. 272)
  - Nilsson's *"sprinkle a little AI"* autobiographical note on 1980s proposal-writing practice (Nilsson, 2010, p. 272)
  - Kahn 1983 *"ripe for exploitation"* framing (another 80s carnival-barker voice, government side) (Nilsson, 2010, p. 286)

  - Cloudflare/OpenAI press release, 13 April 2026: OpenAI (2026)
  - Buchanan-sourced AAAI 1984 trade-show photograph: Nilsson Fig. 21.1, courtesy Bruce Buchanan via *AI Magazine* Vol. 28, No. 4, p. 14 (2005)

- **Full George Johnson passage (Nilsson p. 272, verbatim)**
  - *"Reporting on this increasing interest in 1984, the science writer George Johnson wrote: 'We've built a better brain,' exclaimed a brochure for [an expert system called] TIMM, 'The Intelligent Machine Model.' 'Expert systems reduce waiting time, staffing requirements and bottlenecks caused by the limited availability of experts. Also, expert systems don't get sick, resign, or take early retirement.' Other companies, such as IBM, Xerox, Texas Instruments, and Digital Equipment Corporation, were more conservative in their pronouncements. But the amplified voices of their salesmen, demonstrating various wares [in the 1984 AAAI trade hall], sounded at times like carnival barkers, or prophets proclaiming a new age."*
  - On-slide quote trims to the TIMM brochure text itself; *"carnival barkers, or prophets proclaiming a new age"* lives in speaker voice, not on-slide

- **Sprinkle-a-little-AI (Nilsson p. 272, Nilsson autobio)**
  - Verified verbatim; earlier PC gap (*"more ??? to government sponsors"*) resolves to *"more enticing to government sponsors"*
  - *"Mainly, I thought, they wanted us to 'sprinkle a little AI' on their proposed projects to make them more enticing to government sponsors."*
  - Nilsson is the SRI AI Center director. He is describing 1980s proposal-writing practice from the inside
  - Direct rhyme with 2026 *"AI-native"* branding and the generic *"slap AI on it"* move in current enterprise marketing
  - Deploy verbally when a student points out the 2026 quote is transparently marketing

- **Martins 1984: the contemporaneous inside-critique voice**
  - Martins, L.-F. (1984), former director of RAND Corporation's R&D Program in Information Processing Technology (until 1982), writing from inside the US defence-research establishment, same year as the TIMM trade-show brochure
  - Cited in (Martin, 1986, pp. 56–57)
  - Martins's six factors accounting for expert-systems *"success"*: *"brilliant programmers, the very narrowly defined and/or easy problems worked, generous funding over a long period of time, luck, development of customized tools (without the use of shells) and finally,* ***misleading advertising as to the utility and intelligence of some of the programs****"*
  - 1984 was the year of *both* the TIMM brochure *and* a former-RAND-director publishing the marketing-term critique in print
  - Deploy verbally if a student suggests the *"AI is a marketing term"* critique is only modern hindsight. Martins named it from inside the defence-research establishment, on the record, forty-two years before the 2026 Cloudflare/OpenAI press release

  - **SLIDE 2 CALLBACK:** 1984 primary-source voice confirming the slide 2 thesis

- **Kahn 1983 (Nilsson p. 286)**
  - *"Robert Kahn and the architects of SC believed in 1983 [after the expert systems boom] that AI was ripe for exploitation. It was finally moving out of the laboratory and into the real world … AI would become an essential component of SC; expert systems would be the centerpiece."*
  - Same decade, government side. The carnival barking wasn't only commercial
  - Deploy if discussion pushes back on the commercial-only framing

- **Full Cloudflare/OpenAI press release context**
  - 13 April 2026, OpenAI Global Affairs section
  - Named voices: Dane Knecht (Cloudflare CTO), Rohan Varma (OpenAI Codex product lead)
  - Customer list: Accenture, Walmart, Intuit, Thermo Fisher, BNY, State Farm, Morgan Stanley, BBVA (incumbent enterprise services, not AI-native startups)
  - Specific quantified claims: *"1 million business customers"*, *"3 million weekly active users use Codex"*, *"15 billion tokens per minute"* through OpenAI APIs
  - These specifics do the same rhetorical work as the TIMM brochure's *"reduce waiting time, staffing requirements"*: quantified capability-plus-cost-savings pitch

- **The rhyme made explicit (speaker move, not on-slide)**
  - *"don't get sick, resign, or take early retirement"* *"automatically handle tasks … in a secure, production-ready environment"*
  - Both pitches are labour-replacement framed as capability
  - *"Expert systems reduce waiting time"* *"deploy agents … to perform real work"*
  - *"The Intelligent Machine Model"* *"GPT-5.4"*: branded intelligence as a product
  - The genre hasn't changed. The label has

- **Pedagogical move**
  - Do NOT identify which quote is which before asking the discussion question
  - Let students try to guess. Most will get it right, but the *way* they get it right is the lesson: which markers did they use?
  - *"agents"*, *"GPT-5.4"*, *"production-ready"* read as 2026; *"expert systems"*, *"don't get sick"*, the explicit labour-replacement framing read as 1984
  - Then ask: given how similar the pitches are, what has actually changed?

- **Discussion seeding:** *"One is from 1984. One is from 2026. Which is which, and how can you tell?"*
  - Possible student paths:
    * *"2026 mentions AI agents"* → label has moved from expert-systems to agents. Slide 2 marketing-term thesis paying back

    * *"1984 talks about not getting sick, which feels older"* → 1984 pitch is explicit about labour replacement; 2026 uses euphemism (*"real work"*). Euphemism as evolution in marketing
    * *"Both sound the same"* → that IS the point. Slide 6 Thread 2 (Amara's Law). The promises keep coming back
    * *"GPT-5.4 gives it away"* → brand names date the pitch; the structure doesn't
  - Any answer lands the thesis because the rhyme is genuine

- **Visual assets (still open, per RESUME.md)**
  - Option A (current draft): text blocks side-by-side
  - Option B (upgrade): Nilsson Fig. 21.1 AAAI 1984 trade-show photo (scan needed) + Cloudflare/OpenAI press release screenshot (capture needed)
  - Option B hits harder: visual era-markers (80s trade-show fashion vs modern web design) amplify the rhyme
  - Can upgrade after initial render

- **Callbacks**
  - Slide 2 (marketing term): this slide is the thesis's peak payoff. Same pitch, 42 years apart, different label
  - Slide 6 Thread 2 (Amara's Law): the promise keeps coming back
  - Slide 8 (Feigenbaum): expert-systems boom hype at its peak visual manifestation
  - Slide 9 (FRESH): while the salesmen were at AAAI selling the pitch, the Navy was shelving the actual expert system. Trade-show pitch vs on-the-ground reality

- **Forward-links**
  - Slide 11 (end-of-80s recap): peak 1984 hype → late-80s bust. Amara's Law playing out on the 1980s timescale
  - Slide 14 (Air Canada): the 2026 pitch meets actual liability. The carnival barker sells it; the courtroom tests what was delivered
  - Slide 14 (Bender et al., *Stochastic Parrots*): the academic voice that names what the carnival barkers are doing

- **Budget:** ~6 min (lighter prose; discussion typically runs longer because students try multiple framings)