

The Adoption-Impairment Paradox: General-Purpose GPUs, ASIC Substitution, and Generational AI Adoption

Mark Uni

Pufferfish Artificial Intelligence Research Group

Disclosure. Mark Uni is a researcher at Pufferfish Artificial Intelligence Research Group. He has no employment or consulting relationship with Oracle Corporation, and he holds no position in Oracle Corporation (ORCL) securities. The *Oracle Short Report* is a separate paid product prepared by other members of the group using the results of this research; the author does not prepare or sell that report. This paper contains no investment recommendation, no target price, no options combination, no stop-loss and no trading advice.

Companion papers. This paper is one of three companion studies of the same investment episode. It cross-cites the companion finance paper, *Debt-Financed AI Expansion and Shareholder Value at Oracle* (prepared in the format of *The Journal of Finance*), the companion geopolitics paper on oil, the Strait of Hormuz, inflation and Treasury crowding, and the paid investor volume, the *Oracle Short Report*. Core figures are shared across the four documents: capital expenditure of \$55.7 billion, free cash flow of -\$23.68 billion, a total claim stack of \$457.1 billion, uncommenced leases of \$288.0 billion, a five-year credit default swap spread of 261 basis points (an 8 October 2026 print; later reported prints were higher - see the companion finance paper - and not a current quote), an author-illustrative incremental gross margin of 34% (an assumption, not a figure disclosed in Oracle's filings) against a historical 72%, and a sum-of-the-parts value of \$95.00 per share in the finance paper. This paper asks the technology question.

CRedit authorship contribution statement. Mark Uni: Conceptualization, Methodology, Formal analysis, Investigation, Writing - original draft, Writing - review and editing.

Acknowledgements. The author thanks the members of the Pufferfish Artificial Intelligence Research Group for discussion of the adoption timelines and the impairment arithmetic. All errors are the author's own. The companion finance and geopolitics papers, and the paid investor report, are cited above on the title page; figures shared across the four documents are drawn from the public filings and market data documented in the finance paper.

Corresponding author. Mark Uni, Pufferfish Artificial Intelligence Research Group. E-mail address to be supplied by the author at submission.

Word count. 9,867 words in total, counted as main text plus references, excluding this title page and the running footers (limit for Research Articles: 10,000 words).

Highlights

- AI hardware faces an adoption-impairment paradox in both success and failure
- GPU flexibility options unfixed algorithms; ASICs win once workloads stabilise
- A \$457.1 billion claim stack needs twenty years; two years is impossible
- Four-track supply and grid queues make price competition structural
- Impairment can strike unused hardware: hoarding activates a second channel

Highlights character counts, including spaces (limit 85 each): 76; 78; 72; 67; 74.

Abstract

This paper identifies and formalises an adoption-impairment paradox in artificial intelligence (AI) infrastructure investment. General-purpose graphics processing units (GPUs) are valuable precisely because AI algorithms are not yet fixed: flexibility is an option on algorithmic uncertainty. If research progresses and algorithms stabilise, the same workloads run on custom application-specific integrated circuits (ASICs) - the path taken by the Google Tensor Processing Unit (TPU) and the custom silicon programmes of Broadcom and Marvell - at lower cost and lower power, and the general-purpose fleet is impaired by substitution. If research does not progress and AI does not land at scale, demand never arrives, and the same fleet is impaired by under-utilisation. Only an unmeasured corridor - progress fast enough to create demand, slow enough that installed hardware remains relevant, differentiated enough that competitors do not commoditise price - escapes impairment on this hypothesis, and such a corridor cannot be a default accounting assumption over an assumed useful life of six to ten years. Combining diffusion literature with two channels of impairment - time-based depreciation and utilisation or obsolescence impairment, which can strike unused hardware - the paper derives adoption timelines as a stress test, not a forecast. Applied to a \$457.1 billion claim stack, a two-year landing is arithmetically impossible, a five-year landing fails, and only a twenty-year horizon, close to the fifteen-to-nineteen-year tenor of the underlying leases, approaches plausibility. Supply expanding on four tracks, and a deployment clock set by grid queues, permits and consent, lengthen the wait.

Abstract word count: 248 words (limit 250).

Keywords: technology adoption; diffusion; ASIC; asset impairment; useful life; innovation policy

1. Introduction: investors price a story, not a technology

Many investors in the current AI infrastructure boom do not evaluate the technology. They evaluate a story about the technology: that a general-purpose technology has arrived, that demand for computation is effectively unlimited, that capacity built today will be absorbed tomorrow, and that the assets installed to serve it will earn their cost of capital over long, comfortable useful lives. Each element of the story may eventually prove true. The argument of this paper is that the story, as priced, omits the variable that decides whether the investment succeeds - time, and what time does to the assets while adoption is still on its way.

The omission matters because the investment is no longer small, reversible or equity-financed at the margin. In the case studied across the companion papers, a single firm has committed to a claim stack of \$457.1 billion, including \$288.0 billion of leases not yet commenced, running fifteen to nineteen years, while spending \$55.7 billion on capital expenditure in a single year and reporting free cash flow of -\$23.68 billion. The market already prices part of the risk: the firm's five-year credit default swap spread printed at 261 basis points on 8 October 2026 (higher prints followed; companion finance paper). What credit markets price as a probability, this paper examines as a technological mechanism, which belongs in a journal of innovation studies, not only in a journal of finance.

The research question is distinct from the companion finance paper's. That paper asks whether debt-financed AI expansion destroys shareholder value, and answers with accounts, leases, credit spreads and valuation. This paper asks a prior question: why does AI hardware investment face an adoption-impairment paradox, and how do ASIC substitution, competition and generational adoption determine the value and useful life of AI hardware? The question is prior because the finance paper's valuation rests on assumptions - about useful life, utilisation and the margin new capacity will earn - that are technological assumptions in disguise. If they fail for technological reasons, the valuation fails with them.

Five claims organise the paper. First, adoption of a general-purpose technology is generational, not a step change: it proceeds along an S-curve whose slope is set by habits, trust, skills, regulation and routines, which turn over with cohorts rather than product cycles (Section 2). Second, the hardware bought to serve that adoption carries a substitution risk endogenous to technological progress itself: GPUs are the right tool while algorithms are unfixed, and the wrong tool once they are fixed, because custom ASICs then perform the stabilised workload more cheaply (Section 3). Third, competition ensures that capacity built by every firm arrives at once and is priced against itself, compressing the margin on which break-even arithmetic depends (Sections 4 and 5). Fourth, deployment runs on a slower clock than construction (Section 6). Sections 7 and 8 convert these claims into arithmetic - timelines as a stress test, impairment as a two-channel process. Section 9 states the paradox formally; Sections 10-13 give the historical analogy, policy and management implications, discussion and conclusion.

A note on method. The timelines in Section 7 are scenarios, not forecasts. They answer: if AI lands in two, five, ten or twenty years, what annual profit and revenue would the committed capital require in the meantime? A stress test only has to show which futures the balance sheet can survive.

2. Adoption is generational: the S-curve cannot be financed away

The diffusion of innovations is among the most replicated findings in the social sciences. From Rogers (1962) onward, adoption follows an S-shaped curve: a slow beginning among innovators and early adopters, a steep middle as the innovation crosses into the early and late majority, and a long flattening tail (Rogers, 1962; Bass, 1969). Moore (1991) added that the curve contains a chasm between early adopters and the pragmatic majority, who buy only when a technology is proven, supported and safe - the point at which many general-purpose technologies spend years, and some a generation.

The curve has this shape because adoption is not a purchase decision repeated many times; it is a change of practice. A firm does not adopt AI by signing a contract. It adopts AI by redesigning workflows around it, retraining or replacing staff, renegotiating liability for its errors, satisfying regulators, and persuading the people who must trust its outputs to trust them. Each step is carried out by people whose habits were formed earlier, so habits, trust, skills, regulation and routines change at the speed of cohort replacement. That is a generational clock, not a product cycle, and it does not accelerate because a supplier has borrowed money.

History supports the generational reading. Electrification took decades to reorganise the factory, because the productivity of the electric motor arrived only when factories were redesigned around unit drive rather than the line shaft it replaced. Railways, the automobile and the commercial internet each passed through a construction phase, a collapse, and then a long period of institutional absorption during which the technology delivered what its promoters had promised - frequently to a later cohort of owners, not to the investors who financed the construction (Perez, 2002). As a scale illustration only, the Smithsonian Institution's museum account describes ENIAC as occupying 30 by 50 feet, weighing 30 tons and using about 18,000 vacuum tubes, whereas a modern phone chip packs billions of transistors into a fingernail-sized die (Smithsonian Institution, n.d.). This is the pattern behind the aphorism that history does not repeat itself but rhymes, a line commonly attributed to Mark Twain, though no verified Twain source exists: the technology succeeds, the pioneers' capital does not, and the difference between the two outcomes is the adoption lag.

Two implications follow. First, the adoption horizon for AI infrastructure should be set by analogy to general-purpose technologies - five years as an optimistic lower bound, ten years as a central case, twenty years as the generational case - not by the two-year horizon implicit in market commentary, under which capacity installed today is assumed to be absorbed immediately. Section 7 retains the two-year case to show that it fails on arithmetic grounds as well. Second, the timing is asymmetric. Capital is raised on financial time - a bond is sold in a day, a lease signed in a quarter - while adoption proceeds on generational time. Debt is a claim on future time; adoption is a claim on future behaviour, and behaviour changes more slowly than debt matures. That mismatch, not any judgement about the merit of AI, is the subject of this paper.

3. General-purpose GPUs are an option on unfixed algorithms

To see why the installed hardware is exposed, consider why it was bought. Training and running AI models is a moving target: architectures, numerical formats and the balance between training and inference change, so the operations dominating computation this year may not dominate next year. In that environment the rational accelerator is a general-purpose one: a GPU can be reprogrammed for the next architecture as readily as for the current one, and its premium over specialised silicon - the price of flexibility - is worth paying while algorithms are unfixed. In the language of the innovation literature, the GPU is the dominant design of an unsettled era - the architecture that wins while the technology is still in its fluid phase (Utterback, 1994).

The same literature predicts what happens when the era settles. Once a workload stabilises - once the operations to be performed are known, fixed and repeated at scale - flexibility stops being valuable and efficiency becomes decisive. A custom ASIC, designed to perform exactly those fixed operations and nothing else, performs them at lower cost and lower power per operation than a general-purpose device carrying the overhead of programmability. The substitution is not hypothetical. It is already the stated strategy of the largest buyers of computation themselves: Google's Tensor Processing Unit is a custom ASIC built for stabilised machine-learning workloads, and the custom silicon programmes of Broadcom and Marvell follow the same route. The firms with the most information about where AI workloads are heading are spending on the technology that replaces the general-purpose fleet if those workloads settle.

The substitution has a trigger condition, and the condition is the paradox in embryo. ASIC substitution requires that algorithms stabilise. If they stabilise, AI has by definition progressed to the point at which its workloads are understood and repeatable - the success case - and the general-purpose fleet bought during the fluid phase faces a cheaper, more efficient, purpose-built competitor for every new unit of demand, its economic life cut short regardless of its physical condition. If algorithms never stabilise, the reason will be that AI has not landed: workloads remain experimental because applications remain unproven, and demand at the scale required to absorb the fleet never arrives. Flexibility is then worthless for the opposite reason - there is nothing at scale to be flexible for. Either the technology matures and replaces the fleet, or it does not mature and leaves it idle. Section 9 formalises this two-branch argument; the substitution risk is generated by the technology's own trajectory, in both directions, not by an external shock.

This is a general pattern in the economics of competing technologies. Increasing returns to adoption can lock in an early technology, but they can equally lock it out once a rival architecture accumulates its own installed base, software ecosystem and cost curve (Arthur, 1989; David, 1985). The incumbent's defence - that the installed fleet is large, paid for and integrated - is weakest for computing hardware, because the fleet's value lies in cost per computation, on which a purpose-built rival improves by design rather than by accident.

4. Competition: everyone builds the same capacity at the same time

The break-even arithmetic of the companion papers is, if anything, generous to the investment, because it is monopoly arithmetic: it asks what revenue a single supplier would need if it captured the demand and earned an incremental gross margin of 34% on it. The 34% figure is author-illustrative, not a filing-disclosed margin; it is itself striking, since 34% is barely half the 72% incremental margin the same firm's established business historically earned. A technology paper must correct the monopoly assumption, because the build-out is simultaneous: every hyperscaler and a new class of neocloud providers are building the same capacity, for the same customers, on the same timetable, with the same hardware from the same suppliers.

Simultaneous building has a predictable price consequence: supply planned independently arrives jointly, and identical capacity is then sold by rivals who have each committed their capital in advance and therefore compete on price to fill it, because an idle accelerator earns nothing and a discounted one earns something. Cloud capacity has been commoditised in previous cycles by exactly this mechanism, and AI computation - a standardised unit of work, sold by the hour, with low switching costs at the margin - is more commoditisable, not less. A price war therefore requires no firm to behave irrationally; it requires only that all of them behave identically.

Price competition enters the break-even arithmetic through the margin, and the arithmetic is unforgiving because it is a simple ratio. Required revenue equals required profit divided by the margin earned on that revenue. If the illustrative margin is 34%, required revenue is the values tabulated in Section 7. If competition compresses the margin to 25%, required revenue is multiplied by $34/25$, that is by 1.36. If the margin is compressed to 20%, required revenue is multiplied by $34/20$, that is by 1.70. Nothing else in the calculation changes; the profit the capital demands is fixed by the borrowing, and only the revenue needed to deliver that profit moves, upward, as the margin falls.

Table 1. Margin compression multiplies required revenue

Scenario	Incremental gross margin	Multiplier on required revenue	20-year, 8% case: required revenue
Finance-paper baseline	34%	1.00x	\$136.9B per year
Moderate price competition	25%	1.36x	\$186.2B per year
Severe price war	20%	1.70x	\$232.7B per year

Note: Illustrative scenarios computed from the author-illustrative incremental margin of 34% (not a filing-disclosed figure). The 20-year, 8% required revenue of \$136.9 billion per year is from Table 3. Multipliers: $34/25 = 1.36$; $34/20 = 1.70$.

ASIC substitution and price competition are not two independent risks to be added; they are the same risk in two forms. The rival who builds on custom silicon has a lower cost per computation and can cut price further while still earning a margin, while the owner of the general-purpose fleet, carrying higher unit costs, must match the price or lose the workload. Substitution sets the floor under the price war, which transmits it into the income statement long before any asset is written down.

5. Neoclouds and self-build: supply expands on four tracks at once

Section 4 understates the supply side, because hyperscalers are no longer the only builders. A second class has emerged: the neoclouds, GPU-specialist providers - CoreWeave, Nebius, Crusoe and Lambda are the leading examples - that own or lease accelerator capacity and rent it by the hour, without a hyperscaler's full platform of software, databases and enterprise services. A neocloud sells one product, accelerated computation, in the same standardised unit, the GPU-hour, as everyone else. The neoclouds matter for three reasons. They lower the entry barrier to supplying AI capacity, because a specialist with financing and a power contract can enter without building a platform. They convert a hyperscaler capital decision into a rental market, so capacity can be added by contract as well as by construction. And they are diffusion agents, spreading the technology to customers who would never have built anything, while adding one more independent source of the identical supply of Section 4.

The scale already committed is material, and must be read with a distinction the rest of this paper depends on: capacity in operation is not capacity under contract (Table 2, company-reported and tracker-reported figures).

Table 2. Neocloud scale: capacity in operation and under contract (as reported)

Provider	In operation	Under contract
CoreWeave	1.5 GW	about 4.2 GW
Crusoe	1.0 GW	more than 6 GW
Nebius	about 170 MW	more than 3.5 GW

Note: Company-reported and tracker-reported figures, reproduced as reported; not audited capacity figures. Contracted capacity is a claim on sites, power and equipment largely not yet energised; the gap between the columns is the deployment lag examined in Section 6, and no second-column figure is supply available today.

The gap between the columns is not a reporting defect; it is the phenomenon. Contracting runs ahead of energisation by years, so the supply that will compete for tomorrow's workloads is being committed today, on top of the hyperscaler programmes that dominate the headlines. Any account of future price that counts only hyperscaler construction understates committed supply by a multiple, not a margin.

The price consequence of multi-track supply is not conjecture; cloud computing has run this experiment before (full text: Online Supplement, Section S10). In March 2014 Google cut Compute Engine list prices by 32% and storage prices by 68%; within days Amazon Web Services cut server prices by up to 40% and storage by up to 65%, and Microsoft followed with reductions of a similar order on Azure (Google, Amazon Web Services and Microsoft, 2014; company-announced list prices, not transaction prices). Hedonic, quality-adjusted price indexes for Amazon Web Services (Byrne, Corrado and Sichel, 2018, 2021) fall at average annual rates of about 7% for compute over 2009-2016, more than 11% for database services over 2010-2016 and about 17% for storage over 2009-2016, with declines accelerating to double-digit rates from 2014: commoditisation is the historical norm for standardised cloud resources, not a tail event. Current rate cards show the same dispersion: a B300 accelerator at about \$9.50 per GPU-hour by Nebius against about \$17.80 by Amazon Web Services in October 2026.

Ownership structures multiply the tracks further (full text: Online Supplement, Section S7). Meta's Hyperion campus is held through a joint venture in which Blue Owl holds 80%, with Meta leasing the capacity back: hyperscaler demand, third-party balance sheet. Amazon and Google predominantly self-build and own. Microsoft mixes self-build, leases and neocloud contracts. Oracle, the case of the companion papers, leases campuses built by others - Crusoe, Vantage and STACK among them. Summing the four tracks - self-build, joint venture, landlord lease and neocloud contract - is the only correct way to count future supply, and it is a sum no participant controls: restraint by any one supplier cannot hold the price, because the withheld workload is simply served by another track. Price competition in AI computation is therefore the equilibrium of a diffusion process in which the means of supply have themselves diffused, not a strategic choice firms might avoid.

6. Power, siting and compliance: the deployment clock is set by the grid

The preceding sections dated the problem from adoption and from substitution. There is a third clock, starting earlier than either: the time required to give a data centre power, a lawful site and a social licence to operate. A building shell can be erected in roughly twelve to twenty-four months, but the hardware inside it is useless until the site is energised, and energisation is governed by processes measured in years. Five constraints bind jointly - grid power, fibre, skilled labour, water and permits - and the slowest sets the delivery date. The slowest is normally power and consent; and hardware bought for a site waits for that site, while the technological processes of Sections 3 and 5 continue to run against it.

Grid interconnection is a queue, not a purchase. In the interconnection queue study of Lawrence Berkeley National Laboratory (Rand et al., 2026), 2,061 gigawatts (GW) of capacity was active in queues at the end of 2025; the median time from queue entry to commercial operation for projects that are built is now more than five years, against under two years in the late 2000s; 549 GW had an executed interconnection agreement but was not yet in operation; and completion is the exception: of capacity entering queues between 2000 and 2020, only about 13% had been built by the end of 2025, while about 75% was withdrawn. Withdrawals do not shorten the queue mechanically, because the studies behind each project assume the projects ahead of it. The process is also fragmented rather than national: grid-operator studies, utility network upgrades, state and local permits and environmental review proceed as separate gates, in varying order, with no single authority able to quote a shortest lawful path. Equipment queues run in parallel: large power transformers carry industry-reported lead times of 80-210 weeks for large units (Wood Mackenzie, industry reporting; see Note on sources and Online Supplement, Section S10), against about one year before the current build-out.

Consent is the second gate, and it is not a formality. Community opposition to data centre siting recurs, across jurisdictions, on consistent grounds: local electricity prices, water consumption, noise, air emissions, loss of open land, and feared effects on nearby house prices. For the deployment clock, the perception does the work whether or not the feared effect materialises, because perception is what planning hearings, permit appeals and litigation are made of. The measured evidence on house prices is genuinely divided (full text: Online Supplement, Section S8). An Energy Institute at Haas working paper (Schlauch and Greenhill, 2026; working paper, not peer-reviewed), finds prices about 3% lower within three miles over five years and about 7.2% lower within one mile, with an aggregate effect of \$20.6 billion. A Virginia working paper by Alex Priest (Priest, 2026; working paper, not peer-reviewed), finds an effect that is small and slightly positive, with confidence intervals excluding large declines. A Realtor.com industry analysis (Realtor.com, 2026; industry analysis, not peer-reviewed) of 43 ZIP codes over 2019-2025 finds no meaningful effect on values, while the share of sales within five miles rises from 0.67% to about 1.5%. None is peer-reviewed, and the honest summary is disagreement. What is not in disagreement is the political fact: residents act on the fear, and projects are delayed, conditioned or refused accordingly.

Bypassing the grid does not bypass compliance; it exchanges one gate for another (full text: Online Supplement, Section S8). Behind-the-meter gas generation is now common precisely because it is faster than a queue, but the fast options are the least efficient ones: simple-cycle gas turbines convert fuel to electricity at about 33-37% efficiency, against about 49-54% for combined-cycle plant, so the speed premium is paid in fuel and emissions per megawatt-hour. Air permits bind where grid studies no longer do. The Memphis case, as reported in press and regulatory reporting (xAI Colossus, Memphis, 2025-2026; see Note on sources), is illustrative: aerial counts at xAI's Colossus site showed about thirty gas turbines, and a county air permit (permit 01156-01PC, 2 July 2025 to 2 January 2027, covering fifteen turbines totalling 247.2 megawatts (MW); as reported via the Southern Environmental Law Center and related reporting, not verified from the county permit document itself) was obtained in 2025 only after environmental groups threatened suit over unpermitted operation, and a Senate committee inquiry followed in 2026. Desert and waterside siting trades constraints rather than removing them, and remoteness recreates the connection problem.

Every year a campus spends in a queue, in a hearing or behind an unbuilt substation, algorithms continue to stabilise, custom silicon continues to improve, and rival capacity continues to be energised. Deployment delay is therefore not an additive inconvenience: it lengthens the exposure window inside which the impairment channels of Section 8 operate, and it can strand the sequence entirely - accelerators depreciate in storage while their site waits for power, arriving at energisation already part-way through the economic life substitution was going to shorten. The shell is the fastest part of a data centre; the investment case, which is written as if the shell were the timetable, is dated from the wrong clock.

Update note and cross-reference (29 October 2026). The deployment-clock watch list and the policy-rate anchor are given in full in the Online Supplement (Sections S1 and S2). The anchor is the Federal Open Market Committee statement of 28 October 2026 (Board of Governors of the Federal Reserve System, 2026), which maintained the target range for the federal funds rate at 3-3/4 to 4 percent by a 10-2 vote, Cook and Hammack dissenting in favour of an increase of 1/4 percentage point, with interest on reserve balances set at 3.90 percent, effective 29 October 2026. A December increase of 25 basis points remains a hypothetical scenario only; the full rate grid is in the companion finance paper.

7. Adoption timelines as a stress test, not a forecast

This section converts the adoption horizons of Section 2 into the annual profit and revenue that the committed capital requires under each horizon. The calculation is deliberately simple, so its assumptions are visible. Let P be the claim stack of \$457.1 billion documented in the companion finance paper - debt, lease commitments including the \$288.0 billion not yet commenced, and related obligations. Let r be an annual financing cost of 6%, 8% or 10%, a range anchored by prevailing sovereign yields at its lower end and by the direction of credit spreads at its upper end, of which the firm's 261 basis point five-year CDS spread (8 October 2026 print; later prints were higher) is the market's marker. Let T be the number of years over which the stack must be serviced and recovered, set equal to the adoption horizon. The level annual payment that amortises the stack is the standard annuity formula:

$$\text{Required annual profit} = P \times r / (1 - (1 + r)^{-T})$$

Required annual revenue is that profit divided by the author-illustrative incremental gross margin of 34% (an assumption, not a filing-disclosed figure), since only the gross profit on new revenue is available to service the capital. Interest alone makes the scale visible. At a 7% blended financing cost, the stack carries roughly \$32.0 billion of interest per year, which at the illustrative 34% margin requires about \$94.1 billion of new revenue per year merely to pay the interest - before principal, before further capital expenditure, and against total firm revenue and a cloud infrastructure segment that are both far smaller than the claim. Each further 100 basis points of financing cost adds about \$4.57 billion of annual interest and about \$13.4 billion of required annual revenue. Rising rates, widening spreads and refinancing therefore raise the break-even line mechanically.

Table 3. Required annual profit and revenue by adoption horizon and financing cost

Financing cost	Horizon	Required profit (\$B/yr)	Required revenue at 34% margin (\$B/yr)
6%	2 years	249.3	733.3
6%	5 years	108.5	319.2
6%	10 years	62.1	182.7
6%	20 years	39.9	117.2
8%	2 years	256.3	753.9
8%	5 years	114.5	336.7
8%	10 years	68.1	200.4
8%	20 years	46.6	136.9
10%	2 years	263.4	774.6
10%	5 years	120.6	354.7
10%	10 years	74.4	218.8
10%	20 years	53.7	157.9

Note: Illustrative scenarios, not forecasts. $P = \$457.1B$. Required revenue = required profit / 0.34, before the margin compression of Table 1 and before the utilisation adjustment of Section 8. Figures at 8% are the headline cases used in the text.

Read the 8% column as the paper's central exhibit. If AI lands in two years, the stack requires \$256.3 billion of profit and \$753.9 billion of revenue per year - an order of magnitude beyond the scale of the firm, and beyond any adoption scenario. Two-year landing is therefore not an optimistic case; it is an arithmetically impossible one, and any valuation that implicitly assumes absorption of the new capacity within two years is assuming a number larger than the firm itself. If AI lands in five years - the fastest horizon Section 2 can entertain at all - the requirement is still \$114.5 billion of profit and \$336.7 billion of revenue per year, sustained: a failing lower bound, not a rescue. At ten years the requirement falls to \$68.1 billion of profit and \$200.4 billion of revenue per year: conceivable for a mature, successful AI industry, but requiring that this one firm capture an extraordinary share of it at an uncompressed margin. Only at twenty years does the requirement approach a serviceable scale - and even then only before the adjustments of Sections 4 and 8.

Two caveats must accompany the table, and both cut against careless use in either direction. The first is that the full stack is not due in five years, or in two. Obligations mature on their own schedule, and the leases at the centre of the stack run for fifteen to nineteen years. The short-horizon rows of Table 3 therefore do not describe a payment schedule; they describe a stress test. They show how violently the required annual figures rise as the horizon is shortened. The twenty-year row is the row closest to the lease tenor, and it is the only row that is not absurd - which is another way of saying that the investment works, if it works, on exactly the generational timetable that Section 2 predicts, with no slack for delay or disappointment in between.

The second caveat is that Table 3 is computed at full utilisation and at the uncompressed illustrative 34% margin - the friendliest version of the arithmetic the stated assumptions allow. Section 4 showed that competition multiplies the revenue requirement by 1.36 at a 25% margin and by 1.70 at a 20% margin. Section 8 will show that utilisation below 100% multiplies it again. The twenty-year, 8% requirement of \$136.9 billion per year becomes roughly \$186 billion under moderate price competition, roughly \$233 billion under a severe price war, and about \$273.8 billion at 50% utilisation before any margin compression at all. Every realistic relaxation of the friendly assumptions makes the requirement larger, and none makes it smaller. A break-even that holds only under monopoly pricing, full utilisation and immediate adoption is not a break-even; it is a list of the conditions that must all hold at once.

The table does not claim that AI will take twenty years. It claims that the financing structure behaves as if twenty years were available and two were not, and that anyone underwriting the shorter horizon is underwriting the arithmetic in the upper rows of Table 3. Making the implicit arithmetic explicit is the contribution of a stress test: a reader who believes adoption will be fast is invited to say which row of the table they believe, and to defend the revenue figure in it.

8. Impairment has two channels, and hoarding triggers both

Accounting for long-lived hardware recognises, in principle, that an asset's carrying value must be written down when it can no longer be recovered. In practice, discussion of AI hardware has concentrated on a single, comfortable question: over how many years should the asset be depreciated? The assumption under test in this paper - used throughout as an assumption, and not asserted as any particular firm's verified policy - is a useful life of six to ten years for AI servers and related equipment. That range does enormous work in reported earnings: the annual charge on a tranche, and therefore the profit reported while the market waits for adoption, depends directly and inversely on the life assumed, as Table 4 shows.

Table 4. Useful-life assumptions and the annual depreciation charge (illustrative)

Asset base	10 years	8 years	6 years	5 years	4 years
\$100B tranche, charge per year	\$10.0B	\$12.5B	\$16.7B	-	\$25.0B
\$55.7B single-year CapEx cohort, charge per year	\$5.57B	-	-	\$11.14B	-

Note: Straight-line, zero residual value, illustrative arithmetic only. The \$55.7 billion cohort is the shared capital expenditure figure from the companion papers.

This is the first channel of impairment: time. Even in a world where nothing goes wrong - where the hardware is used, is not superseded, and demand arrives - the choice between a ten-year and a six-year life moves the annual charge on a \$100 billion tranche from \$10.0 billion to \$16.7 billion, and a four-year economic life would carry it to \$25.0 billion. Extending assumed lives is the most powerful lever available to reported profitability during the waiting period, which is why the assumption deserves scrutiny rather than deference. If the true economic life is set by ASIC substitution and generational adoption rather than by physical durability, then a life chosen by reference to how long the machine can run, rather than how long it can earn, overstates earnings in every year of the wait.

The second channel is larger and less discussed: utilisation and obsolescence impairment. An asset is impaired not only when it wears out but when no one uses it, or when a better asset makes its services worth less than its carrying value - and this can happen to hardware that is physically perfect. Three triggers matter. The first is simple under-utilisation: capacity built for demand that has not yet arrived earns nothing while its financing and depreciation continue. Revenue must be earned on the utilised fraction of the fleet, so at 50% utilisation the revenue required to carry the whole fleet doubles: the twenty-year, 8% requirement of \$136.9 billion per year becomes about \$273.8 billion per year, before any margin compression. The second trigger is substitution, from Section 3: the arrival of cheaper ASIC capacity reprices the service the GPU fleet sells, and the fleet's recoverable amount falls even if every machine is running. The third trigger is the most neglected, and it is created by hoarding.

Hoarding is rational behaviour for each buyer and a collective impairment mechanism for all of them. During a supply shortage, each hyperscaler and neocloud has an incentive to order ahead of need and to treat secured supply as a competitive asset in itself. The result is a stock of hardware whose clock runs from the date of purchase, not the date of first use. Computing hardware does not wait patiently. Each new generation of accelerators, and each advance in custom silicon, reduces the value of the stock in the warehouse, whether or not that stock has ever been installed or used. Hardware can therefore become obsolete before it is used at all - in economic terms it arrives already impaired, caused by the passage of technological time during the very adoption lag that made the buyer hoard.

The two channels compound rather than add. Time-based depreciation assumes a smooth decline in value across the assumed life. Utilisation and obsolescence impairment is lumpy, event-driven and correlated across the whole industry at once: when algorithms stabilise, every owner's fleet is repriced in the same period; when adoption disappoints, every owner's utilisation falls in the same period. Diversification across owners does not exist, because the asset, the trigger and the timetable are common. The accounting sequence does not protect the investor either: depreciation is recognised gradually and can be smoothed by extending lives, while impairment is recognised late and arrives as a discrete charge against a balance sheet presented in the meantime as sound. The investor who relied on the depreciation schedule as a map of economic decay is therefore misled in a specific direction: the map shows a gentle slope where the terrain contains a cliff.

For the case running through the companion papers, the implication is concrete. A firm spending \$55.7 billion on capital expenditure in a single year creates a cohort of assets whose annual depreciation charge is \$5.57 billion on a ten-year life and \$11.14 billion on a five-year life - a difference of more than \$5.5 billion of reported profit per year, resting on an assumption about how long general-purpose hardware remains economically relevant in a field whose customers are building its replacement. The question for Section 9 is whether the long-life assumption is safe in either stylised branch. The answer is that it is safe in neither, without further conditions.

9. The adoption-impairment paradox, stated formally

The argument of Sections 2, 3 and 8 can now be stated as a mechanism hypothesis in three steps.

Step 1: the progress branch. Suppose AI research progresses and AI lands at scale. Landing means that applications are proven, workloads are repeatable and the operations are known and fixed. On fixed operations, custom ASICs outperform general-purpose GPUs on cost and power, as the TPU and the Broadcom and Marvell custom silicon programmes already demonstrate. New demand is served disproportionately by ASIC capacity, the price of computation is set by ASIC costs, and the installed fleet is impaired by substitution, its economic life ending well inside its assumed six-to-ten-year accounting life.

Step 2: the no-progress branch. Suppose instead that AI research does not progress to landing: algorithms remain unsettled because the applications remain unproven, enterprise adoption stalls at the chasm, and demand at infrastructure scale never arrives. The general-purpose fleet is safe from ASIC substitution - there is no stabilised workload to build an ASIC for - and worthless for the opposite reason. Capacity built for demand that does not come is under-utilised; hoarded hardware obsolesces unused as generations arrive; and the fleet is impaired by under-utilisation and obsolescence, again well inside its assumed accounting life.

Step 3: the hypothesis. The two branches are stylised poles, not an exhaustive partition of the future, and in both stylised branches the fleet is impaired. Technological progress impairs it by substitution; the absence of progress impairs it by idleness. On this mechanism hypothesis, neither pole by itself supports treating the book value of general-purpose AI hardware as preserved over a full assumed life of six to ten years. A long useful life is not the neutral or default assumption from which a sceptic must argue departure. It is a substantive claim that requires its own defence, because both stylised futures place it under pressure.

Figure 1. The adoption-impairment paradox.

Branch A - Progress: AI lands → algorithms stabilise → ASIC substitution (TPU; Broadcom and Marvell custom silicon) → fleet impaired by substitution.

Branch B - No progress: AI does not land → demand never arrives → utilisation collapses, hoarded stock obsolesces unused → fleet impaired by under-utilisation.

Escape (corridor): progress fast enough to create demand, slow enough that installed hardware stays relevant, differentiated enough that rivals do not commoditise price.

The escape exists, and naming it precisely is part of stating the hypothesis. Book value is preserved if progress creates demand for the capacity yet leaves that hardware relevant while demand arrives, and if the firm's offering is differentiated enough that competitors, including those on cheaper custom silicon, do not compete price below what the fleet's financing requires. Each condition is plausible; their conjunction is a corridor whose width nobody has measured and whose existence is assumed whenever a six-to-ten-year life is written into an account without argument. An assumption holding inside an unmeasured corridor cannot be the default. The corridor may also be wider than a knife-edge: fleets can move to inference, be resold second-hand, gain time from CUDA lock-in, benefit from slow ASIC ramps, sit in hybrid architectures, or be leased rather than owned - routes whose combined width is unmeasured.

The paradox clarifies what would falsify it. It would be falsified by evidence that stabilised workloads run economically on general-purpose hardware against purpose-built rivals; by adoption data showing enterprise absorption of new capacity in two to five years at uncompressed prices; or by utilisation data showing new fleets at levels making the doubling arithmetic of Section 8 irrelevant. Those are empirical questions, stated in advance, with observable answers. Until such observations arrive, the paradox stands as the appropriate prior: in matters of useful life, the optimist, not the sceptic, makes the extraordinary claim. The hypothesis does not depend on the cost of finance: even at a zero financing cost, the progress branch still lowers general-purpose computation prices through ASIC substitution, and the no-progress branch still impairs the fleet through idleness and obsolescence; interest rates change the speed of loss and the tightness of financing constraints, not the mechanism itself.

10. The crypto-GPU precedent: a rhyme, not a repetition

The mechanism of Sections 3 and 9 has run once, in miniature, in cryptocurrency mining - offered strictly as a historical analogy, a rhyme in the sense of Section 2, not as a prediction (full text: Online Supplement, Section S9). While mining algorithms were unfixed, miners used general-purpose Nvidia GPUs, as in Section 3. As algorithms stabilised and volumes grew, ASIC miners performed the fixed operations at a fraction of the energy cost, and GPU mining of those coins became uneconomic - not because the GPUs had worn out, but because a purpose-built rival set the price of the computation they sold. Second-hand GPUs flooded resale markets and hoarded inventory was written down unused: the flexible asset was correct at purchase, and progress destroyed it. The disanalogies make the analogy one of mechanism, not destiny.

11. Policy and management implications

Research Policy asks what follows for policy and management; the paradox yields implications for both, without taking a view on any single firm.

For accounting standards and disclosure. Where useful life is the most consequential assumption in a sector's reported earnings, it should be disclosed and defended as such. Firms extending the depreciable lives of computing hardware should state the technological basis for the extension - why substitution and obsolescence will not shorten economic life - with the same prominence as the earnings benefit. Utilisation of newly installed capacity, and the stock of purchased-but-uninstalled hardware, are facts current disclosure leaves investors to guess. A standard recognising obsolescence impairment as a first-class, forward-looking risk for rapidly substituted assets, rather than a residual recognised after recoverability has failed, would narrow the gap between the depreciation map and the terrain of Section 8.

For corporate management and boards. The paradox reframes capacity commitment as an options problem. Capacity that must be ordered years ahead, financed with debt and leases of fifteen to nineteen years, and justified by adoption that proceeds generationally, should be staged so that commitment follows evidence of absorption rather than preceding it: tranche the build, contract for flexibility where it can be bought, and treat owning general-purpose capacity versus accessing purpose-built capacity as a live strategic choice. Boards should require management to state, in the annuity form of Section 7, which adoption horizon the balance sheet is underwriting, and to show the row for a horizon half as long. A commitment that only works in the twenty-year row should be approved, if at all, as a twenty-year commitment, with financing matched to that tenor.

For innovation and industrial policy. Governments subsidising AI infrastructure are, under the paradox, subsidising whichever branch arrives: stranded general-purpose capacity in one, accelerated substitution in the other. Policy should therefore distinguish between supporting capability - skills, research, grid connection, the generational work of adoption - and supporting installed hardware volume as such. Support tied to utilisation and to demonstrated absorption rewards the slow variable that actually determines success; support tied to capacity installed rewards the fast variable that creates impairment risk. Competition policy has a role as well: simultaneous, debt-financed capacity races among a handful of hyperscalers predictably generate price wars and stranded assets, and the systemic exposure they create is a legitimate object of macroprudential attention.

For investors, by way of method only: this paper offers no investment advice, and this paragraph is a methodological remark, not a recommendation. The framework suggests that the useful questions about an AI infrastructure firm are technological before they are financial: what fraction of the fleet is utilised, what fraction of purchases is installed, on what evidence is the useful life set, and what happens to the accounts if the life is the one ASIC substitution implies? An investor who cannot obtain answers is missing the variable on which the valuation turns.

12. Discussion: boundary conditions and limitations

The paradox is a claim about a class of assets under a class of conditions. It applies to general-purpose computing hardware bought ahead of adoption, financed as a fixed claim, and exposed to a purpose-built substitute. It applies weakly, or not at all, in four situations. First, where the buyer owns the substitute as well as the fleet: a firm that operates both general-purpose and custom-silicon capacity internalises the substitution, and impairment in one line is offset by margin in the other - though this defence is unavailable to a firm whose capacity is predominantly general-purpose and whose custom programme, if any, belongs to its customers. Second, where workloads remain permanently fluid at the frontier while stabilising only in commodity inference: the fleet may then retain a durable training niche, and the paradox bites only on the inference share of capacity - an empirical quantity, and notably the workload expected to dominate at scale. Third, where capacity is genuinely scarce over the whole horizon because power, land or grid connection is the binding constraint (Section 6): a scarce site can hold its price even as the hardware inside it is repriced, though the rent then attaches to the site and the power contract, not to the GPUs. Fourth, where the state absorbs the risk, through subsidy, guaranteed offtake or toleration of the leverage involved (Section 11).

The argument also has limitations. The stress test of Section 7 is deliberately reduced-form: a single claim stack, a single margin, a small set of financing costs, and no modelling of refinancing, asset sales or the option to stop building. Those simplifications suit a stress test, whose purpose is to locate the horizon at which the arithmetic becomes survivable, but they are not a valuation. The utilisation and margin adjustments are applied as independent multipliers, whereas in practice they are correlated - a price war and a utilisation shortfall are likely to arrive together, which would make the combined requirement worse than either multiplier alone, a direction of bias that again leaves the paper's conclusions conservative. The generational adoption claim, finally, is supported by analogy to previous general-purpose technologies and by the diffusion literature, not by a completed econometric study of AI adoption itself; the technology is too young for its S-curve to be estimated. What can be estimated now are the fleet's utilisation, the installed share of purchases, and the realised price of computation, and Sections 8 and 9 state in advance how each series would confirm or falsify the argument.

Placed in the literature, the paradox connects three conversations usually held separately. To the dominant-design and technology-lifecycle literature (Utterback, 1994; Teece, 1986), it adds that the transition out of the fluid phase is itself an impairment event for the assets of the fluid phase, now large enough to appear in credit markets. To the literature on technological revolutions and financial capital (Perez, 2002), it adds a balance-sheet mechanism for the sequence of installation, frenzy, collapse and deployment: generational adoption time mismatched against financial maturity, expressed in an annuity table. To the accounting literature on useful life, it contributes a two-sided test - progress and its absence - that a long-life assumption must pass before it can be treated as neutral.

12.1 Later platforms and portable workloads

12.1.1 Later platforms: space, subsea and tidal power as staged obsolescence

Two proposed successors to the terrestrial campus are sometimes offered as reasons to discount the paradox; neither changes the horizon this paper examines, and both are treated in full in the Online Supplement (Section S3). Orbital computation has been demonstrated only at roughly the one-kilowatt scale; its economics turn on launch cost - competitiveness near \$500 per kilogram and energy parity below \$200 per kilogram, against roughly \$1,500 per kilogram today - putting general parity no earlier than the mid-2030s. Subsea computation worked - Natick (Microsoft, 2024; company-reported) recorded a failure rate of 0.7% against 5.9% on land - yet the programme ended in 2024, because a sealed pod cannot be upgraded at the cadence of GPU generations. Tidal proposals remain at study and permit stage; neither arrives in time to rescue hardware bought today.

12.1.2 Workload portability: why contracted demand does not remove competition

A final cross-reference concerns contracted demand, whose financial modelling belongs to the companion finance paper; the full discussion is in the Online Supplement (Section S6). The technological fact is portability: frontier workloads are built on standard frameworks, are containerised, and move between clouds at switching costs low relative to the price dispersion of Section 5, so a signed backlog records intended demand, not captive demand. Anthropic (S-1 filed 26 October 2026, unpriced; Anthropic, 2026) reported first-half 2026 revenue of \$16.3 billion and cloud commitments of roughly \$450 billion over seven to ten years placed with Google, Amazon Web Services and Microsoft; no Oracle commitment appears in that list. Demand growth and demand capture are different variables.

12.2 Two regimes of growth, and the population underneath adoption

Cross-reference only: the full treatment belongs to the companion papers.

Cross-reference only; the full text is in the Online Supplement (Section S5) and the full treatment in the companion finance and geopolitics papers. Two regimes of growth must be distinguished: liquidity-driven growth can refinance a claim stack; only demographic growth can service it. In the Solow steady state (Solow, 1956), sustained growth in output per worker comes from technological progress, and the support-ratio identity $Y/N = (L/N) \times (Y/L)$ separates the population and working-share margins; a twenty-year break-even (Section 7) is a bet on the second regime, not the first. The generational lag is measured, not assumed: the OECD (2025) reports AI-related training by 46% of those aged 26-35 against under 20% of those aged 55 and over, with over 75% of under-35s finding generative AI useful. The jobs evidence runs both ways: almost 40% of global employment is AI-exposed, about 60% in advanced economies (Cazzaniga et al., 2024); transformation is more likely than replacement for most exposed occupations (ILO and NASK, 2025); workers aged 22-25 in the most exposed occupations show a 13% relative employment decline where AI automates rather than augments (Brynjolfsson, Chandar and Chen, 2025, working paper); and augmentation raises issue resolution by about 14% on average (Brynjolfsson, Li and Raymond, 2025). On fertility, there is no established evidence that AI raises it, and the claim that AI displacement lowers it is a hypothesis only, on which no figure in this paper depends. The demographic corrections are: China in negative growth since 2022 and South Korea in natural decrease since 2020, both with fertility far below replacement (about 1.0 and about 0.7-0.8); India, at about 1.96, still growing on momentum to about 1.69 billion around 2054; China and India together about 34.9% of world population, about one-third, not one-half; and world population rising from 8.2 billion in 2024 to a peak of about 10.3 billion in the mid-2080s (United Nations, 2024). On food, precision-agriculture gains are modelled potentials, not observed increases in global tonnes per hectare; the FAO Food Outlook forecasts cereal output down about 2.0% and wheat down about 3.8% in 2026 (FAO, 2026, forecast, not outcome); the FAO-WMO assessment projects staple yield losses of about 4-10% per degree Celsius of additional warming; no direct effect of data centres on crop yields is claimed, the defensible claim being local competition for electricity, water, land and grid capacity; and data centres consumed about 415 TWh in 2024, about 1.5% of global electricity, rising to about 945-950 TWh by 2030 in the IEA base case (International Energy Agency, 2025).

13. Conclusion: the question is when, and the balance sheet cannot wait

This paper has argued that AI hardware investment faces an adoption-impairment paradox. Adoption is generational: habits, trust, skills, regulation and routines turn over with cohorts, along an S-curve that financing cannot steepen. The hardware bought ahead of that adoption is a general-purpose fleet whose value rests on algorithmic unsettledness. If the technology progresses, algorithms stabilise, custom ASICs - the Google TPU, Broadcom's and Marvell's custom silicon - take the stabilised workloads at lower cost, competition transmits the substitution into price, and the fleet is impaired by replacement. If the technology does not progress, demand never arrives at scale, and the fleet, including hoarded hardware, is impaired by idleness. The annuity stress test shows what the generational timetable costs: at an 8% financing cost, recovery of a \$457.1 billion claim stack requires \$753.9 billion of revenue a year on a two-year horizon, \$336.7 billion on five years, \$200.4 billion on ten, and \$136.9 billion on twenty - before margin compression multiplies those figures by as much as 1.70 and before half-utilisation doubles them again. The twenty-year horizon, closest to the fifteen-to-nineteen-year leases underneath the stack, is the only one the arithmetic does not reject outright.

None of this is an argument that artificial intelligence will fail. The author expects, on the historical evidence assembled here, that it will succeed - as electrification succeeded, as the internet succeeded - by reorganising practice across a generation of users. The argument is about the location of that success in time, and about who owns the assets, and owes the debts, between construction and adoption. Capital can be raised in a quarter; adoption cannot. Debt is a claim on future time, and adoption is a claim on future behaviour, and behaviour changes more slowly than debt matures.

The conclusion can therefore be stated in one sentence. AI may eventually succeed, but the question is when - and a leveraged balance sheet does not have the time that a generational adoption curve requires.

Data availability

All data used in this paper are drawn from public sources: the public filings of Oracle Corporation (Forms 10-K, 10-Q and 8-K, as documented in the companion finance paper) and the published sources cited in the reference list. No proprietary or restricted dataset was used, and no new dataset was created. The illustrative scenario arithmetic in Sections 4, 7 and 8 can be reproduced from the figures stated in the text.

References

- Anthropic, PBC, 2026. Registration statement on Form S-1. Filed with the U.S. Securities and Exchange Commission, 26 October 2026. Company filing.
- Arthur, W.B., 1989. Competing technologies, increasing returns, and lock-in by historical events. *Economic Journal* 99, 116-131.
- Bass, F.M., 1969. A new product growth model for consumer durables. *Management Science* 15 (5), 215-227.
- Board of Governors of the Federal Reserve System, 2026. Statement of the Federal Open Market Committee. Federal Reserve, Washington, DC, 28 October 2026.
- Brynjolfsson, E., Chandar, B., Chen, R., 2025. Canaries in the Coal Mine? Six Facts about the Recent Employment Effects of Artificial Intelligence. Stanford Digital Economy Lab, Stanford University, August 2025. Working paper, not peer-reviewed.
- Brynjolfsson, E., Li, D., Raymond, L.R., 2025. Generative AI at work. *Quarterly Journal of Economics* 140. Journal article.
- Byrne, D.M., Corrado, C.A., Sichel, D.E., 2018. The rise of cloud computing: minding your P's, Q's and K's. NBER Working Paper No. 25188. National Bureau of Economic Research, Cambridge, MA, October 2018. Published 2021 in: Corrado, C., Haskel, J., Miranda, J., Sichel, D.E. (Eds.), *Measuring and Accounting for Innovation in the Twenty-First Century*. University of Chicago Press, Chicago, pp. 519-551.
- Cazzaniga, M., Jaumotte, F., Li, L., Melina, G., Panton, A.J., Pizzinelli, C., Rockall, E., Tavares, M.M., 2024. Gen-AI: Artificial Intelligence and the Future of Work. IMF Staff Discussion Note SDN/2024/001. International Monetary Fund, Washington, DC, January 2024.
- David, P.A., 1985. Clio and the economics of QWERTY. *American Economic Review* 75 (2), 332-337.
- Food and Agriculture Organization of the United Nations, 2026. Food Outlook: Global food commodity market trends face rising geopolitical and weather risks. FAO, Rome, 2026. Forecast, not outcome; see also FAO and World Meteorological Organization, *Extreme Heat and Agriculture* (joint report).
- Google, Amazon Web Services and Microsoft, 2014. Cloud list-price reduction announcements, March 2014. Company announcements, as reported; announced list prices, not transaction prices.
- International Energy Agency, 2025. Energy and AI. IEA, Paris, April 2025.
- International Labour Organization and National Research Institute NASK, 2025. Generative AI and Jobs: A Refined Global Index of Occupational Exposure. ILO, Geneva, May 2025.

References (continued)

- Microsoft, 2024. Project Natick subsea data centre experiment. Company-reported; programme ended in 2024. No peer-reviewed account is cited.
- Moore, G.A., 1991. *Crossing the Chasm: Marketing and Selling High-Tech Products to Mainstream Customers*. HarperBusiness, New York.
- Organisation for Economic Co-operation and Development, 2025. *How Do People Experience New Technologies and Generative AI?* OECD, Paris, 2025.
- Perez, C., 2002. *Technological Revolutions and Financial Capital: The Dynamics of Bubbles and Golden Ages*. Edward Elgar, Cheltenham.
- Priest, A., 2026. Not In My Back Yard! The Effects of Data Centers on Housing Prices. University of Rochester, 27 February 2026. Working paper, not peer-reviewed. No SSRN number is given, because none was found.
- Rand, J., Cheyette, A., Talley, C., Zhang, S., Gorman, W., Wiser, R., Seel, J., Jeong, S., Kahrl, F., 2026. Queued Up: 2026 Edition - Characteristics of Power Plants Seeking Transmission Interconnection As of the End of 2025. Lawrence Berkeley National Laboratory, Berkeley, CA, June 2026. <https://doi.org/10.2172/3376147>.
- Realtor.com, 2026. Americans Aren't Moving Toward Data Centers. Realtor.com Research, 11 August 2026. Industry analysis, not peer-reviewed.
- Rogers, E.M., 1962. *Diffusion of Innovations*. Free Press, New York.
- Schlauch, G., Greenhill, S., 2026. Data Centers Lower Nearby Housing Values. Energy Institute at Haas Working Paper WP 366. University of California, Berkeley, September 2026. Working paper, not peer-reviewed.
- Smithsonian Institution, n.d. ENIAC. National Museum of American History. Museum account.
- Solow, R.M., 1956. A contribution to the theory of economic growth. *Quarterly Journal of Economics* 70 (1), 65-94.
- Teece, D.J., 1986. Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy. *Research Policy* 15 (6), 285-305.
- United Nations, 2024. *World Population Prospects 2024: Summary of Results*. United Nations Department of Economic and Social Affairs, Population Division, New York, July 2024.
- Utterback, J.M., 1994. *Mastering the Dynamics of Innovation*. Harvard Business School Press, Boston, MA.
- Wood Mackenzie, n.d. Transformer lead-time reporting. Industry reporting; no single report title or date was verified. The 80-210 week range is verified as Wood Mackenzie data quoted in the National Infrastructure Advisory Council draft and the North American Electric Reliability Corporation Summer Reliability Assessment 2025; no continuous series exists.
- xAI Colossus, Memphis, 2025-2026. Gas-turbine air-permit proceedings (permit 01156-01PC). Press and regulatory reporting, verified via Southern Environmental Law Center and reporting only, not from the county permit document; no single report is cited - see Note on sources.

Note on sources: the line that history does not repeat itself but rhymes is commonly attributed to Mark Twain, with no verified source in his writings, and is treated here as an aphorism, not a citation. Core evidence in this paper is restricted to official and peer-reviewed sources: the Lawrence Berkeley National Laboratory queue studies (Rand et al., 2026), the NBER study of Byrne, Corrado and Sichel (2018, 2021), the established books and articles in the reference list above, and - via the companion finance paper - Oracle's SEC filings; the Federal Open Market Committee statement of 28 October 2026 (Board of Governors of the Federal Reserve System, 2026) is an official source cited as a cross-reference in Section 7 only. Company figures used in the illustrative scenarios are as documented from public filings and market data in the companion finance paper and are reproduced here without alteration; the full figure list, and the labelling catalogue for all other reported figures, are given in the Online Supplement (Section S4). The incremental gross margin of 34% (against 72% historically) is an author-illustrative assumption, not a figure disclosed in Oracle's filings, and is labelled as such at each use; the adoption timelines, margin multipliers and depreciation charges are illustrative scenario arithmetic computed from those figures, as labelled in the text, and are not forecasts. News reporting is not used as core evidence for any number in this paper.