Investing in Artificial Intelligence: From the Chip to the Application — How Investors Can Take Part in the Second Megatrend

Roland Rupprechter[1]

September 2026

Abstract

No technology has ever moved so much capital in so little time: the four large cloud groups plan to invest around USD 735 billion in 2026 — more than five times their 2023 outlay —, NVIDIA's data-centre business approaches USD 200 billion in annual revenue, and the model providers OpenAI and Anthropic report annualised revenues of more than 40 and 65 billion US dollars respectively. This paper — the second of six on the great megatrends — classifies artificial intelligence as a general purpose technology, maps its value chain from chips through data centres and models to applications, and tests the bubble question against hard numbers.

The empirical picture cuts both ways. On the one hand, controlled studies document sizeable productivity gains — 15 per cent in customer service, 40 per cent time saved in writing tasks, 56 per cent faster programming — and the investments are largely paid out of the current profits of the world's most profitable companies. On the other hand, index concentration has reached historic highs — the ten largest US names make up more than 40 per cent of the S&P 500, more than at the peak of the dotcom bubble —, circular financing between chip makers, cloud providers and model firms is growing, and according to a widely discussed MIT study 95 per cent of corporate pilot projects have yet to deliver a measurable profit contribution.

For investors, the paper names listed companies along the entire chain — exclusively large and mid caps, with business idea and trading venues — and

[1]Roland Rupprechter, Economic Research, R&B Wealth Management, Kreuzgasse 1a, 6850 Dornbirn; Member Austrian Association for Financial Analysis and Asset Management. Email: roland.rupprechter@rbvm.at.
Statement on the use of AI tools: This paper was produced with the selective use of the AI tool Claude (Anthropic). The core elements of the content – including the research question, the objective, the methodology, the data analysis, all calculations, as well as the specialist knowledge and the layout – were produced entirely independently by the author. Under the author's guidance, Claude served as an assistant tool to support research, shorten text, verify derivations and cross-check sources. Some AI-generated text passages were thoroughly reviewed by the author and only incorporated once the content objectives had been met. Finally, the system was used for proofreading (spelling and academic style). The author takes full responsibility for the content of the paper.

applies the fixed ETF checklist: physical replication, fund size of at least EUR 0.2 billion, sustainability classification, costs, the currency question, issuer, UCITS framework and exchange trading. The most important single lesson: whoever holds a world index already owns roughly a third of the AI complex through its heavyweights — every additional position is a deliberate overweight and should be sized accordingly.

Keywords: Artificial intelligence · General purpose technology · Semiconductors · Data centres · Productivity · Thematic ETFs · Index concentration · Megatrends.

JEL codes: G11, O33, O40, L63, L86.

Non-Technical Summary

When the chip maker NVIDIA was worth more than twice all of Germany's listed companies combined in the summer of 2026, one thing was finally clear: artificial intelligence is no longer a technology topic but the dominant capital-market event of the decade. Four groups — Microsoft, Alphabet, Amazon and Meta — are spending around USD 735 billion on data centres in 2026, year after year more than Austria's entire economic output. Whoever owns a world portfolio is already part of this upheaval — knowingly or not.

This paper — the second in a series of six on the great megatrends — first explains what stands behind the buzzword: a general purpose technology like steam power of old and electricity, slow to translate into productivity but fast to turn into infrastructure. It then describes the four layers of value creation — chips, data centres, models, applications —, tests the bubble question against the dotcom years, and answers the investor questions: which shares map the value chain? Which ETFs pass the checklist? And how much AI can a portfolio bear?

The answers in brief: the value chain is fully listed — from the suppliers of chip production through the cloud groups to the power and cooling specialists of the data centres; the paper names around fifty large and mid caps with business and trading venue and deliberately avoids small caps. Among ETFs, several funds pass the checklist — led by Europe's largest AI fund with almost eight billion euros under management. But the big AI winners already sit in the world index with a good third of its weight; AI funds are therefore a sharpening of exposure, not an entry from zero.

The most important warning last: even genuine revolutions can be bought too dearly. Electrification changed everything — and still ruined many of its early shareholders; the internet bubble burst although the internet kept its promise. Today valuations are high, concentration is above the level of 2000, and individual warning signs — circular deals, disappointed corporate pilots — are multiplying. Whoever invests therefore does so spread across the value chain, measured in size, and with a horizon that can absorb a lost decade.

Contents

1 Introduction

In August 2026 the chip maker NVIDIA reported USD 96 billion of revenue for a single quarter — more than the entire company had earned in a full year three years earlier — and was valued at USD 5.25 trillion on the stock market. In the same month the AI company Anthropic reported an annualised revenue rate of USD 65 billion — up from 30 billion in the spring. Numbers of this kind have no precedent in economic history: no product and no technology has ever created such revenues and market values so fast. The question facing every investor is therefore not whether artificial intelligence is a megatrend — but how to take part in it without repeating the mistakes investors have made in every technology wave before.[2]

This paper is the second in a series of six on the great megatrends. The first (Rupprechter, 2026g) dealt with the future of mobility — a field whose most spectacular advances, from the robotaxi to the driverless lorry, are themselves applications of artificial intelligence. The present paper turns to the core: the technology itself, its economics and its value chain. It is written for investors and pursues three aims: to understand what the technology can and cannot do; to judge whether the vast investments are a bubble or a foundational build-out; and to translate the field into single shares and funds.[3]

The academic base is younger than the hype suggests, but solid. What machine intelligence is and how to test it was asked by Turing (1950); the modern breakthrough of learning from data was described by LeCun, Bengio and Hinton (2015); the architecture on which practically all of today's language models rest comes from Vaswani, Shazeer, Parmar, Uszkoreit, Jones, Gomez, Kaiser and Polosukhin (2017). The economic framing is supplied by Bresnahan and Trajtenberg (1995) with the concept of the general purpose technology, by David (1990) with the lesson that such technologies take decades to show up in productivity, and by a young wave of controlled field studies that measure the effect instead of guessing it. This paper stands on that base and cites it throughout.[4]

The paper proceeds as follows. Chapter 2 tells, with due brevity, what artificial intelligence is and why the breakthrough came now. Chapter 3 covers the economics: general purpose technology, productivity evidence, growth effects. Chapter 4 dissects the value chain into its four layers, chapter 5 quantifies the investment boom and its energy question, chapter 6 tests the bubble question against the dotcom years. Chapter 7 presents listed companies along the chain, chapter 8 tests ETFs against the checklist, chapter 9 covers the risks from concentration to regulation. Chapter 10 condenses the lessons for

[2] NVIDIA (2026): quarterly revenue of USD 96.2 billion (second quarter of fiscal year 2027, reported 26 August 2026), market capitalisation around USD 5.25 trillion at the end of August 2026; Anthropic: annualised revenue of USD 65 billion (July 2026).
[3] Rupprechter (2026g), Future Mobility; chapter 5 there covers autonomous driving as an AI application.
[4] Turing (1950); LeCun, Bengio and Hinton (2015); Vaswani et al. (2017); Bresnahan and Trajtenberg (1995); David (1990).

investors, and chapter 11 concludes with an outlook. All market data are as of late August 2026.[5]

2 From Turing to the Transformer: Why the Breakthrough Came Now

2.1 A short history in five leaps

The idea is old, the breakthrough young. Turing (1950) asked whether machines can think and proposed the imitation game as the test: if a human in dialogue can no longer tell whether he is talking to man or machine, the difference is practically moot. Decades of ups and downs followed — rule-based systems, expert software, two “AI winters” of disappointed expectations — until three ingredients came together that transformed the field: very large data sets from the internet, very large computing power from graphics processors, and learning methods that extract patterns from both on their own. LeCun, Bengio and Hinton (2015) summarised this approach — deep learning; for its foundations, Hinton was awarded the 2024 Nobel Prize in physics together with Hopfield.[6]

The decisive leap came in 2017 with a research paper by eight Google researchers: Vaswani, Shazeer, Parmar, Uszkoreit, Jones, Gomez, Kaiser and Polosukhin (2017) introduced the transformer architecture, which processes language through an attention mechanism and — unlike its predecessors — scales almost without limit: more data, more computing, better results, with astonishing regularity. Practically all of today's language models rest on this architecture. The mass market arrived in November 2022, when OpenAI reduced the interface to a conversation with ChatGPT — one hundred million users within two months, the fastest product adoption in history. Since then the race has shifted to so-called agents: systems that do not merely answer but complete multi-step tasks on their own — researching, booking, programming, checking.[7]

Behind the diffusion stands — as with every general purpose technology — a cost curve. The cost of producing a given quality of answer has fallen by orders of magnitude since 2022: through better methods, specialised chips, and competition that passes price cuts on immediately. What required a data centre yesterday runs on a device tomorrow. This deflation of machine intelligence is the true mass-market mechanism — it explains why usage grows faster than any revenue line, and it is the strategic clamp of the business:

---

[5] Methodological note: market capitalisations and fund volumes are snapshots of late August 2026; corporate investment figures follow partly differing definitions (cash capital expenditure versus leases included).

[6] Turing (1950), Mind 59(236), pp. 433-460; LeCun, Bengio and Hinton (2015), Nature 521(7553), pp. 436-444; 2024 Nobel Prize in physics to Hopfield and Hinton.

[7] Vaswani et al. (2017), Advances in Neural Information Processing Systems 30; ChatGPT: launch 30 November 2022, around 100 million users within two months.

whoever wants to earn on intelligence itself must deliver new capabilities faster than the old ones become cheap. Table 1 summarises the milestones.[8]

Table 1: Milestones of artificial intelligence from an investor's perspective

Year	Event	Significance
1950	Turing's imitation game	Conceptual foundation of machine intelligence
2012	Deep-learning breakthrough (ImageNet)	Learning from data beats hand-built rules
2017	Transformer architecture (Vaswani et al.)	Scalable foundation of all language models
2022	ChatGPT (30 November)	Mass market; fastest product adoption ever
2024	Nobel Prizes for AI foundations	Scientific consecration of the field
2025/26	Agents and billion-dollar revenues	From answering to completing work

Source: Turing (1950); LeCun, Bengio and Hinton (2015); Vaswani et al. (2017); own compilation.

2.2 What the systems can do — and what they cannot

For investment purposes a sober picture of capability suffices. Today's models generate and understand language, program text, images and increasingly video at a level that reaches or exceeds average human performance in many office tasks; at the same time they fail in a peculiar way — they occasionally invent content, and their limit does not run along familiar notions of difficulty. Dell'Acqua, McFowland, Mollick, Lifshitz-Assaf, Kellogg, Rajendran, Krayer, Candelon and Lakhani (2023) coined the image of the "jagged frontier" for this: inside the frontier the machine lifts performance substantially, outside it makes things worse — their consultants worked up to a quarter faster and 40 per cent better with AI on suitable tasks, but made 19 percentage points more errors on unsuitable ones. Whoever values the technology should keep both in mind: the real capability and the blurred edge. Economically, Agrawal, Gans and Goldfarb (2018) put it in one formula: artificial intelligence lowers the price of prediction — and almost every decision contains a prediction; cheap prediction makes human judgement more valuable, not obsolete.[9]

The next stage, around which everything turns in 2026, are the agents: systems that accept a brief and carry it out in many steps on their own — planning and booking a journey, finding and fixing a program error, running a research task and writing the report. Technically they combine a language model, access to tools and a memory; economically they shift the offer from information to completed work — and with it the willingness to pay from user subscriptions towards piece rates per finished task. The systems are still error-prone enough that people must check; but the direction explains why model providers

[8] On the continuing cost decline per computing task, the surveys of the Stanford HAI (2026).
[9] Dell'Acqua et al. (2023), Harvard Business School Working Paper 24-013: +12.2 per cent tasks completed, +25.1 per cent speed, +40 per cent quality inside the frontier; −19 percentage points correctness outside; Agrawal, Gans and Goldfarb (2018).

increasingly price by seats and tasks rather than computing minutes — and why the software groups of the application layer are partners and threatened at once.[10]

3 The Economics: General Purpose Technology and Productivity

3.1 General purpose technologies act slowly — then massively

Economically, artificial intelligence belongs to a rare category: the general purpose technologies described by Bresnahan and Trajtenberg (1995) — technologies like steam power, electricity and the computer that can be used everywhere, improve continuously and spawn entire families of follow-on innovation. Such technologies share an inconvenient property: their productivity effect arrives late. David (1990) showed that electrification took roughly four decades to lift factory productivity measurably — factories, processes and occupations first had to be rebuilt around the new technique. Solow (1987) captured the same pattern for IT in his famous sigh that the computer age was visible everywhere except in the productivity statistics. Whoever asks today why the trillions of AI investment have not yet appeared in GDP receives a calm answer from economic history: it was always thus.[11]

How large the eventual effect will be is the field's most contested number. The optimists' case is made by Briggs and Kodnani (2023) of Goldman Sachs: plus seven per cent of world output within ten years, carried by one and a half percentage points of extra productivity growth. The sceptics' case is led by Acemoglu (2025): by his arithmetic the technology will, for the foreseeable future, cut costs in only a small share of all work tasks, which caps the productivity effect at 0.66 per cent over ten years — a tenth of the optimists' figure. Nordhaus (2021), in turn, formulated criteria for a genuine growth acceleration and does not yet find them met in the data. For investors the dispute itself is the message: trillions of value lie between the scenarios, and no price on earth can discount both at once.[12]

A second look at history helps to place today's capital outlay. General purpose technologies have always been accompanied by investment waves that ran ahead of their earnings — the railways of the nineteenth century at times consumed several per cent of a year's British output, and electrification and the telecom networks repeated the pattern. Such waves left two things behind: usable infrastructure that carried decades of growth — and investor losses among those who bought the builders at the crest instead of the later

[10] Classification per provider announcements 2025/26; on the capability frontier again Dell'Acqua et al. (2023).
[11] Bresnahan and Trajtenberg (1995), Journal of Econometrics 65(1), pp. 83-108; David (1990), American Economic Review 80(2), pp. 355-361; Solow (1987), New York Times Book Review of 12 July 1987.
[12] Briggs and Kodnani (2023); Acemoglu (2025), Economic Policy 40(121), pp. 13-58; Nordhaus (2021), American Economic Journal: Macroeconomics 13(1), pp. 299-332.

users. Perez (2002) built her theory of technology waves from this recurring pair. The AI build-out of 2026 — chapter 5 quantifies it — stands in that tradition; the investor question is not whether the infrastructure will be used, but who earns on it: builders, operators or users.[13]

3.2 What the field studies measure

Beneath the macro debate something remarkable has happened quietly: the technology's effect is no longer estimated but measured in controlled experiments — with strikingly consistent results. Brynjolfsson, Li and Raymond (2025) followed more than 5,000 customer-service workers: with AI assistance the number of resolved issues per hour rose by 15 per cent — markedly more among novices and less experienced workers: the machine passes the experience of the best on to the rest. Noy and Zhang (2023) set professionals writing tasks: 40 per cent less time, 18 per cent better grades. Peng, Kalliamvakou, Cihon and Demirer (2023) gave programmers the Copilot assistant: 56 per cent faster completion. Table 2 orders the studies. The common denominator: double-digit productivity gains in precisely the office work that makes up a large share of value creation in modern economies.[14]

Table 2: Controlled studies on the productivity effects of generative AI

Study	Setting	Core result
Brynjolfsson, Li, Raymond (2025)	Customer service, 5,172 workers	+15% resolved issues; novices markedly stronger
Noy and Zhang (2023)	Writing tasks, 453 professionals	−40% time, +18% quality
Peng et al. (2023)	Programming with Copilot	55.8% faster completion
Dell'Acqua et al. (2023)	758 strategy consultants	+25.1% speed, +40% quality — inside the frontier only

Source: Figures as reported by the cited studies; percentages rounded.

Honesty requires the limits of this evidence. The experiments measure bounded tasks over weeks, not whole occupations over years; they measure first effects, which contain novelty and learning; and they measure where output is easy to count — resolved issues, written texts, finished code. How much of it reaches profit-and-loss accounts depends on questions no laboratory answers: is the saved time used productively or competed away? Do all suppliers cut costs at once — and pass the gain to customers in price competition? The history of general purpose technologies suggests the latter: most of the benefit ultimately landed with consumers, not with early adopters. For investors that is no argument against

[13] Perez (2002); on the delay of productivity effects, David (1990).

[14] Brynjolfsson, Li and Raymond (2025), Quarterly Journal of Economics 140(2), pp. 889-942: plus 15 per cent on average; the widely quoted 34 per cent figure for novices stems from the earlier NBER version (Working Paper 31161, 2023) and is no longer reported as a single number in the peer-reviewed final version. Noy and Zhang (2023), Science 381(6654), pp. 187-192; Peng et al. (2023).

the technology — but it is one against assuming that customers' productivity gains translate one for one into suppliers' profits.[15]

Why, then, the gap between such measurements and corporate accounts? A widely discussed study by MIT's NANDA project (2025) found that 95 per cent of in-house AI pilots delivered no measurable profit contribution — a number that went around the world and moved markets. On closer inspection it is less a verdict on the technology than on its adoption: the same study found high success rates where firms bought finished tools instead of building their own, and its definition of success was narrow. The pattern matches David's (1990) lesson exactly: the bottleneck is not the technique but the rebuilding of processes around it. Babina, Fedyk, He and Hodson (2024) show the reward for the rebuilding: firms that invested early and seriously in AI grew measurably faster in revenue and product range than their competitors.[16]

3.3 Labour: displacement of tasks, not of work

The labour question belongs in every honest AI assessment, including one for investors. Eloundou, Manning, Mishkin and Rock (2024) estimate that around 80 per cent of US workers are touched by language models in at least a tenth of their tasks — but exposure is not replacement. History counsels differentiation: Frey and Osborne (2017) once assigned high susceptibility to almost half of US occupations, yet Autor (2015) reminds us why such calculations regularly prove too dark — technology replaces tasks, not occupations; it creates new activities and raises demand elsewhere through income effects. For investors the sober conclusion: the technology's gains will be real, while their distribution — between capital and labour, between firms and customers — is the true open question, and competition settles it faster than forecasts do.[17]

One distribution question concerns investors directly: that between capital and labour. When software takes over work steps, value creation migrates from wage bills to capital income — that is the core of the stock market's enthusiasm and at once its political risk, for societies have historically answered such shifts with taxes, rules and countervailing power. In the short run the effect is barely measurable; in the long run it decides whether today's valuations of the AI winners — which price in a long future of high capital returns — hold. Investors need not answer this question, but they should know that their portfolios assume an answer.[18]

---

[15] On the consumer surplus of technical progress, cf. the discussions in Autor (2015) and Acemoglu (2025).

[16] MIT NANDA (2025), The GenAI Divide; methodological caveats include a narrow success definition and a thin data base, with 67 per cent success for purchased specialist tools; Babina, Fedyk, He and Hodson (2024), Journal of Financial Economics 151, 103745.

[17] Eloundou, Manning, Mishkin and Rock (2024), Science 384(6702), pp. 1306-1308; Frey and Osborne (2017); Autor (2015), Journal of Economic Perspectives 29(3), pp. 3-30.

[18] On the debate over the machine take-over of work, Autor (2015) and Acemoglu (2025).

4 The Value Chain: Chips, Cloud, Models, Applications

4.1 Four layers, four business logics

Whoever wants to invest in artificial intelligence should first grasp that “AI” is not one market but a chain of four very different ones. At the bottom lies the chip layer: the processors themselves, their production, and the machines of that production — an oligopolistic business with extreme entry barriers. Above it lies the infrastructure layer: data centres with their power, cooling and network equipment, and the cloud groups that run them and rent out computing. On the third layer sit the model providers, who turn computing and data into the actual AI systems. At the top, finally, the application layer: software that brings AI into work processes — from the office copilot to driving software. Table 3 orders the layers and their leaders; the business logics could hardly differ more.[19]

Table 3: The four layers of the AI value chain

Layer	What is earned there	Leading listed names
1. Chips	Processors, memory, production, machines	NVIDIA, Broadcom, TSMC, ASML, SK Hynix, Micron
2. Infrastructure	Data centres, power, cooling, cloud rental	Microsoft, Alphabet, Amazon, Vertiv, Eaton
3. Models	AI systems by subscription and usage	OpenAI, Anthropic (private); Alphabet, Meta
4. Applications	AI in work processes and products	Microsoft, Palantir, SAP, Salesforce, Adobe

Source: Own scheme; company assignment per chapter 7.
Notes: The layers overlap: the large platform groups operate on two to four layers at once.

The chain's economic pattern was anticipated by Teece (1986): when an innovation is hard to protect, its profit flows to the owners of the scarce complementary assets. That is exactly what is on display. Models are quickly replicated — the gap between American and Chinese frontier models has shrunk to a few per cent by the Stanford AI Index's measurement — but the processors they run on come four-fifths from a single supplier, made almost exclusively by a single contract manufacturer on machines from a single lithography group. The scarcity sits at the bottom of the value chain — and with it, so far, the profit: NVIDIA earns a gross margin of around 75 per cent on every system sold, while the model providers carry billions in losses despite exploding revenues.[20]

The layers also differ in their capital intensity — for investors half of the valuation question. The chip layer earns high margins on capital tied up in factories that last decades; the infrastructure layer is currently turning from a capital-light software business into a

[19] Layer scheme per common industry practice; on the economics of complementary assets, foundationally Teece (1986).
[20] Teece (1986), Research Policy 15(6), pp. 285-305; Stanford HAI (2026): performance gap of leading US and Chinese models around 2.7 per cent; NVIDIA: GAAP gross margin around 75 per cent (fiscal 2025 and first half of 2027).

capital-heavy plant business — the depreciation of today's data centres will shape the hyperscalers' income statements for years; the model layer burns capital in the hope of later platform rents; and the application layer is classic software with high gross margins and little capital need. Whoever weights the chain in a portfolio thereby chooses a capital profile — from the factory to the licence.[21]

4.2 The chip layer: the needle's eye of the revolution

The chip layer's numbers are the backbone of the whole megatrend, and they deserve their own figure. NVIDIA's data-centre business grew from USD 15 billion in fiscal 2023 to 194 billion in fiscal 2026 — a thirteenfold increase in three years — and reached 164 billion in the first half of fiscal 2027 alone (Figure 1). Behind it stands an ecosystem investors must know: the contract manufacturer TSMC produces nearly all of the world's leading-edge chips; the Dutch equipment maker ASML is the sole supplier of the lithography machines without which that production is impossible; the memory makers SK Hynix, Samsung and Micron supply the high-bandwidth memory that accompanies every processor; and Broadcom designs the hyperscalers' custom chips as a counterweight to NVIDIA dependence. Each link is a quasi-monopoly or tight oligopoly — the chain is as strong as it is valuable.[22]

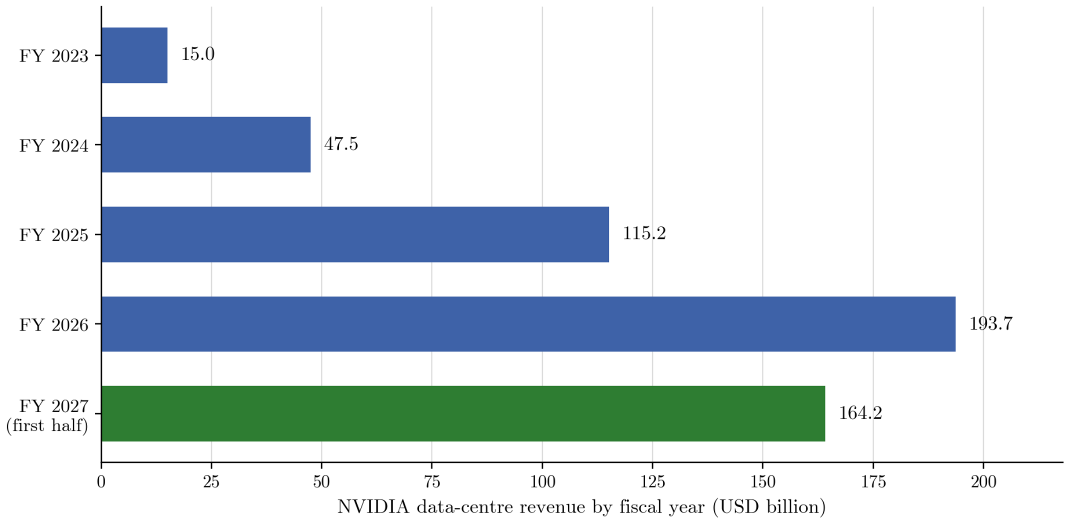

Figure 1: NVIDIA data-centre revenue by fiscal year (the fiscal year ends in late January), in USD billion; green: first half of fiscal year 2027. Source: NVIDIA (2026), quarterly reports; own presentation.

The chip layer carries a valuation observation that surprises many: despite the multiplication of its share price, NVIDIA has become cheaper by classical yardsticks. Its

[21] On the shift in hyperscaler balance sheets: combined investment of USD 735 billion planned for 2026 by four groups (chapter 5).

[22] NVIDIA (2026): data-centre revenues of fiscal years 2023 to 2026: 15.0 / 47.5 / 115.2 / 193.7 billion US dollars; first half of fiscal 2027: 164.2 billion.

price-earnings ratio on expected profits fell from around 40 in mid-2025 to roughly 18 to 24 at the end of August 2026 — earnings grew faster than the price. That separates the market leader fundamentally from the emblem of the dotcom bubble: Cisco traded at 130 to 230 times its earnings in March 2000, depending on the method. This does not disarm the bubble question — chapter 6 poses it at length — but it shifts it: the risk lies less in the price of today's profits than in their durability, should the data-centre build-out one day pause.[23]

The counterweight to NVIDIA's dominance is being built by its largest customers themselves. Alphabet has developed its own processors (TPU) for years and trains its Gemini models largely on them; Amazon has its Trainium chips made and equips, among others, the data centres of its partner Anthropic with them; Microsoft and Meta run their own programmes, and OpenAI has secured AMD as a second supplier complete with an equity option. The big winner of this custom-chip wave is Broadcom, which designs the hyperscalers' bespoke processors and has risen into a valuation league of its own. For investors this qualifies the monopoly question over time: NVIDIA's lead is real, but its biggest customers are visibly working to narrow it — one more reason to map the layer through several names rather than one.[24]

4.3 Models and applications: revenue miracles, profits pending

On the model layer the fastest revenue curves in economic history are forming — with the profit question still open. OpenAI reports an annualised revenue rate above USD 40 billion, Anthropic of 65 billion — up from 30 billion as recently as the spring of 2026 — and Microsoft's disclosed AI business grew from an annual rate of 13 to 37 billion within fifteen months. At the same time, training and running the models consumes computing costs of similar magnitude; the two leading model firms are not listed, and their losses are carried by venture capital and the participating groups. For investors this means two things: access to the model layer runs through the listed parents and partners — Alphabet, Microsoft, Amazon, Meta — and the layer's valuation question has not yet been answered by a market, only by private funding rounds.[25]

The application layer is most tangible where AI moves into existing software and is paid for by subscription: Microsoft's Copilot in the office suite, SAP's assistant in the enterprise system, Adobe's image tools, Salesforce agents in customer business, Palantir's platforms in government and industry. Here the investor rule of general purpose

[23] Price-earnings ratios late August 2026: NVIDIA roughly 18 to 24 on expected earnings (depending on estimate source); Cisco in March 2000: roughly 130 to 230 (depending on method).

[24] AMD-OpenAI agreement of October 2025; Broadcom market capitalisation around USD 1.75 trillion (late August 2026).

[25] OpenAI: annual rate above USD 40 billion (August 2026); Anthropic: 65 billion (July 2026); Microsoft: AI business at a 37 billion annual rate (April 2026; January 2025: 13 billion); OpenAI valuation of USD 852 billion in the March 2026 funding round.

technologies applies: it is not the technology that wins but the access to paying customers; whoever already supplies hundreds of millions of workplaces with software can monetise AI features at little extra cost, while pure newcomers must win every customer dearly. That explains why application profits have so far accrued disproportionately to the established platforms — and why chapter 7 starts there.[26]

What application looks like in practice, three examples from corporate daily life show. In sales, assistant systems summarise customer calls, maintain the records and draft the follow-up — vendors advertise the benefit in saved working hours per salesperson. In software development, programming assistants now co-write a substantial share of the large houses' new code — the Copilot study of chapter 3 measures the effect. And in customer service, language systems answer standard requests in full while people take the exceptions — the field of the Brynjolfsson study. Common to all three is the revenue model: a surcharge per seat and month on software already in place — unspectacular, recurring, and in sum perhaps the most reliable revenue source of the whole field.[27]

5 The Investment Boom and Its Energy Question

5.1 A building programme of historic scale

The build-out's numbers defy comparison. Microsoft, Alphabet, Amazon and Meta together invested USD 140 billion in 2023, 217 in 2024, then 358 in 2025 — and plan around 735 billion for 2026 (Figure 2); adding Oracle and the rest, Morgan Stanley expects a good 800 billion for 2026 and, for 2027, more than a trillion US dollars for the first time, while the market researchers of Dell'Oro see worldwide data-centre spending near the trillion mark as early as 2026. Crucial for the bubble question of the next chapter is the funding side: these sums come overwhelmingly from the running profits of the world's most profitable companies — not, as in the telecom bubble of the millennium, from the debts of barely profitable network operators. Whether they will earn a return is open; that they are affordable is not.[28]

[26] Framing per Teece (1986); on platform economics also Katz and Shapiro (1985), American Economic Review 75(3), pp. 424-440.
[27] Examples per vendor statements by Microsoft, Salesforce and SAP; measurements in chapter 3.2.
[28] Capital expenditure (cash basis) 2023/2024/2025: 140.1 / 217.3 / 357.5 billion US dollars; 2026 plans: Microsoft around 190, Alphabet 195-205, Amazon 220, Meta 130-145 billion; Morgan Stanley (2026), forecast of May 2026; Dell'Oro Group (2026), release of February 2026.

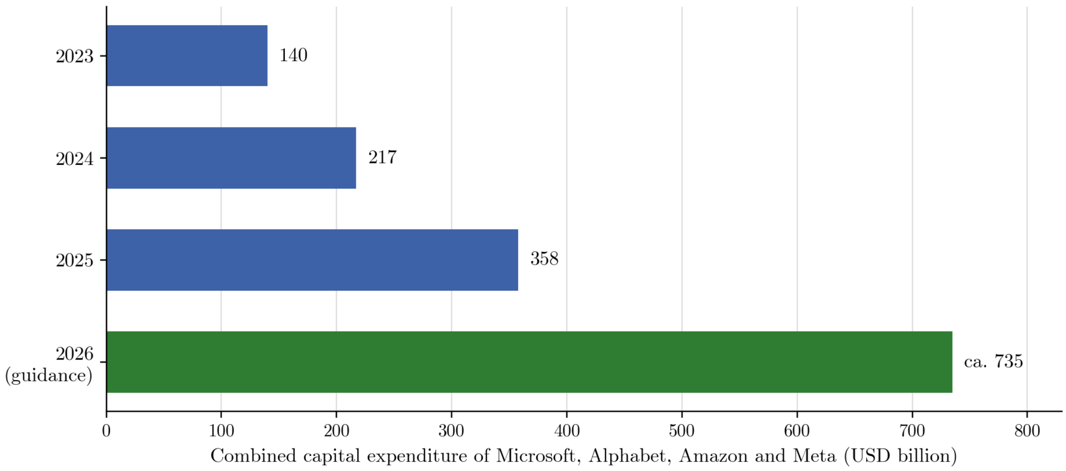

Figure 2: Combined capital expenditure of Microsoft, Alphabet, Amazon and Meta (cash basis), in USD billion; green: company guidance for 2026 (midpoints of ranges). Source: company reports and guidance, as of July 2026; own presentation.

Remarkable is the pace at which the plans themselves age. In July 2026 alone, three of the four groups raised their full-year plans by a combined high double-digit billion amount — Amazon citing, among other things, higher memory prices: a hint that the boom drives its own costs. The funding is shifting too: alongside cash, leasing structures and the first large bond programmes are appearing, and specialised operators such as CoreWeave or Nebius build with debt against pre-contracts from the giants. This is still far from the debt load of the telecom bubble — but it is the direction in which a cash-funded boom would become a credit-funded one. Investors should watch this line: it is the quietest early indicator of chapter 6.[29]

5.2 Power as bottleneck: the physical limit of the digital revolution

The build-out meets a physical limit: electricity. The International Energy Agency puts the power consumption of all data centres at 415 terawatt-hours in 2024 — about one and a half per cent of world consumption — and expects a doubling to around 945 terawatt-hours by 2030, nearly half of it in the United States. The hyperscalers therefore secure power stations the way other firms secure office space: Microsoft is having the retired Three Mile Island reactor restarted for a twenty-year power purchase (target late 2027), Amazon has secured almost two gigawatts of nuclear power, Google orders small modular reactors for the 2030s. For investors a second tier of beneficiaries emerges beside the chips: the suppliers of power and cooling.[30]

[29] Guidance increases: Alphabet to 195-205, Amazon to 220, Meta to 130-145 billion US dollars (each July 2026).

[30] IEA (2025), Energy and AI: 415 terawatt-hours in 2024, around 945 in 2030 (then roughly three per cent of world power, own calculation), US share 45 per cent; Microsoft/Constellation: restart of the Crane reactor (Three Mile Island), target late 2027.

How tight this bottleneck already is, the order books show. The data-centre specialist Vertiv — power supply and cooling — reports a backlog of around USD 15 billion with AI-driven order intake up 80 per cent; GE Vernova has gas turbines above 100 gigawatts under contract and its production slots practically sold out to the end of the decade; Siemens Energy sits on an order book of EUR 162 billion, more than one and a half years of revenue. Such numbers are the sober counterpart to model euphoria: whatever becomes of individual AI applications — transformers, turbines and cooling systems are sold, delivered and paid. The picks-and-shovels rule of the gold rush applies literally: the surest earnings go to whoever supplies the equipment.[31]

For European investors the home role deserves a sober look. The great models, chips and clouds come from America and Asia; Europe's strength lies one layer down — in production equipment (ASML), enterprise software (SAP) and above all the electrical and energy technology without which no data centre goes on line: Siemens Energy, Schneider Electric, ABB and Legrand earn on the build-out on every continent. It is the same division of labour as in the mobility paper: Europe supplies the equipment of the revolution more often than its platforms. A portfolio that maps the megatrend cannot avoid US dollar names — but it finds liquid, dividend-paying links of the same chain in Europe.[32]

How far the build-out has already turned into binding contracts, Table 4 shows with the largest publicly known agreements of 2025/26 — from OpenAI's cloud contract with Oracle through NVIDIA's financing guarantees to the hyperscalers' power-station deals. The list is proof and warning at once: proof, because real capacity is being contracted with real terms; warning, because the same few names appear on both sides of many contracts — the subject of the next chapter.[33]

Table 4: The largest contracts of the AI build-out, 2025/26

Contract	**Scale**	**Content**
OpenAI – Oracle (Sept. 2025)	ca. USD 300bn / 5 years	Cloud computing capacity (Stargate)
NVIDIA – OpenAI (Mar./Aug. 2026)	ca. 30bn stake; up to 105bn guarantee	Funding round share; Ohio data centre
OpenAI – AMD (Oct. 2025)	6 gigawatts of computing	Chips plus equity options for OpenAI
Microsoft – Constellation	835 megawatts / 20 years	Three Mile Island restart (target 2027)
Amazon – Talen	1.9 gigawatts	Nuclear power at the Susquehanna site
Google – Kairos	500 megawatts	Small modular reactors from ~2030

Source: Company announcements and press reports 2025/26; amounts rounded, partly framework intentions rather than firm offtake.

[31] Vertiv: backlog around USD 15 billion, AI order intake +80 per cent (spring 2026); GE Vernova: around 100 gigawatts of gas turbines under contract (Q1 2026); Siemens Energy: order book of EUR 162 billion (third quarter of fiscal 2026).

[32] Cf. the European names in Table 8; on the division of labour, Rupprechter (2026g).

[33] Compilation from public company statements and press reports 2025/26; individual references in Table 4.

6 Bubble or Foundational Build-Out? The Dotcom Comparison

6.1 What argues for a bubble

The bubble question is not polemics but due diligence — and the warners' arguments are serious. First, concentration: the ten largest US names make up more than 40 per cent of the S&P 500, more than ever and clearly more than at the peak of the dotcom bubble, when it was around 27 per cent; the seven large technology names alone weigh a good third of the index. Second, the valuation of the whole market: the cyclically adjusted price-earnings ratio stands at 38 to 40, the second-highest reading in history — exceeded only by March 2000. Third, the circular deals: the chip market leader takes billion-dollar stakes in its largest customer OpenAI and guarantees financing for its data centres, while that customer signs three-hundred-billion contracts with the cloud provider Oracle — revenues that condition one another have played a part in every bubble in history.[34]

Add the sobering news from the application front. The MIT study with its 95 per cent of failed pilots struck a nerve in 2025, and July 2026 delivered a taste of the drop: after disappointing numbers from individual suppliers, chip stocks worldwide lost more than a trillion US dollars of market value within days before prices recovered. Perez (2002) would recognise all of it at once: every technological revolution passed through an installation phase in which capital runs ahead of the technology, expectations overshoot and a bubble forms — which then bursts without refuting the revolution. The question is never whether the technology holds, but whether today's prices already capture the earnings of the day after tomorrow.[35]

A bubble element of its own kind are the private valuations. OpenAI was valued at USD 852 billion in its March 2026 funding round — more than any German company — without ever having published an audited annual report or faced the daily verdict of exchange trading; Anthropic follows with three-digit billion values. Such prices form among few professional buyers with strategic interests — the participating chip and cloud groups are at once suppliers of the companies being valued. This need not end badly; but it means that a sizeable share of AI market value has never been through a bear market. The announced OpenAI listing becomes the litmus test: it brings the largest private valuation in history before the public for the first time.[36]

---

[34] Top-10 share: 40.7 per cent at the end of 2025 against roughly 23-27 per cent in 2000; Magnificent 7: 34.3 per cent; CAPE around 38-40 against 44.2 in March 2000; NVIDIA-OpenAI: stake of around USD 30 billion and financing guarantees up to 105 billion (August 2026); OpenAI-Oracle: around 300 billion over five years.

[35] MIT NANDA (2025): a project report of Project NANDA at the MIT Media Lab (Challapally, Pease, Raskar and Chari), not a peer-reviewed publication; the 95 per cent figure rests on 300 non-randomly selected case studies and measures the absence of documented P&L impact, not technical failure. Chip sell-off of late July 2026: more than a trillion US dollars of market value lost within days, followed by recovery to new highs; Perez (2002).

[36] OpenAI: USD 852 billion valuation (March 2026), listing discussed for 2027.

6.2 What argues against the simple bubble thesis

Against the simple equation “like 2000” stand hard differences, set out in Table 5. Today's heavyweights are not revenue hopes but the most profitable companies in history; they fund the build-out from cash flow rather than debt; their valuations, measured against earnings, are a fraction of the dotcom icons' — the market leader NVIDIA even trades below the market average on expected earnings; and the demand for computing is not a projection but a backlog: the equipment makers' order books are full for years, and the cloud providers report more demand than they can connect. In short: the infrastructure is being built by solvent buyers for paying use — that separates 2026 from 2000 in substance, not merely in degree.[37]

Table 5: The dotcom bubble of 2000 and the AI boom of 2026 compared

Measure	March 2000	Late August 2026
Top-10 share of the S&P 500	roughly 23-27%	40.7% — above 2000
CAPE (smoothed P/E)	44.2 (record)	38-40 (second-highest)
P/E of the emblem	Cisco: 130-230	NVIDIA: 18-24 (expected earnings)
Funding of the build-out	Debt, new issues	Running profits of the platforms
Revenues of the newcomers	small, often fictitious	OpenAI >40, Anthropic 65bn rate
Warning signs	Everything euphoria	Circular deals, pilot disappointment

Source: RBC (top-10 shares); Shiller CAPE series; P/E ranges per chapter 4.2; company statements; own compilation.
Notes: Price-earnings ratios depend on method; ranges rather than point values.

The strongest single piece of counter-evidence is the contracted demand backlog. Microsoft discloses contracted, not yet delivered obligations of USD 678 billion — mostly booked cloud capacity — 84 per cent more than a year earlier; the cloud providers report in unison that the bottleneck is not demand but grid connections and chips. A bubble in the narrow sense — prices without demand — looks different. What remains is the second-derivative risk: prices discount not demand but perpetually growing demand; a mere slowdown would suffice to compress the valuation chain from the equipment makers to the power producers.[38]

6.3 How a turning point would show itself

Because the dispute cannot be settled in theory, a list of observable early indicators helps. First, the hyperscalers' investment plans: the first broad cut in full-year plans would be the signal to which the whole supply chain reacts. Second, the rental prices of computing: if prices per GPU-hour fall across the board, the market tips from a landlord's to a tenant's

[37] Comparison figures in Table 5; Microsoft's contracted backlog (commercial RPO) of USD 678 billion (+84 per cent, July 2026) as an example of the demand overhang.
[38] Microsoft: commercial remaining performance obligations of USD 678 billion, +84 per cent (July 2026).

market. Third, the equipment makers' order books — Vertiv, GE Vernova, Siemens Energy — as the hardest, because contractual, measure. Fourth, the subscription numbers of the applications, which show whether final demand carries. And fifth, the funding mix: if the build-out migrates from cash to debt, the drop grows taller. None of these indicators stood at red at the end of August 2026 — the July setback showed how fast sentiment turns, and the recovery how deep the demand runs.[39]

The honest verdict of the bubble test is: both at once. The foundational build-out is real — chips, power and data centres are needed, paid for and used, and the productivity evidence of chapter 3 is solid. And the exaggeration is real — in the concentration, in the circular structures, in the valuations of the second and third tier, in private model valuations never tested by a public market. Perez (2002) teaches that both truths regularly arrive together: the railways were built AND their shareholders expropriated; the internet came AND the Nasdaq fell 78 per cent. For investors the conclusion is not abstinence but construction: build the position to survive a valuation storm without leaving the trend — chapters 7 to 10 spell that out.[40]

7 Single Names along the Chain

7.1 Selection principles

The principles of the megatrend series apply. First: named are exclusively large and mid caps — above ten billion US dollars of market value, or two to ten billion; small caps remain unnamed, for in the AI field especially the list of crashed hopefuls is long. Second: named is whoever runs a clearly identifiable AI business. Third: naming is stock-taking, not a recommendation to buy or sell. And fourth, Bessembinder's (2018) reminder: since the equity returns of whole decades stem from few winners, picking the right theme guarantees nothing — one spreads across the chain instead of betting on one name. A peculiarity of this field: unlike in mobility, almost all relevant names here are large caps — the chain is already consolidated.[41]

7.2 Chips and production

Table 6 assembles the chip layer. Four observations guide the reading. First, the top: NVIDIA is, at around USD 5.3 trillion, the world's most valuable company; its only serious GPU rival AMD has gained strategic weight through the OpenAI partnership of October 2025. Second, the invisible monopolies: TSMC as maker of nearly all leading-edge chips

[39] Indicator list: own compilation based on chapters 5 and 6.
[40] Nasdaq Composite: decline of roughly 78 per cent from March 2000 to October 2002; Perez (2002).
[41] Bessembinder (2018), Journal of Financial Economics 129(3), pp. 440-457; size classes as in Rupprechter (2026g).

and ASML as sole supplier of the finest lithography are the two bottlenecks through which every AI ambition in the world must pass. Third, memory: high-bandwidth memory has lifted SK Hynix and Micron into the first valuation tier — Micron crossed the trillion mark in 2026. Fourth, the equipment makers behind it all: Applied Materials, Lam Research, KLA and Tokyo Electron earn on every new plant, whoever's chips are made there.[42]

Table 6: Chips and production: large caps in the AI megatrend (late August 2026)

Company	USD bn	Business idea in the megatrend	Trading venue
NVIDIA	5,250	AI processors and software ecosystem (CUDA)	Nasdaq
TSMC	2,165	Contract maker of nearly all leading chips	Taipei; ADR NYSE
Broadcom	1,754	Custom AI chips and network technology	Nasdaq
Samsung Electronics	1,223	Memory, production, electronics	Seoul
Micron	1,053	High-bandwidth memory (HBM)	Nasdaq
SK Hynix	851	HBM market leader for AI processors	Seoul
AMD	760	AI accelerators; OpenAI partner since 2025	Nasdaq
ASML	651	Sole supplier of EUV lithography	Amsterdam; Nasdaq
Intel	473	Processors; contract production in build-up	Nasdaq
Lam Research	378	Production equipment (etch, deposition)	Nasdaq
Applied Materials	366	Broadest semiconductor equipment maker	Nasdaq
Arm	255	Chip architecture licences (server, mobile)	Nasdaq
KLA	229	Measurement and inspection equipment	Nasdaq
Marvell	195	Custom chips and interconnect technology	Nasdaq
Qualcomm	175	On-device AI; new: data-centre chips	Nasdaq
Tokyo Electron	159	Production equipment (Japan)	Tokyo

Source: Market capitalisations (rounded) of late August 2026; own compilation.
Notes: No recommendation to buy or sell; selection without claim to completeness.

7.3 Platforms and software

Table 7 shows the platform and application layer — and with it the companies that are the largest positions of any world portfolio anyway. The four hyperscalers connect all layers: they build the data centres, develop their own chips as a counterweight to NVIDIA, run their own models and sell the applications; Alphabet, with Gemini and its own TPU chips, is the most vertically complete group of the field, Microsoft monetises through Copilot and the OpenAI partnership, Meta funds open models from its advertising business,

[42] Market capitalisations of late August 2026; AMD-OpenAI agreement of 6 October 2025 over six gigawatts of computing plus equity options.

Amazon rents out the most computing through AWS. Behind them stand the application winners: Palantir — the best performer of the S&P 500 in 2025 and by now worth close to half a trillion —, SAP with AI in the enterprise core, Salesforce, ServiceNow, Adobe; alongside them the Chinese groups Tencent, Alibaba and Baidu with model worlds of their own and the familiar political discount.[43]

Table 7: Platforms and software: large caps in the AI megatrend (late August 2026)

Company	USD bn	Business idea in the megatrend	Trading venue
Alphabet	4,193	Gemini models, own TPU chips, cloud, Waymo	Nasdaq
Microsoft	3,813	Azure, Copilot, OpenAI partnership	Nasdaq
Amazon	2,873	AWS cloud, own chips, Anthropic stake	Nasdaq
Meta	1,472	Open Llama models, AI advertising	Nasdaq
Tencent	521	China's super-app with Hunyuan models	Hong Kong
Palantir	448	Deployment platforms for AI in firms, state	Nasdaq
Oracle	435	Cloud for large customers (OpenAI deals)	NYSE
Alibaba	296	Qwen models, China's largest cloud	NYSE; Hong Kong
SAP	256	AI assistant Joule in the enterprise core	Xetra; NYSE
CrowdStrike	224	AI-driven cybersecurity	Nasdaq
IBM	222	watsonx platform, AI consulting	NYSE
Salesforce	211	Agentforce: AI agents in customer business	NYSE
ServiceNow	150	AI in corporate work processes	NYSE
Adobe	116	Generative tools (Firefly)	Nasdaq
Snowflake	114	Data platform with AI services	NYSE
Baidu	33	Ernie models, Apollo Go robotaxi	Nasdaq; Hong Kong

Source: Market capitalisations (rounded) of late August 2026; own compilation.
Notes: No recommendation to buy or sell. Tencent is primarily listed in Hong Kong; US trading only over the counter.

7.4 Infrastructure and energy

Table 8, finally, belongs to the third tier — the suppliers of servers, networks, power and cooling, long the most overlooked storey of the theme. Here stand the server makers Dell and Super Micro, the network specialist Arista and the interconnect maker Amphenol; the electrical groups Eaton, Schneider Electric, Siemens Energy, GE Vernova and Vertiv, whose order books chapter 5 described; the data-centre landlords Equinix and Digital Realty; the power producers Constellation, Vistra and NextEra, beneficiaries of the hyperscalers' hunger for power stations; and the new specialised AI cloud providers CoreWeave and Nebius, both recently risen to large caps. This storey has a double appeal:

[43] Palantir: market capitalisation around USD 448 billion, listed on Nasdaq (transfer from the NYSE in November 2024); 2025 share gain around 130 per cent.

it earns on the build-out regardless of whose models win — and it carries far less weight in the world indices than the platforms, so it genuinely complements a portfolio.[44]

Table 8: Infrastructure and energy: large caps in the AI megatrend (late August 2026)

Company	USD bn	Business idea in the megatrend	Trading venue
Dell Technologies	295	AI servers and storage for enterprises	NYSE
Arista Networks	246	High-speed data-centre networks	NYSE
GE Vernova	243	Gas turbines, grid equipment; slots sold out	NYSE
Schneider Electric	196	Power and cooling for data centres	Paris
Amphenol	196	Connectors and cabling of AI servers	NYSE
NextEra Energy	171	Largest US renewables producer	NYSE
Eaton	162	Power distribution from grid to rack	NYSE
Siemens Energy	146	Grid build-out; order book EUR 162bn	Xetra
Equinix	103	World's largest data-centre landlord	Nasdaq
Vertiv	99	Power supply and cooling; AI specialist	NYSE
Constellation	98	Largest US nuclear fleet (Microsoft deal)	Nasdaq
Digital Realty	69	Data-centre property worldwide	NYSE
Nebius	57	Specialised AI cloud (GPU rental)	Nasdaq
Vistra	46	Power producer with nuclear share	NYSE
CoreWeave	46	Specialised AI cloud; close to NVIDIA	Nasdaq
Legrand	42	Electrical infrastructure in the data centre	Paris
Super Micro	23	AI server systems at short lead times	Nasdaq

Source: Market capitalisations (rounded) of late August 2026; own compilation.
Notes: No recommendation to buy or sell. Deliberately unnamed: small caps in AI software and speech technology.

Just as important is what the tables omit. The two best-known names of the field — OpenAI and Anthropic — are not listed; their reported valuations of several hundred billion US dollars come from private funding rounds, not from daily exchange trading with short sellers and disclosure duties. An OpenAI listing is discussed for 2027 and would be the largest in history; until then, investor access runs only indirectly through Microsoft, NVIDIA, Amazon and Alphabet, which hold stakes in the model firms. Equally absent by design: the small listed AI hopefuls whose market values have fallen below two billion US dollars — they stand as reminders that in the application business distribution decides victory, not technology.[45]

On access, the mobility paper's guidance applies in kind: US names and Xetra stocks trade everywhere; the Asian links of the chain are reached through Seoul, Tokyo and Taipei or, more conveniently, through depositary receipts — TSMC via its NYSE listing; Tencent trades primarily in Hong Kong. Whoever prefers not to manage single names across several currency areas finds the fund alternative in the next chapter. In every case, positions in

[44] CoreWeave: IPO March 2025, around USD 46 billion at the end of August 2026; Nebius: around 57 billion; CoreWeave's planned takeover of Core Scientific failed on the shareholder vote in October 2025.
[45] OpenAI: funding round of USD 122 billion at a valuation of 852 billion (March 2026), listing discussed for 2027; Anthropic: stakes held by Amazon and Alphabet, among others.

Taiwan and Korea carry, beside business risk, the geopolitical concentration risk of chapter 9 — whoever holds them should do so knowingly.[46]

Three cross-observations on the tables. First, geography: the chain runs through the USA, Taiwan, South Korea, Japan and the Netherlands; Europe supplies few but indispensable links in ASML, SAP, Siemens Energy and Schneider. Second, China's AI world is largely separate from the Western one — its own chips, models and applications; whoever wants it maps it deliberately through the platforms of Table 7 and carries the political risk knowingly. Third, unlike in the mobility chapter, the failed small caps are not missing here because the field is kinder, but because its winners grew so large that the chain consists almost entirely of large caps — the consolidation that still lies ahead of mobility is behind the AI chain.[47]

8 ETFs on Artificial Intelligence

8.1 The checklist

For fund selection the series' checklist applies, repeated here only briefly (the full reasoning is in Rupprechter, 2026g): physical replication rather than swap deals; fund size of at least EUR 0.2 billion as protection against closure; a management fee preferably below half a per cent; an Article 8 classification under the EU disclosure regulation as a minimum profile — with the warning, argued in Rupprechter (2026b), not to mistake the label for impact; an established issuer; the UCITS framework; continuous exchange trading. Two peculiarities of the AI field come on top. First: currency-hedged share classes are practically non-existent here too — the dollar bloc dominates the indices, and the currency risk stays with the investor. Second: the warning of Ben-David, Franzoni, Kim and Moussawi (2023) that thematic ETFs are launched on average near peak sentiment and lose roughly 30 per cent risk-adjusted thereafter weighs especially heavily here — more thematic funds than ever were launched on the AI label in 2024/25.[48]

8.2 The market: several funds pass — with very different profiles

Unlike in the mobility field, the AI offering has matured: several funds under European law clear the size threshold comfortably (Table 9). By far the largest is the Xtrackers Artificial Intelligence & Big Data at almost eight billion euros — physical, cheap, Article 8; its profile surprises, though: the largest positions are broad technology and user names up to and including Walmart — it buys the AI users, not only the equipment makers.

[46] Trading venues per Tables 6 to 8; on geopolitics, chapter 9.3.
[47] Cf. the concentration figures in chapter 6.1 and Bessembinder (2018) on return concentration.
[48] Ben-David, Franzoni, Kim and Moussawi (2023), Review of Financial Studies 36(3), pp. 987-1042; Rupprechter (2026b); Rupprechter (2026g), Table 8.

Whoever wants the infrastructure specifically finds a fast-growing specialist in the iShares AI Infrastructure launched in late 2024; whoever wants the chip chain pure takes the established route of the VanEck Semiconductor with more than seven billion euros. In between stand the robotics-and-AI blends of iShares and Amundi and the AI funds of L&G and WisdomTree — all above the threshold, all physical, each with its own emphasis.[49]

Table 9: AI ETFs under European law compared (late August 2026)

ETF (short name)	Assets (EUR m)	Fee p.a.	Replication	SFDR	0.2bn met?
Xtrackers AI & Big Data	7,830	0.35%	physical	Art. 8	yes
iShares Automation & Robotics	4,860	0.40%	physical	Art. 8	yes
L&G Artificial Intelligence	1,900	0.49%	physical	Art. 8	yes
WisdomTree Artificial Intelligence	1,240	0.40%	physical	Art. 8	yes
Amundi MSCI Robotics & AI	1,200	0.40%	physical	Art. 8	yes
iShares AI Infrastructure	1,090	0.35%	physical	Art. 8	yes
ARK AI & Robotics (active)	275	0.75%	physical	Art. 8	yes
Global X Robotics & AI	123	0.50%	physical	Art. 6	no
Invesco AI Enablers	35	0.35%	physical	Art. 8	no

Source: Issuer data and fund databases, late August 2026; Xtrackers: IE00BGV5VN51 (Xetra XAIX); iShares Automation & Robotics: IE00BYZK4552; L&G: IE00BK5BCD43; WisdomTree: IE00BDVPNG13; Amundi: LU1861132840; iShares AI Infrastructure: IE000X59ZHE2; all UCITS-compliant, accumulating, continuously tradable on Xetra and in London, Milan and Zurich.
Notes: No currency-hedged share classes disclosed. Chip alternative: VanEck Semiconductor (IE00BMC38736), EUR 7.3-8.5bn, 0.35%, physical, Art. 8.

Choosing among the passing funds is a choice of emphasis, not of quality. The large Xtrackers fund is the convenient all-in-one, but it overlaps most with a world ETF — whoever buys it partly doubles what he already owns. The infrastructure fund and the semiconductor fund pick out precisely the storeys under-represented in the world index — they complement a world portfolio more effectively, at the price of higher swings. The robotics blends spread most broadly across industrial and technology names and depend least on the Magnificent 7. The active ARK fund, finally, is a manager bet at manager fees. The basic decision remains that of the mobility paper: first the world portfolio, then the measured, deliberate sharpening — not the other way round.[50]

[49] Volumes late August 2026: Xtrackers EUR 7.83 billion; iShares Automation & Robotics around 4.9; L&G around 1.9; WisdomTree 1.24; Amundi 1.20; iShares AI Infrastructure 1.09; VanEck Semiconductor 7.3 to 8.5 billion.

[50] Overlap logic: the five largest names of the Xtrackers fund (Apple, Microsoft, Meta, Walmart, NVIDIA) rank among the largest positions of every world index; cf. chapter 9.1.

Execution, finally, follows the series' craft: all funds named trade continuously on Frankfurt's Xetra and in London, Milan or Zurich; the management fee is deducted daily from fund assets, and purchases add spread and order fee. Building the position in instalments or by savings plan is not a soothing formula here but risk technique: few market segments swing harder around their trend, and July 2026 — a trillion-dollar loss and recovery within weeks — showed what a badly chosen single entry date can cost. The funds' accumulating design reinvests income within the fund; taxation follows the investor's country of residence and belongs in the conversation with the tax adviser.[51]

9 Risks: Concentration, Regulation, Power, China

9.1 The concentration risk starts in one's own portfolio

The most underrated AI risk sits not in California but in one's own portfolio. Because the AI winners are at the same time the largest companies in the world, every world index has long been half an AI index: a good third of an S&P 500 fund and roughly a quarter of a world equity fund sit in the seven large technology names. Whoever additionally buys an AI ETF whose top positions are the same names raises not the diversification but the bet — often unnoticed. The first risk check of an AI investor is therefore an inventory: how much NVIDIA, Microsoft and Alphabet do I already own? Only then is the question meaningful whether and how much should be added. In Markowitz's (1952) terms: the addition must complement the portfolio, not duplicate it — which argues for the infrastructure and chip storeys and against doubling the platforms.[52]

The inventory can be done with on-board means. Whoever holds an S&P 500 fund owns through it roughly a third of Magnificent 7; a world equity fund comes to about a quarter; adding the AI-near names outside the seven — Broadcom, TSMC, ASML and the equipment makers — quickly adds further percentage points. An investor with 80 per cent world index and 10 per cent AI theme fund thus really holds closer to 30 than 10 per cent AI complex. That is no forbidden number — but it should be intended. As the series' rule of thumb: the deliberate overweight beyond the world index stays in the single-digit percentage range of liquid wealth; whoever wants more should write it down and be able to justify it.[53]

[51] Trading venues and fees per issuer data, late August 2026; on the July 2026 setback, chapter 6.1.
[52] Magnificent 7: around 34 per cent of the S&P 500 (end of 2025); Markowitz (1952), Journal of Finance 7(1), pp. 77-91.
[53] Shares: Magnificent 7 around 34 per cent of the S&P 500 (end of 2025); world-index share varies by provider, order of magnitude one quarter.

9.2 Regulation: Europe postpones, America accelerates

Regulation is developing in opposite directions on the two sides of the Atlantic. With its AI regulation the EU created the world's first comprehensive AI statute — in force since August 2024, with staggered duties: bans on certain practices since February 2025, duties for providers of large models since August 2025, transparency duties and full supervision since August 2026. The sharpest obligations for high-risk systems, however, were postponed by more than a year through an amending regulation in the summer of 2026 — Europe's answer to the worry that regulation might cost it the race. The USA went the other way: the new administration repealed the previous executive order in early 2025 and in July 2025 decreed an action plan combining deregulation, infrastructure expansion and export promotion. For investors the divergence itself is the signal: regulatory news moves this sector's prices more than product news — the lesson of Auer and Claessens (2018) from the crypto markets applies unchanged.[54]

9.3 Power, supply chains, geopolitics

Three physical risks complete the picture. Power: if the grids cannot absorb the data-centre build-out, the whole timetable slips — the power-station contracts of chapter 5 are also an admission of this bottleneck. The supply chain: nearly every leading-edge chip in the world passes through Taiwan; an escalation in the Taiwan Strait would be the largest conceivable single shock to the entire megatrend — a risk that cannot be diversified away, only dosed. And geopolitics at large: export controls on leading chips to China have already cost NVIDIA billions in write-downs and split the AI world into two technological blocs — with the remarkable twist that China's models hold a performance gap of only a few per cent despite a fraction of the investment. Whoever invests should know all three risks and treat none as theoretical.[55]

Finally, the earnings risk of the infrastructure itself: data centres age faster than power stations. The processors of one generation are outdated after a few years, their depreciation weighs on future profits, and should AI demand grow merely more slowly than the capacity being built, a tenant's market with falling computing prices would emerge — good for users, bad for landlords. It is the same pattern by which the telecom operators once had to write down their fibre investments while the internet groups prospered on the cheapened network. Investors should therefore think both scenarios: if computing stays scarce,

[54] Regulation (EU) 2024/1689; amending regulation (EU) 2026/1744 (in force 27 July 2026): high-risk duties postponed to December 2027 and August 2028; USA: Executive Order 14179 (January 2025) and AI Action Plan (July 2025); Auer and Claessens (2018).

[55] NVIDIA: around USD 8 billion of charges from China export restrictions in fiscal 2026; Stanford HAI (2026): private AI investment USA 285.9 versus China 12.4 billion US dollars at a model performance gap of only around 2.7 per cent.

builders and landlords win; if it becomes plentiful, the profit moves one storey up to the applications. The spread across the storeys is built against precisely this uncertainty.[56]

10 What Investors Learn from All This

First: general purpose technology means patience — and infrastructure first. The productivity effect arrives, by the lesson of David (1990), with decades of delay, while the infrastructure outlays arrive at once; accordingly the most reliable profits of the first phase have accrued to chips, power and equipment, not to the applications. Second: the return sits with the scarce complementary assets — Teece (1986) remains the field's best compass: models get copied, production monopolies do not. Third: the world index already contains the trend — the most important AI position of most investors is the one they already have; theme funds are a sharpening, and sharpening demands dosage. Ben-David and co-authors (2023) showed what happens when one follows the marketing instead of the dose.[57]

Fourth: bubble and foundational build-out do not exclude each other — one builds for both. Whoever takes the Perez lesson (2002) seriously expects a valuation storm somewhere along the way and constructs the position to ride it out: spread across the storeys of the chain, capped at a written quota, built in instalments, trimmed after strong rises. Fifth: check milestones, not moods. The carrying quantities of this trend are measurable — the equipment makers' order books, data-centre grid connections, the applications' subscription revenues, the productivity studies. As long as these deliver, price falls are opportunities; if they break — if cloud contracts are cancelled and construction stops —, retreat is discipline, not a failure of nerve. Sixth and last: whoever shies from single names is well served by the combination of world portfolio plus one tested fund of the under-represented storeys — the trend demands no more complexity than that.[58]

The series' practical rulebook applies unchanged: write down the target quota before buying; build in instalments or by savings plan so that no single day decides; cut back mechanically to the target quota after strong rises — precisely in a field whose best names can double within a year, the trimming is the only built-in profit-taking that requires no forecast; and review the position against the early indicators of chapter 6.3 rather than against headlines. Whoever additionally repeats the inventory of chapter 9.1 once a year — how much AI sits in the whole portfolio? — keeps the total bet in bounds even when prices quietly enlarge it.[59]

[56] Cf. the fibre parallel of 2000-2003; on the storey logic, chapter 4.1.
[57] David (1990); Teece (1986); Ben-David, Franzoni, Kim and Moussawi (2023).
[58] Perez (2002); checklist and fund comparison in chapter 8; on the instalment logic, Rupprechter (2026g), chapter 10.
[59] Rulebook after Rupprechter (2026g), chapter 10; early indicators in chapter 6.3.

11 Conclusion and Outlook

The opening question — how investors can take part in artificial intelligence without repeating the mistakes of earlier technology waves — can be answered after ten chapters. The technology is a general purpose technology in the sense of the economic literature: real in its effect (the field studies measure double-digit productivity gains), slow in its diffusion (corporate pilots still fail in rows at implementation) and vast in its appetite for capital (USD 735 billion in 2026 from four groups alone). Its value chain is fully listed, almost entirely in large caps; its profits accrue, so far, where the scarcity sits — in chips, production, power and equipment.

On the bubble question the verdict came out differentiated: higher concentration than in 2000, but sounder funding; historically high market valuation, but a market leader whose price-earnings ratio sits below the market average; real circular structures, but also real, paid, contracted demand. The consequence for the portfolio is the architecture of chapters 7 to 10: an inventory of the AI weight already held, then a measured complement through the under-represented storeys — infrastructure, chips — via single names or a tested fund, a written quota, instalments, trimming. Thus one stays invested if the optimists prove right, and solvent if the pessimists get their interlude.

The outlook: three developments deserve watching in the coming years. First, the agents — if AI systems reliably complete multi-step work, value creation shifts from answering to executing, and the application layer could displace infrastructure as the profit centre. Second, the return question — by 2027/28 at the latest, capital markets will demand first answers on whether the trillion-scale investments show up in earnings; the reaction to disappointments will be fierce, the direction of travel probably unchanged. Third, the race of the blocs — America's capital against China's efficiency, with Europe as rule-setter and equipment supplier in between. This paper was the second of six on the great megatrends; the coming papers in the series will order further of these decade-scale shifts for investors.[60]

[60] The series continues; on technology waves once more Perez (2002), on the mobility application of AI Rupprechter (2026g).

References

Preliminary note: Market, investment and fund data without a separate reference are drawn from the reports and guidance of the companies named and from the cited reports of the IEA, Stanford HAI, Morgan Stanley and Dell'Oro; as of late August 2026.

Acemoglu, D. (2025). The simple macroeconomics of AI. Economic Policy, 40(121), 13-58.

Agrawal, A., Gans, J. and Goldfarb, A. (2018). Prediction Machines: The Simple Economics of Artificial Intelligence. Harvard Business Review Press, Boston.

Auer, R. and Claessens, S. (2018). Regulating cryptocurrencies: Assessing market reactions. BIS Quarterly Review, September 2018, 51-65.

Autor, D. H. (2015). Why are there still so many jobs? The history and future of workplace automation. Journal of Economic Perspectives, 29(3), 3-30.

Babina, T., Fedyk, A., He, A. and Hodson, J. (2024). Artificial intelligence, firm growth, and product innovation. Journal of Financial Economics, 151, 103745.

Ben-David, I., Franzoni, F., Kim, B. and Moussawi, R. (2023). Competition for attention in the ETF space. Review of Financial Studies, 36(3), 987-1042.

Bessembinder, H. (2018). Do stocks outperform Treasury bills? Journal of Financial Economics, 129(3), 440-457.

Bresnahan, T. F. and Trajtenberg, M. (1995). General purpose technologies 'Engines of growth'? Journal of Econometrics, 65(1), 83-108.

Briggs, J. and Kodnani, D. (2023). The potentially large effects of artificial intelligence on economic growth. Goldman Sachs Global Economics Analyst, 26 March 2023.

Brynjolfsson, E., Li, D. and Raymond, L. (2025). Generative AI at work. Quarterly Journal of Economics, 140(2), 889-942.

David, P. A. (1990). The dynamo and the computer: An historical perspective on the modern productivity paradox. American Economic Review, 80(2), 355-361.

Dell'Acqua, F., McFowland III, E., Mollick, E., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F. and Lakhani, K. R. (2023). Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality. Harvard Business School Working Paper 24-013.

Dell'Oro Group (2026). AI boom drives data center capex. Dell'Oro, February 2026.

Eloundou, T., Manning, S., Mishkin, P. and Rock, D. (2024). GPTs are GPTs: Labor market impact potential of LLMs. Science, 384(6702), 1306-1308.

Frey, C. B. and Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? Technological Forecasting and Social Change, 114, 254-280.

IEA (2025). Energy and AI. International Energy Agency, Paris, April 2025.

Katz, M. L. and Shapiro, C. (1985). Network externalities, competition, and compatibility. American Economic Review, 75(3), 424-440.

LeCun, Y., Bengio, Y. and Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.

Markowitz, H. (1952). Portfolio selection. Journal of Finance, 7(1), 77-91.

MIT NANDA (2025). The GenAI Divide: State of AI in Business 2025. Massachusetts Institute of Technology, Cambridge, August 2025.

Morgan Stanley (2026). AI market trends. Morgan Stanley Research, May 2026.

Nordhaus, W. D. (2021). Are we approaching an economic singularity? Information technology and the future of economic growth. American Economic Journal: Macroeconomics, 13(1), 299-332.

Noy, S. and Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. Science, 381(6654), 187-192.

NVIDIA (2026). Quarterly and annual reports, fiscal years 2023 to 2027. NVIDIA Corporation, Santa Clara.

Peng, S., Kalliamvakou, E., Cihon, P. and Demirer, M. (2023). The impact of AI on developer productivity: Evidence from GitHub Copilot. arXiv 2302.06590.

Perez, C. (2002). Technological Revolutions and Financial Capital: The Dynamics of Bubbles and Golden Ages. Edward Elgar, Cheltenham.

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 (AI Act). Official Journal of the EU, L series, 12 July 2024.

Rupprechter, R. (2026b). How much green is enough? Green investing between regulation, returns, and impact. Working Paper, R&B Wealth Management.

Rupprechter, R. (2026g). Future Mobility: Electric, autonomous, connected, shared — how investors can take part in the remaking of transport. Working Paper, R&B Wealth Management.

Solow, R. M. (1987). We'd better watch out. New York Times Book Review, 12 July 1987, p. 36.

Stanford HAI (2026). Artificial Intelligence Index Report 2026. Stanford University, Stanford.

Teece, D. J. (1986). Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy. Research Policy, 15(6), 285-305.

Turing, A. M. (1950). Computing machinery and intelligence. Mind, 59(236), 433-460.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. and Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30.

Disclaimer

This working paper serves information purposes only and constitutes neither investment, legal nor tax advice. Its contents are neither an offer nor a recommendation to buy or sell financial instruments; individual companies and funds are named for illustration and expressly not as recommendations to buy or sell. Investments in equities and ETFs are subject to price fluctuations up to substantial losses; past performance and forecasts are not a reliable indicator of future results.

The data and sources used have been selected and checked with the greatest care; no liability is assumed for their accuracy, completeness or timeliness. Market, price and fund data are as of late August 2026 and may have changed materially since. Opinions expressed reflect the personal views of the author at the time of writing and may change without notice.

Any investment decision should be taken on the basis of the investor's individual asset, risk and tax situation and, where appropriate, after consulting qualified professionals. Liability for financial losses arising from the use of the contents of this paper is excluded. Reproduction and distribution, in whole or in part, only with attribution.

The author and R&B Wealth Management hereby declare that the selection of securities was carried out independently and without any payments or benefits from third parties. At no point was there any external influence that initiated this work or affected its content.