



UNIVERSITÀ POLITECNICA DELLE MARCHE
FACOLTÀ DI ECONOMIA “GIORGIO FUÀ”

Corso di Laurea triennale in
DIGITAL ECONOMICS AND BUSINESS

**THE AI MARKET BUBBLE:
AN ECONOMIC ANALYSIS OF SPECULATION AND VALUE**

Relatore:

Prof. Francesco James Mazzocchi

Rapporto Finale di:

Yakeb Kitaw Negussie

S1119322

Anno Accademico 2025/2026

INDEX

Introduction

1 Theoretical Framework of Speculative Bubbles

1.1 Definition and Characteristics of Speculative Bubbles

1.1.1 Lifecycle of Financial Bubbles

1.1.2 Quantitative Indicators of Asset Mispricing

1.2 Historical Analysis: The Dot-com Bubble

1.2.1 Valuation Dynamics during the 1990s Technology Boom

1.2.2 The Role of Venture Capital and Initial Public Offerings

1.3 Information Cascades and Investor Behavior

1.3.1 Herd Behavior in Financial Markets

1.3.2 Investor Sentiment and Speculative Dynamics

1.4 Productivity and Technological Expectations

1.4.1 The Solow Productivity Paradox

1.4.2 Time Lags in Technological Adoption and Measurement Issues

2 The Economics of Artificial Intelligence

2.1 Artificial Intelligence as a General-Purpose Technology

2.1.1 Comparison with Previous Technological Revolutions

2.1.2 Pervasiveness, Spillovers, and Diffusion

2.2 Value Creation and Productivity Mechanisms

2.2.1 Cost Structures and Marginal Productivity Effects

2.2.2 Automation and Efficiency Gains in Cognitive Tasks

2.3 Market Structure and Competitive Dynamics

2.3.1 Market Concentration and the Role of Large Technology Firms

2.3.2 Economies of Scale and Barriers to Entry

2.4 Expectations versus Realized Economic Gains

2.4.1 Time Lags in Technological Diffusion

2.4.2 Constraints on Enterprise Adoption and Implementation

3 Empirical Analysis of the AI Market (2024–2026)

3.1 Capital Expenditure and Infrastructure Investment

3.1.1 Investment Trends among Major Technology Firms

3.1.2 Capital Expenditure and Return on Investment

3.2 Venture Capital Investment and Startup Valuations

3.2.1 Funding Dynamics and Sustainability of High-Valuation Firms

3.2.2 Evolution from Foundational Models to Application Layers

3.3 Stock Market Performance of AI-related Firms

3.3.1 Concentration of Value in Hardware and Infrastructure Providers

3.3.2 Market Reactions to AI-related Announcements

3.4 Profitability and Financial Sustainability

3.4.1 Operating Margins in AI-related Activities

3.4.2 Cash Flow Dynamics and Investment Expansion

4 Bubble Indicators and Market Evaluation

4.1 Price versus Fundamentals

4.1.1 Valuation Metrics and Growth Expectations

4.1.2 Divergence between Market Prices and Intrinsic Value

4.2 Comparative Analysis: AI and the Dot-com Era

4.2.1 Similarities in Market Behavior and Investor Sentiment

4.2.2 Differences in Profitability and Financial Structure

4.3 Identification of Speculative Signals

4.3.1 Market Sensitivity to Efficiency and Cost Shocks

4.3.2 Constraints in Scaling and Diminishing Returns

4.4 Market Scenarios and Future Outlook

4.4.1 Gradual Market Adjustment (Soft Landing)

4.4.2 Market Correction and Devaluation Dynamics

4.5 Limitations of the Analysis

4.5.1 Data Availability and Measurement Constraints

4.5.2 Uncertainty in Assessing AI-driven Productivity

Conclusion

INTRODUCTION

The rapid expansion of artificial intelligence (AI) technologies has become a central force in the contemporary digital economy, significantly reshaping capital allocation across the technology sector. Over recent years, investment flows into AI-related infrastructure, research, and commercial applications have increased at an exceptional pace. This growth is largely driven by expectations of substantial productivity gains that are anticipated to alter competitive dynamics across industries. As a result, major technology firms, venture capital funds, and financial markets have concentrated increasing resources in this field, contributing to rising valuations. At the same time, AI is gradually being reframed from a general-purpose tool into a foundational driver of economic growth and industrial competitiveness. This shift is further reinforced by geopolitical competition over computational capacity and technological autonomy, which in some cases appears to intensify investment beyond traditional return-based considerations.

However, the speed of this expansion raises questions about the relationship between financial expectations and underlying economic outcomes. A growing divergence can be observed between rising capital expenditure particularly in data centres, advanced semiconductors, and computational infrastructure and the comparatively slower monetisation of AI-based products and services. This “monetisation gap” suggests a potential mismatch between market expectations and realised value creation. Similar dynamics have been documented in earlier technological cycles, most notably during the dot-com period, when speculative capital inflows pushed valuations far ahead of earnings fundamentals. The present cycle differs in structure, particularly due to platform-driven market concentration, where a small number of firms may capture disproportionate future gains, potentially reinforcing aggressive investment behaviour.

Against this background, this thesis examines whether the current AI boom is better understood as a speculative bubble or as a rational response to a structural technological shift. The analysis draws on financial bubble theory and behavioural economics, with particular emphasis on information cascades and herd behaviour. As Bikhchandani and Sharma (2001) argue, such mechanisms can amplify market movements when investors rely on the actions of others rather than independent valuation. Within this context, the study also considers whether institutional fear of missing out (FOMO) is contributing to valuation levels that increasingly diverge from observable productivity gains.

At the same time, AI can also be interpreted as a general-purpose technology (GPT), similar to electricity or the internet, with the potential to generate broad productivity improvements across sectors over long time horizons. Historically, however, the diffusion of such technologies is uneven and often delayed, following non-linear adoption patterns. This makes it difficult to clearly distinguish between short-term financial excess and long-term structural transformation. A key challenge is therefore separating the intrinsic economic value of AI from the financial dynamics currently surrounding it, particularly in a context of abundant liquidity and strong speculative positioning.

The empirical analysis focuses on developments between 2024 and 2026, a period marked by the transition from research-driven innovation to large-scale industrial deployment. It examines

capital expenditure patterns among major hyperscale firms, including Microsoft, Alphabet, Meta, and Amazon, which are engaged in extensive infrastructure expansion comparable in scale to previous technological build-outs such as telecommunications. The study also considers venture capital activity in AI startups, where valuations frequently exceed traditional revenue-based benchmarks. In addition, the performance of leading semiconductor firms such as NVIDIA is used as a proxy for ecosystem-wide demand for computational capacity. These indicators are analysed alongside implementation challenges in enterprise adoption, including data quality constraints and high integration costs, to assess whether current valuations are supported by realised productivity gains or primarily by future expectations.

By integrating theoretical and empirical perspectives, the thesis evaluates whether key indicators associated with speculative bubbles such as valuation distortion, capital misallocation, and rapid market expansion are present. It also considers recent efficiency improvements in smaller AI models, which may challenge the dominance of large foundation model providers and reshape market expectations. The objective is to distinguish between a temporary phase of speculative excess and a longer-term structural reconfiguration of the digital economy.

The thesis is structured as follows: Chapter 1 introduces the theoretical framework of speculative bubbles, incorporating behavioural economics and the Solow productivity paradox. Chapter 2 examines AI as a general-purpose technology and its implications for productivity and labour markets. Chapter 3 presents an empirical analysis of market developments between 2024 and 2026 using financial and investment data. Chapter 4 integrates these findings to evaluate bubble dynamics, compares the current cycle with the dot-com crash, and outlines potential future scenarios for the digital economy.

1. Theoretical Framework of Speculative Bubbles

1.1 Definition and Characteristics of Speculative Bubbles

Speculative bubbles emerge when the market price of an asset rises significantly above its fundamental value. Under normal market conditions, asset prices are expected to reflect factors such as expected cash flows, profitability, and broader economic performance. During a bubble, however, valuations become increasingly influenced by investor expectations and the belief that prices will continue rising in the future. Shiller (2015) describes this as irrational exuberance — a state in which collective optimism sustains prices beyond what any rational assessment of fundamentals would support.

Periods of rapid technological innovation are frequently associated with speculative dynamics. When a new technology is perceived as transformative, investors often anticipate substantial long-term returns and direct large amounts of capital toward firms operating in the sector. While part of this growth may be justified by genuine innovation, market valuations can gradually move beyond what firms are realistically able to generate in revenues or profits.

A key feature of speculative bubbles is their self-reinforcing nature. Rising prices attract additional investors, which further increases demand and pushes valuations higher. In these conditions, market participants often pay greater attention to momentum and short-term price appreciation than to underlying business performance. Behavioural factors such as herd behaviour and fear of missing out (FOMO) further intensify speculative activity, as investors increasingly follow prevailing market reaction instead of relying on independent analysis (Bikhchandani and Sharma, 2001).

Another common characteristic of speculative periods is the widening gap between market prices and economic fundamentals. One indicator of this divergence is the presence of unusually high price-to-sales ratios, where companies achieve extremely large market valuations despite relatively modest revenues. A similar pattern can be observed in the decline of value relevance, where traditional accounting indicators such as earnings or book value become less effective in explaining stock prices (Shiller, 2015).

Speculative environments are also often accompanied by analyst forecast bias. During periods of optimism, analysts may produce overly positive earnings projections that reinforce investor confidence and contribute to additional price increases. In the current AI market, similar dynamics can be observed through what may be described as a monetisation gap. Large technology firms continue investing heavily in computational infrastructure and AI development, while the revenues generated by many AI-related products remain comparatively limited. This imbalance raises questions about whether current market valuations are supported by realised economic performance or by expectations of future growth that may prove difficult to sustain.

Overall, speculative bubbles are characterised by excessive optimism, behavioural pressures, and a weakening relationship between asset prices and underlying economic fundamentals.

Understanding these mechanisms is essential when evaluating whether current AI-related valuations reflect long-term value creation or a period of speculative excess.

1.1.1 Lifecycle of Financial Bubbles

Speculative bubbles generally develop through a sequence of identifiable stages. According to Carlota Perez's framework(2002), technological revolutions are often accompanied by cycles of financial expansion and speculative investment before reaching broader economic maturity.

The first stage is the installation phase, during which a new technology begins attracting attention and investment. Investors recognise the potential of the innovation and increasingly direct capital toward firms associated with its development. In recent years, Artificial Intelligence has increasingly been viewed in this way, often compared to the early commercial expansion of the internet during the 1990s.

The second stage is commonly described as the frenzy phase. During this period, market optimism intensifies and valuations rise rapidly as investors begin projecting extremely high future growth. Prices increasingly reflect expectations rather than current economic performance, creating conditions associated with what Shiller (2015) calls irrational exuberance.

The third stage involves market correction or the bursting of the bubble. At this point, investors begin recognising that valuations are no longer sustainable, leading to falling prices and higher market volatility. Firms that relied primarily on speculative expectations rather than viable business models are often the most vulnerable during this adjustment period.

The final stage is the deployment phase, which occurs after the correction. During this stage, the underlying technology achieves broader economic adoption and begins generating more measurable productivity gains and long-term value creation across industries.

1.1.2 Quantitative Indicators of Asset Mispricing

Although speculative bubbles are difficult to identify with certainty in real time, economists rely on several indicators to assess whether asset prices may be significantly overvalued.

One widely used indicator is the presence of abnormal price-to-sales ratios, where firms achieve exceptionally high market valuations relative to their revenues. This can be expressed as:

$$P/S = \text{Market Capitalisation} / \text{Annual Revenue}$$

When P/S ratios rise far above historical sector averages, this is typically interpreted as a signal that valuations are driven by expectation rather than current performance. During the dot-com bubble, many internet firms traded at P/S ratios exceeding 20x to 30x — levels that proved impossible to sustain once investor sentiment reversed (Morris and Alam, 2008).

Another important measure is declining value relevance, which occurs when traditional accounting variables such as earnings and book value lose explanatory power in relation to stock prices. This can be measured through the R^2 of a regression of stock returns on accounting fundamentals (Shiller, 2015):

$$P = \alpha + \beta_1(\text{EPS}) + \beta_2(\text{BV}) + \varepsilon$$

A declining R^2 over time — meaning earnings per share (EPS) and book value (BV) explain less and less of the variation in stock price (P) — is a statistical signal that the market is pricing assets on narrative rather than on fundamentals.

The concept of a monetisation gap is particularly relevant in technology-driven markets. This refers to situations in which investment expenditures grow substantially faster than the revenues generated by the technology itself. In the AI sector, major firms continue allocating enormous resources toward infrastructure, computing power, and model development, while many AI-based products have yet to generate stable or proportional profits.

Another indicator of speculative activity is analyst forecast bias. During periods of strong market optimism, analysts may systematically overestimate future earnings and growth potential, contributing to further increases in investor confidence and market valuations.

Taken together, these indicators provide a useful quantitative framework for evaluating whether current AI market valuations are supported by realistic expectations of future value creation.

1.2 Historical Analysis: The Dot-com Bubble

The dot-com bubble of the late 1990s and early 2000s remains one of the clearest examples of a technology-driven speculative bubble. During this period, investors displayed exceptional optimism toward internet-related firms and digital technologies. The rapid expansion of the internet created expectations of profound economic transformation, leading to massive inflows of capital into technology companies and startups (Morris and Alam, 2008).

The internet was widely viewed as the foundation of a “New Economy,” encouraging investors to prioritise future growth potential over present profitability. As a result, many firms with limited revenues and weak earnings achieved extremely high market valuations.



Figure 1: NASDAQ Composite Index, January 1995 – December 2004. Source: Macrotrends (2025)

One of the defining characteristics of the dot-com bubble was the growing separation between stock prices and economic fundamentals. Traditional valuation approaches based on earnings, cash flows, and profitability gradually became less influential as speculative expectations intensified. This environment contributed to excessive optimism and rapid increases in technology stock prices, as illustrated in Figure 1 above.

The dot-com period remains highly relevant when analysing the current AI market because both episodes share several important characteristics, including strong technological optimism, rapidly rising valuations, and speculative investor behaviour. At the same time, there are also important structural differences, particularly regarding the profitability, scale, and financial strength of today's leading technology firms.

1.2.1 Valuation Dynamics during the 1990s Technology Boom

During the dot-com era, market valuations increasingly reflected speculative expectations rather than underlying financial performance. Many investors adopted momentum-based strategies, purchasing technology stocks primarily because prices were already rising rapidly.

Traditional accounting measures gradually lost importance, while alternative metrics such as website traffic, user growth, and page views became central to valuation narratives. Some firms also relied heavily on pro-forma reporting to present a more favourable picture of their financial position (Morris and Alam, 2008).

This period demonstrated a significant decline in value relevance, as the relationship between stock prices and accounting fundamentals weakened considerably (Shiller, 2015). Abnormal price-to-sales ratios became increasingly common, with many internet firms achieving valuations far above those of established companies with substantially stronger revenues and profits.

Financial analysts also contributed to speculative dynamics by issuing highly optimistic earnings forecasts and growth projections, reinforcing analyst forecast bias throughout the market.

1.2.2 The Role of Venture Capital and Initial Public Offerings (IPOs)

Venture capital funding and initial public offerings played a central role in the expansion of the dot-com bubble. Large amounts of capital flowed into internet startups, allowing many firms to scale rapidly despite weak profitability and uncertain business models.

IPO activity accelerated significantly during the late 1990s, and many newly listed technology firms experienced dramatic price increases immediately after entering public markets. For many investors, IPOs became associated more with short-term speculative gains than with long-term investment opportunities.

The case of eToys.com became one of the most widely cited examples of speculative enthusiasm during this period. Despite lacking a sustainable path to profitability, the company experienced a rapid surge in market value following its IPO (Morris and Alam, 2008).

Conflicts of interest within financial institutions further amplified speculative behaviour. Analysts frequently promoted technology companies that were clients of the same investment banks employing them, reinforcing optimism and contributing to herd behaviour within financial markets (Bikhchandani and Sharma, 2001).

The period also involved substantial capital misallocation. Enormous investments were directed toward internet infrastructure and telecommunications projects based on highly optimistic assumptions regarding future demand. This contributed to a widening monetisation gap between investment levels and the revenues ultimately generated by many technology firms.

The dot-com bubble illustrates how technological innovation can simultaneously generate genuine economic progress and speculative excess. Although the internet eventually transformed the global economy, financial markets significantly overestimated both the speed and profitability of that transformation.

1.3 Information Cascades and Investor Behavior

Speculative movements in financial markets are often reinforced by information cascades and broader behavioural dynamics. An information cascade occurs when investors base their decisions primarily on the actions of others rather than on independent evaluations of economic fundamentals. As Bikhchandani and Sharma (2001) explain, this happens because market participants assume that others possess superior information, leading them to imitate prevailing behaviour instead of conducting objective analysis.

These dynamics contribute to a growing disconnect between market prices and real economic performance. As increasing numbers of investors follow the same trend, prices may continue rising even when evidence of sustainable profitability remains limited. In the current AI market, this process is strengthened by strong expectations surrounding the future potential of Artificial Intelligence, as well as institutional fear of missing out. Many firms and investors continue allocating large amounts of capital toward AI-related projects partly out of concern that failing to participate could leave them at a competitive disadvantage in a potentially transformative technological shift.

1.3.1 Herd Behavior in Financial Markets

Herd behaviour occurs when investors follow broader market trends rather than relying on independent analysis. During speculative periods, this tendency can contribute to collective capital misallocation, as investment decisions become increasingly influenced by sentiment and price momentum instead of underlying financial fundamentals.

The dot-com bubble offers a clear historical example of herd behaviour in practice. Throughout the late 1990s, many investors purchased technology stocks primarily because prices were rising rapidly, reinforcing the belief that continued growth was inevitable. Momentum-driven investment strategies gradually displaced more traditional valuation approaches based on profitability, revenues, and cash flows.

Financial analysts also played a significant role in reinforcing speculative enthusiasm. In many cases, analysts issued highly optimistic forecasts for technology companies that were simultaneously clients of the investment banks employing them. These conflicts of interest contributed to increasingly positive market sentiment and strengthened the perception that internet-related firms would continue expanding indefinitely.

The experience of eToys.com illustrates how herd behaviour can amplify both rapid market expansion and equally severe corrections. Shortly after its IPO, the company experienced a sharp increase in market value despite lacking a sustainable business model. Once investor sentiment deteriorated, however, the company collapsed into bankruptcy — demonstrating how quickly cascade dynamics can reverse (Morris and Alam, 2008).

1.3.2 Investor Sentiment and Speculative Dynamics

Investor sentiment plays a central role in the formation and expansion of speculative bubbles. In many cases, market expectations are shaped less by underlying economic conditions than by previous price movements. Rising asset prices create optimism among investors, encouraging additional capital inflows that contribute to further price increases. Over time, this process can develop into a self-reinforcing cycle capable of sustaining speculative growth for prolonged periods.

One important feature of these dynamics is asymmetric expectation updating. During expansionary phases, investors often react more strongly to positive information while discounting or underestimating negative signals (Bikhchandani and Sharma, 2001). As valuations continue rising, confidence in future growth becomes increasingly detached from observable economic performance.

Speculative behaviour is further intensified when short-term trends are projected far into the future. During bubble periods, investors frequently prioritise anticipated future growth over present profitability and financial sustainability. As a result, market valuations may become progressively disconnected from measurable earnings, productivity gains, and broader economic fundamentals.

In the context of the current AI market, investor sentiment continues to play a major role in sustaining elevated valuations. Expectations surrounding future productivity improvements and technological transformation have encouraged substantial investment activity, even though many AI-related products and services have yet to generate stable or proportionate financial returns.

1.4 Productivity and Technological Expectations

The rapid growth of investment in Artificial Intelligence is largely driven by expectations that the technology will generate substantial productivity improvements across multiple sectors of the economy. Many firms and investors anticipate that AI will transform production processes, reduce operational costs, and improve efficiency in both industrial and service-based activities.

At the same time, a significant gap remains between these expectations and the measurable productivity outcomes currently visible in economic data. Although AI technologies have advanced rapidly in recent years, evidence of broad productivity acceleration remains limited (van Ark, 2016). This divergence between anticipated gains and observed economic performance has become a central issue in debates surrounding the sustainability of current AI-related valuations.

1.4.1 The Solow Productivity Paradox

The Solow Productivity Paradox refers to the observation that technological progress may become widespread throughout the economy without producing immediate improvements in measured productivity. Economist Robert Solow captured this idea in his well-known statement that “you can see the computer age everywhere but in the productivity statistics.” (Triplett, 1999).

Despite major advances in digital technologies over recent decades, global productivity growth has remained relatively modest. Triplett (1999) provides empirical evidence that productivity growth did not accelerate in line with information technology investment during the 1980s and early 1990s and examines the methodological reasons why this divergence occurred. Van Ark (2016) identifies a similar pattern in more recent digital investment cycles, suggesting that the paradox is recurring rather than a one-off anomaly.

Although AI may possess significant long-term economic potential, current productivity indicators do not yet point toward a dramatic transformation of economic performance. This does not necessarily imply that AI lacks value. Rather, it suggests that the economic benefits associated with the technology may take considerably longer to materialise than current financial markets often assume.

1.4.2 Time Lags in Technological Adoption and Measurement Issues

One explanation for the gap between technological investment and measured productivity is the existence of long adoption cycles. Historically, major technological revolutions have often required decades before generating widespread economic benefits. According to theories of technological diffusion, innovations typically pass through an installation phase followed by a deployment phase. During the installation period, markets are frequently characterised by experimentation, disruption, and speculative investment. More substantial productivity gains usually emerge later, once the technology becomes broadly integrated across industries and business processes.

Another possible explanation is the mismeasurement hypothesis. Some economists argue that the benefits generated by digital technologies are difficult to capture through traditional economic indicators such as GDP or labour productivity statistics. Many AI-related improvements—including convenience, digital accessibility, and enhanced user experience—may create substantial consumer value while generating relatively limited measurable market transactions.

Productivity growth also depends heavily on complementary investments in intangible assets such as employee training, organisational restructuring, and managerial adaptation. Brynjolfsson and McAfee (2014) show that firms generating the highest returns on technology investment are those that simultaneously invest in these organisational capabilities — and that building such capabilities takes time. Consequently, the economic impact of AI adoption may remain only partially visible in the short term, even while firms continue investing aggressively in the technology.

2. The Economics of Artificial Intelligence

Before evaluating whether the current AI investment cycle is a speculative bubble, it is necessary to understand what kind of economic phenomenon AI actually is. This matters because the bubble question and the technology question are not independent: if AI genuinely is a general-purpose technology, it is comparable to electricity or the internet, then a period of overinvestment followed by correction would not disprove its long-run economic significance. If, on the other hand, AI's economic impact is narrower or slower arriving than its financial market valuation implies, then the current investment cycle contains a fundamental misjudgement that no subsequent deployment phase will fully correct. This chapter examines AI as an economic phenomenon, focusing on its role as a general-purpose technology, the mechanisms through which it may generate productivity gains, the competitive market structure emerging around it, and the widening gap between what financial markets currently price and what economic data can presently verify.

2.1 Artificial Intelligence as a General-Purpose Technology

The classification of AI as a general-purpose technology is not just a label; it carries specific analytical implications. Bresnahan and Trajtenberg (1995) define general-purpose technologies as innovations characterised by three properties: pervasiveness across sectors, improvement over time, and the capacity to generate innovation complementarities. That is, to enable new innovations in the sectors that adopt them. For instance, electricity satisfied all three conditions, as did the internal combustion engine and, more recently, the internet. AI appears to satisfy them too, it is already deployed across healthcare, financial services, logistics, manufacturing, legal services, and consumer products; its capabilities are improving rapidly; and it is generating new product and service categories that would not exist without it (Agrawal, Gans, and Goldfarb, 2019).

This classification matters for the bubble analysis because the history of general-purpose technologies reveals a consistent pattern. The period in which a new GPT begins to be widely recognised as transformative tends to attract capital flows that anticipate the full deployment phase of the technology before that phase has arrived.

2.1.1 Comparison with Previous Technological Revolutions

Each wave of industrial transformation has altered the relationship between technology and labour in a distinct way. The first industrial revolution replaced human physical power with steam; the second extended to electricity, enabling the factory system and mass production; the third introduced digital computation, which also brought about automation. What makes the current AI-driven shift sometimes described as Industry 4.0 structurally different from its predecessors is the nature of the tasks being automated. Previous technological revolutions primarily augmented or replaced physical labour and routine information processing. AI, is beginning to automate cognitive and decision-making tasks that were previously assumed to

be the exclusively done by human judgment. This is a meaningful shift, even if its full economic implications have not yet materialised in the data (Lu and Zhou, 2021).

The 1990s productivity paradox provides the most directly applicable historical precedent. During that decade, investment in information and communication technology accelerated sharply while measured aggregate productivity growth remained weak. The eventual resolution of the paradox in the mid-to-late 1990s productivity acceleration is often used as evidence that the gains from digital investment simply took time to materialise, and that patience was rewarded (Triplett, 1999). The same argument is being made about AI today, and it is not unreasonable. The caution required is that the productivity acceleration of the 1990s was not a direct proof of every investment made during that period: many of the firms that made those investments did not survive to participate in the gains.

2.1.2 Pervasiveness and Diffusion

The pervasiveness of AI across sectors is already empirically observable in ways that were not true of earlier GPT cycles at comparable stages of development. Speech recognition, image classification, natural language processing, drug discovery, autonomous navigation, and personalised recommendation systems are each operating at commercial scale, in multiple sectors, simultaneously. This scale of application is one reason why the GPT classification seems more secure for AI than it did for, say, the internet in 1995, when the commercial applications were still highly concentrated in a small number of sectors and business models.

The diffusion of AI, however, deserve more careful attention than the pervasiveness evidence alone provides. Economic research on GPT diffusion consistently finds adoption cycles of twenty-five to fifty years from initial commercialisation to broad economic integration (Perez, 2002). Electricity was commercialised in the 1880s but did not generate measurable aggregate productivity effects until the 1920s, a lag of roughly forty years. The internet became commercially available in the early 1990s, yet its effects on aggregate productivity only became clearly visible in the data around the mid-2000s a gap of roughly fifteen years. If AI follows a similar trajectory, which is not guaranteed but is historically plausible, then there is a real mismatch between how financial markets are currently pricing AI returns and what the historical evidence actually suggests. Some argue AI will diffuse faster because the channels for adoption are more immediate; others point out that the organisational changes required to realise AI's benefits are, if anything, more complex than those required by the internet. Either way, it is hard to reconcile a twenty-year diffusion horizon with equity valuations that seem to assume strong earnings growth within the next five years.

2.2 Value Creation and Productivity Mechanisms

Understanding how AI creates economic value rather than just assuming it does is important here because the whole bubble question hinges on it. If the value creation mechanisms are real and already working, then the investment levels start to look more justified. If they are theoretical or have narrower in scope than the investment narrative suggests, then the

speculative case gets stronger. Three mechanisms matter most for this analysis: the way AI is reshaping cost structures for firms that adopt it, the unusual economics of digital AI products where marginal cost is effectively zero, and the automation of cognitive tasks that is starting to shift how income is distributed between labour and capital.

2.2.1 Cost Structures and Marginal Productivity Effects

One of the more concrete economic shifts that AI has enabled is the conversion of data infrastructure from a fixed cost into a variable one. Firms that previously needed to own servers and on-site computing hardware can now rent equivalent capability through cloud services, paying only for what they use. In principle this lowers the barrier to entry for AI-enabled businesses. The catch is that it simultaneously deepens the dependence of the entire AI ecosystem on the small number of hyperscalers who own the underlying infrastructure so while access is democratised, economic power is increasingly concentrated. A related dynamic is the near-zero marginal cost of serving additional users once a model is trained. This sounds like a competitive advantage, and in terms of scale economics it is. But competitive markets with zero marginal cost products tend to drive prices toward zero as well the same mechanism that made it so difficult for internet companies to generate sustainable profits from content in the 2000s (Lu and Zhou, 2021). If the outputs of competing AI models become similar enough that users choose on price, the high valuations currently assigned to AI model providers become very hard to justify regardless of how large their user bases are.

The data volume question complicates this further. More training data does improve model performance, but with diminishing returns beyond a certain scale, architectural innovation matters more than accumulating additional data. This makes data advantages less durable than they appear. The clearest evidence came in early 2025, when DeepSeek released a model matching frontier performance at a fraction of the usual training cost. NVIDIA lost roughly \$600 billion in market capitalisation in a single day. That reaction tells you something important about the current investment cycle: assets priced on solid fundamentals do not move like that because of one laboratory's publication.

2.2.2 Automation and Efficiency Gains in Cognitive Tasks

The replacement of Cognitive tasks with AI is one of the most debated aspects of its economic impact. The distinction between AI making workers more productive and displacement where AI replaces workers entirely matters enormously for the economic picture, but the ratio between the two is still unclear and likely varies significantly by sector. What is relatively clearer is the distributional implication: where displacement does occur, labour's share of value-added declines and capital's share increases. This is consistent with what Chapter 3 shows empirically the financial returns from the current AI cycle are heavily concentrated at the infrastructure and hardware layers, not distributed across the workforce. The broader question of what this means for consumption-driven economic growth is beyond the scope of this thesis, but it is worth

mentioning as a background condition that shapes the long-run market for AI products (Lu and Zhou, 2021).

The shift in skill demand that accompanies this is often discussed in optimistic terms as AI takes over analytical tasks, human workers move toward judgment, communication, and creativity, which are harder to automate (Brynjolfsson and McAfee, 2014). That is probably the right long-run direction. The short-run picture is not as clear, because skill transitions take time and displaced workers do not automatically acquire more valuable skills. What is perhaps most striking, though, is that if AI were already displacing significant volumes of cognitive labour at scale, aggregate productivity data in the most AI-exposed sectors should be starting to reflect it. As of 2026, that signal remains weak. That does not mean there are no gains but what it means is the market is at this moment, pricing productivity improvements that the data has not yet shown us.

2.3 Market Structure and Competitive Dynamics

The market structure that AI is producing is characterised by a tension that runs throughout the economic analysis of the current cycle: AI simultaneously has the potential to democratise access to productive capability and to concentrate market power in a small number of dominant platforms. Understanding which of these tendencies is currently dominant — and which is likely to prevail in the medium term is essential for evaluating both the sustainability of current AI valuations and the broader economic implications of the technology's diffusion.

2.3.1 Market Concentration and the Role of Large Technology Firms

The concentration of AI infrastructure in a handful of large technology firms is actually one of the clearest ways in which it differs from the dot-com era. In the late 1990s, the companies driving the bubble were largely new entrants with little or no revenue. Today, the dominant players are established incumbents with some of the strongest balance sheets. Microsoft, Alphabet, Meta, and Amazon all entered the AI investment cycle from positions of financial strength, which means they can manage the costs of large-scale infrastructure build-out in ways that smaller competitors simply cannot.

This level of concentration creates what economists call superstar firm dynamics essentially, a market where a small number of dominant companies capture most of the profits while everyone else is left competing for much smaller margins (Varian, in Agrawal et al., 2019). In AI markets, this happens for several reasons. First, training large models gets cheaper per unit as you scale up, so bigger firms have a natural cost advantage. Second, companies like Microsoft and Google have spent years accumulating data that smaller competitors cannot easily replicate.

2.3.2 Economies of Scale and Barriers to Entry

There are barriers to entry related to the AI market, but not all have equal durability. For instance, talent constraint the shortage of researchers and engineers capable of building and deploying advanced AI systems is genuinely binding right now, but this kind of scarcity resolves as educational programmes catch up. It is a short-run problem, not a permanent one.

Data access presents a more complex issue. Firms controlling large proprietary datasets often possess meaningful advantages, particularly in narrow applications requiring highly specialised information. At the same time, two developments weaken the assumption that data advantages alone guarantee long-term dominance. First, the performance gains associated with increasing data volumes eventually diminish. Second, the emergence of large foundational models accessible through APIs allows startups to build products without developing expensive proprietary models from scratch. This explains why AI startup activity remains strong despite the extreme concentration of infrastructure ownership. In practice, access to AI capability has become relatively accessible, while ownership of the underlying infrastructure remains highly concentrated (Agrawal, Gans, and Goldfarb, 2019).

2.4 Expectations versus Realized Economic Gains

The gap between what AI investment is financially pricing and what economic data currently shows is the central analytical problem of this thesis. It is important to be precise about what this gap does and does not imply. It does not imply that AI will fail to deliver significant economic gains. The technology is real, the applications are real, and the historical pattern of GPT adoption suggests that patience is warranted. What it does imply is that there is a meaningful probability that near-term returns will disappoint relative to current financial market assumptions, and that this probability is not adequately reflected in current valuations. This section examines the evidence for the gap and the structural reasons it is likely to persist in the near term.

2.4.1 Time Lags in Technological Diffusion

The monetisation gap in the current AI cycle is visible in the most basic financial data. The four hyperscalers deployed approximately \$356 billion in capital expenditure in 2025, primarily on AI infrastructure, while the revenue being generated directly by AI products remains a fraction of that. Infrastructure investment always precedes revenue generation — that is normal in capital-intensive industries. What is less normal is the scale of the gap and how long it is persisting despite rapid cloud revenue growth.

Perez (2002) offers the most convincing explanation for why this happens. The installation phase of any major technological cycle is characterised precisely by investment running ahead of monetisation capital flows in on the basis of anticipated future gains, and infrastructure is built that will not reach full utilisation for years. The dot-com telecoms build-out is the obvious parallel: the fibre optic networks laid in the late 1990s were not fully utilised until the mid-

2000s, by which point most of the companies that built them had gone bankrupt. The infrastructure survived. The investors did not always.

What makes the current cycle harder to read than the dot-com precedent is that the gap is not uniform. Microsoft's \$392 billion commercial remaining performance obligation represents real contracted near-term demand, not speculative positioning. AWS growing at 28 percent on a \$128 billion revenue base with expanding margins is not a bubble narrative. The speculative component sits above these operational realities in the valuation multiples applied to future earnings, in the prices assigned to AI-native startups, and in assumptions about how long current competitive advantages will last.

2.4.2 Constraints on Enterprise Adoption and Implementation

The supply-side investment story dominates most AI commentary, but the demand side is arguably more revealing. Only around 20 percent of executives currently qualify as serious AI adopters, while 28 percent report that their firms simply lack the technical capability to implement AI at scale (van Ark, 2016). That statistic is striking given the hundreds of billions being spent on infrastructure it suggests that the mass-market enterprise adoption needed to justify current investment levels has not yet materialised.

Part of the reason this gap is slow to close is that adopting AI successfully is not just a matter of buying access to it. Brynjolfsson and McAfee (2014) show that the firms generating the highest returns from technology investment are those that also invest in what they call complementary organisational capital workforce retraining, process redesign, data governance, and management capability. These things take years to build and cannot be purchased off-the-shelf. AI is no different.

The practical barriers make this worse. Most enterprises hold data that is too messy and fragmented to feed directly into AI systems. Integrating AI with legacy infrastructure is expensive and disruptive. Regulatory uncertainty around data privacy and algorithmic accountability adds further hesitation, particularly in sectors like financial services and healthcare where compliance risk is high. The overall picture is one where the gap between infrastructure availability and actual enterprise productivity impact is likely to be considerably longer than the investment narrative assumes which matters directly for the bubble assessment in Chapter 4.

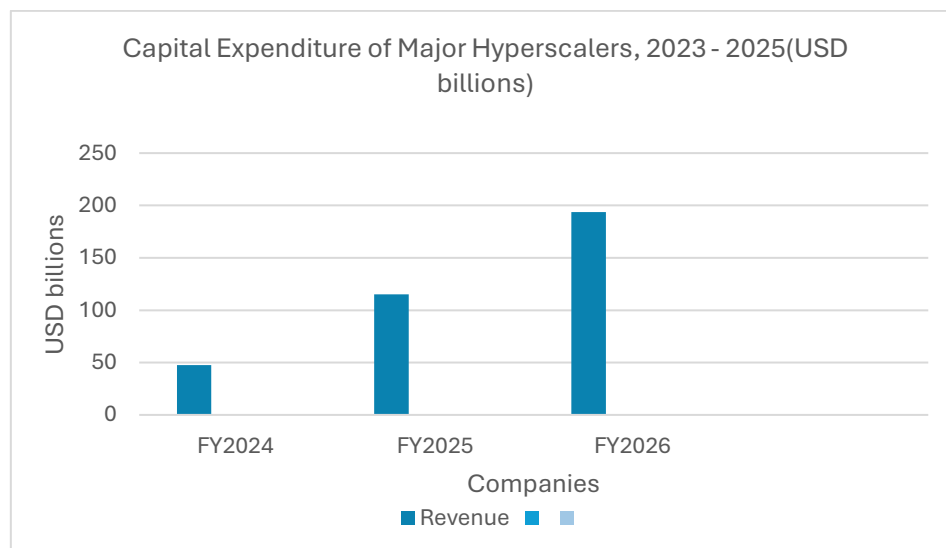
3. Empirical Analysis of the AI Market (2024–2026)

The theoretical frameworks discussed in Chapters 1 and 2 identify several financial conditions commonly associated with speculative bubbles, including capital expenditure growth that significantly exceeds revenue growth, valuations detached from near-term earnings, and an expanding gap between infrastructure investment and measurable productivity outcomes. This chapter examines whether these characteristics are visible in the AI market between 2024 and 2026 through an analysis of financial reports, earnings calls, and regulatory filings from Microsoft, Alphabet, Meta, Amazon, and NVIDIA.

3.1 Capital Expenditure and Infrastructure Investment

3.1.1 Investment Trends among Major Technology Firms

Between 2023 and the first quarter of 2026, the four dominant hyperscalers undertook one of the most aggressive capital expenditure cycles in the history of the technology sector. The scale of this deployment is visible in Figure 2 below.



[Figure 2: Capital Expenditure of Microsoft, Alphabet, Meta and Amazon — 2023, 2024, 2025 (USD billions). Bar chart to be inserted here. Data sourced from annual financial reports.]

Microsoft's additions to property and equipment rose from \$28.1 billion in FY2023 to \$44.5 billion in FY2024 (+58 percent) and \$64.6 billion in FY2025 (+45 percent). In Q1 FY2026 alone, total capital expenditure including finance leases reached \$34.9 billion an annualised rate exceeding \$130 billion with CFO Amy Hood confirming that full-year FY2026 growth would exceed FY2025 (Microsoft FY26 Q1 Earnings Transcript, 2025). Against this backdrop, FY2025 revenues grew 15 percent to \$281.7 billion, with Azure growing 34 percent and accelerating to 40 percent in Q1 FY2026. The investment is partly anchored in contracted demand: Microsoft's commercial remaining performance obligation reached \$392 billion at the end of Q1 FY2026, up 51 percent year-on-year, including a \$250 billion Azure services commitment from OpenAI.

Alphabet's capital expenditure rose from \$32.3 billion in 2023 to \$52.5 billion in 2024 (+63 percent) and \$91.4 billion in 2025 (+74 percent) the steepest proportional increase among the four hyperscalers (Alphabet 10-K, 2025). Total revenues grew from \$307.4 billion in 2023 to \$402.8 billion in 2025 (+15 percent), while Google Cloud revenues grew from \$43.2 billion to \$58.7 billion (+36 percent), accelerating to \$20.0 billion in Q1 2026 a 63 percent year-on-year increase. Capital expenditure as a share of revenue rose from approximately 10.5 percent in 2023 to 22.7 percent in 2025, more than double its prior level.

Meta pivoted from its efficiency approach in 2023 to one of the sector's most aggressive expansions. Capital expenditures rose from \$28.1 billion in 2023 to \$39.2 billion in 2024 (+40 percent) and \$72.2 billion in 2025, with full-year 2026 guidance raised to \$125–145 billion — partly due to higher GPU component pricing (Meta Q1 2026 Earnings Release). Q1 2026 revenues of \$56.3 billion grew 33 percent year-on-year at a 41 percent operating margin, suggesting that Meta's advertising business continues to generate the cash required to fund the programme, at least for now.

Amazon's cash capital expenditures rose from \$48.1 billion in 2023 to \$77.7 billion in 2024 (+61 percent) and \$128.3 billion in 2025 (+65 percent), with Q1 2026 spending of \$43.2 billion representing a 78 percent year-on-year increase (Amazon Annual Report, 2025). AWS revenue grew from \$90.8 billion in 2023 to \$128.7 billion in 2025, reaching \$37.6 billion in Q1 2026 described by CEO Andy Jassy as the fastest AWS growth in fifteen quarters. Amazon also disclosed a \$15 billion investment in OpenAI's Series C, alongside commitments from OpenAI and Anthropic to consume multiple gigawatts of Amazon's Trainium AI chip capacity.

Viewed in aggregate, combined capital expenditure across the four firms in 2025 alone totalled approximately \$356 billion roughly 22 percent of their combined revenues, more than double the approximately 10 percent ratio of 2020–2021. Capital expenditure is growing at rates substantially in excess of revenue growth across all four firms, which is the first observable signal of the monetisation gap identified in Chapter 2.

3.1.2 Capital Expenditure and Return on Investment

The scale of current infrastructure spending raises an important question regarding whether these investments are generating adequate financial returns. The available evidence presents a mixed picture. At Microsoft, the Intelligent Cloud segment which absorbs a significant share of AI-related infrastructure spending—generated revenues of approximately \$106.3 billion in FY2025, representing annual growth of 21 percent (Microsoft FY2025 Annual Report). However, costs increased at a faster pace than revenues, placing pressure on operating margins as infrastructure expansion accelerated. A similar pattern emerged at AWS. Although the segment maintained operating margins in the range of approximately 35–37 percent, profitability declined modestly as capacity expansion costs outpaced near-term revenue growth (Amazon Annual Report, 2025). Google Cloud produced the clearest evidence that AI infrastructure spending may eventually improve profitability at scale. Operating income increased from approximately \$6.1 billion in 2024 to \$13.9 billion in 2025, implying margin expansion from roughly 14 percent to 24 percent (Alphabet 10-K, 2025). This suggests that

cloud-based AI services may become more economically efficient once deployment reaches sufficient scale.

The free cash flow situation appears more constrained. Amazon's free cash flow declined from approximately \$38.2 billion in 2024 to roughly \$11.2 billion in 2025, largely because the increase in capital expenditure absorbed most of the company's operating cash generation (Amazon Annual Report, 2025). Similar dynamics are visible across the broader hyperscaler sector, where infrastructure investment is consuming an increasingly large share of operating cash flow. These developments do not necessarily indicate financial instability. However, they raise an important analytical question: whether firms are rationally investing ahead of future demand or whether the sector is building infrastructure capacity faster than sustainable monetisation can justify. This issue becomes central to the broader evaluative framework developed in Chapter 4.

3.2 Venture Capital Investment and Startup Valuations

3.2.1 Funding Dynamics and Sustainability of High-Valuation Firms

Alongside hyperscaler capital expenditure, the AI investment cycle has been characterised by an unprecedented concentration of venture capital in AI-native companies. OpenAI represents the clearest example of this trend. Microsoft committed approximately \$13 billion in investment, while external market estimates placed OpenAI's valuation between \$150 billion and \$300 billion despite the company remaining unprofitable (Microsoft FY2025 Annual Report). Additional investments from Amazon further strengthened the financial interdependence developing across the AI ecosystem. Increasingly, hyperscalers are simultaneously acting as investors, infrastructure providers, strategic partners, and competitors within the same market structure.

The broader venture capital landscape tells a similar story. Many generative AI companies are being valued primarily on what their models can do rather than on what they are actually earning, at revenue multiples that would have been considered extreme even during the dot-com era. Whether these valuations hold up depends on a fairly specific condition being met: that AI applications develop the kind of winner-takes-most dynamics seen in earlier digital markets, where dominant platforms built durable advantages through data and network effects. If that does not happen — if the application layer turns out to be competitive and fragmented, with model capabilities becoming increasingly similar across providers — then the pricing power needed to justify current valuations simply will not materialise.

3.2.2 Evolution from Foundational Models to Application Layers

A defining feature of the 2024–2026 period has been the migration of commercial value from foundational model capabilities toward the application and workflow integration layer, driven partly by commoditisation of base model performance through open-source releases including Meta's Llama series. Microsoft's commercial narrative has centred on Copilot — described by

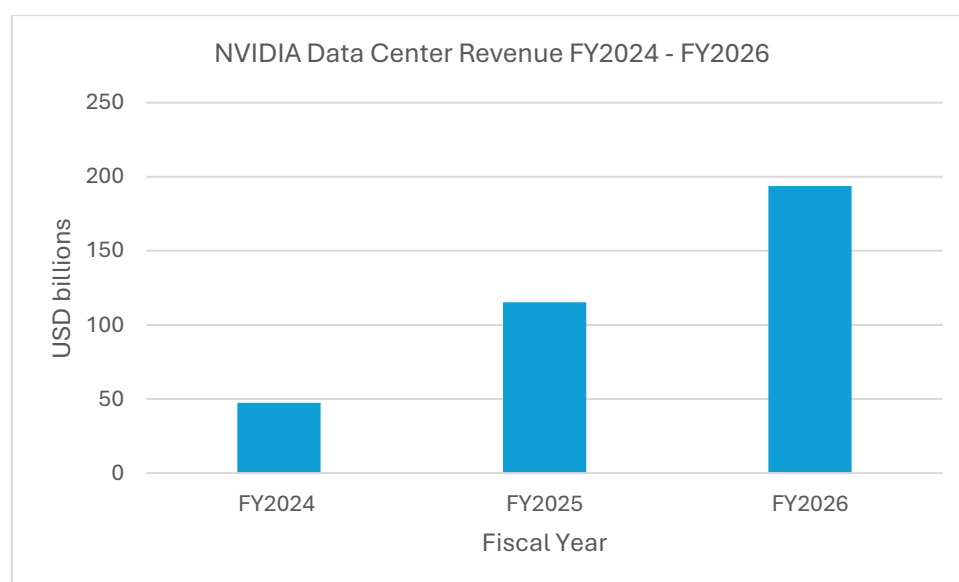
CEO Satya Nadella as 'the new UI for AI' which bundles Azure AI services with Microsoft 365 into integrated enterprise workflows, leveraging the company's existing distribution advantage across hundreds of millions of enterprise users (Microsoft FY2024 Shareholder Letter, 2024). Alphabet has pursued a parallel strategy through Gemini integration into Google Workspace, with Google Cloud revenue growth accelerating to 63 percent year-on-year in Q1 2026 (Alphabet 10-Q, Q1 2026).

Amazon adopted a more vertically integrated strategy centred around proprietary hardware. Its Graviton, Trainium, and Nitro chip businesses expanded rapidly, reflecting efforts to reduce dependence on NVIDIA hardware while strengthening long-term infrastructure economics (Amazon Q1 2026 Earnings Release). Meta's AI strategy remained closely linked to advertising optimisation: average revenue per person reached \$15.66 in Q4 2025, up from \$14.25 in Q4 2024, providing a partial observable signal of advertising ROI on AI investment (Meta Q1 2026 Earnings Presentation).

3.3 Stock Market Performance of AI-related Firms

3.3.1 Concentration of Value in Hardware and Infrastructure Providers

The most analytically surprising feature of the AI cycle between 2024 and 2026 has been the concentration of financial returns in the hardware layer, most notably NVIDIA. As illustrated in Figure 4 below, NVIDIA's Data Center revenues grew from \$47.5 billion in FY2024 to \$115.2 billion in FY2025 (+142 percent) and \$193.7 billion in FY2026 (+68 percent), representing approximately 90 percent of total revenues of \$215.9 billion.



[Figure 3: NVIDIA Data Center Revenue FY2024–FY2026 (USD billions). Bar chart to be inserted here. Data sourced from NVIDIA Annual Reports 2025–2026.]

Net income reached \$120.1 billion in FY2026 at an operating margin of approximately 60 percent margins more consistent with a software platform than a hardware manufacturer, reflecting the pricing power accumulated through NVIDIA's CUDA software ecosystem and

the switching costs it creates (NVIDIA Annual Report, 2026). This concentration of returns is more consistent with a bottleneck rent during a supply-constrained transition period than with the broad-based productivity gains expected from a genuinely transformative general-purpose technology in its diffusion phase (Perez, 2002).

The hyperscalers themselves also trade at valuations incorporating substantial AI-related optionality. Current market pricing reflects expectations that AI capabilities will generate future earnings materially above existing growth trajectories. These assumptions remain inherently speculative because they depend on uncertain adoption rates, competitive outcomes, and long-term monetisation dynamics. Alphabet's Q1 2026 results provided partial support for this thesis. Operating income increased approximately 30 percent year-on-year alongside revenue growth of roughly 22 percent, demonstrating operating leverage despite rapidly rising infrastructure investment (Alphabet Q1 2026 10-Q). The critical question, however, is whether this trajectory can remain sustainable if capital expenditure continues accelerating.

3.3.2 Market Reactions to AI-related Announcements

Markets have consistently rewarded forward-looking AI investment commitments with positive initial price responses even when associated with near-term earnings dilution, suggesting investors are pricing AI infrastructure as a long-duration strategic option. The disclosure of Microsoft's \$392 billion remaining performance obligation was received as demand validation, as contracted obligations reduce the speculative component of capital expenditure by converting optionality into committed revenue.

The most informative single market event in the study period, however, was the DeepSeek episode of early 2025. When a Chinese laboratory released a model matching frontier AI performance at a reported training cost of approximately \$6 million compared with the hundreds of millions assumed necessary NVIDIA lost roughly \$600 billion in market capitalisation in a single trading session. The speed and scale of that reaction is itself a speculative signal: assets priced on solid fundamentals do not lose 20 percent of their value because of one laboratory's publication. The equally rapid recovery, as the market reframed DeepSeek as evidence of broader AI adoption rather than a threat to the GPU investment thesis, illustrates the asymmetric sentiment updating described by Bikhchandani and Sharma (2001) positive information is weighted heavily, negative information is absorbed and reinterpreted.

3.4 Profitability and Financial Sustainability

3.4.1 Operating Margins in AI-related Activities

The profitability profile of AI-related activities in 2024–2026 is sharply divergent across value chain layers. At the hardware layer, NVIDIA achieved an operating margin of approximately 60 percent in FY2026, reflecting near-monopolistic pricing power in AI training hardware and the switching costs embedded in the CUDA ecosystem. At the cloud infrastructure layer, AWS sustained an operating margin of approximately 35–37 percent, Microsoft's Intelligent Cloud

reached approximately 42 percent, and Google Cloud expanded from approximately 14 percent in 2024 to 24 percent in 2025 — the latter being the most concrete evidence available that AI infrastructure investment can generate improving unit economics over time.

Meta's advertising margins illustrate a structurally different dynamic. The Family of Apps operating margin of approximately 47 percent in Q1 2026 reflects the high incremental profitability of AI enhancements to an advertising platform where marginal revenue carries minimal marginal cost — a fundamentally different economic model from cloud infrastructure provision. The question of whether AI-related margins will converge toward a sector equilibrium depends on whether durable competitive moats — CUDA lock-in, enterprise customer relationships, proprietary data — can resist the commoditisation pressure that the DeepSeek episode briefly made visible.

3.4.2 Cash Flow Dynamics and Investment Expansion

Free cash flow provides one of the clearest indicators of whether the current AI investment cycle is financially sustainable. Across the hyperscaler sector, operating cash flow continues to grow strongly, but capital expenditure is increasing at an even faster pace. Microsoft generated approximately \$136.2 billion in operating cash flow during FY2025 while spending \$64.6 billion on capital expenditure, leaving free cash flow of roughly \$71.6 billion. Alphabet followed a similar pattern, generating approximately \$164.7 billion in operating cash flow against capital expenditure of \$91.4 billion (Alphabet 10-K, 2025). Although both firms remain highly cash generative, infrastructure spending is consuming a significantly larger share of cash flow than in previous years. The pressure was most visible at Amazon. Free cash flow declined from \$38.2 billion in 2024 to approximately \$11.2 billion in 2025, driven almost entirely by the sharp increase in infrastructure investment (Amazon Annual Report, 2025). While this does not indicate immediate financial weakness, it demonstrates the extent to which operating cash generation is now being redirected toward AI-related infrastructure expansion.

Another important development is the increasing reliance on finance leases and long-term contractual obligations across the sector. These arrangements reduce immediate cash outflows while simultaneously creating substantial future commitments. Alphabet, for example, disclosed approximately \$232.7 billion in long-term contractual obligations tied to infrastructure, energy agreements, and supply commitments extending through 2030 (Alphabet Q1 2026 10-Q). Taken together, the evidence suggests that financial returns from the current AI investment cycle remain highly concentrated. NVIDIA has captured an outsized share of sector profitability, while hyperscalers continue allocating enormous amounts of capital toward sustaining infrastructure growth.

The broader analytical question is whether this investment phase represents a temporary installation period typical of earlier technological revolutions, eventually followed by wider productivity gains and sustainable monetisation, or whether the sector is experiencing a more persistent imbalance between infrastructure expansion and achievable long-term returns. This issue becomes central to the evaluation developed in Chapter 4.

4. Bubble Indicators and Market Evaluation

The previous chapters have developed the tools needed to address the central question of this thesis. Chapter 1 established what speculative bubbles look like their lifecycle, quantitative indicators, and the behavioural mechanisms that sustain them. Chapter 2 examined AI as an economic phenomenon and identified the structural gap between what financial markets are pricing and what the data can currently confirm. Chapter 3 documented that gap empirically. This chapter brings those threads together, evaluating where speculative dynamics are most visible, how the current cycle compares to the dot-com era, and what conditions would determine whether the outcome is a gradual adjustment or a more disruptive correction.

4.1 Price versus Fundamentals

NVIDIA is in many ways the most revealing data point in the entire AI cycle. The company earned \$120 billion in net income in FY2026 at an operating margin of approximately 60 percent audited financial results, not speculative projections (NVIDIA Annual Report, 2026). The issue is not that the earnings are fictional, but that the equity valuation applied to them embeds assumptions about how long a 60 percent hardware margin can be sustained. Margins at that level reflect pricing power more consistent with a monopoly position than a competitive market, and monopoly positions in technology hardware have historically proved more temporary. The DeepSeek episode of early 2025 in which NVIDIA lost roughly \$600 billion in market capitalisation in a single session following one laboratory's research publication demonstrated briefly but dramatically how sensitive those assumptions are to efficiency improvements.

The picture is somewhat different for the four hyperscalers. Microsoft, Alphabet, Meta, and Amazon are all profitable, have real contracted demand, and are growing their cloud businesses at rates most companies would never achieve. The question is not whether they are making money — they clearly are. The question is whether the extra premium that markets are assigning to them because of AI is actually justified. Put simply, each of these companies is currently trading at a price that assumes AI will generate significantly more earnings in the future than their existing businesses would produce on their own. That is essentially a bet on a future that has not yet arrived, which is where the speculative element enters the picture.

4.1.1 Valuation Metrics and Growth Expectations

The declining value relevance phenomenon from section 1.1.2 is partially visible in current AI markets. Traditional metrics price-to-earnings ratios, free cash flow yield are being supplemented by metrics more favourable to the AI narrative: total addressable market estimates and contracted infrastructure commitments. Some carry genuine analytical content: Microsoft's \$392 billion remaining performance obligation represents real contracted demand (Microsoft FY26 Q1 Earnings Transcript, 2025). Others are more loosely connected to near-term earnings. Reaching for metrics that support a bullish case when conventional metrics do

not is one of the more reliable signals of speculative dynamics (Shiller, 2015). The compression of Amazon's free cash flow from \$38.2 billion in 2024 to \$11.2 billion in 2025 driven entirely by capital expenditure acceleration is a concrete example of a financial outcome the investment narrative did not adequately foreground (Amazon Annual Report, 2025).

4.1.2 Divergence between Market Prices and Intrinsic Value

Price and fundamental value are not uniformly divergent across the AI market the gap varies significantly by layer of the value chain. At the operational layer, the evidence is largely reassuring Azure growing at 40 percent, Google Cloud expanding its operating margin from 14 to 24 percent in a single year, AWS delivering 28 percent growth at 35 percent margins. These are measured results. The divergence becomes more pronounced in the forward-looking multiples applied to those results, and starkest at the AI startup layer, where companies are valued at revenue multiples with no historical precedent in enterprise software. As Shiller (2015) identifies, the mechanism sustaining prices above fundamental value is self-reinforcing: each investor's response to price momentum contributes to a dynamic that can persist longer than sceptics expect and reverse faster than optimists anticipate.

4.2 Comparative Analysis: AI and the Dot-com Era

4.2.1 Similarities in Market Behavior and Investor Sentiment

The structural similarities between the two cycles are real. In both cases, a genuine technological shift attracted capital flows exceeding what the near-term commercial landscape could justify, and valuations became progressively disconnected from conventional earnings metrics as investors adopted alternative frameworks (Morris and Alam, 2008). The information cascade dynamic described by Bikhchandani and Sharma (2001) is arguably more pronounced now than during the dot-com period, because AI is genuinely difficult to evaluate independently. Most institutional investors lack the technical background to assess whether a model's capabilities justify the infrastructure investment being made to develop it. In that environment, following what other sophisticated actors are doing is individually rational — which is exactly the mechanism that produces cascade dynamics.

4.2.2 Differences in Profitability and Financial Structure

The most important structural difference is the financial profile of the companies at the cycle's centre. The dot-com bubble was built primarily on companies with minimal revenues and absent earnings. The current AI cycle is built on some of the most profitable companies in corporate history Microsoft earned \$128 billion in operating income in FY2025, and Meta maintained a Family of Apps margin of approximately 47 percent while accelerating capital expenditure toward \$125–145 billion for 2026. This means the correction risk is primarily a valuation risk a repricing of equity multiples rather than an insolvency risk. If a correction

occurs, its consequences fall on equity holders rather than spreading through the economy via corporate failures, which is a materially less destructive form of adjustment.

4.3 Identification of Speculative Signals

4.3.1 Market Sensitivity to Efficiency and Cost Shocks

The DeepSeek episode of early 2025 is the most informative market event in the study period for bubble assessment. When a Chinese laboratory released a model matching frontier AI performance at a reported training cost of approximately \$6 million compared with the hundreds of millions assumed necessary NVIDIA lost roughly \$600 billion in market capitalisation in a single session. Assets priced on robust fundamentals do not move like that because of one publication. The equally rapid recovery, as the market reframed DeepSeek as evidence of broader AI adoption rather than a GPU demand threat, illustrates the asymmetric sentiment updating described by Bikhchandani and Sharma (2001): negative information is absorbed and reinterpreted rather than fully incorporated into valuations.



[Figure 4: NVIDIA Corporation (NVDA) Stock Price, January 2024 – March 2026. The sharp decline in January 2025 corresponds to the release of the DeepSeek R1 model. Source: Yahoo Finance (2026).]. Bar chart to be inserted here. Data sourced from annual financial reports.]

4.3.2 Constraints in Scaling and Diminishing Returns

The diminishing returns to data accumulation discussed in section 2.2.1 represent a structural constraint becoming more empirically visible over time. Performance improvements from

successive frontier model generations have required exponentially larger compute investments while delivering more incremental observable improvements — consistent with the theoretical expectation of decreasing returns. The demand-side constraints compound this: only approximately 20 percent of executives currently qualify as serious AI adopters, while 28 percent report lacking the technical capability to implement AI at scale (van Ark, 2016). The infrastructure being built by the hyperscalers is, in structural terms, running ahead of the enterprise market's current absorption capacity.

4.4 Market Scenarios and Future Outlook

4.4.1 Gradual Market Adjustment (Soft Landing)

The soft landing scenario requires three conditions to hold simultaneously: enterprise AI adoption accelerating from roughly 20 percent toward 50 to 60 percent within three to five years; cloud AI operating margins sustaining or improving as utilisation increases; and competitive dynamics gradually eroding NVIDIA's pricing power through maturing custom silicon programmes. None of these is implausible. Google Cloud's margin expansion from 14 to 24 percent in a single year demonstrates that AI cloud economics can improve rapidly at scale (Alphabet 10-K, 2025). Microsoft's \$392 billion remaining performance obligation with a two-year weighted average duration provides near-term demand visibility that supports the rational investment reading (Microsoft FY26 Q1 Earnings Transcript, 2025).

4.4.2 Market Correction and Devaluation Dynamics

The correction scenario does not require AI to fail as a technology — only that the timeline to broad monetisation extends beyond what current valuations assume. The historical precedent of GPT adoption timelines — twenty-five to fifty years for full economic integration (Perez, 2002) — suggests that the optimistic timeline currently embedded in AI equity prices is the exception rather than the rule. The most likely trigger is not a single dramatic event but a gradual accumulation of evidence that AI revenue growth is decelerating relative to infrastructure investment. Markets tend not to reprice gradually in response — they maintain the prevailing narrative until disappointing data forces a discontinuous reassessment that moves faster and further than fundamentals alone would imply. A correction would be most acute at the startup and venture capital layer, where companies lack the financial resources of the hyperscalers to sustain a prolonged period of below-expectation monetisation.

4.5 Limitations of the Analysis

4.5.1 Data Availability and Measurement Constraints

The empirical analysis relies primarily on financial disclosures from publicly listed companies, which are comprehensive but selective. Granular data on AI-specific revenues — the disaggregation of Azure or AWS figures into AI-driven versus conventional cloud workloads

is not publicly available, limiting the precision with which the monetisation gap can be measured. The private market dimension adds further opacity: the financial positions of OpenAI and Anthropic, whose commercial trajectories underpin much of the contracted demand justifying hyperscaler capital expenditure, are not publicly reported with the same granularity as listed company filings.

4.5.2 Uncertainty in Assessing AI-driven Productivity

The most fundamental limitation is that the productivity question at the heart of the bubble assessment cannot be resolved while the cycle is still in progress. As Triplett (1999) showed in relation to the Solow paradox, GPT productivity effects can be real and eventually large but arrive after a lag long enough that original investors did not always benefit. That pattern is consistent with both the bubble scenario and the rational investment scenario, which means productivity uncertainty alone cannot distinguish between them. What can be said with reasonable confidence is that the current AI cycle contains both justified investment and speculative excess operating simultaneously at different layers of the value chain. The operational earnings are real; the valuation multiples embed assumptions that historical GPT diffusion patterns do not obviously support; and the market's sensitivity to efficiency shocks reveals a fragility that pure fundamentals analysis would not predict. Whether the balance tips toward soft landing or correction depends on variables enterprise adoption rates, hardware competition, and AI capability development that remain genuinely uncertain.

CONCLUSION

This thesis set out to examine whether the current wave of investment in artificial intelligence represents a speculative bubble. The analysis suggests that the answer is more complicated than a simple yes or no. AI is clearly a technology with genuine economic potential, and there is substantial evidence of real investment, demand, and revenue generation. At the same time, some parts of the market show characteristics commonly associated with speculative periods. The distinction therefore depends largely on which part of the AI value chain is considered and, importantly, on the time horizon used to evaluate the returns on current investment.

Chapter 1 established the theoretical framework used throughout the thesis. The analysis drew on Shiller's (2015) concept of irrational exuberance, Perez's (2002) distinction between the installation and deployment phases of technological revolutions, and the theory of information cascades developed by Bikhchandani and Sharma (2001). Together, these approaches provide a useful way of understanding how expectations can reinforce one another and influence investment decisions. They also highlight several characteristics that are relevant to the current AI market, including strong price momentum, a weakening relationship between market valuations and traditional fundamentals, large differences between investment and current monetisation, and the tendency of investors to respond differently to positive and negative information.

Chapter 2 examined the economic characteristics of AI and the argument that it should be considered a general-purpose technology. Based on the definition provided by Bresnahan and Trajtenberg (1995), AI displays several characteristics associated with GPTs: it can be applied across a wide range of sectors, its capabilities continue to improve, and it can support further innovations. The chapter identified three important mechanisms through which AI can create economic value: changes in firms' cost structures, the very low marginal cost of digital AI products, and the automation or augmentation of cognitive tasks. These mechanisms demonstrate that the economic case for AI is not purely speculative. However, their long-term effects remain difficult to measure. In particular, the chapter identified a significant adoption gap. A relatively small proportion of firms have so far integrated AI deeply into their operations, while organisational changes such as employee training, process redesign, and data restructuring require considerable time (van Ark, 2016; Brynjolfsson and McAfee, 2014). Therefore, the technology itself appears well established, but the speed at which its broader economic benefits will emerge remains uncertain.

Chapter 3 provided the main empirical evidence for the analysis. The scale of current infrastructure investment is particularly notable. Microsoft, Alphabet, Meta, and Amazon collectively invested approximately \$356 billion in capital expenditure during 2025, equivalent to around 22 percent of their combined revenues. This represents a substantial increase compared with the levels observed several years earlier. Such spending can be interpreted in two different ways. On one hand, it reflects a rational attempt by firms to build the infrastructure required for a technology expected to become increasingly important. On the other hand, the scale of investment raises questions about whether future demand and monetisation will develop quickly enough to justify the capacity being constructed.

The financial performance of NVIDIA provides another important element of the analysis. The company generated approximately \$120 billion in net income and maintained an operating margin close to 60 percent in FY2026. These results demonstrate that the AI investment cycle has already generated substantial economic returns for firms occupying critical positions in the supply chain. At the same time, the concentration of these returns suggests that the benefits of the current investment cycle are not yet distributed evenly throughout the wider economy. Instead, a significant proportion of the value is currently being captured by companies controlling scarce computing infrastructure and specialised hardware.

The market reaction to DeepSeek in early 2025 further illustrated the sensitivity of AI-related valuations to changes in expectations. Following the release of a model that challenged assumptions about the amount of computing resources required to achieve frontier-level performance, NVIDIA experienced a loss of approximately \$600 billion in market capitalisation in a single trading session. The subsequent recovery demonstrated that investors could rapidly reinterpret the same technological development in different ways. Rather than necessarily undermining the broader AI investment thesis, lower computing costs could also be interpreted as potentially increasing future AI adoption. Nevertheless, the episode demonstrated how strongly current valuations remain connected to expectations about technological efficiency, infrastructure requirements, and future demand.

Taken together, the findings indicate that the AI investment cycle between 2024 and 2026 contains both economically justified investment and speculative elements. These two characteristics can exist at the same time. At the operational level, there is evidence of genuine demand, contracted cloud revenues, increasing AI-related revenues, and improving profitability in some segments. These factors make it difficult to characterise the entire AI market as being based on purely speculative expectations.

The situation becomes different when valuation is considered. The prices assigned to some AI-related companies, particularly AI-native startups, incorporate expectations of very rapid adoption, high future revenues, and the persistence of current competitive advantages. These assumptions are considerably more uncertain. Historical evidence from previous general-purpose technologies suggests that technological diffusion can take decades, while financial markets currently appear to place substantial emphasis on much shorter periods of future growth (Perez, 2002). This creates a potential mismatch between the economic timeline of technological adoption and the financial timeline embedded in current valuations.

For this reason, the findings of this thesis are better understood as evidence of a **layered investment cycle** rather than a simple distinction between a “real” market and a “bubble.” The underlying technology is real and has already produced measurable economic activity. The question is instead whether the prices currently assigned to future AI earnings adequately reflect the uncertainty surrounding adoption, competition, infrastructure utilisation, and productivity gains. In this respect, speculative behaviour may exist alongside genuine technological progress rather than replacing it.

Several limitations should also be acknowledged. First, the study period ends in early 2026, meaning that the current AI investment cycle is still developing. A longer observation period

would provide stronger evidence regarding whether current infrastructure spending eventually generates proportional economic returns. Second, publicly available financial reports do not consistently separate AI-related revenues and costs from broader cloud, software, advertising, or hardware activities. This makes it difficult to calculate the monetisation gap with complete precision. Third, the financial positions of important private companies, including OpenAI and Anthropic, are not disclosed with the same level of detail as those of publicly traded firms. Finally, the productivity effects of AI cannot yet be assessed conclusively. As the experience of previous technological revolutions demonstrates, the relationship between technological investment and measured productivity can only become fully clear over time (Triplett, 1999).

Several questions therefore remain open. Will enterprise adoption accelerate sufficiently to justify the infrastructure currently being constructed? Will cloud AI margins increase as utilisation rises, or will stronger competition eventually place downward pressure on prices? Will the development of custom AI chips reduce dependence on NVIDIA, or will NVIDIA maintain its position through its hardware and software ecosystem? Most importantly, will AI generate the broad productivity improvements currently incorporated into many long-term investment expectations?

The answers to these questions will ultimately determine how the current investment cycle is understood. If adoption expands rapidly and AI generates substantial productivity and revenue gains, much of today's investment may eventually appear as an early response to a major technological transformation. If adoption remains slower than expected and the returns on infrastructure investment fail to materialise, some of the current valuations may instead prove to have incorporated excessive expectations. At the time of writing, the available evidence does not provide sufficient grounds to resolve this question definitively. What can be established is that the current AI cycle combines a genuine technological transformation with significant financial expectations, making it particularly important to distinguish between the economic value of AI itself and the prices that financial markets are currently willing to pay for its anticipated future benefits.

REFERENCES

Academic Sources

- Agrawal, A., Gans, J. and Goldfarb, A. (eds.) (2019), *The Economics of Artificial Intelligence: An Agenda*, University of Chicago Press, Chicago. ISBN: 978-0-226-61333-8.
- Bikhchandani, S. and Sharma, S. (2001), 'Herd Behavior in Financial Markets', *IMF Staff Papers*, International Monetary Fund, vol. 47, no. 3, pp. 279–310.
- Bresnahan, T.F. and Trajtenberg, M. (1995), 'General Purpose Technologies: Engines of Growth?', *Journal of Econometrics*, vol. 65, no. 1, pp. 83–108.
- Brynjolfsson, E. and McAfee, A. (2014), *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*, W.W. Norton & Company, New York.
- Lu, Y. and Zhou, Y. (2021), 'A Review on the Economics of Artificial Intelligence', *Journal of Economic Surveys*, vol. 35, no. 4, pp. 1045–1072. DOI: 10.1111/joes.12422.
- Morris, J.J. and Alam, P. (2008), *Analysis of the Dot-com Bubble of the 1990s*, Working Paper, Department of Accounting, Kansas State University / Kent State University, June 2008.
- Paine, J. (2026), 'Speculative Bubbles and Control Theory — An Endogenous Approach', *System Dynamics Review*, Research Article (Open Access). DOI: 10.1002/sdr.70021.
- Perez, C. (2002), *Technological Revolutions and Financial Capital: The Dynamics of Bubbles and Golden Ages*, Edward Elgar Publishing, Cheltenham.
- Shiller, R.J. (2015), *Irrational Exuberance*, 3rd edn., Princeton University Press, Princeton.
- Triplett, J.E. (1999), 'The Solow Productivity Paradox: What do Computers do to Productivity?', *The Canadian Journal of Economics / Revue canadienne d'Économique*, vol. 32, no. 2, pp. 309–334.
- van Ark, B. (2016), 'The Productivity Paradox of the New Digital Economy', *International Productivity Monitor*, no. 31, pp. 3–18.
- Varian, H. (2019), 'Artificial Intelligence, Economics, and Industrial Organization', in Agrawal, A., Gans, J. and Goldfarb, A. (eds.), *The Economics of Artificial Intelligence: An Agenda*, University of Chicago Press, Chicago, pp. 399–419.

Primary Financial Sources

- Alphabet Inc. (2025), Annual Report 2025 (Form 10-K), filed with the U.S. Securities and Exchange Commission, fiscal year ended December 31, 2025.
- Alphabet Inc. (2026), Quarterly Report Q1 2026 (Form 10-Q), filed with the U.S. Securities and Exchange Commission, quarter ended March 31, 2026.
- Amazon.com Inc. (2025), Annual Report 2025, Amazon.com Inc., Seattle.
- Amazon.com Inc. (2026), First Quarter 2026 Earnings Release, April 29, 2026.

Meta Platforms Inc. (2024), Annual Report 2024 (Form 10-K), filed with the U.S. Securities and Exchange Commission, fiscal year ended December 31, 2024.

Meta Platforms Inc. (2026), Q1 2026 Earnings Presentation, investor.atmeta.com, April 2026.

Meta Platforms Inc. (2026), First Quarter 2026 Results — Exhibit 99-1 Earnings Release, April 2026.

Microsoft Corporation (2025), Annual Report FY2025, Microsoft Corporation, Redmond.

Microsoft Corporation (2025), FY26 Q1 Earnings Call Transcript, October 2025.

Microsoft Corporation (2024), FY2024 Shareholder Letter, Microsoft Corporation, Redmond.

NVIDIA Corporation (2025), Annual Report FY2025, NVIDIA Corporation, Santa Clara.

NVIDIA Corporation (2026), Annual Report FY2026, NVIDIA Corporation, Santa Clara.

Websites and Online Sources

Macrotrends (2025), *NASDAQ Composite — 45 Year Historical Chart*, available at: www.macrotrends.net.