Partners Universal Multidisciplinary Research Journal (PUMRJ)
Volume: 03 Issue: 03 | April - May 2026 | ISSN: 3107-5789 | www.pumrj.com

The Hidden Economics of Artificial Intelligence: Why Compute Costs, Token Accounting, and Human Intuition Are Rewriting the AI Value Equation

Dr.A.Shaji George[1], Dr.Nataliia Siranchuk[2]

[1]Independent Researcher, Chennai, Tamil Nadu, India.
[2]Associate Professor, Professor at the Department of Primary Education, Borys Grinchenko Kyiv, Metropolitan University, Kyiv, Ukraine.

Abstract – There's a quiet revolution taking place in the AI sector. The news comes as top chipmakers' executives have been weighing in on the costs of compute, with some of the world's top high-performance AI research teams now spending more on compute than payroll. This article explores the hidden cost structure of enterprise AI and how token based pricing causes accounting distortions and when human labor is the less expensive and more intelligent choice. The analysis is based on lessons learnt from cloud infrastructure billing trends, the trajectory of costs of inferences and workforce productivity studies to distinguish between "real" and "imagined" savings. It says that the AI bubble, as claimed, is not about to burst, but nor is AI as claimed, "superior to human labor. The paper provides a task by task decision framework, a five step operational roadmap for leaders and a forward looking perspective on how efficiency improvements in hardware and model design could change the equation in the next 3-5 years. The main point is that the only way to extract value from AI is to match the workflow with either the worker or machine or human that has the greatest return on investment in dollars.

Keywords: AI Economics, Cost Efficiency, Human vs AI, Workforce Automation, Compute Costs, Task Allocation, AI Deployment, Scaling Laws.

1. INTRODUCTION

In the past few months, a strange fact has been trickling through the technology news. The highly paid scientists and engineers who built the models had been quietly outstripped by the expense of the computer power required for one of the largest semiconductor companies' specialized research team. The comment ricocheled through the big business press and sparked numerous analyses of the strength of the AI economy's underpinnings. The statement is significant because it burst into a convenient myth. The prevailing narrative of the last few years is that AI is almost invariably more cost-effective, more efficient, and more scalable than people. That belief has led to software budgets in the boardroom that are enormous. It has been used to make workforce decisions. But when viewed in detail, the numbers depict a more nuanced picture.

In this article, we'll take a close look at the true economics of enterprise AI. It examines the meaning behind the frontier cost statement, the hidden aspects of token based billing, and the remaining true price advantages of human workers. It also offers practical tools, which any organisation can apply immediately, such as a task allocation framework and a series of operational priorities. The idea isn't to disparage AI. This technology is game-changing and will stay that way. The goal is to enable readers to cut through the noise, down onuate results properly, and ultimately make choices that wouldn't be rejected by a chief financial officer or tossed aside while looking down a marketing presentation.

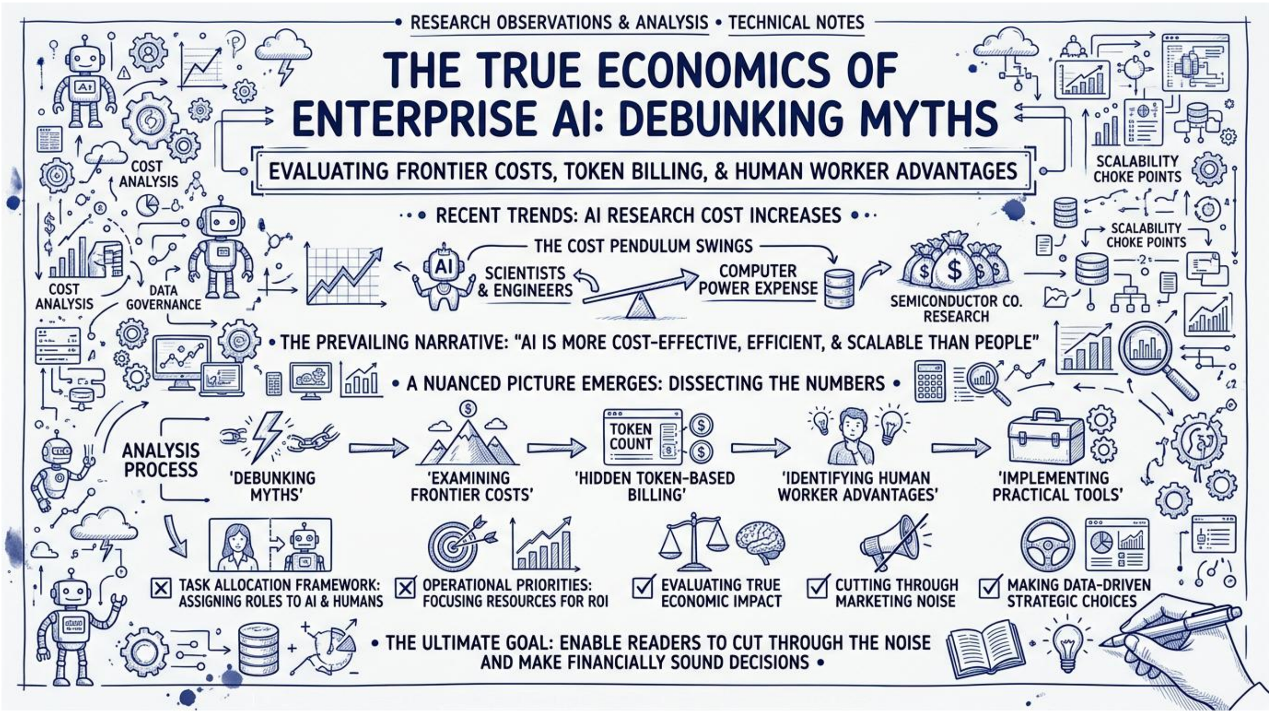

Fig -1: True Economics of Enterprise AI Debunking Myths

2. OBJECTIVES

The research here has a number of related aims.

1. First, it's intended to simplify the understanding of the recent and widely talked about assertion that the cost of compute is greater than the cost of the personnel in some advanced AI environments. As with most quotes, context is important, and a lot of the commentary on that quote has detached it from the context.
2. Second, it aims to open the black hole in the accounting books that occurs with token-based billing options, where large amounts of money may enter and leave internal reviews without real oversight.
3. Third, it assesses the entire cost stack of an AI system and compares these costs with the costs of a human knowledge worker performing the same work, in an honest side-by-side comparison.
4. Fourth, it offers a decision framework which can be used at individual workflow level by leaders to assign work to the most cost effective worker.
5. Fifth, it asks and answers the question of 'when is a burst bubble not a burst bubble, but rather a healthy correction or the beginning of a long efficiency curve.'
6. Sixth, it provides an organizational framework of operational moves that can be made in the short-term to realize true value from AI without settling for commodity output at high prices.

3. READING THE FRONTIER STATEMENT IN CONTEXT

The provoking remark was limited in scope, but it has been assumed to be general. The speaker was

Volume: 03 Issue: 03 | April - May 2026 | ISSN: 3107-5789 | www.pumrj.com

talking about a team which trains and fine tunes very large neural networks as part of their daily work. Thousands of specialized processors are required to be running for weeks at a time to perform that work. The model, a frontier size, can cost tens of millions of dollars in electricity and hardware amortization for a single training run. In that sort of setting, even researchers who earn hundreds of thousands of dollars a year are a meager line item compared to their electric bill.

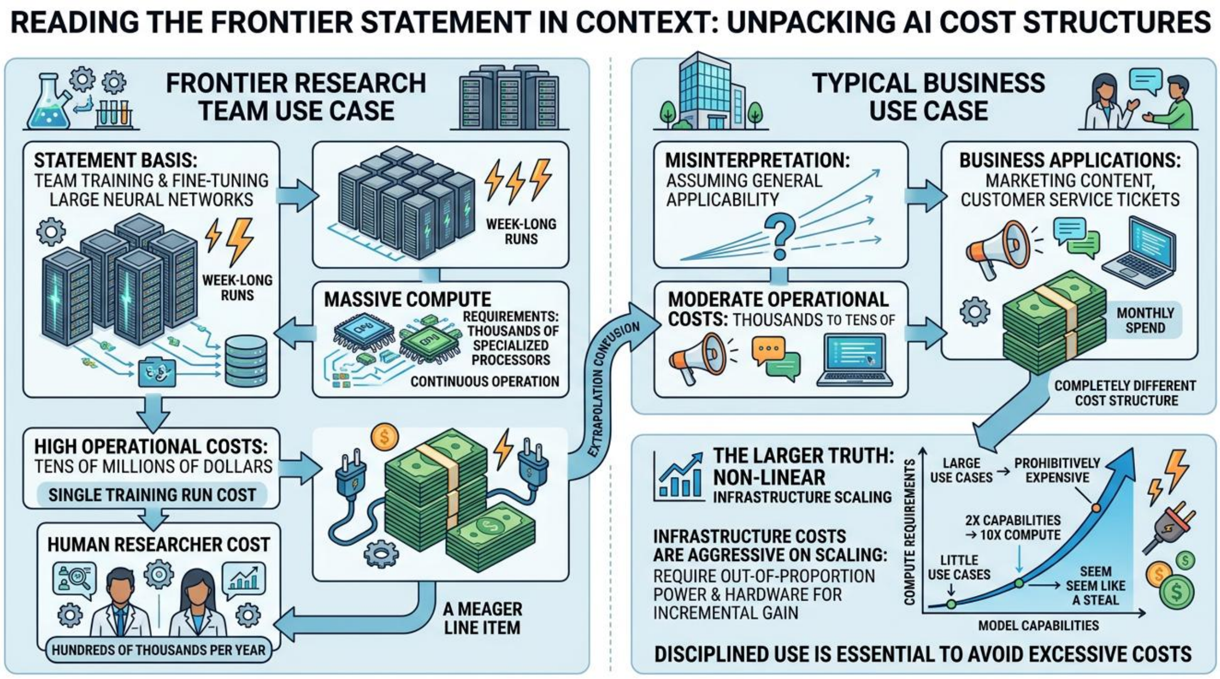

Fig -2: Reading the Frontier Statement in Context Unpacking AI Cost Structures

It got confusing when readers began to extrapolate the statement to typical business use cases. A midsize business leveraging AI for generating marketing content or assisting customer service tickets has a completely different cost structure. Its spending is in thousands or tens of thousands of dollars each month, not millions of dollars each week. The numbers are different for that company. However, the change is not the one that most people think of. The frontier comment is a larger truth, and one to be clung to. Infrastructure costs are not linear with the increase in AI workloads. They're aggressive on scaling because the latest and largest ones require an out-of-proportion amount of power and hardware to provide an increment of power. Doubling the capabilities of a model typically takes ten times the compute. This bending is what makes the economics trickier. The little AI use cases seem like they're a steal. If not used in a disciplined way, large use cases of AI can be prohibitively expensive.

4. THE TOKEN ACCOUNTING ILLUSION

The typical pricing model for today's AI generative services is token based billing. Token Text fragment corresponding to 3/4 of a word. An enterprise typically subscribes to a big block of tokens, perhaps billions, when it enters a contract with a big cloud AI provider and then employees use the tokens every day.

At first glance this appears neat. In real practice it poses a minor accounting issue. Tokens are not a tangible product. These are created by the provider as and when needed, at virtually no cost at all, just like a central bank can expand the monetary base by making a few entries on a ledger. As tokens are intangible, it's common for internal finance teams not to have a proper cost assigned to them when assessing workflows. Tokens are stored within a bulk contract with time prepaid and simply count down there are no departmental cost charges.

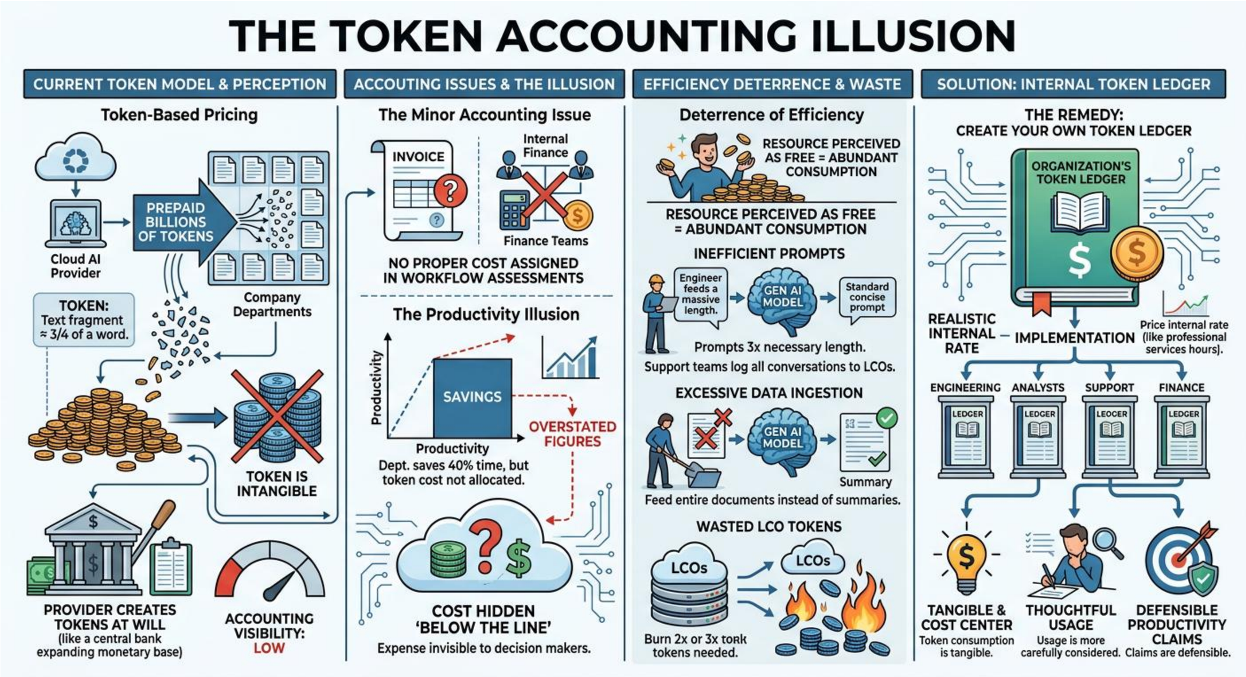

Fig -3: The Token Accounting Illusion

The result is that productivity is not a reliable measure. A department may boast that the use of AI reduced the time they spent on a task by 40%. That savings figure is overstated if the cost of the token was not originally allocated to the department. The actual expense was below the line in the cloud infrastructure, where they were not seen by decision makers. Such a trend also deters efficiency. If the resource is free, man consumes it abundantly. Engineers construct prompts three times the length they are necessary. Rather than a summary, analysts feed entire documents into models. Support teams log all conversations to LCOs that burn through twice or three times as many tokens as they need.

The remedy is simple in theory, but tough in execution. All organizations that are using Generative AI on a scale should have their own token ledger. The cost of the use of the tokens by each of the teams should be priced at a realistic internal rate like that used to allocate billable hours to professionals in professional services. Token consumption becomes tangible and visible as a cost center, usage is now more thoughtfully done, productivity claims are now defensible.

5. THE TRUE COST STACK OF AN AI WORKER

There's a place where a human worker must be seated, there's a computer to work on, there's a reasonable amount of pay, there's benefits and there is a workplace that allows concentration and cooperation. Such costs are known and can be booked in the standard financial systems. The cost of an AI system is much more intricate and much of it is not paid in cash by the end-user.

Volume: 03 Issue: 03 | April - May 2026 | ISSN: 3107-5789 | www.pumrj.com

The entire stack includes land, building, regulatory overhead and the actual data center. It comprises specialised processors which decrease in value every eighteen to twenty four months as newer generation processors are produced. It encompasses the power used while training and making predictions, which is quite substantial in some places and has begun to impact national grid planning. It consists of the industrial scale cooling systems necessary to avoid thermal damage that may be chilled water systems or immersion cooling. It contains networking gear to enable processors to coordinate between racks. It comprises software licenses, security overhead, and compliance tooling. It involves the engineers and technicians that are necessary to run the system 24/7.

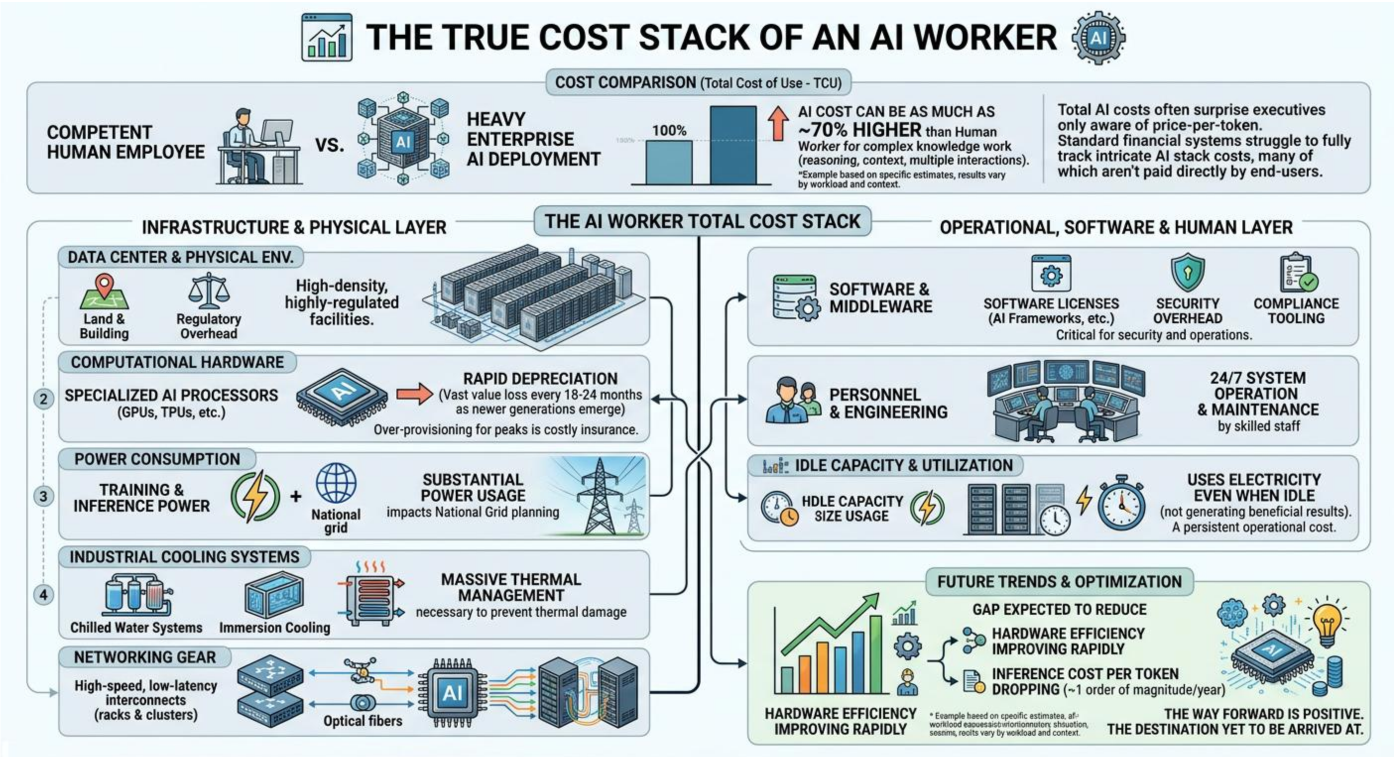

Fig -4: The True cost Stack of an AI Worker

All of this is in addition to an AI system that never sleeps. Using electricity even when it is not generating a beneficial result. There is still a cost associated with idle capacity. Over provisioning is often necessary at peak usage times to maintain response times and thus some of the hardware is costly insurance and not productive infrastructure. The total expenses for an AI to do knowledge work can be a surprise for executives, who are only told a price per token. In fact, independent estimates of the costs of heavy enterprise AI deployments have estimated the total cost of use to be as much as seventy percent higher than the total cost of use of a competent human employee, especially if the work requires reasoning, context, or multiple back and forth interactions. The aforementioned gap is expected to be reduced. Hardware efficiency is improving rapidly and the inference costs per token has dropped by about one order of magnitude each of the past few years. The way forward is positive. The destination is yet to arrive.

6. WHERE HUMANS STILL WIN ON PRICE

Partners Universal Multidisciplinary Research Journal (PUMRJ)
Volume: 03 Issue: 03 | April - May 2026 | ISSN: 3107-5789 | www.pumrj.com

After being a part of an organization for a year, a person has something that's truly hard to replicate in software. Other scientists refer to it as "tacit knowledge". Some refer to it as institutional memory or instinct. The labels vary. It's the same substance. The unspoken, tacit, or unformulated knowledge about how things really work in some context.

An experienced worker will be able to tell who needs the extra care on a Monday morning. At the time when a supplier is on the verge of falling behind, a veteran operations manager can feel it, weeks before they can even see a problem in the numbers. A proficient nurse derives a patient's condition from cues not covered in any structured set of data. All this knowledge is not recorded. It cannot be copied to a prompt at all. It resides within the individual, and develops over time, through real-life experience.

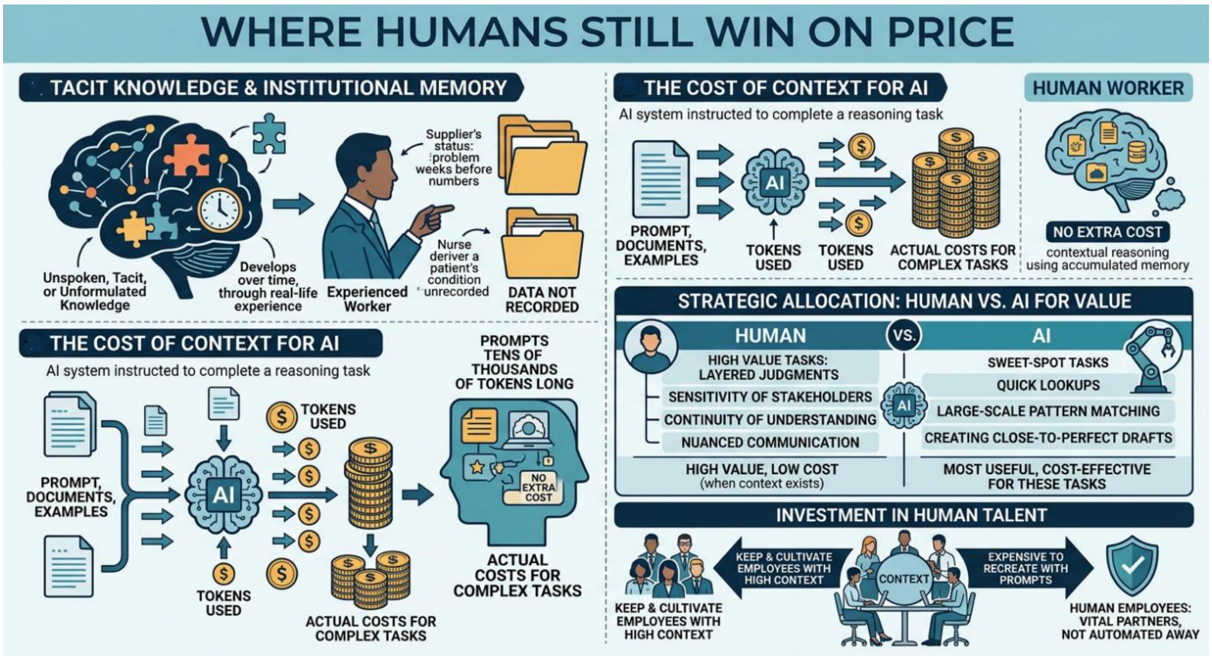

Fig -5: Human Still Win on Price

Each time an AI system is instructed to complete a reasoning task in a given context, this context must be explicitly created by means of prompting, documents, examples, etc. Tokens are used up for each level of context. The prompts used for tasks are often tens of thousands of tokens long, resulting in actual costs for complex tasks. A human does the same contextual reasoning at no extra cost and it's memory accumulated through normal course of business.

This results in a distinct pattern in the way operations are run. Humans are more effective and less expensive when the work requires layered judgments, sensitivity of stakeholders, continuity of understanding or nuanced communication. AI is most useful and cost-effective when the task is to quickly look something up, for large-scale pattern matching, or creating close-to-perfect drafts that are then refined. This does not mean there's anything wrong with AI. It's an argument for being thoughtful with allocation. The best human employees in the next 10 years won't be those that are anti-AI. These will be the ones with a lot of context, and it will be very expensive to recreate with prompts. Those that see this will put investment into keeping and cultivating such employees; they aren't going to go away simply because someone says we can automate them.

Partners Universal Multidisciplinary Research Journal (PUMRJ)
Volume: 03 Issue: 03 | April - May 2026 | ISSN: 3107-5789 | www.pumrj.com

7. A PRACTICAL FRAMEWORK FOR THE AI OR HUMAN DECISION

It's not often a useful question to ask whether there's a general reduction in cost with AI. Whether AI is cheaper or not to perform a task in a specific context. This is a framework (tested in a number of industries) for that task level decision.

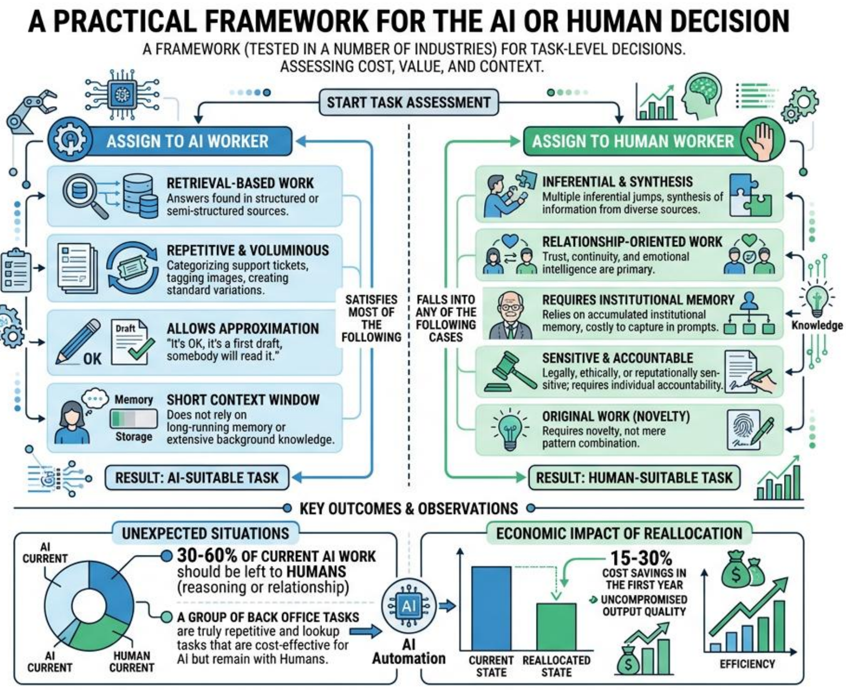

Fig -6: A Practical Framework for the AI OR Human Decision

When the task satisfies most of the following, then AI is deserving of the job. The work is retrieval based, that is, the answer can be found in a structured or semi structured source and must be retrieved. The work is repetitive and voluminous, like categorising support tickets, tagging images or creating variations of standard communication. The work will allow for approximation "it's OK, it's a first draft, somebody will read it". The work is in a short context window, that is, the task does not rely on long running memory, or on a lot of background knowledge.

The task should be done by a human if it falls into any of the following cases. The work is inferential, with a number of inferential jumps, and the synthesis of information from different sources. It is a relationship orientated work, and the quality of trust, continuity and emotional intelligence is more important than speed. Requires institutional memory, which is too costly to be used in prompts. The work is legally, ethically, or reputationally sensitive and requires a named individual to be held accountable for the work. The work should be original (novelty) rather than being the mere combination of patterns.

When organizations use this framework in good faith, they often find themselves in the following two unexpected situations. First, there is a middle ground, between 30% and 60% of the work that is currently given to AI should be left to a human as it is reasoning or relationship. Second, a group of back office tasks that, while now executed by human beings, would be vastly more cost-effective if they were automated to AI They are truly repetitive look-up tasks that haven't been automated simply because they have been done this way for so long. By simply assigning tasks to the right worker, the reallocation exercise can save 15-30% in costs in the first year without compromising on output quality.

Partners Universal Multidisciplinary Research Journal (PUMRJ)
Volume: 03 Issue: 03 | April - May 2026 | ISSN: 3107-5789 | www.pumrj.com

8. CURRENT TRENDS SHAPING THE AI COST LANDSCAPE

There are a number of developments that are changing the economic landscape of enterprise AI, and it's crucial to understand them to inform any AI strategy.

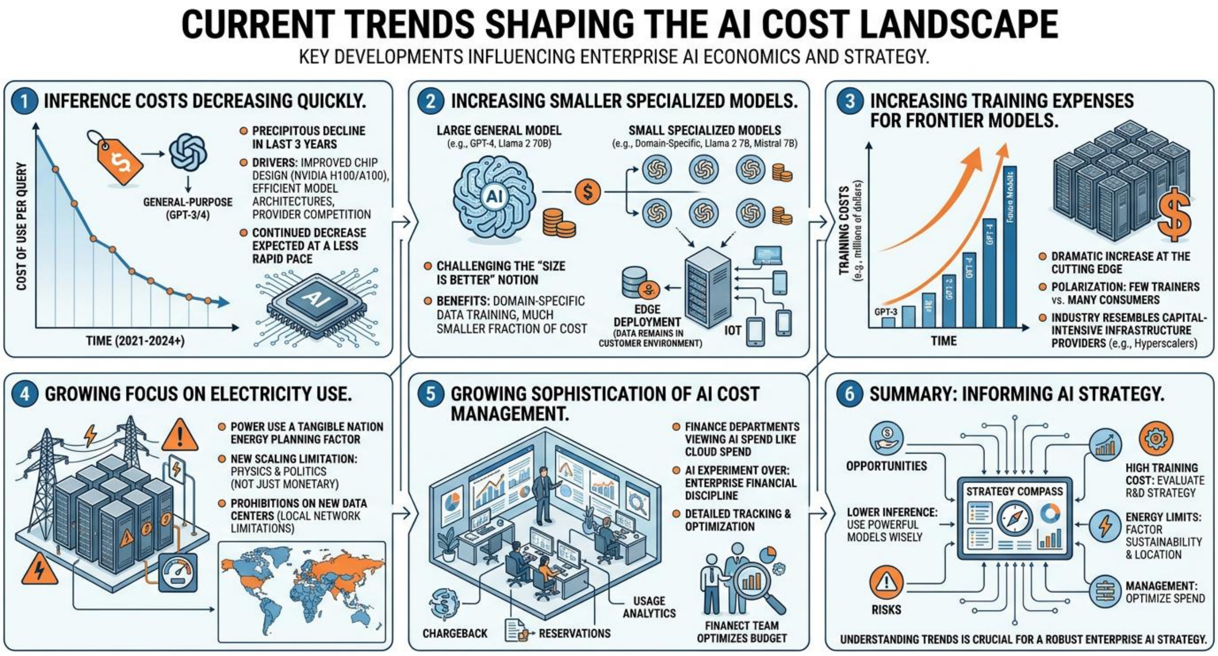

Fig -7: Current Trends Shaping The AI Cost Landscape

The first trend is that inference costs are decreasing quickly. The cost of using a powerful general purpose model to execute a query has declined precipitously over the last 3 years. The reason is the improvement in the chip design, the efficiency of the models' architecture and the competition among the providers. Inference costs will be and are supposed to keep decreasing, albeit at a less rapid pace as the easy cost reductions are realized.

The second trend is the increasing number of smaller specialized models. Many enterprise applications use cases can benefit from a small model trained on domain specific data over a large general model and require a much smaller fraction of the cost to run. The fact is that this is occurring quietly and is threatening the notion that size is the better option. It's also helping to increase deployments at the edge, where data does not leave the customer's environment.

Another trend is the increasing expenses of training frontier models. The cost of inference is decreasing, but at the cutting edge the cost of training is increasing dramatically. This is leading to the polarization of a small number of organizations which can train frontier models and a large number of organizations, which are consuming them. The economics of the industry are beginning to be more like other capital intensive industries such as, for example, a small number of large companies that offer infrastructure for a far larger community of users.

The fourth trend is the increasing focus on the use of electricity. The power use of data centers has become a tangible factor in nation energy planning. There are a number of areas where there is now a prohibition on new data centers because local networks are unable to service their needs. This is a new type of scaling limitation for AI one that has nothing to do with the monetary aspect. It's a matter of physics and politics.

The next trend is the growing sophistication of AI cost management within companies. Charg back, reservations and usage analytics are all part of the way finance departments are beginning to view AI spend in the same way that they view cloud spend as a whole. The AI experiment is over.

9. WHY ANALYSTS REMAIN DIVIDED ON THE BUBBLE QUESTION

Market observers have diverging views of the meaning of the current cost pressures. Some interpret the signals as being a sign of a potential correction. They cite the "tsunami" of capital investments in data centers, enterprise adoption lagging hyped expectations and the fact that many companies have struggled with converting AI trials into tangible economic impact. They say that many AI PoC (proof-of-concept) projects are never used in production and the capital efficiency ratio of the firms that are most ardently investing in AI infrastructure has eroded.

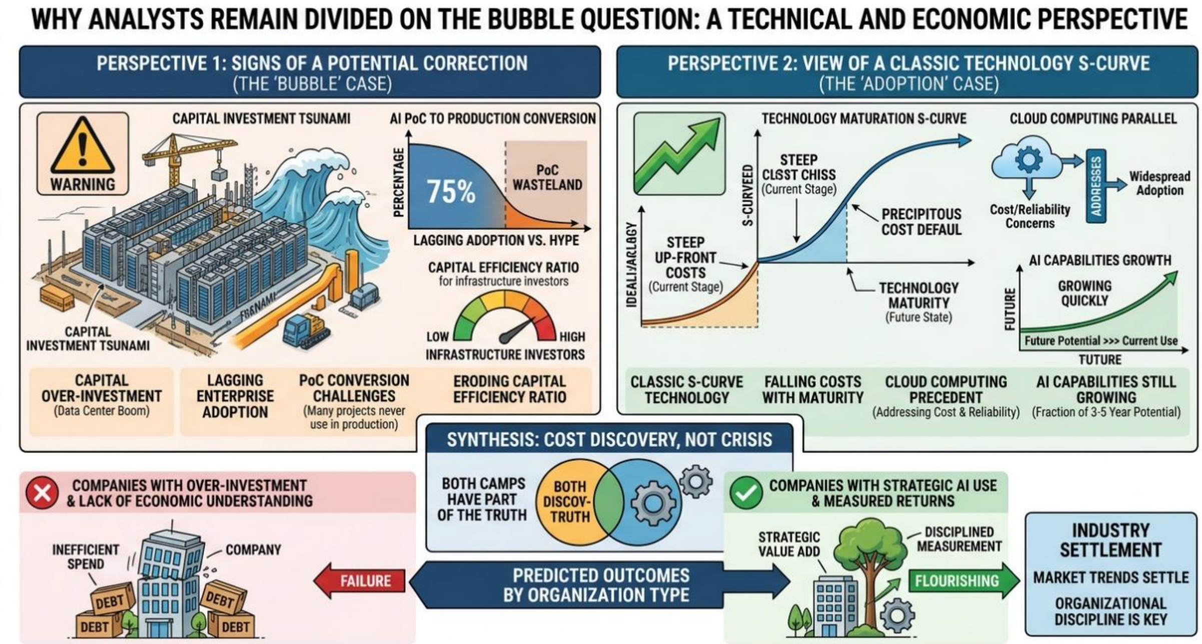

Fig -8: Analysts Remain Divided on the Bubble Question

There are many observers who interpret the same signals very differently. They recognize that they see a classic S curve technology, with up-front costs appearing to be steep yet falling off precipitously as the technology becomes more mature. They reference cloud computing, the past experience of technology, and how cost and reliability concerns were addressed to make the technology widely adopted. They say that AI capabilities are still growing quickly, and that what's possible today, is just a fraction of what will be possible in 3-5 years.

There are reasons for both positions. The best guess is that both camps have part of the truth. This is a time of cost discovery, not a cost crisis, in the industry. There will be some companies which will fail due to their over investment in the AI infrastructure with a lack of understanding of the economics. Others will flourish as they have been able to strategically use AI only in the areas where it really adds value. Where the market trend settles and how disciplined individual organisations are in measuring their own returns

Partners Universal Multidisciplinary Research Journal (PUMRJ)
Volume: 03 Issue: 03 | April - May 2026 | ISSN: 3107-5789 | www.pumrj.com

will be more important than the outcome of the trend.

10. HISTORICAL PARALLELS WORTH STUDYING

There are a number of parallels in the current moment of AI that can provide valuable lessons.

This was the trend that followed with the introduction of enterprise software in 1990s and early 2000s. Early adopters are paying huge amounts of money to implement solutions that in many cases never promised to bring back the returns that were expected. There were consultants that got rich, giving advice on deployments that didn't work. Over time, the mix of improved products, well-defined implementation approaches, and more controlled procurement processes resulted in true value. Those that survived were the organizations that didn't succumb to all the vendor hype, and they were focused on measurable results.

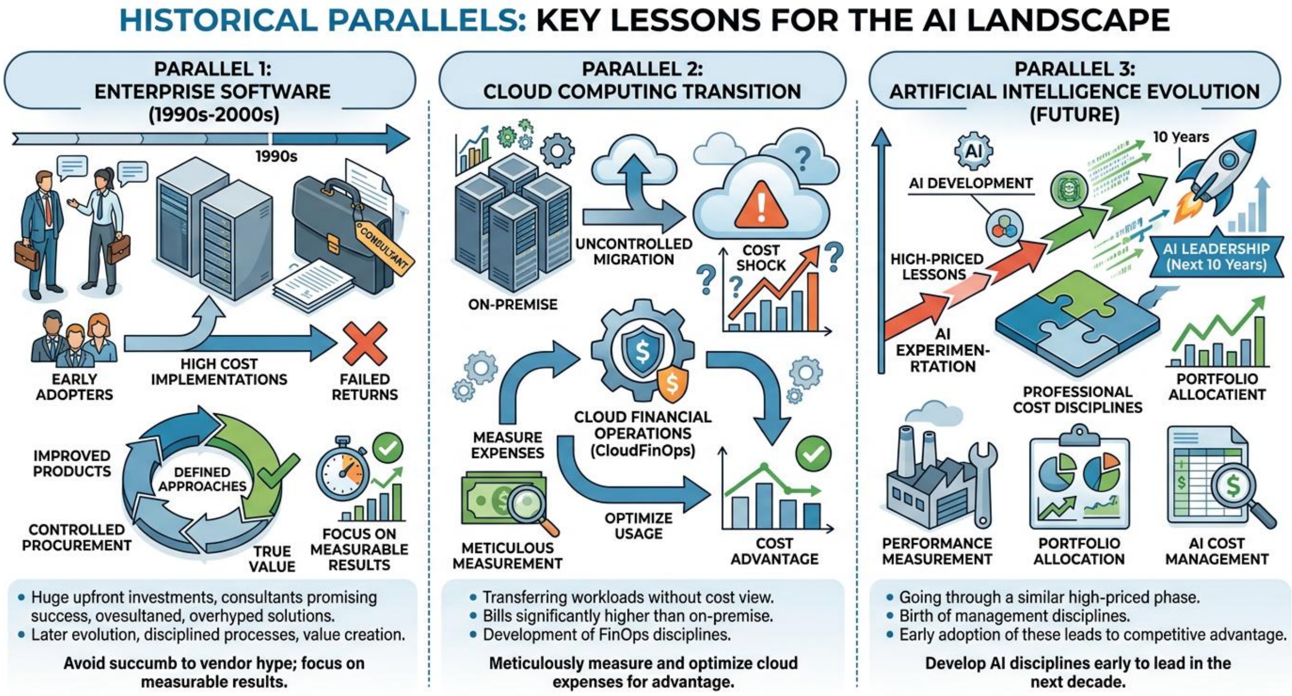

Fig -9: Key Lessons for the AI Landscape

The other parallel is the cloud computing transition. When cloud services first came into the world, many businesses were simply transferring workloads to the cloud without giving a second thought, which invariably led to cost shock. The bills came in that were much, much more than the on premise option. A field known as cloud financial operations evolved over time, the focus of which was to meticulously measure and optimize cloud expenses. The companies that adopted this field got the advantage from the cloud. Those companies that didn't turn out to be overpaid for 10 years.

AI will probably go through a similar development. There will be a series of high-priced lessons, and eventually professional disciplines of cost management, performance measurement and portfolio allocation will be born. Whoever does so first will be first and foremost in the next ten years.

Volume: 03 Issue: 03 | April - May 2026 | ISSN: 3107-5789 | www.pumrj.com

11. CHALLENGES FACING ORGANIZATIONS TODAY

There are a number of challenges that are hampering organisations' efforts to get AI economics right in practice.

The first challenge is measurement. Of the current enterprise accounting systems, few are built to monitor the use of tokens or model performance or AI productivity increases. It takes time to retrofit them, and is politically unpalatable, as it may reveal former assertions to have been exaggerated.

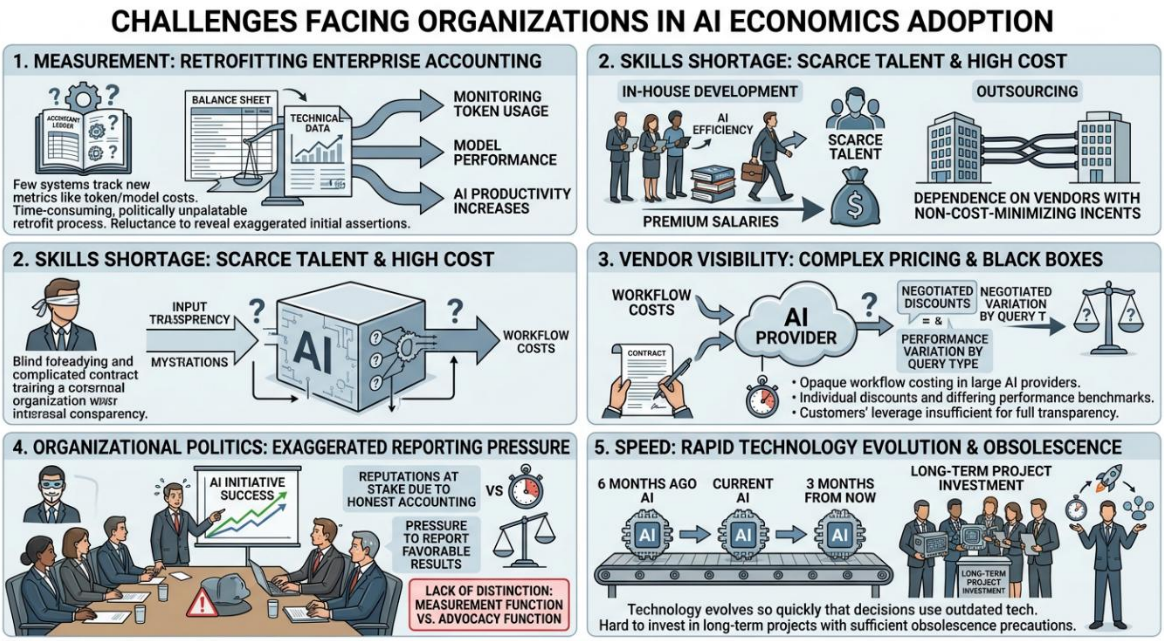

Fig -10: Challenges Facing Organizations in AI Economics Adoption

Skills shortage is the second issue. The individuals who know how to effectively utilize AI at a reasonable cost are uncommon and costly. Many organizations attempt to develop this talent in house, offering some premium salaries for people that end up leaving. The ones that outsource it are at risk of becoming dependent on vendors where the incentives may not be for cost minimisation.

The third challenge is lack of visibility from vendors. It's not always easy to find out what exactly a specific workflow costs in the large AI providers. The pricing is complex, the discounts are negotiated individually, and the performance is different for each type of query. Independent benchmarking is still in its infancy and customers' leverage is not yet sufficient to require complete transparency.

The fourth challenge is politics within the organization. AI initiatives often have executives with their reputations on the line behind them. Those reputations can be tarnished by honest cost accounting, and this puts pressure on reporting favorable results instead of accurate results. Groups which don't distinguish between the 'measurement function' and the 'advocacy function' will always make the wrong decisions.

The Fifth Challenge is speed. Technology is evolving so rapidly, that 6 months ago you are making decisions based on outdated technology. Thus, it is not easy to invest in long term projects without taking precautions for obsolescence.

Volume: 03 Issue: 03 | April - May 2026 | ISSN: 3107-5789 | www.pumrj.com

12. FUTURE PROSPECTS AND THE THREE TO FIVE YEAR VIEW

In terms of the future, there are some reasonable scenarios that will alter the costs of AI:

Hardware efficiencies will likely continue to improve at the rate of or faster than past semiconductor history. The new AI-focused chip architectures are yielding significant performance-per-watt improvements. This implies that the cost of useful output must continue to decrease with an increase in raw power use.

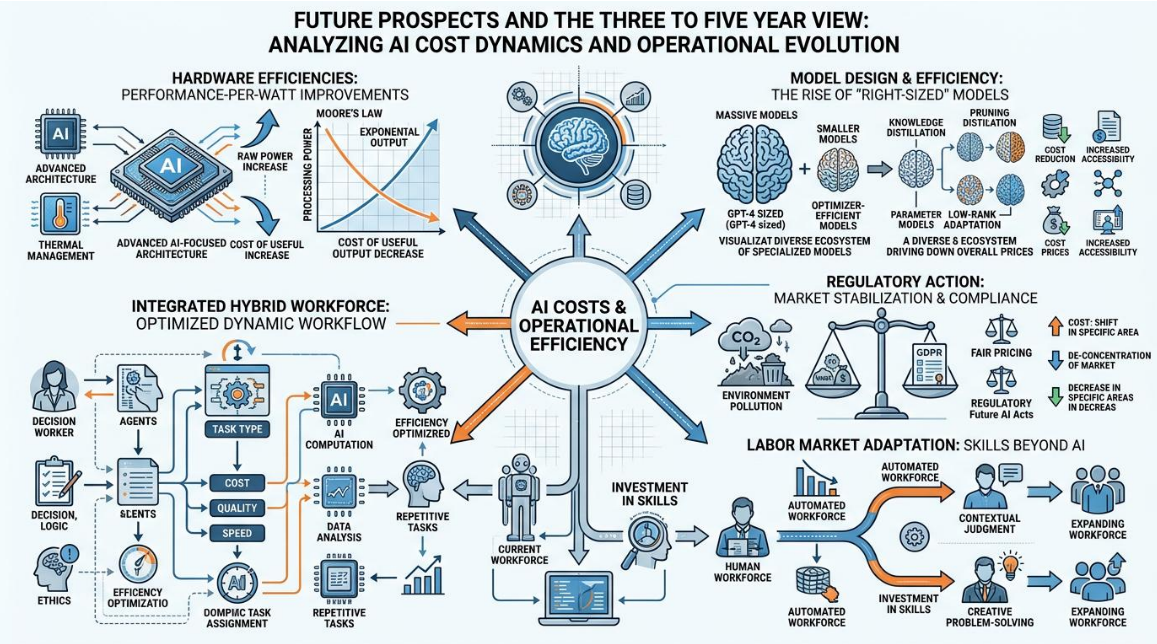

Fig -11: Analyzing AI Cost Dynamics and Operational Evolution

The design of the models is becoming more efficient. Methods enabling the models to fit in with a small number of parameters in comparison to the performance of much bigger models are maturing. This means that the most expensive and largest models don't necessarily reign supreme. There's a diverse ecosystem of right sized models coming which will drive down the price.

The market will probably be altered by regulatory action, too. AI infrastructure pollutes the environment, and there is a lack of fair pricing and concentration risk, and governments are starting to pay attention to these issues. These interventions can result in an increase or decrease in costs in certain areas.

Labor markets among human beings will adapt, too. AI will automate some roles, leading to a decrease in their numbers. Others will grow due to an increase in the worth of contextual human judgment. However, education systems lag behind these changes and that opens space for those who invest in skills that go beyond the scope of AI.

Most intriguing is the idea of what could truly be a hybrid workforce, where the AI and human workers are managed in one system, both knowing what the cost, quality and speed of each task is, and dynamically route work to the best mix. Such an operational sophistication is not very common at the moment, but it could be standard practice in the next 5 years.

13. FIVE OPERATIONAL MOVES FOR LEADERS

No matter where the market trend goes, there are five definite steps any leader can take now to place the organisation in a good light.

The initial step is to audit team usage of tokens and usage per workflow. This reveals hidden consumer habits which manipulate productivity claims. The audit should be conducted every 3 months and results should be openly communicated with finance leadership.

The second step is to establish a cost per outcome metric for each workflow using AI. It should be measured in dollars per completed task not dollars per seat or dollars per hour. This reframing puts teams in a position to ask themselves, "Is that AI creating value or just eating up resources.

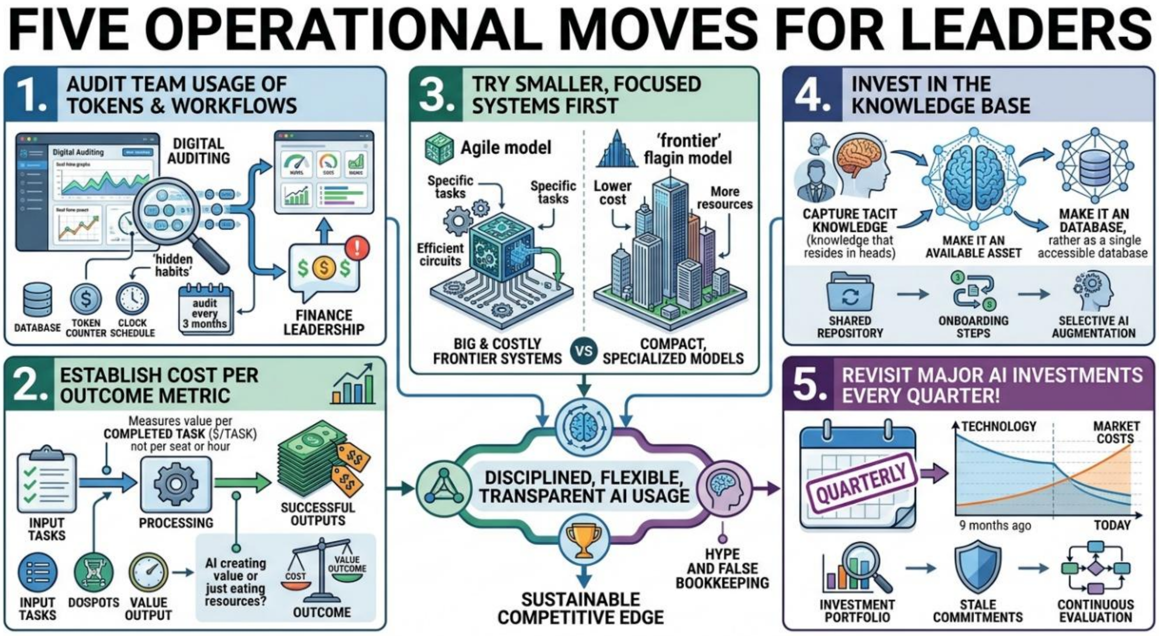

Fig -12: Five Operational Moves for Leaders

The third step is to try out smaller, more focused systems first before moving on to the big and costly frontier systems. There are lots of things that perform wonderfully on smaller models that are a lot less expensive than the flagship models. It can translate to a lot of money saved and for end users, there may be no noticeable difference in performance.

The fourth step is to invest in the knowledge base of people. Capture the knowledge that resides in the heads of a few people and not let it get locked in that place but make it available as an asset and not a single point of failure. This investment also makes it easier to onboard new staff members as well as providing a base for selective AI augmentation.

Revisit all major AI investments every quarter! Cost is moving so quickly that what seemed like a good choice nine months ago might seem like a bad choice today. Standing review process ensures that stale commitments are not built up.

Combined, these five moves result in discipline, flexibility, and transparent organization around the use of AI investments. It turns out to be a sustainable competitive edge in a marketplace where many are working off of hype and false bookkeeping.

14. THE ETHICAL DIMENSION OF THE COST EQUATION

The business model outlined in this article can serve as a guide for companies to help them decide where to use AI and where to preserve their human employees. However, at the macro level rational choices have macro implications. Any serious discussion of an AI economy needs to consider the moral implications of these decisions.

The first is job losses. Real consequences for humans doing jobs that are deemed to be cheaper with AI. Others will be given new tasks of greater value. Others will be released. The International Labour Organization (ILO) and the World Economic Forum (WEF) have published research indicating that over time AI will generate new jobs but in the interim, there will be a mismatch in the job market, with clerical and administrative, and some creative roles coming under the most pressure. Leaders who use the cost framework, but not a related investment in retraining and transition support, are getting efficiency at the social cost that will be paid by their communities.

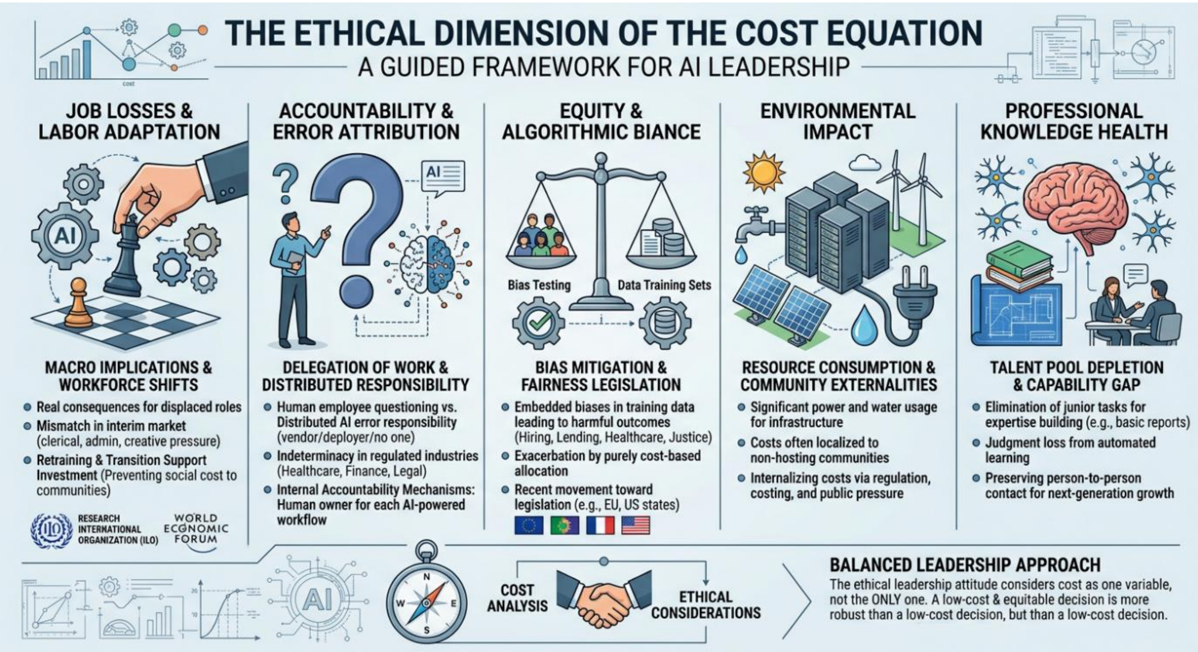

Fig -13: The Ethical Dimension of the Cost Equation

The second aspect is accountability. As work is delegated from human to AI, it will be more difficult to attribute errors. A human employee can be questioned, corrected, and held accountable for his or her error. An AI system which fails generates a more distributed responsibility with the vendor or deployer, or no one at all. In principle, this indeterminacy is already being experienced by industries subject to highly specific regulatory requirements, including health care, financial and legal services. Leaders cannot take it for granted that legal frameworks will rapidly prove to be up to speed. They should put in place internal accountability mechanisms, which means a human owner for each AI powered workflow.

Equity is the third consideration. AI systems embed the biases found in their training sets, and biases can lead to outcomes that harm certain groups in hiring, lending, access to health care, and criminal justice. Such harms can in fact be exacerbated by a purely cost based allocation scheme, if the bias testing is not considered as a "non-negotiable" element of deployment. There has been a recent movement towards legislation on algorithmic fairness in several jurisdictions, such as the European Union and some US states. Saving money by not doing so is short-term as regulatory penalties and reputational consequences follow.

The fourth factor is the environmental impact. Large scale AI infrastructure requires significant amounts of power and water, and these costs often lie on communities that did not select to host these data centers. The environmental costs should be considered when the leaders make their cost decisions – not to be fashionable, but rather because more and more, the environmental costs are internalized via regulation, costing and public pressure.

The fifth is professional knowledge's long term health. Organizations that eliminate tasks that younger professionals used to do to gain expertise may end up depleting the talent pool unintentionally. If a junior analyst doesn't do a basic report, because the AI does it for them, they may not learn to do a basic report and the judgment that comes from it. Saving money today equals capability gap tomorrow. Rational management will leave some person-to-person contact points, not so that they are of maximum economic value in the short term, but so that they continue to promote the growth of the next generation.

The system presented in this article is still valid. Its ethical dimension involves that the attitude of leadership is to consider cost as one of the variables and not the only one. A low-cost and equitable decision is more robust than a low-cost decision.

15. CONCLUSION

The frontier, which kicked off the current debate, does not necessarily represent a bursting bubble. It's a sign of an industry that is maturing. Senior leaders are for the first time being forced to question whether the intelligence that they are purchasing is representative of the level of value that they are obtaining. Answer depends on the task and that is the crux of the whole discussion. What it comes down to is there is no justification for AI maximalism or AI skepticism. In certain instances, AI can be more cost-effective than human effort, and in other cases, human effort can be more cost-effective than AI. The answer lies in computing as a precious source and in charging people for this use and allocating each computing task to the worker silicon or biological who obtains the greatest benefit per dollar. The organisations that develop this discipline will perform better than the enthusiasts and the cynics - they will be making decisions based on information, the others based on assumptions. The machines are indeed powerful and will become even more powerful. It's not as though humans are not smart and flexible enough to deal with the hype cycle. Those who will know how to call upon each one in a timely manner is the real advantage in the next decade.

REFERENCES

[1] Amini, L. (2022). Research directions in AI algorithms and systems. From Artificial Intelligence to Brain Intelligence. https://doi.org/10.1201/9781003338215-2
[2] Arya, A., & Mittendorf, B. (2007). Interacting supply chain distortions: The pricing of internal transfers and external procurement. The Accounting Review, 82(3), 551-580. https://doi.org/10.2308/accr.2007.82.3.551
[3] George, A., George, A., T.Baskar, & Pandey, D. (2021a). XDR: The evolution of Endpoint Security Solutions -Superior extensibility and analytics to satisfy the organizational needs of the future. Zenodo (CERN European Organization for Nuclear Research). https://doi.org/10.5281/zenodo.7028219
[4] Borowiack, E., Sadowska, E., Nowak, A., & Brozek, J. (2023). MSR61 AI support reduced screening burden in a systematic review with costs and cost-effectiveness outcomes (SR-CCEO) for cost-effectiveness

modeling. Value in Health, 26(6), S288. https://doi.org/10.1016/j.jval.2023.03.1596

[5] George, D., George, A., & Dr.T.Baskar. (2023). SD-WAN Security Threats, Bandwidth Issues, SLA, and Flaws: An In-Depth Analysis of FTTH, 4G, 5G, and Broadband technologies. Zenodo (CERN European Organization for Nuclear Research). https://doi.org/10.5281/zenodo.8057014

[6] Cronin, I. (2026). Scaling training with infrastructure and distributed systems. Building and Training Generative AI Models. https://doi.org/10.1007/979-8-8688-2332-9_6

[7] George, D., Dr.T.Baskar, Srikaanth, P. B., & Dr.M.M.Karthikeyan. (2025). Building resilient API security through a Five-Dimensional Framework for data breach prevention in modern digital ecosystems. Zenodo (CERN European Organization for Nuclear Research). https://doi.org/10.5281/zenodo.15862111

[8] Formosa, P., Rogers, W., Griep, Y., Bankins, S., & Richards, D. (2022). Medical AI and human dignity: Contrasting perceptions of human and artificially intelligent (AI) decision making in diagnostic and medical resource allocation contexts. Computers in Human Behavior, 133, 107296. https://doi.org/10.1016/j.chb.2022.107296

[9] George, D., Dr.T.Baskar, & Srikaanth, P. B. (2025). Bridging the Security Skills Gap: A comprehensive framework for developing application security competencies in modern software engineering. Zenodo (CERN European Organization for Nuclear Research). https://doi.org/10.5281/zenodo.15616416

[10] George, D., Dr.T.Baskar, & Dr.M.M.Karthikeyan. (2026). Cloud Security Architecture: A comprehensive guide to zero trust, governance, and operational resilience. Zenodo (CERN European Organization for Nuclear Research). https://doi.org/10.5281/zenodo.19551592

[11] Milchanowski, K. L. (2026). Economic justification and return on investment. Return on Intelligence. https://doi.org/10.4324/9781003749899-5

[12] George, D., Dr.S.Sagayarajan, Baskar, D., & Pandey, D. (2024). Assessing the security and privacy implications of India's DigiYatra initiative. Zenodo (CERN European Organization for Nuclear Research). https://doi.org/10.5281/zenodo.14599297

[13] Takkalapally, D. (2024). Deepperfnet: Ai-powered predictive workload forecasting and autonomous resource scaling. International Journal of AI, BigData, Computational and Management Studies, 5, 222-234. https://doi.org/10.63282/3050-9416.ijaibdcms-v5i2p122

[14] George, D., & Dr.T.Baskar. (2025). Security and privacy comparison of Arattai, WhatsApp, and WeChat: India's messaging app landscape and digital sovereignty. Zenodo (CERN European Organization for Nuclear Research). https://doi.org/10.5281/zenodo.17483067

[15] Ahn, C. M., Sim, W., An, M., & Hill, C. E. (2026). Dream work in the united states and korea: Self-oriented and relationship-oriented themes of female college students.. Dreaming. https://doi.org/10.1037/drm0000342

[16] Author, N. G. (1979). General basin modeling for site suitability. draft report 1. baseline data. https://doi.org/10.2172/5601968

[17] George, A., George, A., T.Baskar, & Pandey, D. (2021b). XDR: The evolution of Endpoint Security Solutions -Superior extensibility and analytics to satisfy the organizational needs of the future. Zenodo (CERN European Organization for Nuclear Research). https://doi.org/10.5281/zenodo.7028219

[18] Cao, B., Cai, D., & Lam, W. (2025). Infiniteicl: Breaking the limit of context window size via long short-term memory transformation. Findings of the Association for Computational Linguistics: ACL 2025. https://doi.org/10.18653/v1/2025.findings-acl.595

[19] George, D. (2024a). When Trust Fails: Examining Systemic Risk in the Digital Economy from the 2024 CrowdStrike Outage. Zenodo (CERN European Organization for Nuclear Research). https://doi.org/10.5281/zenodo.12828222

[20] George, D. (2026d). IEC 62443 Wireless Security: Deploying OT wireless controllers in industrial factory networks. Zenodo (CERN European Organization for Nuclear Research). https://doi.org/10.5281/zenodo.19428491

[21] Kanumuri, S., & Zeier, M. (2024). Cloud cost visibility. Scaling Cloud FinOps. https://doi.org/10.1007/979-8-8688-0388-8_2

[22] George, D. (2026e). Multi-Vendor firewall strategy: IT, OT, and edge networks. Zenodo (CERN European Organization for Nuclear Research). https://doi.org/10.5281/zenodo.19630402

[23] Kasimala (2024). Computational cost and output quality analysis of prompt engineering methods in AI systems. Journal of Information Systems Engineering & Management. https://doi.org/10.52783/jisem.v9i4.166

[24] George, D. (2026c). Architectural Convergence in Security Operations: a technical framework for AI-Augmented Threat Detection, Automated response, and Organizational cyber resilience. Zenodo (CERN European Organization for Nuclear Research). https://doi.org/10.5281/zenodo.19986642

[25] Lavin, L. (2025). Large load management and grid planning. https://doi.org/10.2172/2589486

[26] George, D. (2026b). Architectural Convergence in Security Operations: a technical framework for AI-

Augmented Threat Detection, Automated response, and Organizational cyber resilience. Zenodo (CERN European Organization for Nuclear Research). https://doi.org/10.5281/zenodo.19986642

[27] Liang, S., & Cai, Y. (2025). Challenges of generative AI on human–ai interaction and collaboration. Human-AI Interaction and Collaboration. https://doi.org/10.1017/9781009587877.012

[28] George, D. (2024b). Personal privacy at risk: The security threats of sharing boarding passes online. Zenodo (CERN European Organization for Nuclear Research). https://doi.org/10.5281/zenodo.14503012

[29] George, D. (2025a). An exploratory study of friendship marriage and its role in redefining partnership for economic security and personal autonomy in modern society. Zenodo (CERN European Organization for Nuclear Research). https://doi.org/10.5281/zenodo.17137271

[30] Liu, J., Li, L., Xiang, T., Wang, B., & Qian, Y. (2023). TCRA-LLM: Token compression retrieval augmented large language model for inference cost reduction. Findings of the Association for Computational Linguistics: EMNLP 2023. https://doi.org/10.18653/v1/2023.findings-emnlp.655

[31] George, D. (2025d). Sanchar Saathi Digital Security versus Civil Liberty in India 's Smartphone Era. Zenodo (CERN European Organization for Nuclear Research). https://doi.org/10.5281/zenodo.17838468

[32] Lwango, A. (2025). Original reasearch work. https://doi.org/10.55277/researchhub.ao86m1g3

[33] George, D. (2026a). Self-Driving Networks: AI automation for Enterprise IT. Zenodo (CERN European Organization for Nuclear Research). https://doi.org/10.5281/zenodo.19335608

[34] Mamidisetti, G., Makala, R., & Anilkumar, C. (2020). RETRACTED ARTICLE: A novel access control mechanism for secure cloud communication using SAML based token creation. Journal of Ambient Intelligence and Humanized Computing, 15, 147 - 147. https://doi.org/10.1007/s12652-020-02427-8

[35] George, D. (2025c). Cyber resilience in an AI-Driven world: a Strategic framework. Zenodo (CERN European Organization for Nuclear Research). https://doi.org/10.5281/zenodo.18002783

[36] Neier, A. (2012). Accountability. The International Human Rights Movement. https://doi.org/10.23943/princeton/9780691135151.003.0011

[37] George, D. (2025b). Digital Watermarking in Cloud Environments for Copyright Protection: A Comprehensive review. Zenodo (CERN European Organization for Nuclear Research). https://doi.org/10.5281/zenodo.17726895

[38] Niznikowski, T., Sadowski, J., & Mastalerz, A. (2018). The effectiveness of different types of verbal feedback on learning complex movement tasks. Human Movement, 14(2), 148-153. https://doi.org/10.2478/humo-2013-0009

[39] Oard, D. W., & Hackett, P. G. (1997). Document translation for cross-language text retrieval at the university of maryland. https://doi.org/10.6028/nist.sp.500-240.clir-umd

[40] Peters, M. O. (1949). Supervising the experienced worker. Social Casework, 30(5), 188-195. https://doi.org/10.1177/104438944903000503

[41] Sadykhov, R., Goodell, G., De Montigny, D., Schoernig, M., & Treleaven, P. (2023). Decentralized token economy theory (detect): Token pricing, stability and governance for token economies. Frontiers in Blockchain, 6. https://doi.org/10.3389/fbloc.2023.1298330

[42] Sakib, M. S., & Sun, Y. (2024). From cooking recipes to robot task trees – improving planning correctness and task efficiency by leveraging llms with a knowledge network. 2024 IEEE International Conference on Robotics and Automation (ICRA). https://doi.org/10.1109/icra57147.2024.10611369

[43] Shonder, J. (2013). Beyond guaranteed savings: Additional cost savings associated with ESPC projects. https://doi.org/10.2172/1086643

[44] Turek, D., & Radgen, P. (2021). Optimized data center site selection—mesoclimatic effects on data center energy consumption and costs. Energy Efficiency, 14(3). https://doi.org/10.1007/s12053-021-09947-y

[45] Walujo, D. A. (2018). INFLUENCE OF REPETITIVE WORK ON WORK PRODUCTIVITY ON DEPARTMENT OF PACKAGING. Tibuana, 1(1), 34-40. https://doi.org/10.36456/tibuana.1.1.1589.34-40

[46] Zhang, K., & Shasha, D. (1997). Tree pattern matching. Pattern Matching Algorithms. https://doi.org/10.1093/oso/9780195113679.003.0014

[47] (2004). Reasoning and task environments. Psychology of Reasoning. https://doi.org/10.4324/9780203506936-4

[48] (2013). The tragedy of the commons. Parks and Carrying Capacity. https://doi.org/10.2307/jj.41379835.6

[49] (2014). Tacit knowledge and language. Tacit Knowledge. https://doi.org/10.4324/9781315729886-11

[50] (2015). Quantitative data analysis: Making sense of inferential stats. The Social Work Student's Research Handbook. https://doi.org/10.4324/9781315730622-24

[51] (2022). Andrew lovett-barron discusses job roles, responsibilities, and automation. Better Research through Automation. https://doi.org/10.4135/9781529604948.n1

[52] (2023). Moore's law, n.. Oxford English Dictionary. https://doi.org/10.1093/oed/6338591394
[53] (2026). 6 what does AI cost?. AI for Business. https://doi.org/10.1071/9781486321438_ch06
[54] Adabala, S. K. (2024). Cloud migration: Overcoming challenges and ensuring successful transition. Journal of Artificial Intelligence & Cloud Computing, 3(1), 1. https://doi.org/10.47363/jaicc/2024(3)e220
[55] Alan, A., & Gelmez, K. (2025). Can AI assess design students' projects? discussing the future of assessments with a proof of concept. LearnXDesign 2025. https://doi.org/10.21606/drslxd.2025.136
[56] Fischer, B. (2019). Capital-intensive industries in newly industrializing countries. Routledge. https://doi.org/10.4324/9780429033377
[57] Frank, R., Schumacher, G., & Tamm, A. (2023). Cloud transformation – how the public cloud is changing businesses. Cloud Transformation. https://doi.org/10.1007/978-3-658-38823-2_9
[58] H Zeiner-Gundersen, D. (2022). Project management shall use artificial intelligence (AI) driven algorithms when addressing project costs and risks. Current Trends in Engineering Science (CTES), 2(01), 1-2. https://doi.org/10.54026/ctes/1010
[59] Karlsson, J. F., & Moshfegh, B. (2005). Investigation of indoor climate and power usage in a data center. Energy and Buildings, 37(10), 1075-1083. https://doi.org/10.1016/j.enbuild.2004.12.014
[60] Kommuri, R. S. (2025). AI integration for sustainable finance data analyst interdisciplinary: Framework optimized resource management and cost efficiency. 2025 IEEE International Conference on Emerging Trends in Computing and Communication (ETCOM). https://doi.org/10.1109/etcom66606.2025.11436692
[61] Liu, Y., & Zhang, Z. (2025). Unlocking operational efficiency: How AI human capital investment enhances data processing efficiency?. Economics Letters, 247, 112147. https://doi.org/10.1016/j.econlet.2024.112147
[62] Mamillapalli, S. K. (2025). AI innovations in cloud financial operations: A framework for the future. Journal of Artificial Intelligence & Cloud Computing, 4(1), 1. https://doi.org/10.47363/jaicc/2025(4)e259
[63] Pasupuleti, M. K. (2025). Ai-driven quantum chip design: Integrating semiconductors, nanotechnology, and machine learning. https://doi.org/10.62311/nesx/rb610
[64] Russell, J. (2022). The 1990s and 2000s. Management and Labor Conflict. https://doi.org/10.4324/9780429294938-5
[65] Sakly, H., Guetari, R., & Kraiem, N. (2025). Scaling AI capabilities and establishing a roadmap for sustainable growth in healthcare. Scalable Artificial Intelligence for Healthcare. https://doi.org/10.1201/9781003480594-8
[66] Thillairajah, V., Gosain, S., & Clarke, D. (2007). Realizing the promise of RFID. Enterprise Architecture and Integration. https://doi.org/10.4018/978-1-59140-887-1.ch018
[67] Vemulapalli, V. K. C. (2025). Choosing the right model for enterprise AI applications: A comprehensive analysis of small language models versus large language models in enterprise architecture. European Modern Studies Journal, 9(5), 275-284. https://doi.org/10.59573/emsj.9(5).2025.25
[68] (2016). Data center equipment energy consumption. Making Your Data Center Energy Efficient. https://doi.org/10.1201/b11255-8
[69] (2020). Projected costs of generating electricity 2020. OECD. https://doi.org/10.1787/a6002f3b-en
[70] (2024). Securing AI model weights: Preventing theft and misuse of frontier models. https://doi.org/10.7249/rra2849-1
[71] (2025). Comparing training ammunition options for firing ranges: Life-cycle costs, market trends, and purchasing strategies. https://doi.org/10.7249/rra3020-2
[72] (2026). The platforms that power ai -first organizations. Architected Intelligence, 361-386. https://doi.org/10.1002/9781394374175.ch12
[73] Adedokun, A. (2024). Global AI regulatory landscape challenges, trends, and future outlook. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4855369
[74] Ayeni, O. (2025). Advanced multi-phase project management frameworks: Optimizing ai-driven decision-making, risk control, and efficiency.. International Journal of Research Publication and Reviews, 6(3), 330-348. https://doi.org/10.55248/gengpi.6.0325.1117
[75] Bartel, A., & Sicherman, N. (1995). Technological change and the skill acquisition of young workers. https://doi.org/10.3386/w5107
[76] Björnson, E., Hoydis, J., & Sanguinetti, L. (2017). Massive MIMO networks: Spectral, energy, and hardware efficiency. Foundations and Trends® in Signal Processing, 11(3-4), 154-655. https://doi.org/10.1561/2000000093

[77] Cousineau, C., Herger, N., & Dara, R. (2025). Transparency requirements across AI legislative acts, frameworks and organizations: Shaping a sample transparency card. AI and Ethics, 5(4), 4255-4267. https://doi.org/10.1007/s43681-025-00725-5
[78] Crawford, K., & Paglen, T. (2021). Excavating AI: The politics of images in machine learning training sets. AI & SOCIETY. https://doi.org/10.1007/s00146-021-01162-8
[79] Dull, D., & Tracy, A. (2026). Ai-driven sustainability. AI-Powered Sustainability. https://doi.org/10.4324/9781003715689-5
[80] ENDOUWARE,, E. (2026). Human AI collaboration tools for boosting productivity of library staff in academic institutions. Rivers State University Journal of Librarianship, Information and Contemporary Studies (RSU-JOLICS). https://doi.org/10.61955/xrcars
[81] Fives, A. J. (2008). Human flourishing: The grounds of moral judgment. The Journal of Value Inquiry, 42(2), 167-185. https://doi.org/10.1007/s10790-008-9098-8
[82] Irawan, V. E., & Paramita, A. S. (2026). Sentiment analysis of ai-driven job automation in indonesia: Integrating LSTM, bertopic, and time-series forecasting for job displacement risk prediction. Journal of Applied Informatics and Computing, 10(2), 1806-1817. https://doi.org/10.30871/jaic.v10i2.12335
[83] Kastl-Riemann, G. (2025). Regulation of generative AI speech: An EU perspective. The Oxford Handbook of the Foundations and Regulation of Generative AI. https://doi.org/10.1093/oxfordhb/9780198940272.013.0022
[84] Kozma, R., Wang, F. Y., Wan, Y., & Garai, V. (2025). IEEE SMCS convenes 2025 executive workshop to define strategic AI initiatives. The International Journal of Intelligent Control and Systems. https://doi.org/10.62678/ijics202509.10274
[85] Srivastava, A. (2024). Impact of AI/ML on job market and skills set and health industry. ESP Journal of Engineering & Technology Advancements, 4(3), 122-126. https://doi.org/10.56472/25832646/jeta-v4i3p111
[86] (2024). Keynote talk: Rethinking network fabric architectures for distributed AI/ML platforms. 2024 17th IEEE/ACM International Workshop on Network on Chip Architectures (NoCArc). https://doi.org/10.1109/nocarc64615.2024.10749926
[87] (2025). Enhancing in-house U.S. government AI talent. https://doi.org/10.7249/wra3882-1
[88] (2025). Bridging the AI skills gap. https://doi.org/10.1787/66d0702e-en