Business and Management Sciences
International Quarterly Review

SPECIAL ISSUE
In memoria del Prof. Gianluca Colombo

"Proprietà, governance e management oggi: le sfide della complessità e dell'innovazione strategica per imprese e studiosi/e"

Race with the machine: can Artificial Intelligence help managers fight their limits?

Francesco M. Sacco

Pavia, December 31, 2025
Volume 16 – N. 4/2025 – SPECIAL ISSUE
DOI: 10.13132/2038-5498/16.4.1361-1381

www.ea2000.it
www.economiaaziendale.it

PaviaUniversityPress

Electronic ISSN 2038-5498
Reg. Trib. Pavia n. 685/2007 R.S.P.

Vol. 16-4/2025 - DOI: 10.13132/2038-5498/16.4.1361-1381

Race with the machine: can Artificial Intelligence help managers fight their limits?

Francesco M. Sacco

Researcher
of Digital Economy and Management.
Department of Economics.
Insubria University, Varese, Italy

Corresponding Author:

Francesco M. Sacco

francesco.sacco@uninsubria.it

Cite as:

Sacco, F. M. (2025). Race with the machine: can Artificial Intelligence help managers fight their limits? E*conomia Aziendale Online*, *16*(4), 1361-1381.

Section: *Refereed Paper*

SPECIAL ISSUE 2025

Received: November 2025
Published: 31/12/2025

ABSTRACT-SOMMARIO

Intelligence, whether biological or artificial, is fundamentally constrained by energy. This paper explores the parallel *energy taxes* paid by human evolution and modern Artificial Intelligence (AI), arguing that both systems are shaped by the thermodynamic need to optimize information processing within strict metabolic limits. While the human brain solved this constraint by evolving energy-saving heuristics, manifesting today as cognitive biases, AI is currently overcoming its energy bottlenecks through massive scaling and architectural efficiency. Managerial decision making is persistently vulnerable to cognitive bias, yet many debiasing interventions remain modest in impact because they target surface-level judgment errors rather than the structural constraints that make heuristics metabolically and organizationally attractive. This conceptual paper reframes cognitive bias as a constraint-induced default emerging from (i) evolutionary canalization of survival-oriented responses and (ii) the efficiency topology of neural networks that favors low-cost, fast, "good-enough" inference over globally optimal computation. In parallel, contemporary artificial intelligence systems face their own energy and scaling constraints, but they also offer a new design opportunity: AI can function as a metacognitive scaffold that externalizes and stress-tests managerial assumptions through structured dialogue. Building on discourse-oriented strategy research, we argue that AI is most valuable not as an oracle that replaces judgment, but as a dialogical partner that systematically generates counter-arguments, alternative stakeholder perspectives, and probabilistic scenario distributions to induce reflective "System 2" pauses in high-stakes workflows. Because AI can also introduce automation bias, hallucinations, and training-data bias, effective debiasing must be treated as a socio-technical design problem involving transparency, contestability, and governance. We conclude by outlining managerial design principles and a research agenda for testing when and how AI-supported dialogue improves decision quality under uncertainty.

L'intelligenza, biologica o artificiale, è fondamentalmente vincolata dall'energia. Questo articolo esplora, in parallelo, le *tasse energetiche* pagate dall'evoluzione umana e dall'Intelligenza Artificiale (IA) moderna, sostenendo che entrambi i sistemi sono plasmati dalla necessità termodinamica di ottimizzare l'elaborazione delle informazioni entro limiti metabolici rigorosi. Mentre il cervello umano ha risolto questo vincolo evolvendo euristiche di risparmio energetico, oggi manifestate come bias cognitivi, l'IA sta attualmente superando i suoi colli di bottiglia energetici attraverso una scalabilità massiccia e un'efficienza architettonica. Il processo decisionale manageriale è costantemente vulnerabile al bias cognitivo, eppure molti interventi di *debiasing* rimangono modesti nell'impatto perché mirano a errori di giudizio superficiali piuttosto che ai vincoli strutturali che rendono le euristiche metabolicamente e organizzativamente attraenti. Questo articolo concettuale riformula il bias cognitivo come un predefinito indotto da vincoli che emerge da (i) la canalizzazione evolutiva delle risposte orientate alla sopravvivenza e (ii) la topologia dell'efficienza delle reti neurali che favorisce un'inferenza a basso costo, veloce e "abbastanza buona" rispetto al calcolo globale ottimale. Parallelamente, i sistemi di intelligenza artificiale contemporanei affrontano i propri vincoli energetici e di scala, ma offrono anche una nuova opportunità progettuale: l'IA può funzionare come un'impalcatura metacognitiva che esternalizza e mette sotto stress le assunzioni manageriali attraverso un dialogo strutturato. Basandoci sulla ricerca strategica orientata al discorso, sosteniamo che l'IA sia più preziosa non come oracolo che sostituisce il giudizio, ma come partner dialogico che genera sistematicamente controargomentazioni, prospettive alternative degli stakeholder e distribuzioni di scenari probabilistici per indurre pause riflessive "System 2" in flussi di lavoro ad alto rischio. Poiché l'IA può anche introdurre bias di automazione, allucinazioni e bias di addestramento-dati, il debiasing efficace deve essere trattato come un problema socio-tecnico di progettazione che coinvolge trasparenza, contestabilità e governance. Concludiamo delineando i principi di progettazione manageriale e un'agenda di ricerca per testare quando e come il dialogo supportato dall'IA migliora la qualità delle decisioni in condizioni di incertezza.

Keywords: Artificial intelligence, Cognitive bias, Debiasing, Decision making, Strategy, System thinking

1 – Introduction. The energy constraint of intelligence

It looks paradoxical that artificial intelligence and humans evolved having both solved the same energy constraint.

The energy constraint for artificial intelligence (AI) was initially unclear. From the 1960s until 2012, the computing power used to train leading machine learning systems grew in line with Moore’s law: performance doubled about every 24 months. Around 2005, however, power dissipation constraints pushed chip manufacturers to increase the number of CPU cores rather than further raising clock speeds (Russell & Norvig, 2003/2021). But after 2012, the trend accelerated dramatically. That year, a team from the University of Toronto, led by Geoffrey Hinton and his students Alex Krizhevsky and Ilya Sutskever, presented a model called AlexNet (Krizhevsky et al., 2012) to the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), the world's largest computer vision competition. It was a deep convolutional neural network (CNN), not the first, but the first time the model was trained using two gaming GPUs, NVIDIA GeForce GTX 580s (about $500 each). These cards, designed for gaming, offered parallel processing power at a fraction of the cost of supercomputers at the time, reducing the training time to just five days and thus making the experiment realistically possible.

Between 2012 and 2018, computing power for state-of-the-art AI models increased by roughly a factor of 300,000, when a 2-year doubling period would yield only a 7x increase, which is equivalent to a doubling every 100 days (Amodei & Hernandez, 2018). As a result, a model that required a full day to train in 2014 could be trained in just two minutes by 2018 (Ying et al., 2018).

That did not happen without consequences. Its counterbalance was a dramatic shift in power density, followed by a shift in power demand. A single server rack, a metal frame designed to hold servers in a standardized vertical stack, that required 5-10 kilowatts (kW), today running the latest AI chips (like Nvidia’s GPUs), needs 50-100 kW, 10 times more power than a traditional server rack. Twenty years ago, a large data centre might have consumed 20 megawatts (MW) of power. Today, the industry is planning gigawatt-scale campuses that consume as much power as a nuclear power plant or a mid-sized city. But the heat generated by these AI chips is so intense that traditional air conditioning is no longer sufficient. Facilities are redesigning their entire infrastructure to accommodate liquid cooling systems, which are more efficient but require new plumbing and designs. Nevertheless, the appetite for new data centres is so huge that, as shown by a McKinsey research, up to $5.2 trillion in investment could be required by 2030 to keep pace with demand (Stylianou et al., 2025).

However, the energy problem for data centres is not only the cooling but also receiving enough energy from the electric network. This insatiable demand is colliding with the limits of power grids, and the available grid power has become the single most significant constraint, forcing big tech companies to hunt for stranded power or build their own generation capabilities. For this reason, tech giants are restarting dormant nuclear plants or funding new ones to guarantee 24/7 carbon-free power, building data centres in the neighbourhoods to ensure enough power with short electrical connections, and avoiding the grid. Microsoft signed a deal to restart the Three Mile Island nuclear plant, and Meta bought a nuclear plant in Illinois, Amazon purchased a data centre campus connected to the Susquehanna nuclear plant, and Google invested in a small modular reactor (SMR) with Kairos Power.

The demand for energy is so crucial that it could be the deciding factor in the global competition between the USA and China for dominance in the AI landscape. A model like ChatGPT-4 can consume over 463,000 MWh per year (equivalent to 44,000 US homes). With global data centre electricity consumption projected to double by 2030, the availability and cost of power will command the pace of AI progress. Because of the fast-rising demand, wholesale electricity prices near US data centres have surged up to 267% in the last five years, while investment in large-scale US wind and solar projects fell in the first half of the year, driven by shifting regulations and the removal of certain subsidies by the Trump administration. On the other side, China is aggressively expanding its power grid to support industrial and AI growth. Last year, it added 356 GW of renewable capacity (277 GW solar, 80GW wind), far exceeding the total capacity of the US. Besides, it is building high-voltage lines to transport cheap renewable energy from inland regions (Inner Mongolia solar, Sichuan hydro) to coastal tech hubs. Finally, local Chinese governments offer preferential electricity rates to tech giants like Alibaba and Tencent. The consequence is that cheap, abundant energy allows Chinese companies to run less-efficient domestic chips (like Huawei's) for more extended periods at a lower cost, effectively neutralizing the hardware performance gap with Nvidia's processors, which are the performance leaders. It looks like we are witnessing another iteration of historical cycles in which technological superpowers rise on the back of cheap energy (Britain with coal; the US with oil and hydro), while energy constraints scale faster than transistor efficiency. Therefore, the victor in the AI race could be the nation that can keep AI models running most affordably, rather than the one with marginally better hardware (Yoon, 2025).

Similarly, in natural history, homo sapiens emerged among other species as dominant from an evolutionary standpoint, being able to solve the energy bottleneck more effectively. According to Herculano-Houzel (2009, 2016), a neuroscientist, the human brain has about 86 billion neurons, with about 16 billion in the cerebral cortex and 69 billion in the cerebellum. Each billion neurons costs roughly 6 kcal a day to run. While biological scaling laws suggest that larger animals should possess larger brains, humans present a striking anomaly. Despite gorillas being up to three times our size, the human brain is roughly three times larger than theirs. This distinctiveness extends to metabolic cost. The human brain consumes approximately 500 kilocalories daily, claiming 25% of the body's total energy budget while comprising just 2% of its mass. By contrast, a mouse brain consumes only 8% of its bodily energy while accounting for about 1% of its body mass. Thus, the human brain is biologically unique not just in size, but in how energetically costly it is to maintain. Gorillas and orangutans, for example, with bodies in the 50–100 kg range and about 33 billion brain neurons, are right on the edge of what an 8-hour-per-day average feeding schedule can sustain. During periods of fruit scarcity, this foraging time is insufficient to meet metabolic demands, resulting in weight loss. Doubling their number of neurons would require additional feeding time, which they simply do not have. Thus, as stated by Herculano-Houzel, ape brains do not stop growing because they were on the unlucky side of evolution, but because the energy requirements made more neurons unaffordable in such large, muscular bodies living on raw food. Herculano-Houzel (2016, p. 184) summarizes this as a trade-off:

> *A primate can't have both a very large body and a very large number of neurons: it is either brains or brawn — and great apes seem to have 'chosen' brawn.*

There are viability zones for all the living species, affordable combinations of body mass and brain neurons that can be supported on a diet. Beyond those zones, animals would starve. So the difference is not that humans or apes have a special brain design. The difference is how many neurons each lineage can afford to maintain, especially in the cortex. Upright walking made our australopithecine ancestors (4 million years ago) more efficient long-distance foragers: walking on two legs costs about four times fewer kilocalories than knuckle-walking. They became food gatherers rather than food pickers, able to roam long distances and bring back food, unlike modern great apes, which remain limited to picking food. Homo erectus (by 2 to 1.9 million years ago) evolved longer legs, elastic tendons, a large gluteus maximus, a nuchal ligament, and other traits that made endurance running energetically cheap and efficient. Archaeological evidence suggests early Homo erectus hunted large animals (e.g., wildebeest, kudu), requiring coordinated group hunting and thus favouring more cortical neurons for planning, cooperation, and social cognition. But the decisive step was cooking with fire (1.5 to 1.0 million years ago). The controlled use of fire allowed cooking, massively increasing the number of kilocalories extracted per hour of feeding. Herculano-Houzel's calculations show that a 70 kg primate with 86 billion neurons, like a contemporary human, would need 9-10 hours a day of foraging on a raw primate diet, not viable in the wild, and well above the 8-hour practical limit observed in great apes. But if food could be predigested outside the body (cutting, pounding, then heating), our ancestors could obtain more calories in less time, freeing them from the energetic ceiling that keeps apes small-brained. The result was rapid brain growth in our ancestors, while apes stagnated. Fossil-based estimates, combined with primate scaling rules, suggest that in the last 1.5 million years, our brain nearly tripled in mass and added about 57 billion neurons (Figure 1), while great apes remained around 30–33 billion neurons.

So, Herculano-Houzel concludes that the human brain differs from ape brains because we are primates that escaped the raw-food energy constraint via cooking (supported by earlier changes such as bipedalism and hunting), and thus could continue adding cortical neurons, whereas apes could not.

This biological difference mirrors the current trajectory of artificial intelligence. Just as the human brain evolved into an energetic luxury, modern AI has become a similarly disproportionate consumer of the electrical grid. The shift from standard CPUs to power-hungry GPUs parallels the evolutionary leap from the mouse's metabolically cheap brain to the human's expensive one. In both domains, the emergence of higher intelligence comes at a cost. It imposes an exponential energy tax.

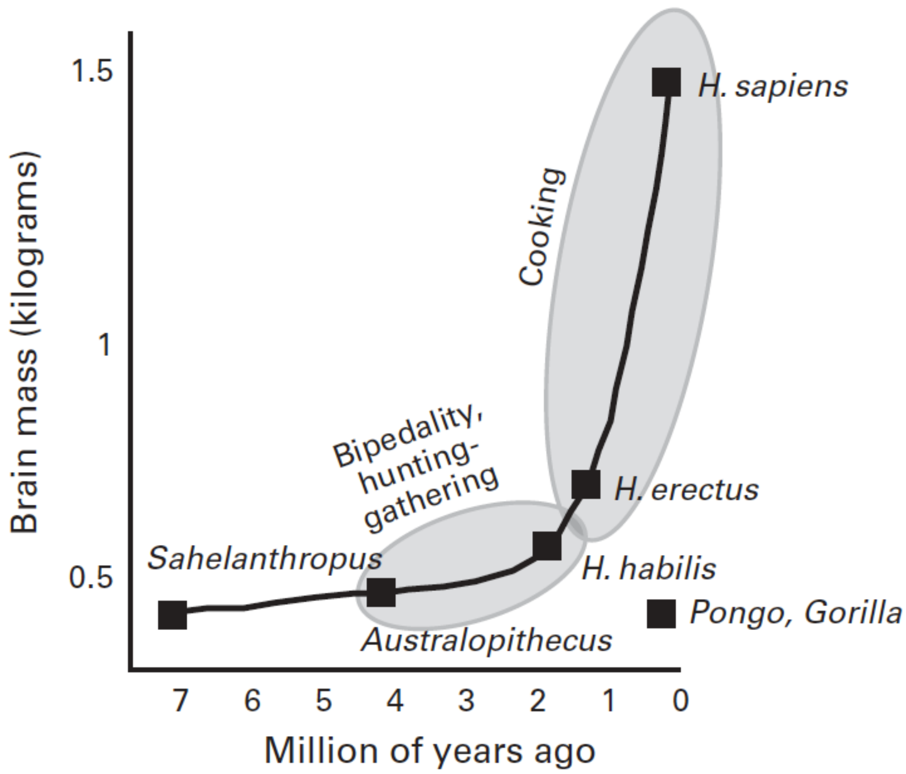

Figure 1 – Rapid increase of brain mass in the homo lineage in the last 1.5 million years coincides with the invention of cooking, probably by homo erectus (Herculano-Houzel, 2016, p. 192)

Whether it is a primate requiring cooked food to fuel its 86 billion neurons, or a tech giant requiring nuclear reactors to fuel a gigawatt-scale data centre, the fundamental constraint remains identical: intelligence is, first and foremost, a thermodynamic challenge. Intelligence, whether biological or artificial, is fundamentally constrained by the ability to acquire and manage enough energy. Evolution, whether natural or technological, is not just about architectural improvements, developing better brains or better chips, but about overcoming an energy limit.

Both the human brain and AI server clusters are large-scale, distributed information-processing infrastructures whose capabilities and limits are determined less by raw compute elements and more by architecture: how the system allocates a constrained energy budget to computation, communication, memory/state storage, and heat removal.

A state-of-the-art AI cluster used to train models like ChatGPT-4 has 100,000 Nvidia H100 GPUs, each with 80 billion transistors, for a total of 8 quadrillion transistors, communicating over copper at a significant fraction of the speed of light (300 million m/s). It needs to work an average of 150 MW of power, the equivalent of 125,000 homes, but concentrated on a much smaller surface, creating enormous problems of heat management.

On the other hand, the brain is made up of about 86 billion neurons that do not communicate through a passive conductor like a copper wire. In the cerebral cortex, there are about 0.15 quadrillion synapses, and a typical cortical neuron participates in about 7,000 synaptic connections. But a neuron is not a processor, it is an excitable electrochemical system that maintains transmembrane ionic gradients via ion pumps and generates electrical signals primarily through voltage-gated ion channels that produce and propagate action potentials. Its maximum signal speed is just 120 m/s (Tsubo & Kurokawa, 2018). The brain as a whole needs the equivalent of a 20W light bulb, even when we lie still in complete darkness, doing nothing physically. Contrary to the assumption that high-load cognitive tasks spike energy consumption, the brain operates on a constant metabolic budget with only tiny fluctuations (<5%) for conscious tasks. Rather than increasing total power draw during complex problem-solving, energy resources are diverted to active processing centres while simultaneously being downregulated in other areas. This suggests an architecture designed for constant readiness, akin to a car with the engine always running at full speed, always ready to action, where the vascular system serves the dual critical role of logistical supply chain and thermodynamic cooling mechanism for a system permanently operating at maximum capacity (Dosenbach et al., 2025; Raichle & Mintun, 2006).

Hardware in a computer is immutable after fabrication. It has fixed wiring and programmable interconnects. Learning in AI is a software abstraction where virtual weight matrices are updated mathematically. In the brain, hardware is *fluid*. Synaptic plasticity physically alters the structure of the connections. The brain can physically rewire itself in real time in response to stimuli but it can also degrade gracefully: substantial neuron loss can sometimes be compensated by plasticity and redundancy, within limits. On the contrary, computers are extremely reliable per operation but can fail catastrophically in case of hardware failure.

In a computer, there is a distinct separation between processing units (CPU/GPU) and memory (DRAM/SRAM). This creates the Von Neumann Bottleneck, in which energy and time are primarily consumed by moving data back and forth rather than by computation. In a biological brain, memory and processing are colocated. The synapse functions as both the storage unit and the processing gate. Memory is not retrieved, it is the physical pathway through which information flows.

Computers rely on binary logic (0 or 1) driven by a global clock (synchronous) and are deterministic; the same input always guarantees the same output. Signals travel near the speed of light but are serialized. Brains operate on mixed-signal processing (analog/spiking). Analog because dendritic integration sums voltage potentials continuously, but also digital-like, because the axon hillock fires an all-or-nothing action potential (spike) only when a threshold is breached, but neurotransmitter release is probabilistic. Release probability varies widely by synapse type, history, and conditions. Many central synapses have relatively low release probabilities on the order of ~20%-30% in some contexts, but reported ranges span from very low (<5%) to high (>90%).

In a processor, any single logic gate output can control between 2 and 10 logic gates, while each neuron can activate between 1,000 and 10,000 synapses. This enormous difference needs special computational processes. In a processor, to process complex associations as in AI, information must pass through many sequential layers of logic gates. This approach excels at algorithmic precision but struggles with pattern matching, where context is distributed across millions of data points. In the brain, a single neuron can broadcast a feature detection (e.g., a rotating ball playing tennis) to thousands of downstream processing centers simultaneously (motion detection, object recognition, memory recall). This massive interconnectivity allows the brain to solve ill-posed problems, such as recognizing a face in a shadow, via holistic parallel voting rather than sequential logic.

In conclusion, brain and AI systems differ primarily in *where* bottlenecks lie and *how* they manage them. Both are "always-on" conversion machines: they continuously convert energy into structured signal processing, and the dominant constraint is how efficiently they can move energy in and out as heat. Both systems can be modeled as weighted directed graphs that transform inputs into outputs by propagating signals along links. Intelligence emerges from the dynamics of that propagation and from adaptation of link strengths (learning). In both brains and machines, performance is constrained by the economics of communication: latency, bandwidth, and energy per bit/event moved. But digital hardware aims for deterministic correctness per operation, with reliability engineered via design margins or error detection and correction. Brains accept noisy components and achieve reliability through population coding, redundancy, and adaptation. Because Individual neurons are variable and sometimes unreliable, so the brain gets stable behavior by using many neurons in parallel (population coding), building in overlap and backups (redundancy), and continuously recalibrating connections and activity levels (adaptation). However, this evolutionarily selected architecture is not a perfect match with our complex, transforming, and symbolic society.

2 – How AI works

Given these differences between AI and our biological brain, the development of the AI industry has a long way to go before reaching the same level of efficiency and effectiveness that nature and evolution have developed for our central nervous system. This effort is sustained by investments of a magnitude never seen in other fields, with commitments of about $1.4 trillion over the next 8 years for OpenAI alone (Bort, 2025). This unprecedented financial backing is largely being deployed to win the challenge of reaching Artificial General Intelligence (AGI), a level of capability at which AI could understand, learn, and apply intelligence to solve any task a human can perform. According to some leading Big Tech experts, between 2026 and 2035, AI would reach AGI.

In the meantime, the recent surge in the capabilities of large language models (LLMs) is due to the software breakthrough represented by the *transformer architecture*, which replaced recurrence with attention-based sequence modelling and enabled efficient large-scale training (Vaswani et al., 2017). Yet, beyond the architectural shift that made modern LLMs feasible, it is helpful to distinguish three complementary "scaling" mechanisms that continue to increase model performance and shape the future trajectory of AI: pre-training scaling, post-training scaling, and test-time scaling.

2.1 – *Pre-training scaling and empirical scaling laws*

Pre-training scaling, often characterized as a *brute-force approach*, involves the massive expansion of a model's foundational resources: data, computation, and parameter count. Analytically, one might liken this process to constructing a university library: increasing parameters expands the library's capacity (adding shelves and rooms); increasing data fills those shelves with knowledge; and increasing computation represents the labor and time required to organize these vast resources. Once this training phase concludes, the model's core structure is fixed.

The relationship between these inputs is governed by empirical scaling laws (Kaplan et al., 2020), which state that model performance improves predictably as resources grow. This predictability has transformed large-scale model development into an engineering science rather than a speculative art (Hoffmann et al., 2022). Yet, this scaling imperative erects high barriers to entry. The need for tens of thousands of GPUs and proprietary data restricts competition to major technology firms and heavily funded entities such as OpenAI and Anthropic.

While scaling laws helped transform frontier-model development into a more engineering-driven discipline (because they allow performance forecasting under alternative resource allocations), pre-training scaling also produces substantial barriers to entry the LLMs arena. The requirement to co-scale compute, data, and model size tends to privilege organizations with access to large capital expenditures, sustained compute supply chains, and large proprietary or well-curated datasets, conditions that are difficult to replicate for smaller actors. Moreover, concerns have intensified regarding the long-run sustainability of this regime:

1) High-quality human-produced data is increasingly scarce relative to demand (the so-called "data wall")
2) Training costs for frontier models exhibit steep growth patterns, raising both economic and environmental questions about the continued brute-force scaling (Epoch AI, 2024; Stanford HAI, 2025; Villalobos et al., 2022).

These constraints motivate the view that future competition may depend less on raw scale alone and more on methods that extract additional capability from already-trained models.

2.2 – *Post-training scaling: adaptation, alignment, and instruction-following behavior*

Post-training scaling refers to capability gains obtained after base pre-training, by specializing or aligning the model through additional optimization stages, often under the umbrella of instruction tuning and alignment. Rather than expanding the model's general knowledge via another large pre-training run, post-training aims to shape the model's behaviour (for example, improving instruction-following, helpfulness, harmlessness, or domain specificity) using curated datasets and human preference signals.

Canonical post-training methods include supervised fine-tuning (SFT) on instruction–response pairs (Chung et al., 2024; Wei et al., 2022) and reinforcement learning from human feedback (RLHF), where a reward model is trained on preference comparisons and then used to optimize the policy model (Christiano et al., 2017; Ouyang et al., 2022; Stiennon et al., 2020). In practice, these approaches can be substantially more compute-efficient than full pre-training. Still, they remain resource-intensive in different ways: they require high-quality annotation pipelines, robust evaluation and safety processes, and access to users or expert labellers to generate preference data at scale (Ouyang et al., 2022). Consequently, the comparative advantage may shift from economies of scale in computing alone toward economies of experience, where organizations with large user bases and mature data-production and evaluation infrastructures can iterate faster and align models more effectively.

2.3 – *Test-time scaling: allocating compute during inference to improve reasoning quality*

A third driver of capability growth is test-time scaling, i.e., techniques that allocate additional inference-time computation to improve output quality without changing the model weights. Conceptually, this strategy treats inference as an optimization problem under a latency–accuracy trade-off: the system spends more *thinking time* (and therefore more compute) to generate better answers. Mechanisms include prompting or decoding strategies that elicit intermediate reasoning steps, such as chain-of-thought prompting (Wei et al., 2022), ensemble-like sampling and selection procedures such as self-consistency (Wang et al., 2022), and structured search over multiple reasoning trajectories such as Tree of Thoughts, which generalizes linear reasoning into a branching and backtracking process (Yao et al., 2023). Recent surveys propose frameworks for organizing this design space and clarifying what exactly is being scaled at test time, like samples, search depth, tool calls, or deliberation steps (Zhang et al., 2025).

Because test-time scaling increases per-query costs, it can reinforce competitive asymmetries: actors with more substantial margins, better infrastructure, and more stable user relationships can more readily absorb higher inference costs and present *tiers* of response depth (e.g., fast vs. deliberative modes), whereas latency budgets and marginal serving costs may constrain smaller providers.

2.4 – *Beyond brute-force scaling*

Importantly, the three scaling mechanisms are not substitutes; they are largely complementary. Improvements in base models (pre-training) raise the ceiling for what can be achieved via alignment and specialization (post-training) and can also make inference-time deliberation more effective (test-time scaling). Conversely, bottlenecks in any one component (e.g., data scarcity, compute limits, or weak post-training pipelines) can cap the returns available from the others. Taken together, the combined scaling regime can plausibly generate compounding progress in frontier capability. However, it also tends to concentrate advantage among actors that can simultaneously mobilize compute, data, organizational expertise, and distribution.

In response to the practical limits of continued pre-training scaling, research increasingly emphasizes approaches that improve architectural efficiency, enhance algorithmic efficiency, and enable more autonomous data generation and broaden model capabilities via multimodality, retrieval, and agentic interaction.

A central line of work modifies the Transformer-based backbone to reduce marginal compute costs while maintaining or improving quality. Sparse Mixture-of-Experts (MoE) architectures implement conditional computation by routing tokens to a small subset of specialized *expert* sub-networks, dramatically increasing total parameter capacity while keeping active computation relatively bounded (Fedus et al., 2022; Shazeer et al., 2017). Contemporary implementations demonstrate that MoE can provide strong performance–efficiency trade-offs; for example, Mixtral 8×7B, the Frech LLM, uses a router that activates only a subset of experts per token (Jiang et al., 2024). Major industrial systems have also publicly emphasized MoE for efficiency and long-context performance. Google explicitly describes Gemini 1.5 Pro as adopting a MoE architecture and reports contexts up to one million tokens (Google, 2024). It is interesting to note that this line of research ultimately replicates the human brain's functional organization, which has specialized areas for certain functions that are activated only for specific tasks.

A second frontier addresses the dual constraints of rising data costs and the scarcity of high-quality data by leveraging synthetic data generation. In the Self-Instruct framework, a model generates instruction–response pairs that are subsequently filtered and used to train improved instruction-following behaviour, reducing reliance on expensive human-authored datasets (Wang et al., 2022). More broadly, iterative refinement strategies, in which a model produces an output, critiques it, and then revises it, aim to approximate some of the benefits of continual human feedback at lower cost (Madaan et al., 2023; Shinn et al., 2023). At the same time, the literature cautions that unconstrained reliance on synthetic data can induce failure modes (e.g., degradation under self-generated training distributions), motivating careful data governance, filtering, and evaluation (Shumailov et al., 2024). Here too, the parallel between synthetic data generation and the role that dream activity has for the human mind is striking (Walker, 2017).

A third research direction broadens what foundation models can perceive and do, rather than merely increasing the amount of parameterized knowledge. Frontier systems increasingly aim for native multimodality, training a single model end-to-end across text, vision, and audio, rather than attaching separate modality modules post hoc. OpenAI's GPT-4o is described as a single model trained across text, vision, and audio modalities, while OpenAI also presents GPT-5 as a unified system with multimodal capabilities among its core features (OpenAI, 2025). Google similarly positions Gemini 1.5 as a multimodal model and emphasizes long-context multimodal understanding (Google, 2024). It should be emphasized that multimodality, which was a great achievement for LLMs, is innate in the human mind and offers a significant advantage over AI in complex reasoning, especially when it requires the abstract contemporary representation of objects, concepts, sounds, music, contexts, and situations.

Finally, a major frontier concerns agency and grounding. Agentic architectures treat the LLM not only as a text generator but as a planner that can select actions, call tools, and interact with environments, digital or physical, toward goals. Some works developed the coupling between reasoning traces and action selection (Yao et al., 2022). In robotics and embodied settings, research highlights the importance of grounding language in sensorimotor experience and feasible actions; examples include SayCan, which connects language to robotic affordances (Ahn et al., 2022), and RT-2, which investigates transferring web-scale vision–language knowledge into robotic control policies (Zitkovich et al., 2023). A parallel line of work develops embodied multimodal language models that integrate perceptual and state inputs into a language-model backbone to support grounded decision-making (Driess et al., 2023). In this way, AI can have access to social learning, as babies in a group, exploring and organizing social actions, or to perceptive learning, creating links between words and percepts, as it is natural for a newborn.

Taken together, contemporary AI progress is driven by both continued scaling of foundational resources (compute, data, parameters) under empirically regular scaling laws and a shift toward efficiency-enhancing and capability-expanding innovations that can yield substantial gains without proportionate increases in training cost. These forces can be mutually reinforcing: better base models facilitate better alignment and more effective inference-time deliberation, while architectural and algorithmic efficiencies make additional scaling more economically feasible.

However, these dynamics can create a cumulative advantage. Merton's "Matthew effect" characterizes how early advantage in scientific recognition and resources can compound over time, increasing inequality in opportunities and capabilities (Merton, 1968). Analogously, in AI development, initial advantages in compute access, data pipelines, user distribution, and organizational expertise can translate into faster iteration cycles and superior models, which then attract additional capital and talent. This feedback structure closely resembles the systems archetype often described as "success to the successful," associated with Peter Senge's systems thinking tradition (Senge, 1990). The underlying logic echoes the biblical formulation: «For to everyone who has, more will be given, and he will have abundance, but from the one who has not, even what he has will be taken away» (Matthew 25:29).

From this point of view, the competition between LLMs seems like an accelerated evolutionary competition in which the solutions that are relatively more suitable must be selected, taking into account the energy constraints and the system's own resources.

3 – How the brain works

As AI witnesses it, we try to develop computers often using a metaphor of the brain that resembles more a computer, designed from scratch with a master plan, than what the brain really is, the evolutionary result never reengineered of a natural selection. From this perspective, the brain is more like an old house that has been renovated three times over millions of years. Evolution never tore down the old structure. It just built new, fancy additions on top of the old foundations. This means humans have deep, ancient instincts running alongside modern logic, and they do not always agree. As explained by the Nobel laureate François Jacob "Evolution does not produce novelties from scratch. It works on what already exists […] It is a process of tinkering" (Jacob, 1977, p. 1164). Evolution is not an engineer, who plans for the future, it is a bricoleur that manages with what is at hand.

The human brain is the result of millions of years of bricolage. It is a *kludge*, a messy accumulation of adaptations built on top of one another, where new structures are forced to work through old ones. But it is no longer the idea of the *Triune Brain* model (MacLean, 1990), which divides the brain into three evolutionary layers (the basal reptile substrate, the limbic layer of early mammals, and the neocortex of late mammals) like an onion. Mainly, modern evolutionary developmental biology and connectomics have falsified it. The brain is not stratified like geological layers, it is a highly integrated, co-evolved network. Evolution modified the entire organism. When the cortex expanded in mammals, the older structures (such as the brainstem and basal ganglia) did not remain frozen in time. They co-evolved and re-wired to support the new complexity. While the genes building these structures are highly conserved and similar to ancestors, the structures themselves have been heavily modified. The human amygdala and basal ganglia are intricately wired into the prefrontal cortex. You cannot isolate *emotional circuits* from *rational* ones, they form a single, continuous loop of information processing (Striedter, 2005).

Emotions do not live in the paleomammalian layer (the limbic system), separate from the rational neocortex. The *lizard brain* (basal ganglia) is heavily involved in learning and habit formation, not just instinct. Conversely, the *rational cortex* is essential for constructing emotions. Neuroscientist Lisa Feldman Barrett has demonstrated that emotion is a whole-brain construction, not a reflex from a specific ancient region (Barrett, 2016). We cannot make cold, rational decisions just by shutting off the old brain. Patients with damage to emotional centers often cannot make decisions at all because they lack the biological value signals (valence) to weigh options (Damasio, 1994). For Barrett (2016), emotions are not *noise* or primitive impulses that the rational mind must suppress. Instead, it frames emotion as functional and informational, they are a signaling mechanism. Emotions are heuristic shortcuts that prioritize attention and resources (e.g., fear facilitates risk avoidance; empathy facilitates social cohesion). Rather than competing with logic, emotions provide the value system (valence) necessary to weigh options. Without emotional input, decision-making creates paralysis, not rationality. Reductionist models, like the triune brain, are easy to teach, but they lead to flawed strategies in education and management because they ignore the brain's massive interconnectivity. A sophisticated understanding of human behaviour requires acknowledging the brain as a complex system where higher and lower functions are inextricably linked.

From this perspective, intuition is a high-speed processing and *gut feelings* are not primitive reptile signals. They are rapid, cortex-integrated statistical predictions based on prior learning. The basal ganglia (the so-called reptile) are actually the brain's habit engine, executing complex patterns efficiently to save energy. Besides, stress breaks the brain's functional network, not the layer of different evolutionary parts. High stress doesn't just uncover the lizard, it disrupts the connectivity between the prefrontal cortex and the rest of the network. The goal of a leader is not to suppress the old brain, but to maintain the connectivity that allows the whole system to function.

The brain operates through two mutually exclusive functional architectures, creating a fundamental trade-off between execution and strategy. When the brain engages in goal-directed behavior, it activates the Action-Mode Network. This state is the neurological equivalent of operational excellence: it drives task initiation, sustains focus on external goals, and processes real-time feedback, such as errors or pain. Crucially, to maximize performance, this mode actively suppresses internal distractions and self-referential processing. In the absence of external demands, the brain reverts to the Default-Mode Network (DMN). This is the center for strategic processing, including handling memory consolidation, emotional regulation, and future-oriented simulation. These two networks function in a *yin-yang* relationship of anti-correlation. Structurally, the brain cannot fully engage in deep reflection and high-intensity execution simultaneously: when one network activates, the other is inhibited (Dosenbach et al., 2025).

Looking from this perspective, toward the future of management, neuroplasticity, the brain's ability to reorganize itself, is a crucial resource to draw from. The architecture of the brain is not fixed. Understanding this allows for new approaches in mental health and cognitive enhancement. We are not prisoners of our evolutionary past, through training and environmental design, we can harness the brain's capacity for adaptation.

Finally, the idea that the brain reacts to the world in a stimulus-response relationship is no longer the prevailing view. The brain is a prediction engine. Its primary evolutionary function is allostasis, regulating the body's energy budget. It predicts metabolic needs before they arise. What we call *rational decisions* are often just high-level justifications for these metabolic predictions. We cherish the belief that our decisions flow top-down: we analyse data, form a conclusion, and then act. Neurobiology suggests the inverse is often true. If the brain primarily regulates allostasis, when the body enters a deficit or surplus state, the cortex, our rational center, is tasked not with making the decision but with creating a plausible narrative to justify the biological impulse. This phenomenon is known as *confabulation* or *post-hoc rationalization* (Danziger et al., 2011; Steffen et al., 2022).

A proof of metabolic interference in rational decision-making is the famous study on parole judges (Danziger et al., 2011). Researchers analyzed over 1,000 parole rulings made by experienced judges. The findings were astonishing. In the morning, with judges rested and energized, prisoners had a 65% chance of being granted parole. Before lunch, with judges naturally low of glucose and with a high allostatic load, the possibility of parole dropped to nearly 0%. After lunch, with the allostatic budget restored, the chance jumped back up to 65%. The judges, when asked why they denied parole at 11:45 AM, would point to the law, the severity of the crime, or the prisoner's behaviour. They would give a rational justification. But there is an alternative explanation. They denied parole because "No" is the energy-saving default option for a metabolically depleted brain, "No" is safe while "Yes" requires a complex risk assessment.

To understand why reason is often a slave to *metabolism*, it must accept the hierarchy of neural evolution: the brain's only non-negotiable goal is keeping the organism alive. This requires energy efficiency. The brain constantly monitors the body (heart rate, glucose, cortisol). This monitoring is called *interoception*. These internal measurements are summarized into a general feeling called affect (feeling pleasant or unpleasant, agitated or calm). The conscious mind (the neocortex) does not have direct access to your blood sugar levels or your adrenal glands. It only feels the affect, it feels *bad* or *agitated*. When a manager, for example, is sleep-deprived and facing a high-risk merger his/her brain predicts a massive energy expenditure it cannot afford. The allostatic response is that brain triggers a withdrawal response (avoidance) and creates an unpleasant, anxious affect. Your cortex receives this *unpleasant signal*. It searches the environment for a cause. It looks at the merger data. The cortex concludes: "I have a bad feeling about this deal. The results projections look too volatile." To correct this error, the brain must deploy emergency funds (cortisol/adrenaline). Doing this occasionally is fine. Doing it chronically, typical for a manager navigating a volatile, uncertain, complex, and ambiguous environment, creates a structural deficit. Understanding and avoiding such contingencies is not a medical issue, it is a managerial problem.

4 – Cognitive bias and advances in neurology

The modern empirical tradition defines biases as systematic deviations from classical rationality and statistical norms in judgment under uncertainty (e.g., availability, representativeness, anchoring). It documents their robustness across tasks and contexts. In organizational behaviour and strategic management, the dominant paradigm follows the Kahneman and Tversky tradition (Kahneman, 2011; Kahneman et al., 2021; Tversky & Kahneman, 1974): System 1 (fast, intuitive) and System 2 (slow, analytical) interact in normal decision making and in daily life, but biases emerge when System 1 misfires in complex environments or situations. However, to truly understand the persistence of these biases in high-stakes management decision-making, it is necessary to look below the psychological layer to the biological substrate (Haselton & Buss, 2000). From that perspective biases could be interpreted as the phenotypic result of a brain architecture shaped by two forces:

1) The conservation of developmental programs that ensure survival in ancestral environments.

2) The topology of neural networks, which favours *small-world efficiency* over exhaustive computation.

The first force, identified by the *evolutionary developmental biology*, posits that evolution does not just select for adult traits, but shapes the developmental trajectories (ontogeny) that build those traits. One example is the concept of canalization, introduced by C.H. Waddington (Waddington, 2014). It suggests that developmental pathways are robust and resistant to perturbation. In the context of cognition, evolution has canalized specific neural responses to environmental stimuli like a ball rolling down a valley (the epigenetic landscape). The depth of the valley represents the strength of the genetic constraint. Cognitive biases, such as *loss aversion*, are deep valleys. They are not random errors, they are developmentally entrenched responses to scarcity. In an ancestral environment, the cost of missing a meal was manageable, but the cost of becoming a meal was terminal. Thus, the developmental program constructs a limbic system that weighs negative stimuli significantly heavier than positive ones (Tooby et al., 1992) (Tooby & Cosmides, 1992). Another example is the social proof bias, a cognitive bias where people look to the actions and behaviours of others to determine the correct way to act, especially in uncertain situations, assuming the majority knows best, leading them to conform to group norms or popular choices, like choosing a crowded restaurant over an empty one. The developmental priority is placed on social cohesion circuitry because isolation meant death. But for the modern manager, this biological history creates a mismatch. The developmental program builds a brain optimized for the Pleistocene savanna, not the contemporary boardroom, where the biological imperative to align with the tribe overrides the intellectual imperative to analyse financial risk objectively.

The second force is *connectomics*, the study of the brain's structural and functional connectivity, its wiring diagram. The physical reality of the connectome imposes strict constraints on decision-making.

As we have analysed before, the human brain consumes approximately 20% of the body's metabolic energy despite representing only 2% of its mass (Raichle & Gusnard, 2002). Evolution has exerted immense pressure to minimize wiring cost (the length and number of connections). To save energy, the connectome is organized as a *small-world network*. This topology features high local clustering, with specialized processing, and short path lengths, with quick communication via *hub* nodes.

Cognitive bias can be described as a function of this network topology. Calculating a globally optimal solution (rationality) requires recruiting vast, distributed networks across the cortex, which is metabolically expensive and slow. Conversely, a heuristic (bias) relies on strong, short-range connections between *rich-club hubs* (e.g., the amygdala and the ventromedial prefrontal cortex). For example, confirmation Bias is connectomically efficient. It utilizes existing, well-trodden neural pathways (strong synaptic weights). To process contradictory information requires the formation of new synaptic patterns or the inhibition of established ones, a process that is metabolically costly and cognitively slower (Sporns, 2016) and can be described as:

$$Cost_{rational} \gg Cost_{heuristic}$$

In connectomics, "*neurons that fire together, wire together*", an example of Hebbian learning (Hebb, 1949/2002). Over a career, a manager's connectome physically reinforces the pathways of successful past decisions. When a new, disruptive innovation appears, the manager's brain is physically wired to route data through the old, efficient pathways. We call this *experience*, but biologically, it is a connectomic bias against novelty.

For management scholars, acknowledging the biological basis of biases suggests that *de-biasing training* is often ineffective because it fights against millions of years of canalized development and the physical laws of network efficiency. But the problem is not merely thinking poorly: we are running high-friction software (knowledge) on hardware (brain) designed for low-friction survival. Understanding the connectome and its evolutionary history moves the conversation from blaming the decision-maker to designing organizational solutions that mitigate these biological constraints. Using AI to assist decision makers could be a viable solution.

5 – Dialogue, AI and debiasing, a research perspective

In his research, Gianluca Colombo proposed in 2004 a research program to improve strategic management by analysing the discourses that feed strategic thinking and action, focusing on how organisational change happens through conversation and dialogue (Colombo, 2004).

In his view, firms change through sequences of thoughts and actions whose medium is conversational, so dialogue becomes a primary lever for strategic change. He then extends "dialogue" to two arenas:

– inside organizations, among managers and organizational actors,

– inside the scientific community, among strategy scholars.

In both arenas, dialogue has two inseparable dimensions:

– an innovative dimension (creating and diffusing actionable knowledge),

– a political dimension (selecting actions in firms, and selecting theories in academia).

Colombo is not treating discourse as "just communication," but as the tool that produces knowledge and power or selection operates. Colombo reframes "grammar" as a gateway to rhetoric and logic (the classical trivium), then expands it into *cognitive grammar*, drawing on cognitive linguistics and Chomskyan ideas about grammar. Musso et al. (2003) provided physical evidence that the human brain is indeed biologically *hardwired* to process specific linguistic structures (hierarchies) and ignore others, validating Chomsky's core biological claim.

In the same way, neurology explains that rhetoric is not just ornamentation but a mechanism that hacks the brain's decision-making centres (Jack & Appelbaum, 2018). It works because it triggers somatic states (bodily feelings), and neuroscientific studies show that *pathos* (emotional appeal) activates the amygdala and insula, regions critical for emotional processing and salience. Without this emotional *tag*, data remains abstract and does not drive action. As noted in persuasion research, "*emotions create movement and action*," whereas logic only provides a foundation. Besides, the rhetorical concept of *ethos* (credibility) has a specific neural correlate in the *ventromedial prefrontal cortex*. This area is responsible for judging trustworthiness and social value. When a manager uses rhetoric to build a "we" identity (as Colombo suggests with the "inclusive we" vs. "them"), they are engaging the listener's social brain, lowering defence mechanisms. Finally, as Colombo emphasises the importance of metaphors in strategic conversation, neuroscience confirms that metaphors are not processed as abstract symbols but as physical simulations. When we hear a metaphor like "grasping a concept," the brain activates the premotor cortex responsible for hand movements, just as if we were physically grasping an object. This "embodied cognition" makes rhetorical language felt physically, creating a more profound cognitive imprint than abstract logical propositions.

Neurology can also explain why logic (logos) is often insufficient for persuasion, despite being essential for verification (Falk et al., 2010). Logical reasoning relies heavily on the dorsolateral prefrontal cortex, which handles working memory and executive function. This is a high-energy, cold-processing system that fatigues easily. Biological studies suggest that humans primarily make decisions based on emotional or rhetorical inputs (limbic system) and use logic only afterwards to justify those decisions to themselves and others. This aligns perfectly with Colombo's philosophical argument: logic judges based on truth or falsity (non-contradiction), but it is *sterile* because it cannot assess utility or happiness. Neurologically, a brain damaged in the emotional centres can remain perfectly logical but become incapable of making simple decisions because they lack the *rhetorical value signal* of what is good or bad for them.

However, on the opposite side, neuroscience cannot fully explain:

- linguistic and semiotic analysis (how meaning is structured),
- institutional and cultural context (who can credibly speak, to whom, and why),
- power and politics (what counts as "reasonable" in a given organisation).

But this is precisely the point discourse-oriented strategy research makes in Colombo's perspective: strategic talk is evaluated not only by "truth" but by utility, felicity, and actionability in context.

Starting from this, AI could be used not merely as an analytical tool, but as a *neuro-cognitive exoskeleton* that compensates for the biological limitations of the managerial brain. However, this requires treating AI as a dialogical partner that challenges our mental models, rather than an oracle that replaces them.

To understand how to debias, we must first define what we are debiasing. As (Colombo, 1993, 1997, 2004) argues, strategic management is not the objective discovery of a pre-existing reality (positivism), but a *constructed universe* built through language and social interaction.

Managers rely on *cognitive maps* (Colombo, 1998) to navigate complexity that can be improved using concepts and tools of system thinking (Coda & Mollona, 2003; Mella, 2014, 2017/2025; Mollona, 2000). But in the day-by-day activity of companies, these maps are prone to ossification because the human brain prefers *cognitive parsimony* (minimizing effort) over accuracy. So, they need a mechanism to disrupt these static constructions. Colombo suggests *strategic conversation* as a tool for *transdisciplinary coherence*. Today, AI provides the technological infrastructure to scale this conversation and be a metacognitive scaffold. Recent research indicates that AI can act as an external prefrontal cortex, explicitly targeting the neural mechanisms responsible for cognitive bias. Neuroscience shows that the brain creates prediction errors only when confronted with undeniable conflicting data. However, managers often unconsciously filter out this data (confirmation bias). But generative AI agents can be programmed to act as a *Devil's Advocate* or as a representative of a diverse set of stakeholders. A study on multi-agent AI systems demonstrated that when AI agents simulated clinical team dynamics (e.g., one agent acting as a sceptic), diagnostic accuracy increased from 0% to 76% by effectively debiasing the initial human judgment (Ke et al., 2024). This confirms that AI can induce the *reflection* Colombo deems essential for learning, forcing the manager to confront the "dialogic" tension between their vision and conflicting evidence.

In a similar way, AI can mitigate overconfidence with a calibration effect. While AI can theoretically debias, it can also amplify overconfidence if managers treat it as a verification tool for their gut instinct. Keding and Meissner (2021) propose that AI is most effective when designed to actively trigger, also in a management dialogue, cognitive dissonance or reflection, creating a *calibration effect*. But overconfidence often stems from the *illusion of validity* (Lovallo & Kahneman, 2003) (Flyvbjerg, 2021), where managers mistake the coherence of a story for its probability. Research suggests that *reflection-inducing prompts* from AI systems can significantly reduce decision-making errors by forcing a System 2 (slow, analytical) pause in the workflow. By simulating thousands of scenarios (Monte Carlo simulations via AI), managers are forced to see the probabilistic distribution of outcomes, breaking the deterministic illusion of their vision.

By contrast, AI can help reduce cognitive load too. Biologically, unbiased decision-making is metabolically expensive. It requires the dorsolateral prefrontal cortex to inhibit the amygdala's emotional impulses. By automating data synthesis and pattern recognition, AI reduces the *cognitive burden* on frontline practitioners and can facilitate management dialogues. This *offloading* preserves the manager's neural energy for the high-level task of judgment and ethical reasoning, domains where human nuance is irreplaceable.

While AI debiases human errors, it introduces its own risks that must be managed to maintain the validity of the strategic discourse. There is a proven tendency for humans to over-rely on algorithmic advice, viewing it as objective truth, a phenomenon known as *automation bias*. This is dangerous because AI models can hallucinate or inherit *dataset bias* from their training corpora.

Colombo warns against *symptomatic actions* that address surface problems rather than structural causes. If managers use AI only to answer immediate questions without understanding the underlying causal maps, they risk reinforcing a self-destructive prophecy where the AI amplifies the organisation's existing blind spots (e.g., groupthink bias in AI outputs). But a methodologically assisted use of AI can avoid this trap and force management to focus its attention on the most relevant issues.

6 – Conclusions

The "race with the machine" is not a zero-sum game of substitution, but a convergent evolution of two different solutions to the same thermodynamic problem: how to process information in a complex world.

We must abandon the management myth of the purely rational, logical decision-maker. Modern neuroscience confirms that the human brain is a metabolically frugal engine that prioritizes survival over accuracy, defaulting to "No" and relying on emotional tags to make decisions. Cognitive biases are not software bugs we can simply "learn" our way out of; they are the hardware architecture of a system designed to save energy.

Cognitive bias in management should be understood less as a removable "error" and more as a predictable consequence of constraints. Human judgment is produced by a biological system optimized for energetic efficiency and survival, not for exhaustive computation. Evolutionary-developmental canalization makes some responses robust and "default," while the small-world efficiency of neural connectivity privileges fast, low-cost inference and re-use of established pathways over the costly construction of new ones. In organizational settings, where uncertainty, time pressure, and political stakes are high, these constraints do not disappear; they become amplified. As a result, purely educational or motivational debiasing approaches often struggle because they ask individuals to reliably override biologically efficient defaults without changing the decision environment.

At the same time, artificial intelligence does not offer a simple escape from bias. Managers can miscalibrate trust in algorithmic advice (automation bias), and generative models can hallucinate or reproduce training-data distortions. The central claim is therefore not that AI eliminates bias, but that it can externalize it into a system that is, in principle, inspectable and contestable. By shifting part of the inference process from the opaque internal dynamics of the human brain to an explicit conversational artifact, prompts, counter-arguments, assumptions, scenario distributions, and decision logs, organizations can create a practical surface on which bias can be interrogated, compared, and improved over time. This resonates with the view that strategic management is constructed through discourse and that organizational change is shaped by dialogue as both an innovative and political process.

The managerial implication is a design principle: AI should be deployed less as an “answer machine” and more as a dialogical partner that reliably introduces constructive friction into strategic conversations. Concretely, AI is most promising when configured to (i) generate structured dissent (e.g., Devil’s Advocate roles, stakeholder simulations), (ii) enforce consideration of base rates and counterfactuals, (iii) expand the option set and articulate competing causal maps, and (iv) surface uncertainty through probabilistic reasoning and scenario distributions rather than single-point forecasts. In this way, AI can help managers confront disconfirming evidence that would otherwise be filtered out and can reduce cognitive load by offloading synthesis while preserving human responsibility for value-laden judgment.

Finally, the integration of AI allows us to revive Gianluca Colombo’s vision of the manager as a *researcher-practitioner*. In this new paradigm, the manager does not hoard answers but curates questions. By using AI to stimulate strategic conversations and trigger the calibration effects that force us out of our comfort zones, we can finally overcome the evolutionary constraints that have limited human decision-making for millennia. We are no longer prisoners of our Pleistocene past; with the right "neuro-cognitive exoskeleton," we can construct a new, hybrid rationality.

The manager of the future, in this view, is less a passive consumer of algorithmic outputs and more a researcher-practitioner who uses AI to continuously test, challenge, and reconstruct strategic theories in action.

7 – Bibliography

Ahn, M., Brohan, A., Brown, N., Chebotar, Y., Cortes, O., David, B., Finn, C., Fu, C., Gopalakrishnan, K., & Hausman, K. (2022). Do as I can, not as I say: Grounding language in robotic affordances. *arXiv preprint arXiv:2204.01691*.

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

Barrett, L. F. (2016). *How emotions are made: The secret life of the brain*. Houghton Mifflin.

Bort, J. (2025). Sam Altman says OpenAI has $20B ARR and about $1.4 trillion in data center commitments. *TechCrunch*(November 6).

Christiano, P. F., Leike, J., Brown, T., Martic, M., Legg, S., & Amodei, D. (2017). Deep reinforcement learning from human preferences. *Advances in neural information processing systems*, *30*.

Chung, H. W., Hou, L., Longpre, S., Zoph, B., Tay, Y., Fedus, W., Li, Y., Wang, X., Dehghani, M., & Brahma, S. (2024). Scaling instruction-finetuned language models. *Journal of Machine Learning Research*, *25*(70), 1–53.

Coda, V., & Mollona, E. (2003). Il Governo della dinamica strategica. *Finanza, marketing, produzione*(1).

Colombo, G. (1993). Da Atena a Hermes: pensare e agire la strategia. *Economia e management*(5).

Colombo, G. (1997). L’impresa nella complessità. *Sviluppo e Organizzazione*(143 maggio-giugno).

Colombo, G. (1998). Mappe cognitive. In V. Coda, G. Invernizzi, & M. Rispoli (Eds.), *Enciclopedia dell’impresa UTET (EDI), vol. Strategia aziendale*.

Colombo, G. (2004). La dimensione discorsiva del management. *Working Paper Università dell’Insubria - Facoltà di Economia*(2004/8).

Damasio, A. R. (1994). *Descartes’ error: emotion, reason, and the human brain*. Avon Books.

Danziger, S., Levav, J., & Avnaim-Pesso, L. (2011). Extraneous factors in judicial decisions. *Proceedings of the National Academy of Sciences*, *108*(17), 6889–6892.

Dosenbach, N. U. F., Raichle, M. E., & Gorpdon, E. M. (2025). The brain's action-mode network. *Nature Reviews Neuroscience*, *26*(3), 158–168. https://doi.org/10.1038/s41583-024-00895-x

Driess, D., Xia, F., Sajjadi, M. S., Lynch, C., Chowdhery, A., Wahid, A., Tompson, J., Vuong, Q., Yu, T., & Huang, W. (2023). Palm-e: An embodied multimodal language model.

Epoch AI. (2024). *The data wall: data scarcity analyses and training cost trends*. Epoch AI.

Falk, E. B., Rameson, L., Berkman, E. T., Liao, B., Kang, Y., Inagaki, T. K., & Lieberman, M. D. (2010). The neural correlates of persuasion: A common network across cultures and media. *Journal of cognitive neuroscience*, *22*(11), 2447–2459.

Fedus, W., Zoph, B., & Shazeer, N. (2022). Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity. *Journal of Machine Learning Research*, *23*(120), 1–39.

Flyvbjerg, B. (2021). Top Ten Behavioral Biases in Project Management: An Overview. *Project Management Journal*, *52*(6), 531–546. https://doi.org/10.1177/87569728211049046

Google. (2024). Introducing Gemini 1.5, Google's next-generation AI model. *Google Blog*.

Haselton, M. G., & Buss, D. M. (2000). Error management theory: a new perspective on biases in cross-sex mind reading. *Journal of personality and social psychology*, *78*(1), 81.

Hebb, D. O. (2002). *The organization of behavior: A neuropsychological theory* (2 ed.). Psychology press. (Original work published 1949)

Herculano-Houzel, S. (2009). The human brain in numbers: a linearly scaled-up primate brain. *Frontiers in human neuroscience*, *3*, 857.

Herculano-Houzel, S. (2016). *The Human Advantage: A New Understanding of How Our Brain Became Remarkable*. The MIT Press.

Hoffmann, J., Borgeaud, S., Mensch, A., Buchatskaya, E., Cai, T., Rutherford, E., Casas, D. d. L., Hendricks, L. A., Welbl, J., & Clark, A. (2022). Training compute-optimal large language models. *arXiv preprint arXiv:2203.15556*.

Jack, J., & Appelbaum, L. G. (2018). "This is your brain on rhetoric": Research directions for neurorhetorics. *Fifty Years of Rhetoric Society Quarterly*, 247–275.

Jacob, F. (1977). Evolution and tinkering. *Science*, *196*(4295), 1161–1166.

Jiang, A. Q., Sablayrolles, A., Roux, A., Mensch, A., Savary, B., Bamford, C., Chaplot, D. S., Casas, D. d. l., Hanna, E. B., & Bressand, F. (2024). Mixtral of experts. *arXiv preprint arXiv:2401.04088*.

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Kahneman, D., Sibony, O., & Sunstein, C. R. (2021). *Noise: a flaw in human judgment*. Little, Brown and Company.

Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., & Amodei, D. (2020). Scaling laws for neural language models. *arXiv preprint arXiv:2001.08361*.

Ke, Y., Yang, R., Lie, S. A., Lim, T. X. Y., Ning, Y., Li, I., Abdullah, H. R., Ting, D. S. W., & Liu, N. (2024). Mitigating cognitive biases in clinical decision-making through multi-agent conversations using large language models: simulation study. *Journal of Medical Internet Research*, *26*, e59439.

Keding, C., & Meissner, P. (2021). Managerial overreliance on AI-augmented decision-making processes: How the use of AI-based advisory systems shapes choice behavior in R&D investment decisions. *Technological Forecasting and Social Change*, *171*, 120970. https://doi.org/https://doi.org/10.1016/j.techfore.2021.120970

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *25*.

Lovallo, D., & Kahneman, D. (2003). Delusions of success. *Harvard Business Review*, *81*(7), 56–63.

MacLean, P. (1990). *The Triune Brain in Evolution: Role in Paleocerebral Functions*. Plenum Pub Corp.

Madaan, A., Tandon, N., Gupta, P., Hallinan, S., Gao, L., Wiegreffe, S., Alon, U., Dziri, N., Prabhumoye, S., & Yang, Y. (2023). Self-refine: Iterative refinement with self-feedback. *Advances in neural information processing systems*, *36*, 46534–46594.

Mella, P. (2014). *Teoria del controllo: Dal systems thinking ai sistemi di controllo* (4. ed.). Franco Angeli.

Mella, P. (2025). *The Combinatory Systems Theory: A Powerful Theory for Understanding, Modeling and Simulating Collective Phenomena* (2nd ed.). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-86946-4 (Original work published 2017)

Merton, R. K. (1968). The Matthew effect in science: The reward and communication systems of science are considered. *Science*, *159*(3810), 56–63.

Mollona, E. (2000). *Analisi dinamica dei sistemi aziendali*. Egea.

Musso, M., Moro, A., Glauche, V., Rijntjes, M., Reichenbach, J., Büchel, C., & Weiller, C. (2003). Broca's area and the language instinct. *Nature neuroscience*, *6*(7), 774–781.

OpenAI. (2025). Introducing GPT-5. *OpenAI blog*.

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., & Ray, A. (2022). Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, *35*, 27730–27744.

Raichle, M. E., & Gusnard, D. A. (2002). Appraising the brain's energy budget. *Proceedings of the National Academy of Sciences*, *99*(16), 10237–10239.

Raichle, M. E., & Mintun, M. A. (2006). Brain work and brain imaging. *Annual Review of Neuroscience*, *29*(1), 449–476.

Russell, S., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach* (4 ed.). Pearson. (Original work published 2003)

Senge, P. M. (1990). *The fifth discipline* [La quinta disciplina]. Doubleday.

Shazeer, N., Mirhoseini, A., Maziarz, K., Davis, A., Le, Q., Hinton, G., & Dean, J. (2017). Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. *arXiv preprint arXiv:1701.06538*.

Shinn, N., Cassano, F., Gopinath, A., Narasimhan, K., & Yao, S. (2023). Reflexion: Language agents with verbal reinforcement learning. *Advances in neural information processing systems*, *36*, 8634–8652.

Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R., & Gal, Y. (2024). AI models collapse when trained on recursively generated data. *Nature*, *631*(8022), 755–759. https://doi.org/10.1038/s41586-024-07566-y

Sporns, O. (2016). *Networks of the Brain*. MIT press.

Stanford HAI. (2025). *AI Index Report 2025*. Stanford.

Steffen, P. R., Hedges, D., & Matheson, R. (2022). The Brain Is Adaptive Not Triune: How the Brain Responds to Threat, Challenge, and Change. *Front Psychiatry*, *13*, 802606. https://doi.org/10.3389/fpsyt.2022.802606

Stiennon, N., Ouyang, L., Wu, J., Ziegler, D., Lowe, R., Voss, C., Radford, A., Amodei, D., & Christiano, P. F. (2020). Learning to summarize with human feedback. *Advances in neural information processing systems*, *33*, 3008–3021.

Striedter, G. F. (2005). *Principles of brain evolution*. Sinauer associates.

Stylianou, N., Learner, S., Bradshaw, T., Uddin, R., Bott, I., Nevitt, C., Clark, D., & Joiner, S. (2025). Inside the relentless race for AI capacity. *Financial Times - Special Report*(July 31).

Tooby, J., Cosmides, L., & Barkow, J. H. (1992). The psychological foundations of culture. *The adapted mind: Evolutionary psychology and the generation of culture*, *19*(1), 1–136.

Tsubo, T., & Kurokawa, M. (2018). Verification of the effect of the axon fluid as a highly dielectric medium in the high-speed conduction of action potentials using a novel axon equivalent circuit. *Biophysics and Physicobiology*, *15*, 214–228. https://doi.org/10.2142/biophysico.15.0_214

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, *185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

Villalobos, P., Sevilla, J., Heim, L., Besiroglu, T., Hobbhahn, M., & Ho, A. (2022). Will we run out of data? An analysis of the limits of scaling datasets in machine learning. *arXiv preprint arXiv:2211.04325*, *1*, 1.

Waddington, C. H. (2014). *The strategy of the genes*. Routledge.

Walker, M. (2017). *Why We Sleep: Unlocking the Power of Sleep*. Scribner.

Wang, X., Wei, J., Schuurmans, D., Le, Q., Chi, E., Narang, S., Chowdhery, A., & Zhou, D. (2022). Self-consistency improves chain of thought reasoning in language models. *arXiv preprint arXiv:2203.11171*.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Le, Q. V., & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, *35*, 24824–24837.

Yao, S., Yu, D., Zhao, J., Shafran, I., Griffiths, T., Cao, Y., & Narasimhan, K. (2023). Tree of thoughts: Deliberate problem solving with large language models. *Advances in neural information processing systems*, *36*, 11809–11822.

Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K. R., & Cao, Y. (2022). React: Synergizing reasoning and acting in language models. The eleventh international conference on learning representations,

Ying, C., Kumar, S., Chen, D., Wang, T., & Cheng, Y. (2018). Image classification at supercomputer scale. *arXiv preprint arXiv:1811.06992*.

Yoon, J. (2025). What if the AI race isn't about chips at all? *Financial Times*(November 10).

Zhang, Q., Lyu, F., Sun, Z., Wang, L., Zhang, W., Hua, W., Wu, H., Guo, Z., Wang, Y., & Muennighoff, N. (2025). A Survey on Test-Time Scaling in Large Language Models: What, How, Where, and How Well? *arXiv preprint arXiv:2503.24235*.

Zitkovich, B., Yu, T., Xu, S., Xu, P., Xiao, T., Xia, F., Wu, J., Wohlhart, P., Welker, S., & Wahid, A. (2023). Rt-2: Vision-language-action models transfer web knowledge to robotic control. Conference on Robot Learning,