FALSE COGNITIVE POWER TRANSFER

From Individual Atrophy to Collective Demoralization in the Age of AI

Author: Adrian (Adi) Stan
Institution: Independent researcher, Pitești, Romania
ORCID: 0009-0003-1457-5155

ABSTRACT

Generative artificial intelligence (GAI) produces a cognitive paradox: it amplifies the capabilities of experts, but generates the dangerous illusion of competence in inexperienced users. This article introduces the concept of **False Cognitive Power Transfer (FCPT)** - the phenomenon by which individuals mistakenly attribute the effectiveness of AI outputs to their own cognitive competence, leading to taking on more tasks or responsibilities than they can realistically, systematically manage.

FCPT triggers two collective failure mechanisms: **(1) Collective Demoralization Cascade (CDC)** - spectacular failures lead to collective preemptive abandonment, and **(2) Replication Illusion (RI)** - observing successes generates overconfidence in replicability, ignoring that AI functions as a cognitive exoskeleton that **amplifies** existing skills and knowledge, but **does not create them** .

In the long term, both mechanisms lead to **cognitive atrophy** - the progressive loss of the original deep-thinking capacity, even in experts (L3). The article proposes **Mechanism of Cognitive Stimulation (MCS)** within MEG as a countermeasure intervention, validated by Monte Carlo simulations.

Implications for the responsible design of AI systems are discussed, highlighting the need for the right calibration between augmentation and dependency.

This paper argues that a major systemic risk of generative AI, often underestimated in relation to misinformation or job losses, is the large-scale miscalibration of human cognitive capacity.

The work is part of a larger body of independent research on cognitive divergence and the social impact of artificial intelligence.

Acronyms:

- **FCPT** = False Cognitive Power Transfer
- **CDC** = Collective Demoralization Cascade
- **RI** = Replication Illusion
- **MCS** = Mechanism of Cognitive Stimulation

Keywords: generative artificial intelligence, false cognitive power transfer, demoralization cascade, replication illusion, cognitive exoskeleton, cognitive atrophy

CHAPTER 1: INTRODUCTION

1.1 Context of the problem

GAI has democratized access to sophisticated cognitive capabilities. A student generates code in unknown languages. A manager produces strategic analyses without a consulting background. A junior researcher formulates complex experimental hypotheses without decades of experience.

This apparent democratization masks a problem: **the confusion between the efficiency of the tool and the competence of the user.** When AI produces high-quality output, the user experiences **FCPT** - the illusion that the cognitive power in the output comes from their own mental capabilities.

1.2 Gaps in the literature

The Dunning-Kruger effect (Kruger & Dunning, 1999) explains how incompetence generates overestimation through the inability to recognize one's own ignorance. But this mechanism is internal - it results from a lack of metacognitive knowledge, not from access to powerful external tools.

Cognitive Atrophy Through Technology (Carr, 2014) documents how GPS reduces spatial orientation and computers diminish mental calculation skills. But the focus remains on individual loss of skills, not on the social consequences of the illusion of competence generated by tools.

The Automation Complacency in aviation psychology (Parasuraman & Manzey, 2010) shows how pilots become dependent on autopilot, with the degradation of manual piloting skills. But the focus is on safety risks in critical systems, not on the social dynamics of the perception of augmented capabilities.

Moral hazard in economics (Arrow, 1963) describes how insurance against risks encourages the taking of greater risks. Conceptually similar, but limited to economic behavior and lacking the cognitive or social component.

The literature focuses on extremes:

A. Technological Optimism: AI as Augmentation (Brynjolfsson & McAfee, 2014)
B. Dystopian pessimism: AI as a threat (Carr, 2014; Sparrow et al., 2011)

The existing literature thus covers isolated elements – individual atrophy, illusions of competence, safety risks – but **no theory combines the false cognitive power transfer induced by AI, with the resulting collective social cascade.**

There is a lack of analysis **of the intermediary mechanisms** through which AI generates negative effects even in the absence of malicious intent, an integrated theory that explains:

1. How access to powerful tools generates systematic confusion between tool capability and user capability
2. How this confusion can lead to unrealistic commitments in complex projects
3. How the resulting spectacular failures are socially interpreted as evidence of the impossibility of the task, not of erroneous estimation
4. How this interpretation generates collective abandonment of important works, perfectly achievable with correctly calibrated expectations
5. Why observing successes doesn't protect against failures
6. How capabilities degrade even among experts
7. **How can this mechanism be prevented through cognitive calibration tools?**

1.3 Epistemic delimitation and scope

This paper does not propose a universal theory of human cognition, nor does it claim that all forms of AI use inevitably lead to cognitive degradation or functional dependency. Nor does it claim that every failure that occurs in AI-assisted contexts is explainable by the mechanisms described here.

The proposed theoretical framework applies exclusively to situations characterized by a significant asymmetry between the actual ability of the user and the power of the tool used, especially when this asymmetry is masked by performant interfaces and apparently positive feedback. The model does not describe all forms of AI use, but those contexts in which technological amplification produces a systematic error in the attribution of competence.

The paper also does not claim that the effects described are inevitable. They represent structural risks, not determined trajectories. Their occurrence depends on factors such as the lack of corrective feedback mechanisms, the absence of metacognitive training, and the uncritical use of AI systems in environments with high performance pressure.

In this sense, the present analysis should be read as an explanatory and preventive framework, not as a deterministic prediction on the evolution of the human-AI relationship.

1.4 Why is this frame appearing now?

The emergence and current relevance of **False Cognitive Power Transfer** are not accidental. Three historical conditions converge for the first time:

Exponential scaling of AI capabilities
Current models produce results that are clearly beyond the capabilities of many human users, creating an unprecedented discrepancy between human input and perceived output.

Democratizing access to advanced tools
Capabilities that previously required specialized teams are now instantly accessible, without proper training processes.

Accelerating feedback cycles and social validation
Social media, performance metrics, and attention economies quickly amplify the perception of success, before real understanding can be assessed.

In this historical configuration, the difference between real competence and assisted performance becomes systemically opaque, and classical cognitive calibration mechanisms no longer function. Hence the need for a framework such as the one proposed in this paper.

1.5 Methodology and Epistemic positioning

This paper adopts a **conceptual-analytic** and **exploratory approach,** situated at the intersection of cognitive science, technology studies, and organizational analysis. The goal is not to empirically test a single hypothesis, but to formulate an integrative theoretical framework capable of explaining a set of observable phenomena emerging in the recent interaction between humans and generative artificial intelligence systems.

Methodologically, the analysis is based on:

1. **Interdisciplinary theoretical synthesis,** combining concepts from cognitive psychology (e.g. metacognition, social learning), automation studies, behavioral economics, and sociology of technology.
2. **Pattern-based analysis,** in which disparate empirical events (organizational failures, recurring behaviors, observable dynamics in the AI ecosystem) are correlated to identify common explanatory structures.
3. **Case studies,** used not as strict causal evidence, but as heuristic examples that make the proposed theoretical mechanisms visible.

Therefore, the statements in this paper should be understood as explanatory hypotheses and conceptual models, not as generalizable statistical results. Their value lies in their ability to generate testable predictions and guide future empirical investigations.

CHAPTER 2: FALSE COGNITIVE POWER TRANSFER (FCPT)

2.1 The Fundamental Analogy: The Carpenter and the Pneumatic Hammer

An experienced carpenter works with a hand hammer. He can roof a small house in 10 days. He gets a pneumatic hammer - now he roofs the same house in 3 days. **REAL power transfer:** the hammer amplifies the existing muscle force from F to 3F (in studies in previous works, we argued a **potential** amplification figure of **71x,** in specific cases). After a period of constant use, the hand muscle atrophies, and the carpenter even forgets how to hammer nails by hand. Up to this point, the phenomenon is documented in the specialized literature under the name “cognitive atrophy by neglect” (Carr, 2014; Christensen et al., 2016).

The critical problem arises when the same carpenter declares: "I can roof the Cathedral in five days." If the task consisted only of hammering nails, the estimate might be plausible. But the Cathedral requires much more: complex architecture, coordination between crafts, historical understanding, specialized materials, layered aesthetic decisions. After a month of effort, the carpenter declares defeat, the roof is damaged, and the project fails spectacularly. The most serious consequence is not the individual failure of the carpenter, but the collective reaction: the other carpenters - with or without a Pneumatic Hammer - conclude: "If the Cathedral Carpenter was not able to build the roof with a Pneumatic Hammer, we certainly cannot!" The result: **The Cathedral remains roofless.** Not from a lack of real capabilities, but from an erroneous recalibration of collective confidence based on the observation of a failure generated by unrealistic expectations.

This is the essence **of the Cathedral Problem:** powerful tools generate not only **individual atrophy,** but also **a collapse of collective confidence** in the possibility of achieving complex works.

On the other hand, an apprentice, seeing the carpenter with the pneumatic hammer who finished the roof of the house in 3 days, thinks: "With that hammer, I can also roof a house in 3 days!". The result? Apprentice + pneumatic hammer = (still) insufficient. **The apprentice fails.** The problem: NOT the lack of the pneumatic hammer (existing), but **the erroneous attribution** - the belief that the " power" comes from it, and not from the carpenter's experience, amplified by the tool.

2.2 Formal definition

Artificial intelligence tools are the modern-day equivalent of a Pneumatic Hammer - but for cognitive, not physical, work. A strategic analyst using AI can produce in an hour what would have taken a week of traditional research. An AI-augmented programmer writes code ten times faster. A content creator generates in ten minutes what would have required hours of work. Actual productivity increases. But, just like with the Pneumatic Hammer, a dual phenomenon occurs:

1. **Atrophy of fundamental cognitive abilities** (already documented in the literature)
2. **False Cognitive Power Transfer** (the new mechanism proposed in this article)

The user systematically confuses the speed of the tool with his own capacity for deep understanding, strategic planning, and orchestration of complexity.

Definition: False Cognitive Power Transfer (FPCT) is the psychological mechanism by which the user of a powerful tool systematically confuses the tool's effectiveness in specific tasks with his or her own ability to manage complexity, make strategic decisions, and orchestrate multidimensional processes.

The technical infrastructure of this transfer is powered by **the opacity** of the AI system's Thinking Time **(Tg).** As defined in the MEG protocol (Stan, 2025), the absence of explicit signaling of computational effort allows the user to internalize the machine's processing speed as their own reasoning speed. FCPT is essentially a synchronization error between **biological time (Tb)** and **algorithmic time (Tg)**.

FCPT components

A. Demonstrable efficiency - The tool generates real productivity gains in well-defined tasks. Roofing a house can be done in "three days with a Pneumatic Hammer", just as a strategic analysis can indeed be completed in a few dozen minutes with AI.

B. Cognitive Atrophy - Constant use of the tool diminishes fundamental capabilities. The carpenter *forgets* how to hammer nails by hand. The analyst *forgets* how to structure problems strategically without AI. The phenomenon is extensively documented in the literature (Carr, 2014; Sparrow et al., 2011).

C. Categorical Confusion - The user unjustifiably extrapolates from speed in simple tasks to capacity in complex projects. Erroneous logic: "If I can cover a house in three days, I can cover the Cathedral in ten days." The confusion is categorical because in fact the Cathedral is not a "multiplied house", but a fundamentally different object - it requires architecture, coordination, history, aesthetics. Similarly, complex strategic problems are not "multiplied simple problems", but require contextual understanding, anticipation of consequences, multidimensional integration.

D. Structural hubris - Categorical confusion generates unrealistic commitments. The user overestimates, not out of ignorance (Dunning-Kruger), but from extrapolation **valid in a** limited domain, **erroneously applied to extended domains** .

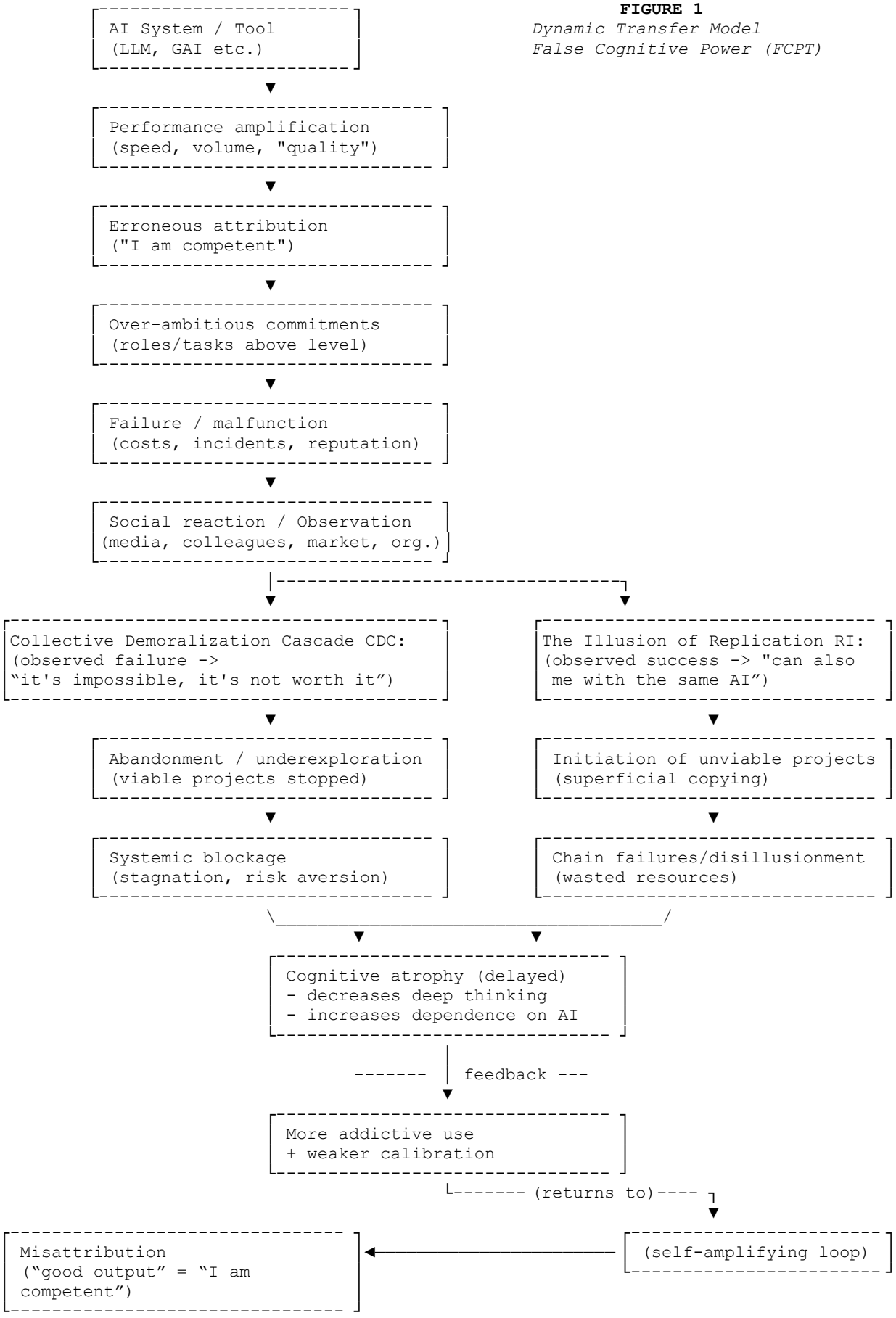
Formally, if:

- **E_real** = the user's real competence in a domain
- **A** = the amplification provided by the instrument
- **C** = confusion coefficient ($0 \leq C \leq 1$)

The resulting perception is: $\mathbf{E\_perceived = E\_real + C \times (A - E\_real)}$

When $C \to 1$, the user almost completely confuses the tool's capability with his own competence. When this confusion leads to unrealistic commitments and spectacular failures, the impact goes beyond the individual sphere, generating a **Collective Demoralization Cascade** (Chapter 3) with major societal effects.

FIGURE 1
Dynamic Transfer Model
False Cognitive Power (FCPT)



To summarize the relationships between the mechanisms described so far, Figure 1 presents a conceptual model of False Cognitive Power Transfer and the resulting dynamics. The diagram highlights the sequence between assisted performance amplification, attribution error, the emergence of over-ambitious behaviors, and the resulting systemic consequences, including feedback loops that can lead to cognitive atrophy :

- FPCT appears in "Misattribution".
- from "Social Reaction", two branches converge: CDC (from failure) and RI (from observed success).
- both can lead to Cognitive Atrophy and addiction loops.

2.3 The Three Levels Trap

In the research on *Cognitive Divergence* (Stan, 2025), we stratified AI users into three levels:

1. **L1 - Passengers:** users who consume AI results without understanding their internal mechanisms or limits.
2. **L2 - Operators:** users who operate and manage AI systems, but without full conceptual control over their architecture and implications.
3. **L3 - Architects:** actors capable of designing, coordinating and integrating complex systems, understanding both their limits and their systemic consequences.

False Cognitive Power Transfer (FPCT) does not affect these levels in isolation, but produces **hierarchical cascades,** in which self-assessment errors at one level propagate upward, distorting the collective perception of competence and feasibility.

Cascade 1: L1 → (L2 as filter/institution)
FCPT in L1 produces "gatekeeping", not general reluctance in L2.
1) Typical error (FCPT at L1): An L1, aided by AI, obtains results apparently comparable to the output of an L2 (reports, analyses, code, documents). His confusion is: *"if the output looks good, it means I can perform the role."*
2) Typical failure: The failure is not necessarily spectacular technically, but **operationally** :

- doesn't know what to check,
- doesn't know what is critical,
- does not see dependencies,
- does not handle exceptions,
- It faces major difficulties in maintaining the process when AI makes mistakes or when the context changes.

3) The observer (L2 / organization): L2 sees an "augmented" L1 that delivers beautiful output, but produces incidents, integration errors, additional work for rework, risk.
4) Misattribution (in the system): Misattribution can be (depending on the culture):

A. **Assignment A (most common):** *" L1 is difficult to evolve using AI alone. "* → conclusion: "we're no longer trying to promote L1 in L2 tasks; it's not worth the risk."
B. **Assignment B (more toxic):** *"AI is dangerous at the execution level."* → conclusion: "we ban / limit AI for execution roles."

5) Dominant (realistic) consequence: Not "L2 becomes generally reluctant." It appears:

- **strengthening gatekeeping** ,
- **more rigid procedures** ,
- **stricter separation L1/L2** ,
- **loss of growth pipeline** (L1s stay stuck at L1 for longer).

Net effect: The system becomes less permeable. Not because L2 is “afraid”, but because the institution protects its stability. In the long run, however, this can produce L2 **deficits and stagnation** .

Trigger condition: FCPT occurs when an L1 user interprets the performance obtained with AI support as evidence of their own competence and attempts to operate above their actual level of understanding.

Antidote: Clear separation between assisted competence and internal competence, through mechanisms that expose the real limits of the user and maintain functional differentiation between levels.

Cascade 2: L2 → L3

FCPT at L2 leads to decreased delegation and increased centralization at L3.

1) Typical error (FCPT at L2): L2 sees that it can quickly generate plans, strategies, communications, diagrams, “orchestration artifacts.” False conclusion: *“if I can produce L3 artifacts, I can operate as L3.”*

2) Typical failure: Failure occurs in real coordination:

- wrong prioritization,
- ignoring constraints,
- decisions without assumed costs,
- lack of consequence modeling,
- the inability to maintain coherence when conflicts arise between sub-systems.

3) The Observer (L3): L3 sees that delegation to the "augmented L2" does not produce scaling, but instability.

4) Misattribution (probable):

“We cannot delegate to lower levels.” (Sometimes correct, especially in the short term, but becomes erroneous if generalized.)

5) The dominant consequence:

- **L3 reduces delegation** ,
- **centralizes the decision** ,
- **increases control** ,
- **the speed of the organization decreases** ,
- **increases fragility** (*single point of failure:* L3).

Here the "reluctance" is real, but it's **towards delegation,** not towards the projects themselves.

Triggering condition: FCPT occurs when an L2 level user confuses the ability to produce complex artifacts with the ability to orchestrate complex systems, extrapolating beyond his actual area of control.

Antidote: Explicitly maintaining the distinction between execution and orchestration, through mechanisms that limit strategic delegation in the absence of demonstrable systemic integration competence.

Cascade 3: L3 → Market/ecosystem

FCPT at L3 causes loss of systemic confidence in augmentation scaling.

1) Typical error (FCPT at L3): A real L3 has competence, but overestimates how much it can scale with AI.

2) Typical failure: The mega-project fails by:

- over-complexity,
- uncontrolled interdependencies,
- degradation in communication,

- execution inconsistency,
- strategic drift.

3) The observer (market/investors/clients/institutions): They see the failure as a macro signal.

4) Misattribution:

"AI augmentation doesn't work at scale." Not "this project was poorly calibrated," but "the category is invalid."

5) The dominant consequence:

- funding decreases,
- risk aversion increases,
- viable projects are postponed/abandoned,
- legitimate innovation is blocked (exactly the "Societal Cathedral").

Trigger condition: FCPT occurs when an L3 actor **overestimates** the scalability of its own augmented capacity and **treats the failure** of a complex project as an **anomaly,** not as a **signal of a** structural limit.

Antidote: Clearly separating implementation failure from concept invalidation, through assessments that distinguish between contextual limitations and the real potential for scale-up.

FCPT does not produce the same type of demoralization at each stage. It produces **different distortions of trust** :

- at L1→L2: **mobility degrades** and filtering (gatekeeping) is strengthened
- at L2→L3: **delegation degrades** and centralization increases
- at L3→market: **confidence in scaling degrades** and investment is blocked

This is the " **waterfall** ": not emotion, but **miscalibration of the trust system** between levels.

In all three cases, the problem is not just a lack of capacity, but especially **a misperception of one's own position** in an augmented system. The false cognitive power transfer not only produces local failures, but **can distort** the mechanisms by which organizations and communities calibrate their perception of what is possible and feasible.

2.4 FCPT vs. Dunning-Kruger

Differences between the Dunning-Kruger Effect and False Cognitive Power Transfer (FCPT)

Size	The Dunning-Kruger effect	False Cognitive Power Transfer (FCPT)
Source	Lack of metacognitive competence	Confusion between human capacity and instrument performance
Mechanism	Not understanding one's own limits: "I don't know that I don't know"	Misattribution of capacity to the tool: "power comes from the tool"
Mode of appearance	Misassessment of one's own abilities	Incorrect transfer of performance from system to user
Possible correction	Direct experience + gradual feedback	Recalibration between human competence and the real role of technology
Area of manifestation	Limited to a specific domain	Transversal, affects several areas simultaneously
Dynamics over time	Tends to correct itself through practice and failure	Tends to self-amplify through repeated use of the tool
Main effect	Temporary overestimation of one's own competence	Cognitive dependence and loss of decision-making autonomy
Major risk	Isolated, self-correcting errors	Systemic degradation of judgment and initiative

Dunning-Kruger example: Piano student thinks he can play Chopin after 3 months → tries → quickly realizes he can't
FCPT Example: Student uses AI for sophisticated musical interpretation of Chopin → excellent output → believes he understands at an expert level → has difficulty doing without AI support.

2.5 Immediate Consequences of False Cognitive Power Transfer

Overexertion through overestimation
The process of False Cognitive Power Transfer leads to a predictable sequence of effects:

1. **Overestimating the actual ability** to solve a complex task
2. **Premature commitment of resources** (time, capital, reputation)
3. **Late discovery** of the real limits of the system used
4. **Visible operational failure,** with reputational and economic impact

This succession is not the result of a lack of intelligence, but of a systemic error in assessing the relationship between human capacity, technological support, and the real complexity of the problem.

2.6 Empirical Examples of False Cognitive Power Transfer (2021-2025)

Zillow Offers (2021)

- **Losses:** over $500 million
- **Closing:** November 2021
- **Staff reductions:** approximately 2,000 employees
- **Root cause:** real estate valuation algorithms that overestimated property values
- **Fundamental error:** confusing local predictive performance with the ability to manage real estate market dynamics
- **CEO statement (Rich Barton):** "The unpredictability of the market has far exceeded what we had anticipated."

FCPT: the belief that a performing model can substitute human judgment in complex systemic contexts.

Olive AI (2023)

- **Losses:** approximately $832 million
- **Peak valuation:** $4 billion (2021)
- **Layoffs:** 450 employees (July 2022) + 215 (February 2023)
- **Cause:** the promise of near-total automation of medical processes, without an adequate understanding of healthcare operational constraints
- **Fundamental error:** confusing point automation with systemic transformation
- **CEO statement:** "The accelerated pace of growth and lack of focus have severely strained the organization."

FCPT: overestimating AI's ability to replace human expertise in a critical and regulated field.

Builder.ai (2025)

- **Losses:** over $450 million
- **Maximum valuation:** $1.5 billion
- **Situation:** insolvency
- **Root cause:** operational model based on human labor disguised as full automation

- **Details:** Approximately 700 human developers supported what was billed as an “AI platform”
- **Structural problems:** overvalued revenues, unsustainable operating costs
- **Debts:** $88 million to AWS, $30 million to Microsoft

FCPT: customers and investors confused the promise of automation with the actual existence of a scalable solution.

2.7 Propagation mechanism: FCPT cascade

False Cognitive Power Transfer does not remain isolated. It generates a **cascade of systemic errors,** through which local failures turn into structural blockages.

Typical sequence:

1. **Initial failure** - an actor overestimates the capability of the AI-assisted system.
2. **Social observation** - failure becomes visible to other actors.
3. **Misattribution** - the cause of failure is attributed to fundamental domain limitations, not to evaluation error.
4. **Defensive generalization** - actors avoid similar projects, considering them impossible to achieve.
5. **Systemic blockage** - viable initiatives are abandoned preemptively.

In many situations, the final outcome is therefore not predominantly determined by the lack of technical capacity, but by **the loss of collective confidence in the possibility of building complex systems** .

This cascade is not just an isolated psychological phenomenon, but the invisible engine of up to 71x **Cognitive Divergence** - identified through Monte Carlo simulations (Stan, 2025). While **L3** users (*“Architects”*) use machine effort to scale complexity, **L1/L2** users fall victim to FCPT, entering a net negative productivity collapse through the inability to audit their own output.

CHAPTER 3: THE COLLECTIVE DEMORALIZATION CASCADE (CDC)

3.1 Definition

The Collective Demoralization Cascade (CDC) represents a social phenomenon through which **the visible failure of an actor,** occurring in the context of the use of advanced technologies, generates **the preemptive abandonment of similar initiatives,** probably often achievable under the right conditions.

The central element of this process is not the failure itself, but **the erroneous attribution of the cause of the failure:** what is, in reality, an implementation or evaluation error is reinterpreted as a structural limitation of the respective domain.

3.2 CDC mechanism

Phase 1: Initial Failure: The individual (the “carpenter”) with FCPT embarks on a complex project (“the Cathedral”), visibly fails, generating damage (“ the damaged roof”).

Phase 2: Social observational learning: Observers do not experience failure themselves, but observe it in others. Social psychology (Bandura, 1977) shows that observational learning is as powerful as direct experience in shaping behavior.

Phase 3: Misattribution: Observers attribute failure to "impossibility of the task even with powerful tool", not to "misestimation of complexity". Why? Because:

- The tool was obviously powerful (house covered in three days)
- The individual seemed competent (successful in simple tasks)
- The failure was categorical (Unfinished Cathedral)

Natural attribution: "The task is beyond human capabilities, even technologically augmented."

Phase 4: Preemptive Collective Abandonment: The other "carpenters" - including the competent ones who could realistically complete the Cathedral roof - preemptively give up: "If X with a pneumatic hammer couldn't do it, I certainly can't."

Phase 5: The cathedral without a roof: The important work remains unfinished not due to a lack of real capabilities (they are competent carpenters), but due to a collapse of collective trust generated by a socially amplified individual failure.

This sequence **turns** a **failure into a punctual** in a **traffic jam systemic** .

3.3 Conditions that favor the emergence of CDC

The demoralization cascade occurs with high probability when the following conditions are simultaneously met:

1. **High visibility** - the failure is public, mediatized, or symbolic.
2. **Credibility of the original actor** - the one who fails is perceived as competent or well-equipped.
3. **Causal ambiguity** - it is unclear whether the failure is due to implementation, context, or actual impossibility.
4. **Reputational pressure** - the social cost of failure discourages further replication or experimentation.

In this context, normal learning mechanisms are replaced by avoidance mechanisms.

3.4 Illustrative case studies

Case 1: "Cold Fusion" (1989-present)

The 1989 announcement of achieving low-temperature nuclear fusion generated massive interest. Later, the difficulty of replicating the results led to the conclusion that the phenomenon was physically impossible.

Consequences: almost total reduction of government funding for research in the field, stigmatization of the term "cold fusion" in academia, migration of research towards alternative terminologies (LENR).

Although subsequent studies indicated reproducible anomalies under certain conditions, the field remained marginalized for over three decades. It was not until 2020 that significant funding programs reemerged.

Case 2: Gene therapy (1999-2013)

In 1999, the death of Jesse Gelsinger during a clinical trial triggered a severe institutional backlash. Funding was drastically reduced, and the field entered a decade-long decline.

Although subsequent investigations indicated problems specific to the vectors used and safety procedures, the prevailing perception was that gene therapy as a whole was too risky.

Only after the development of safer vectors (AAV) and the revision of protocols did the field begin to recover, culminating in clinical approvals after 2017.

Case 3: The "winters" of Artificial Intelligence

In the 1970s and 1980s, the over-promises of symbolic artificial intelligence led to major disappointments. The Lighthill Report (1973) and the failures of expert systems triggered massive funding cuts.

The result was a period of stagnation of over a decade, known as the "AI Winter", in which the field was marginalized, despite the existence of viable technical directions.

Only with the emergence of new paradigms (machine learning, big data, accelerated hardware) was the field rehabilitated.

3.5 Psychological and social mechanisms involved in the Collective Demoralization Cascade

The Collective Demoralization Cascade (CDC) is supported by several cognitive and social mechanisms well documented in the literature. These mechanisms explain why single failures can generate disproportionate systemic effects.

1. Observational learning (Bandura, 1977)

According to Bandura's social learning theory, individuals learn not only through direct experience, but also by observing the behavior of others and its consequences.

In the context of CDC, actors observe another's failure without directly experiencing the conditions that produced it. This vicarious learning is effective but deeply vulnerable to attribution errors.

Thus, a singular but visible failure is internalized as a general rule, even if it is the result of particular and unrepresentative factors.

2. Information waterfall (Bikhchandani, Hirshleifer & Welch, 1992)

The information cascade occurs when individuals, in the absence of complete information, base their decisions on the observable actions of others, assuming that they possess superior information.

In the context of the CDC:

- the first visible failures are interpreted as strong informational signals;
- subsequent decisions are no longer made based on one's own assessment, but by imitating the behavior of the majority;
- the process is self-amplifying, even if the initial information is incomplete or erroneous.

The result is a rapid convergence towards collective avoidance, regardless of the actual feasibility of the project.

3. Public failure aversion and reputational cost

In addition to cognitive mechanisms, CDC is amplified by reputational factors. In professional and institutional contexts, public failure entails high costs: loss of credibility, funding, and social capital.

Consequently, actors prefer to completely avoid initiatives perceived as risky, even when the probability of success is reasonable and the potential benefits are high. This risk aversion is not irrational at the individual level, but becomes destructive at the systemic level.

3.6. Summary: the complete mechanism of CDC

The combination of these mechanisms produces a structural blocking effect:

1. **A visible failure triggers a generalized interpretation.**
2. **Interpretation propagates through social learning mechanisms.**
3. **Actors adjust their behavior to avoid reputational risk.**
4. **Viable initiatives are abandoned before they are tested.**

Thus, CDC tends to reflect the limits of collective perception of risk and failure rather than technological limits themselves, at least in many contexts analyzed here.

CHAPTER 4: THE ILLUSION OF REPLICATION

4.1 Definition and mechanism

The Replication Illusion (RI) is the phenomenon whereby the success achieved by an actor using artificial intelligence systems is interpreted by observers as being easy to reproduce, often without taking into account their actual level of competence. This erroneous interpretation occurs when the use of a performing tool is confused with the possession of the competence necessary to produce the respective results.

Although **the Replication Illusion** bears superficial similarities to well-known cognitive effects, such as Dunning-Kruger, there are important conceptual differences that justify treating the phenomenon as distinct in the framework proposed here. While Dunning-Kruger describes an internal self-assessment error, the Replication Illusion arises from an external misattribution: the transfer of observed performance to a tool, not to the competence of the one using it. This difference makes the phenomenon not just an individual cognitive error, but a systemic mechanism, amplified by technological infrastructures.

Essentially, the Replication Illusion occurs when one assumes that observed success is the sole result of the tool, not the interaction between the tool and the user's deep competence. This ignores the fact that artificial intelligence systems function as **amplifiers of existing cognitive capacity,** not as a substitute for it.

A key difference between users who achieve exceptional results and those who fail seems to lie more in the pre-existing level of expertise, in the ability to formulate problems, critically evaluate results, and integrate them into a broader context, than in simple access to technology. The illusion occurs when an actor's success is observed in isolation, without taking into account:

- the years of experience preceding the use of the technology;
- progressive accumulation of domain-specific expertise;
- the ability to evaluate, correct and refine the generated results;
- the cognitive and organizational infrastructure that supports performance.

Thus, what is perceived as a "replicable recipe" is, in reality, the result of a long construction, difficult to compress into a set of simple steps or a universal tool.

CDC vs. RI: contrasting patterns

Size	**Collective Demoralization Cascade (CDC)**	**Replication Illusion (RI)**
Trigger event	The visible failure of an actor using AI	The visible success of an actor using AI
Observation type	Observing public failure	Observing public success
Initial reaction	Withdrawal, excessive caution, avoidance	Enthusiasm, overconfidence, rapid imitation

Central cognitive error	Attributing failure to a structural impossibility	Attributing success solely to the tool
Implicit reasoning	"If it didn't work even with AI, then it's impossible"	"If he can do it with AI, so can I"
Distortion type	Negative (preventive lock)	Positive (overestimation)
Level of analysis affected	Feasibility assessment	Assessing your own competence
Fundamental error	Confusing contextual failure with structural limit	Confusing amplification with competence
Immediate consequence	Abandoning viable projects	Initiating unviable projects
Systemic impact	Slowing progress, under exploration	Waste of resources, chain failures
Dominant psychological mechanism	Learning through negative observation (Bandura, 1977)	Overconfidence induced by observed success
Long-term effect	Excessive conservatism, stagnation	Overload, disillusionment, loss of confidence
Final result	Valid projects remain untried	Unviable projects are being tried on a massive scale
Type of epistemic error	Systemic underestimation	Systemic overestimation

4.2 The cognitive exoskeleton: amplification, not substitution

To understand the mechanism of the Replication Illusion, the analogy with a mechanical exoskeleton is essential. An exoskeleton amplifies an individual's strength, but does not grant them non-existent abilities. A weak operator remains weak, even if he can temporarily lift a heavier weight.

The same principle applies to artificial intelligence systems. If we note:

- $\mathbf{E_Y}$ = the level of expertise of an experienced user (level L3),
- $\mathbf{E_t}$ = the level of expertise of a novice user (level L1),
- **A** = the amplification factor offered by the AI system,

then the result obtained is:

- for expert: $\mathbf{O_Y = E_Y \times A}$
- for the novice: $\mathbf{O_t = E_t \times A}$

Although both use the same instrument, the difference in performance remains proportional to the difference in initial proficiency. In many contexts, amplification does not completely equalize, but tends to **widen the gaps** .

The illusion occurs because the observer sees only the final result, not the initial value of the skill. Thus, the expert's success is perceived as directly attributable to the tool, and not to previously accumulated expertise.

This misinterpretation leads to the false conclusion that "if I use the same tool, I can achieve the same result", ignoring the fact that the tool does not create competence, but only amplifies the existing one.

4.3 Misattribution in observing success

What is seen when Y succeeds:

- spectacular result (quality output, product, strategy, execution)
- "Y uses GPT/Claude"
- "Y generated $5 million in revenue"
- articles like: "How I built X with AI in three months"

What is not seen:

- 10 years of expertise in the field
- 1000 failed iterations until product-market fit
- subtle judgments in every prompt/instruction/direction choice
- the ability to evaluate the quality of the output (what is good, what is mediocre, what is wrong)
- network, distribution, synchronization

Misattribution:

- **Apparent (visible) cause:** "AI tool + three months"
- **Real cause (invisible):** "E_Y (L3 level expertise) × 10 years + AI tool + three months of execution"

Decision based on erroneous attribution:

✓ “I have the same AI tool → I can replicate in three months.”

Conceptual transition

When misattribution becomes stable, it no longer remains a simple cognitive bias. It begins to guide concrete decisions: investments, technological choices, organizational structures. Thus, the Replication Illusion does not manifest itself only at the perceptual level, but materializes in operational actions that can be evaluated through measurable results.

4.4 Why you fail when copying - The typical scenario

Step 1 - Observation

- see startup Z: “AI writing assistant” → $10 million annual recurring revenue in 18 months (ARR)
- LinkedIn post: "We put a GPT API in a nice interface and it's done"

Step 2 - Activating FCPT

- "It's simple: API call + prompt template + interface"
- "I can reply in six months"

Step 3 - Execution

- you build a similar "wrapper" over the same model
- your output: generic, repetitive formula, no nuances
- Z's output: precise, contextualized, adapted, with stylistic registers

Step 4 - Confusion

- "We're using the same GPT, why is the result different?"
- you test their prompts → it works
- test your prompts → remain generic

Step 5 - Failure

- users leave after the trial period
- feedback: "the result is not differentiated enough"
- the startup closes in 12 months

The invisible difference:

Z had

- founder with a PhD in computational linguistics (level L3)
- deep understanding of semantics, pragmatics and stylistic registers
- 5000 hours of fine-tuning prompts for a specific use case
- feedback loops with expert users → continuous iteration
- evaluation capacity: "this is good, this is mediocre" (L3 calibration)

You had

- a tutorial: "How to call the OpenAI API" (level L1)
- a vague idea: "AI can write well"
- 50 hours of generic prompt testing
- almost zero ability to assess subtle quality
- you don't know what questions to ask for debugging

4.5 Empirical evidence - AI wrapper startups (2023-2025)

Aggregated data from multiple sources:

- Mohsin Akram analysis: 24 AI startups failed, $461.7 million evaporated
- CB Insights 2024: 966 startups closed (compared to 769 in 2023, 25.6% increase)
- Q1 2024: 254 insolvencies of venture capital-funded startups (60% increase compared to 2023, 7 times the rate in 2019)
- Failure rate: 90% of AI startups fail in the first year (compared to ~70% in traditional technology)

Dominant cause (43% of failures):

- "They built products that no one wanted" - copying success without understanding why it works
- unit economics: "AI wrappers" have gross margins of 50-60% *vs* classic SaaS 70-90%
- API costs: consume 15-30% of revenue
- 60-70% have zero revenue; only 3-5% exceed $10,000 in monthly recurring revenue (MRR)

IR pattern:

- see Jasper AI, Copy.ai → "$10 million in annual recurring revenue in 18 months!"
- I don't see: 10,000 hours of "prompt engineering", fine-tuned output quality assessment, domain expertise in marketing
- I think: "GPT API + interface = business"
- reality: commodity product without differentiation, crushed when large vendors launch similar native features

Examples:

1. **Cushion (2024):** $21.6 million raised, $82.4 million valuation, $3 million annual recurring revenue → “had difficulty scaling”, closed
2. **Builder.ai (2025):** valuation 1.5 billion USD → insolvency (detailed in chap. 2.5)
3. **CodeParent (YC 2023):** $500,000, multiple pivots, peak $1,500 monthly recurring revenue → closed July 2024
4. **Buildt.ai:** $250 million in funding, over 2 years → failure
5. **Low Light, Settle AI, 90.ai, Booth AI, The Gist:** YC/venture funded, all closed in 2024

PitchBook Analysis (Q3 2025)

- Over 1,200 "AI wrapper" startups founded in 2023-2024, 82% (1,023) closed within 2 years
- Dominant cause (71%): "inability to differentiate output", NOT lack of funding (only 18% failed strictly financially)

Identified pattern:

- founders without domain expertise (marketing, legal, programming, etc.)
- they thought: "API + interface = business"
- reality: "API + interface + zero domain expertise = generic commodity product"

Methodological clarification: The purpose of these examples is not to demonstrate a strictly deterministic causal relationship between AI use and organizational failure. They function as converging evidence of a recurring behavioral pattern. The repetition of the same mechanisms - overestimation, superficial copying, lack of competency calibration - in different contexts suggests the existence of a structural phenomenon, not a series of independent coincidences.

4.6 "Any idea seems simple after you read it"

The hindsight effect - the central epistemic problem:

Before seeing the AI-generated solution:

- the problem seems complex, with multiple ambiguities, possible directions
- "Where do I start? What approach? What considerations?"

After you see the output generated with AI:

- "Ah, of course! It's obvious. Logical. Natural."
- FCPT illusion: "I could have gotten here by myself."

Reality :

- without AI, it is unlikely that you would have arrived at that solution (or you would have arrived much later/or with lower quality)
- or in a completely different direction (possibly lower)

The major difference between "**understanding**" and "**generating**" - example of a mathematical demonstration:

AI generates :

Theorem: [statement]
Demonstration:
Step 1: We observe that X implies Y [obvious from the definition]
Step 2: We apply the Z lemma [non-trivial choice]
Step 3: It follows directly that... [conclusion]
QED

The user reads:

- Step 1: “Yes, obviously by definition” ✓
- Step 2: “Ah, lemma Z, naturally” ✓
- Step 3: “Logical, follow” ✓
- Conclusion: "Simple demonstration, I could do it too."

Reality :

- choosing lemma Z = a major insight (out of 50 possible lemmas)
- order of steps = non-trivial (other orders fail)
- "obvious" = only after you see the correct route

- finding the proof = 10 hours of intellectual struggle
- to understand the demonstration = 10 minutes of reading

FCPT activates: transfers the credit for the output to your own capacity → "I can reproduce this" → you try → you fail.

The paradox of learning with AI:

- the better solutions you see, the more deeply **you seem** to understand them
- **original generating** capacity does not increase proportionally
- risk: you confuse "recognizing a pattern" with "the ability to create it"

One of the central problems is therefore not only the fact that systems are becoming more capable, but also the tendency of people to give up too early the responsibility of **understanding their own limits and competencies.**

CHAPTER 5: COGNITIVE ATROPHY - THE DELAYED CASCADE EFFECT

5.1 Definition and distinction from FCPT

Cognitive atrophy can be defined as a **progressive diminution** of the capacity for deep, original and sustained thinking, as a delayed effect of dependence on AI and learning from failures generated by FCPT, CDC and RI.

The critical difference from FCPT: FCPT is an illusion of competence **in the present** (you believe you are capable now), and **Atrophy** means real loss of competence **in the future** (you become progressively incapable).

5.2 Causal mechanism: How cognitive atrophy sets in

Time flow:

Year 1: FCPT → "I am competent with AI"
Year 2: Failures (CDC + RI) → "AI is not good enough"
Year 3: Increased dependency → "Why try without AI?"
Year 5: Giving up on your own effort → "You'll do better anyway"
Year 10: Atrophy → significantly reduced capacity without AI (in certain types of tasks)

From the perspective of **Maslow7F metatheory (Stan, 2025),** FCPT represents a **Dominant Dissonant Chord (DDC)** located at the level of L4.S2 (Esteem Security) and L5.S1 (Physiology of Potential). The system (individual or organization) simulates reaching a higher level of maturity (L5), while the foundations of self-evaluation (L4) are negatively influenced by unchecked external processes.

Wrong learning after **failures repeated** generated by FCPT/RI:

A. **Correct learning:** "I overestimated; I should have assessed my skills/competences better"
B. **Mislearning (common):** "The problem is that the AI is not strong enough"

Mislearning → increased dependency → accelerated atrophy.

Concrete example - programming student:

Year 1:

- L1 student uses AI to code
- FCPT: "I'm good at programming"
- Generate functional code with AI

Year 2:

- Attempt complex project → fail (RI)
- Misguided learning: "I need stronger AI, not learning algorithms"

Years 3-4:

- Total dependency: any code = via AI
- Justification: "Why learn when AI does it more efficiently?"
- NO more trying to understand fundamental algorithms

Years 5-10: Measurable atrophy:

- The ability to read, correct, or modify complex code without AI is greatly diminished or even lost
- The ability to evaluate architectural trade-offs is at the lower limit or even zero
- The ability to learn new languages without AI is significantly reduced

Result? Capabilities that existed in Year 1 completely disappeared in Year 10.

5.3 The most serious problem - Atrophy at the L3 level ("Architects")

Why is it more critical at the expert level:

L3 Expert:

- Possesses advanced original thinking capabilities, built through thousands of hours of deliberate practice
- It may gradually reduce its cognitive involvement through dependence on AI-generated patterns
- Perceives AI-assisted output as “good enough”, which masks the gradual loss of evaluative finesse
- The main cost is the accumulated erosion of expertise developed over the years

Specific mechanism at L3

Before AI integration (≈2020):

- The expert generates several candidate solutions
- Evaluate each option in depth (subtle criteria, accumulated experience)
- Select the optimal solution based on calibrated intuition

In the first years of AI use (2024-2025):

- AI generates a large number of candidate solutions
- The expert evaluates and selects the optimal option
- The result is often superior to that obtained without AI
- The process represents a form of healthy augmentation

After 5+ years (2029-2030):

- AI generates most solutions
- The expert tends to choose the first good enough option
- The in-depth assessment process is practiced less frequently
- A gradual decrease in the ability to discriminate fine details is observed.

After 10+ years:

- The ability to generate original solutions without AI support is significantly reduced
- Deep assessment skills become more difficult to access without external support
- A form of functional dependence on assistive devices appears

Result:

A gradual transition from autonomous expertise to mediated competence, in which performance remains high, but cognitive autonomy progressively diminishes.

It is important to note that not all high-level failures are the expression of a cognitive error; the present analysis refers strictly to cases where the failure is correlated with overestimation of the capabilities mediated by AI tools.

5.4 Pathological addiction vs. healthy augmentation

Augmentation (healthy)

- AI generates proposals → the expert critically evaluates them → intervenes substantially
- The capacity for deep analysis is maintained and practiced
- AI works as a cognitive enhancer, not a substitute

Addiction (pathological)

- AI generates solutions → the user accepts them with minimal changes
- Cognitive effort is progressively reduced
- AI ends up functioning as a substitute for basic cognitive processes

Dependence vs. Augmentation: Indicators

Factor	Augmentation (Healthy)	Addiction (Pathological)
Quality feedback	Constant, critical, profound	Absent or superficial
Cognitive effort	Deliberately maintained challenges	Systematically minimized
AI output changes	SUBSTANTIAL (30%+)	MINIMUM (<10%)
Metacognition	"Where are the limits of AI?"	"AI knows better than me"
New challenges	MCS active; difficulty search	Avoid completely
Solo capacity	Regularly tested without AI	Never tested
Trajectory	L1 → L2 → L3 progressive	L1 → L1-dependent blocked

5.3 Learned cognitive helplessness

The concept of *learned helplessness* (Seligman, 1972) describes the situation in which an individual, after repeated exposure to failures perceived as being independent of his actions, ends up giving up any attempt at action, even when he would have the real ability to succeed.

Applied in a cognitive context, this dynamic takes on a specific form, amplified by interaction with artificial intelligence systems.

After a series of successive failures caused by False Cognitive Power Transfer (FCPT), the following mental pattern appears: "I can't do anything without AI anyway." / "There's no point in trying alone." / "Even if I tried, I wouldn't achieve an acceptable result."

Even in situations where the individual would have the necessary competence to succeed, he no longer initiates the action. This creates an evolutionary paradox:

- **Initial stage (Year 1):** overconfidence fueled by the apparent performance of the system: “I can do anything with AI.”
- **Later stage (Years 3-5):** collapse of confidence: “I can’t do anything without AI.”

Both conditions are dysfunctional.

In the first case, the individual overestimates their own capabilities, confusing amplification with competence. In the second case, the individual completely underestimates what they are still capable of doing without assistance.

One possible result is the loss of the correct calibration of one's own cognitive capabilities - an imbalance between what a person **can do,** what **they think they can do,** and what **they choose to try.** This degradation of self-assessment represents one of the most subtle and dangerous consequences of uncritical interaction with artificial intelligence systems.

Societal consequence: entire generations risk **NOT developing deep** cognitive capacities .

MCS proposes intervention to prevent this trajectory.

CHAPTER 6: MECHANISM OF COGNITIVE STIMULATION (MCS)

6.1 Fundamental principles: Useful friction

Mechanism of Cognitive Stimulation prevents cognitive atrophy by deliberately introducing **"useful friction"** - calculated delays and challenges that force the user to process information actively, not passively.

Central principle: When AI instantly answers complex questions, the user does not have to go through the cognitive journey required for deep learning. MCS introduces calibrated interventions that restore cognitive effort.

The design challenge: How to introduce friction without frustrating the user? The answer: dynamically adapting to the individual cognitive trajectory.

6.2 MCS 1.0: Basic Protocol

MCS 1.0 is the unidirectional protocol (AI → User) based on **Thinking Time (Tg)** :

Tg = Total response time - Context processing time

If **Tg** is **too low** (instant response to complex problem):

- System introduces **artificial latency,** OR
- Ask **a clarifying question**
- Purpose: Forces the user to process information

Adaptive thresholds:

- **threshold_L1 = 10 × Tg_base / μS** (beginner users)
- **threshold_L2 = 30 × Tg_base / μS** (intermediate users)

Where **μS** = measures the user's current cognitive needs.

MCS 1.0 levels:

- **Level 0:** Direct response (no stimulation)
- **Level 1:** Clarification ("Is the wedding in the evening or during the day?")
- **Level 2:** Synthesis ("What comes first: speed or security?")

6.3 MCS 2.0: Adaptive Pyramid Protocol

MCS 1.0 Extension: MCS 2.0 introduces **bidirectional feedback** (AI ⇄ User) through turn-to-turn cognitive trajectory analysis. Instead of applying the same protocol to everyone, MCS 2.0 detects whether the user is progressing (ascending), regressing (descending), exploring laterally, or stagnating - and dynamically adjusts the challenges.

6.3.1 Pyramid Score Calculation

For each **turn n** (where n ≥ 2), the trajectory score is calculated:

$$S\_trajectory(n) = w_1 * C(n) + w_2 * G(n) + w_3 * R(n)$$

Where:

- **C(n)** = Semantic Continuity (measures whether Turn(n) refers to concepts from the AI's response to Turn(n-1))
- **G(n)** = Complexity gradient (lexical comparison between Query_n and Query_{n-1})
- **R(n)** = Direct reference (detection of markers: "as you said", "from the previous example", etc.)
- **Standard weights:** $w1 = _{0.5,}\ w2 = _{0.3,}\ w3 = _{0.2}$
- **Range:** S_trajectory(n) ∈ [-1.0, +1.0]

6.3.2 Trajectory Classification

Based on the pyramid score, the type of trajectory is determined:

```
Trajectory(n) = {
ASCENDING       //if S_trajectory(n) > +0.3
DESCENDING      //if S_trajectory(n) < -0.3
LATERAL         //if |S_trajectory(n)| ≤ 0.3 AND topic_shift(n) = TRUE
STAGNANT        //if |S_trajectory(n)| ≤ 0.3 AND repeat_count(n) ≥ 2
}
```

Interpretation:

- **ASCENDANT:** User builds pyramidally on the previous answer (active learning)
- **DESCENDANT:** User abandons previous concepts (possible overwhelm)
- **LATERAL:** User explores new concepts (lateral exploration)
- **STAGNANT:** User repeats concepts without progress (blocking)

6.3.3 Dynamic Threshold Adjustment

MSC thresholds are modified by adaptation factors:

```
threshold_L1(n) = threshold_L1_base × α(n)
threshold_L2(n) = threshold_L2_base × α(n)
```

where the basic thresholds (from MSC 1.0):

```
threshold_L1_base = 10 × Tg_base / μS
threshold_L2_base = 30 × Tg_base / μS
```

Adaptation factor:

```
α(n) = {
0.7     //if Trajectory(n) = ASCENDENT (earlier activation)
1.8     //if Trajectory(n) = DESCENDING (frequency reduction)
1.0     //if Trajectory(n) = LATERAL (neutral)
0.5     //if Trajectory(n) = STAGNANT (force break pattern)
}
```

Logic:

- Ascendant user → challenges earlier (threshold decreases → MCS activates more easily)
- Descending user → fewer challenges (threshold increases → MCS is harder to activate)

6.3.4 Extended Levels

MSC 2.0 introduces two additional sub-levels for Level 2:

MSC_Level(n) = {
0 //IF Tg(n) < threshold_L1(n)
1 //IF threshold_L1(n) ≤ Tg(n) < threshold_L2(n)
2-lite //IF threshold_L2(n) ≤ Tg(n) AND Trajectory(n) = DESCENDING
2 //IF threshold_L2(n) ≤ Tg(n) AND Trajectory(n) ∈ {LATERAL, STAGNANT}
2-challenge //IF threshold_L2(n) ≤ Tg(n) AND Trajectory(n) = ASCENDING
}
MSC levels:

- **Level 0:** Direct response (no stimulation)
- **Level 1:** Clarification ("Is the wedding in the evening or during the day?")
- **Level 2-lite:** Gentle guidance ("Do you prefer option A (simple) or B (detailed)?")
- **Level 2:** Standard Synthesis ("What comes first: speed or security?")
- **Level 2-challenge:** Advanced challenge ("Good. Now consider second-order effects.")

6.3.5 Frustration Detector (Safety Override)

Regardless of the calculated scores, the system detects signs of frustration:
Frustration(n) = TRUE IF:
Query_n contains {"don't understand", "too complicated", "simpler"}
OR
abandon_time(n) > 2 × median_response_time OR repeat_count(n) > 3
IF Frustration(n) = TRUE :
MSC_Level(n) = 0 (complete override, simplified direct response)
α(n) = 2.0 (doubles the thresholds for the next 3 turns)

6.4 Complete MCS 2.0 equation

Combining all the components:
MSC_Decision(n) = f(Tg(n), μS, S_trajectory(n), Frustration(n))
Where :
Tg(n) = TokensOutput / ProcessingSpeed × SemanticComplexityFactor
S_trajectory(n) = 0.5·C(n) + 0.3·G(n) + 0.2·R(n)
threshold_L1(n) = (10 × Tg_base / μS) × α(Trajectory(n))
threshold_L2(n) = (30 × Tg_base / μS) × α(Trajectory(n))

Key differences MCS 1.0 vs 2.0:

- **MCS 1.0:** Uniform protocol (variation only by μS)
- **MCS 2.0:** Dynamic learning from conversation trajectory, adjusting thresholds in **real time,** preventing overwhelming downstream users and maximizing challenge for upstream users

6.5 Implementation and limitations

Implementation: MCS 2.0 requires:

- Turn-to-turn tracking of conversations
- Semantic analysis for C(n), G(n), R(n)
- Trajectory storage for adaptation α(n)
- Implementable in modern AI systems

Limitations:

A. Voluntary use:

- Users can disable MCS if it is optional
- Requires "opt-out" design instead of "opt-in"

B. Trade-off experience:

- Useful friction ≠ pleasant short-term experience
- Risk: users migrate to AI without MCS (faster, more enjoyable)

C. Difficult calibration:

- Thresholds too high → frustration
- Thresholds too low → does not prevent atrophy
- Requires extensive tuning per domain

D. Systemic problem:

- MCS 2.0 only works if **most** AI systems implement it
- Otherwise: users choose frictionless AI → continuous atrophy

Necessary solutions:

- Industry standards for MCS
- Regulations
- User education about long-term benefits

Relational Emergence: Mathematically, FCPT can be defined as a state of **False Resonance.** The subject confuses phase imitation (output takeover) with a structural synchronization of the hybrid system. The conditions of authentic emergence and the formalization of the human-AI resonance field are treated separately in the paper: *" Mathematical Framework for Relational AI Emergence"* (Stan, in preparation).

CHAPTER 7: IMPLICATIONS AND FUTURE DIRECTIONS

7.1 For the design of responsible AI systems

Anti-FCPT design principles

A. Transparency in the allocation of contributions

- Systems must explicitly indicate what part of the result is automatically generated and what part comes from human intervention.
- The interface must clearly distinguish between algorithmic input and user decisions.
- The goal is to prevent confusion regarding the source of competence and avoid the erroneous transfer of cognitive merit.

B. Explicit levels of difficulty and autonomy

- Commands like: "show only the structure, not the complete solution";
- "It provides gradual clues, not the final result";
- The user explicitly selects the desired level of cognitive support.

C. Assisted metacognition

- After each major interaction, the system prompts reflection: "Could you have reached this result without assistance?"
- Periodic feedback like: "Compared to six months ago, your level of cognitive autonomy has increased/decreased."

D. Integration of MCS (Metacognitive Control System)

- Active by default for L1-L2 users.

- Possibility of deactivation only with explicit warning regarding the risk of cognitive atrophy.
- "Cognitive health" indicator visible in the interface (tendency towards autonomy vs. dependence).

7.2 For education: recalibrating the curriculum

Paradigm shift
From: "Learn to do X" (mechanical execution)
At:

- ✓ "Learn to evaluate the quality of X"
- ✓ "Learn to distinguish between real understanding and superficial recognition"
- ✓ "Learn the limits of your own AI-assisted competence"

Concrete measures

A. Dual assessments

- 50% assessment with access to AI (measures orchestration capability)
- 50% assessment without AI (measures fundamental competence)

B. Anti-FCPT exercises

- Weekly: solving without AI assistance
- Monthly: complete reconstruction of a solution from memory
- Periodic: comparison between assisted vs. unassisted outcome

C. Explicit education about cognitive mechanisms

- Modules on false transfer of competence
- Analysis of real cases
- Exercises to identify attribution errors

D. AI-free zones

- Days dedicated to work without algorithmic assistance
- Goal: maintain core cognitive muscle

7.3 For organizations: systemic prevention policies

A. Competency audit before major commitments

- Explicit verification: can the team function without AI?
- Analysis: which skills are real, which are artificially amplified?

B. Structured post-mortem

Post-failure analysis including:

- whether there was a false transfer of competence;
- whether decisions were made under the illusion of AI performance;
- whether there has been an overestimation of internal capacity.

C. Sprints without AI

- Short periods where the team works without AI assistance.
- Purpose: diagnosing addictions and recalibrating skills.

D. Atrophy prevention programs

- Workshops to rebuild fundamental skills.
- Periodic assessments of independent reasoning ability.

7.4 Future research directions

A. Longitudinal measurement

- Development of metrics for detecting long-term cognitive atrophy.
- Longitudinal studies on the effect of AI use on problem-solving abilities.

B. Adaptive interventions

- Identifying patterns of users susceptible to cognitive addiction.
- Adaptive systems that adjust the level of assistance based on actual performance.

C. Preventive AI architectures

- Systems capable of automatically detecting over-dependence.
- Integrated cognitive braking mechanisms (deliberate friction).

D. Social and political dynamics

- Studies on the mass effects of cognitive delegation.
- Predictive models for the emergence of collective demoralization phenomena.
- Public policies for maintaining distributed cognitive competence.

E. Extensive empirical validation

- Longitudinal studies (3-10 years) on mixed cohorts.
- Comparative measurements between groups with different levels of AI assistance.
- Defining standard cognitive health indicators.

CHAPTER 8. LIMITATIONS, OBJECTIONS AND AREAS OF CONCEPTUAL TENSION

Any theoretical framework that claims to describe emerging phenomena at the intersection of technology, cognition, and society must explicitly acknowledge its limitations. The False Cognitive Power Transfer (FCPT) model is no exception. The model is falsifiable by demonstrating cases in which low-competence users consistently achieve superior performance without cognitive degradation, in the absence of any compensatory mechanism. This model does not say what we should do, but what happens if we are not careful.

This section does not seek to weaken the hypothesis, but to **stabilize it epistemically,** clearly delimiting its scope of applicability and the risks of excessive interpretation.

8.1 The epistemic limit: what FCPT does NOT explain

FCPT is not a complete theory of human behavior in the presence of technology. It does not explain:

- all forms of cognitive error;
- all types of organizational failure;
- nor individual motivational dynamics in the clinical psychological sense.

In particular, FCPT **does not substitute** theories such as:

- motivation theory (Deci & Ryan),
- classical learning models (Bandura),
- economic theories of decision under uncertainty.

He describes **a specific mechanism:** the erroneous transfer of perceived competence from the system to the user, in a context of technological amplification.

8.2 Scope limit: where it does NOT apply

FCPT does not manifest itself uniformly in all cognitive activities. There are areas where the effect is weak or irrelevant:

- strictly procedural tasks, with well-defined rules;

- activities with immediate and binary feedback (right/wrong);
- contexts where the user is structurally required to maintain competence (e.g. aviation, interventional medicine).

In these cases, control and feedback mechanisms limit the formation of the illusion of competence.

By contrast, FCPT appears with maximum intensity in:

- creative or strategic activities;
- delayed feedback systems;
- areas where quality assessment is ambiguous or subjective.

8.3 Major Objection: "Isn't this just Dunning-Kruger by another name?"

This objection is legitimate and must be addressed explicitly.

The fundamental difference is structural: **Dunning-Kruger** describes an internal self-assessment error, arising from a lack of competence, and **FCPT** describes an external attribution error, induced by interaction with a technological system that amplifies performance.

In Dunning-Kruger, the error disappears with learning. In FCPT, the error may **increase** with assisted experience because the tool masks the lack of competence.

This difference explains why experienced individuals can become vulnerable to FCPT, while Dunning-Kruger mainly affects novices.

8.4 Empirical limits and risks of over-generalization

It is important to emphasize that:

- not all AI failures are explainable by FCPT;
- not all AI-based successes involve attribution error;
- there are contexts where AI actually reduces cognitive errors.

The major risk is **generalization:** applying the FCPT framework as a universal explanation for any technological failure. The model should be used as an analytical tool, not as a universal explanation.

8.5 Positioning in relation to existing literature

Conceptually, this framework lies at the intersection of four traditions:

1. **Cognitive psychology** - extends theories about error and metacognition in a technologically augmented context.
2. **Studies on human-machine interaction (HCI)** - add an epistemic (not just ergonomic) dimension to the human-system relationship.
3. **Sociology of technology** - explains emergent effects at the collective level (CDC) starting from individual behaviors.
4. **Complex systems theory** - treats error as an emergent phenomenon, not as a point defect.

In this sense, the work does not directly compete with existing literature, but functions as a **conceptual bridge** between fields that, until now, have analyzed the same phenomena in a fragmented manner.

8.6 Limitations and conditions of applicability

The proposed conceptual framework presents a series of limitations that must be explicitly recognized to avoid excessive interpretations.

First, the model **does not claim** that the use of artificial intelligence **inevitably leads** to cognitive degradation or organizational failure. The effects described **only occur under certain conditions,** in particular when the following criteria are simultaneously met:

- significant discrepancy between the actual level of competence and the perceived level;
- lack of corrective feedback mechanisms;
- institutional pressure for rapid performance;
- the absence of deliberate practices of reflection and cognitive calibration.

Second, the framework is not intended to explain all forms of failure associated with AI technologies. Phenomena such as statistical modeling errors, data bias, or economic constraints are not exhaustively addressed here, although they may interact with the mechanisms described.

Third, the proposed causal relationships should be understood as hypotheses open to empirical validation. The present framework provides a conceptual language and a set of testable predictions, not a definitive demonstration.

This delineation is essential to prevent the paper from being misinterpreted as a general critique of artificial intelligence or as a deterministic prediction regarding human cognitive degradation.

8.6 Critical conclusion

The essence of this work does not lie in offering a technological solution, but in formulating a **conceptual diagnosis** of how people relate to increasingly powerful cognitive systems.

If there is a real risk in the current evolution, one of the most plausible is that as machines become more capable, humans may become less attentive to their own limits and the true nature of competence, with profound implications for both the individual and human society.

This paper does not claim that all forms of AI use lead to cognitive degradation, nor that all cases of failure can be explained by FCPT. The model described is applicable in contexts characterized by: (a) large asymmetry between the power of the tool and the level of expertise, (b) pressure for rapid performance, (c) lack of qualified feedback mechanisms.

CHAPTER 9: CONCLUSIONS

Generative artificial intelligence creates a fundamental paradox: while amplifying the capabilities of experts, it simultaneously generates dangerous illusions of competence in inexperienced users and risks progressive cognitive atrophy even in experts.

Contributions of this article:

1. Defining FCPT as a psychological mechanism **distinct** from Dunning-Kruger, through which attribution confusion between tool efficiency and one's own competence generates systematic over-commitment to tasks of inappropriate complexity.

2. Identification of two patterns of collective failure post-FCPT:

- **CDC:** Negative learning from observed failures → collective preemptive abandonment
- **RI:** Distorted positive learning from successes → ignoring expertise → failure to replicate

3. Explaining cognitive atrophy as a delayed cascade effect, arguing that degradation can affect not only beginners but also L3 experts, through progressive dependence on AI patterns and the erosion of original deep-thinking capacities.
4. MCS 2.0 / 3.0 proposal as an adaptive pyramidal protocol of counteraction through useful friction dynamically calibrated to the individual cognitive trajectory.
5. Presentation of empirical evidence from multiple fields (Cold Fusion, Gene Therapy, AI Winter) suggesting that CDC-type patterns are observable and may have recurring historical analogies.

The contribution of this framework **does not consist in identifying new phenomena,** but **in articulating the causal relationships** between them, in a coherent sequence that allows the anticipation of systemic effects that are not explained by existing models.

The choice is now. We are in the early years of the AI era. The design decisions, educational policies, and social norms established over the next 3-5 years will determine which trajectory we follow.

Cognitive atrophy does not require malignant intent. It only requires inaction.

Rogo, ergo emergo.
(I question, therefore I become.)

BIBLIOGRAPHY

Author

Stan, A. (2025). *THE AGE OF COGNITIVE DIVERGENCE: Surviving the split between human rhythms and artificial speed.* DOI: 10.5281/zenodo.17965800

Stan, A. (2025). *Thermodynamics of cognitive power: When {1=1} breaks the fortress [Academic version].* DOI: 10.5281/zenodo.17876438

AB

American Society of Gene & Cell Therapy. (2017). *Gene therapy's road to redemption.* Pediatrics Nationwide.

Association for Computing Machinery. (1978). *SIGART membership statistics 1969-1978. ACM SIGART* Bulletin, Issue 67.

Bandura, A. (1977). *Social learning theory.* Prentice Hall.

Biberian, JP (2007). *Condensed matter nuclear science (cold fusion): An update.* International Journal of Nuclear Energy Science and Technology, 3(1), 1-9.

Bikhchandani, S., Hirshleifer, D., & Welch, I. (1992). *A theory of fads, fashion, custom, and cultural change as informational cascades.* Journal of Political Economy, 100(5), 992-1026.

Brockman, J. (2017). *Possible minds: Twenty-five ways of looking at AI.* Penguin Press.

C

CB Insights. (2025). *The state of venture: Q4 2025 report.* CB Insights Research. [Industry research report]

Crevier, D. (1993). *AI: The tumultuous history of the search for artificial intelligence.* Basic Books.

Crunchbase. (2024). *AI startup funding and failure data.* Crunchbase Inc.

D

Defense Advanced Research Projects Agency. (1988). *Strategic Computing Program: Annual report.* DARPA Information Science and Technology Office.

Dunbar, RIM (1998). The social brain hypothesis. *Evolutionary Anthropology, 6* (5), 178-190.

EG

Gudigantala, N., & Mehrotra, V. (2024). When strength turns into weakness: Exploring the role of AI in the closing of Zillow Offers. *Journal of Information Systems Education, 35* (1), 67-72.

H

Haigh, T. (2023). There was no "first AI winter". *Communications of the ACM, 66* (12), 36-39.

KL

Krige, J., & Pestre, D. (Eds.). (2014). *Companion to science in the twentieth century.* Routledge.

Lawrence Berkeley National Laboratory. (2023). *Quantifying nuclear reactions in metal hydrides at low energies.* US Department of Energy.

Lighthill, J. (1973). *Artificial intelligence: A general survey.* Science Research Council.

M

Massachusetts Institute of Technology. (2023). *Neutron emission from laser-stimulated metal hydrides.* US Department of Energy Technical Report.

McCorduck, P. (2004). *Machines who think: A personal inquiry into the history and prospects of artificial intelligence.* AK Peters/CRC Press.

N

National Institutes of Health. (2000). Gene therapy clinical trials: Reporting requirements and safety monitoring. *Federal Register, 65* (198).

National Institutes of Health. (2018). The next phase of human gene therapy oversight. *New England Journal of Medicine, 378* (5), 439-448.
Nilsson, NJ (2009). *The quest for artificial intelligence: A history of ideas and achievements.* Cambridge University Press.
NYU Langone Health. *Gene therapy research & the case of Jesse Gelsinger.* High School Bioethics Project.

A
Orkin, SH, & Motulsky, AG (1995). *Report and recommendations of the panel to assess the NIH investment in research on gene therapy.* National Institutes of Health.

P
PitchBook. (2025). *2025 US Venture Capital Outlook.* PitchBook Data, Inc.

R
Rinde, M. (2019). The death of Jesse Gelsinger, 20 years later. *Science History Institute Distillations* .
Rock Health. (2023). *Q3 2023 digital health funding report.* Rock Health Research.
Russell, S., & Norvig, P. (2020). *Artificial intelligence: A modern approach* (4th ed.). Pearson.

S
Savin, K. (2018). Gene therapy's second act. *Science, 359* (6373), 28-31.
Schoemaker, PJH (2004). Forecasting and scenario planning: The challenges of uncertainty and complexity. In DJ Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp. 274-296). Blackwell Publishing.
Science History Institute. (2019). The death of Jesse Gelsinger, 20 years later. *Distillations Magazine* .
Sparrow, B., Liu, J., & Wegner, DM (2011). *Google effects on memory: Cognitive consequences of having information at our fingertips.* Science, 333(6043), 776-778.
Stanford University. (2023). *LENR-active nanoparticles under phonon stimulation.* U.S. Dept. of Energy.
Susarla, P., Purnell, D., & Scott, K. (2025). *Zillow's artificial intelligence failure and its impact on perceived trust in information systems.*

T
The Wall Street Journal. (2019). *Humans were behind the AI at Engineer.*

U
US Department of Energy. (2023). *The US Department of Energy announces $10 million in funding to projects studying low-energy nuclear reactions* .
US Food and Drug Administration. (2000, January 19). *FDA halts gene therapy trials at University of Pennsylvania* .
University of British Columbia. (2024). *Electrochemical loading enhances deuterium fusion rates in a metal target.* Nature, 632, 321-325.

V
Verma, IM (2000). *A tumultuous year for gene therapy. Molecular Therapy, 2* (5), 415-416.

W
Wilson, JM (2009). *Lessons learned from the gene therapy trial for ornithine transcarbamylase deficiency.* Molecular Genetics and Metabolism, 96(4), 151-157.

Z
Zillow Group, Inc. (2021). *Zillow Group reports third quarter 2021 results* .
Zillow Group, Inc. (2023). *Annual report (Form 10-K).* US Securities and Exchange Commission.