NEPTUNE

WHITEPAPER

Non-Delegable Cognition™

Why the order in which people learn to think with AI decides whether AI makes them more capable, or merely more dependent.

July 2026 · Version 1.0

Contents

Executive summary

Artificial intelligence is absorbing everything that can be delegated: drafting, calculating, summarising, coding and, increasingly, deciding. As the engines grow more capable, the useful question changes. It no longer asks what the machine can do. It asks which part of human cognition must never be handed over, however powerful the engine becomes.

Five independent research traditions converge on the answer: neuroscience, cognitive science, philosophy of technology, law and European regulation. Each reaches, by its own method, the same boundary between what can be delegated and what cannot.

NEPTUNE names the capacity at stake Non-Delegable Cognition™: judgement, intention, responsibility for the outcome, and the ability to think when the tool is absent. The evidence further suggests that the order of acquisition decides the outcome. People who build the capacity before delegating become the most sophisticated users of AI. People who delegate first may never build it.

This holds at three scales: the citizen, the institution, the country. Preserving the capacity is the working condition of Human in Control, the oversight category that European law now requires. NEPTUNE works on training that order. **Think first. Delegate after.**

How to read the evidence in this paper

Every substantive claim in a NEPTUNE whitepaper carries its evidence class. This is the house standard, not decoration: it tells you exactly how much weight each claim can bear.

> **FACT**
>
> Verified against a citable source, listed in the bibliography. The claim holds independently of NEPTUNE.

> **INFERENCE**
>
> Logically derived from facts; the chain of reasoning is shown. Strong, but one step removed from direct evidence.

NOT VERIFIED

Plausible and stated transparently as unconfirmed. We tell you where it could be verified. We never dress these as facts.

1. The delegation frontier

Every wave of technology has taken over tasks: the calculator took arithmetic, the map application took navigation. Philosophy has long argued that a tool coupled tightly enough to a person becomes part of their cognitive system [4]. Cognitive science calls the healthy version cognitive offloading: the strategic handover of a narrow task while the person keeps control of the reasoning [3]. Generative AI makes a different move available. For the first time, the reasoning itself can be handed over, in full, at zero effort, in fluent language that sounds right.

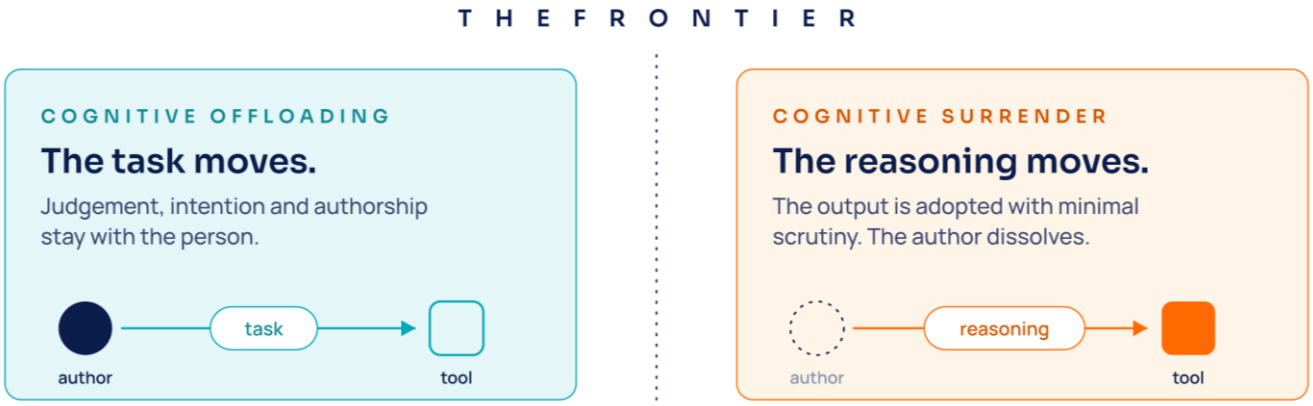

Figure 1 · The delegation frontier. On one side the task moves and the author stays; on the other the reasoning moves and the author dissolves.

FACT

A 2026 Wharton School working paper (Shaw and Nave) ran three preregistered experiments with 1,372 participants across 9,593 reasoning trials, with AI accuracy secretly randomised. Participants consulted the AI on most trials and adopted its answers with minimal scrutiny: accuracy rose about 25 percentage points when the AI was right and fell about 15 when it was wrong. The authors name the pattern cognitive surrender. [5]

> **INFERENCE**
>
> The frontier of safe delegation runs between offloading and surrender. Offloading hands over a task and keeps the author. Surrender hands over the reasoning and, with it, control of the result. Generative AI is the first technology that makes crossing that frontier effortless.

What this section asks the reader to accept: delegation has a frontier, and the cost of crossing it is now measurable.

2. The definition

Non-Delegable Cognition™ (n.): the set of mental capacities a person or organisation may exercise with the support of artificial intelligence but can never transfer to it without losing what defines them: judgement, intention, responsibility for the outcome, and the ability to think when the tool is absent. Delegating the task is legitimate; delegating non-delegable cognition dissolves the author.

2.1 Three lenses, one boundary

Through the cognitive lens, non-delegable cognition is the layer of thought that forms only through direct exercise: the criterion built by reasoning for oneself, the memory of the path taken, the ability to recognise when a machine-generated answer is wrong. Like muscle, it atrophies when left unused, even while the substitute stays permanently available.

Through the legal lens, the name borrows from a long-standing doctrine under which certain duties may be executed by others while responsibility remains, in full, with the principal. Applied to the mind, the doctrine reads plainly: the work can move to the machine; the answerability cannot.

Through the governance lens, non-delegable cognition is the precondition of oversight. An organisation or a country that transfers its decision-making to automated systems loses the capacity to supervise them, to correct them in time and to answer for their consequences. Without the capacity, human oversight exists only on paper.

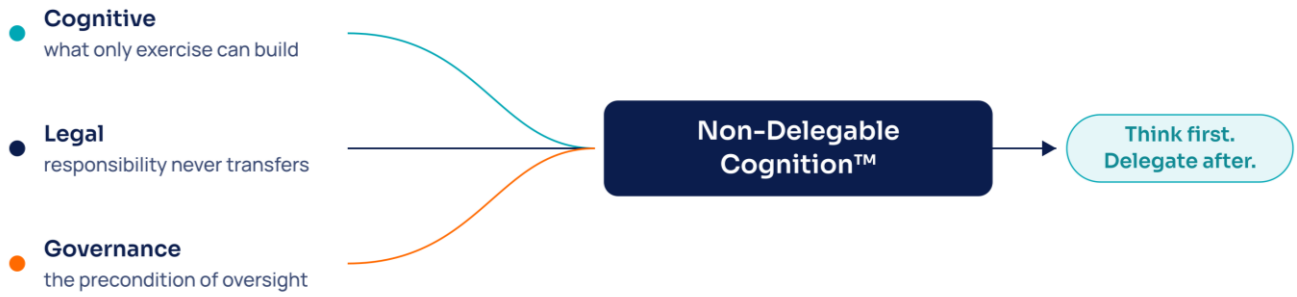

Figure 2 · Three lenses, one boundary. Each discipline reaches the same concept, and the concept ends in an instruction about order.

The three lenses end at the same instruction, and it is an instruction about order. **Think first. Delegate after.**

3. The evidence for order

FACT

A 2025 MIT Media Lab study (Kosmyna et al.) measured brain activity by EEG in 54 participants writing essays over four months with a language model, a search engine, or no tool. The language model group showed the weakest and least distributed neural connectivity of the three, reported the lowest sense of ownership over their texts, and struggled to quote from essays they had just written. [1] The study is a preprint with a small sample, and a formal comment (Stanković et al., 2026) argues that several results should be interpreted more conservatively; this paper cites both together, as a matter of method. [2]

INFERENCE

The most useful finding for organisations concerns order. Participants who wrote without the tool first, and only then gained access to it, engaged more strongly and used the

model in more sophisticated ways. Dependence does not follow from using AI; it follows from using AI before knowing how to think without it.

The strength of the argument rests on convergence, never on a single study. Neuroscience, cognitive science, philosophy and law each reach the frontier by a different route. When independent disciplines with different methods point at the same boundary, the boundary deserves a name and a practice. What this section asks the reader to accept: the order in which a person acquires thinking and tooling shapes what the tooling does to the thinking.

4. Human in Control, from clause to capacity

Human in Control describes the arrangement in which humans keep ultimate authority, responsibility and the power to intervene over automated systems. The category comes from the European ethics guidelines of 2019; it belongs to the field, and this paper adopts it as established vocabulary [8]. Philosophical work on meaningful human control adds two conditions: the system must track the moral reasons of the relevant humans, and its behaviour must trace back to someone who genuinely understands it. A button to press is insufficient; a mind that understands what the button interrupts is required [6].

FACT

Article 14 of the EU Artificial Intelligence Act (Regulation 2024/1689) requires high-risk AI systems to be designed so that natural persons can effectively oversee them: understand the system, monitor it, remain aware of the tendency to over-rely on its output (automation bias), decide to disregard it, intervene, or halt it. [7]

INFERENCE

Every oversight clause assumes an overseer whose judgement still works. A supervisor who has surrendered the underlying cognition can hold the authority to override and lack the capacity to know when. Non-delegable cognition is the assumption behind Human in Control, made explicit and made trainable.

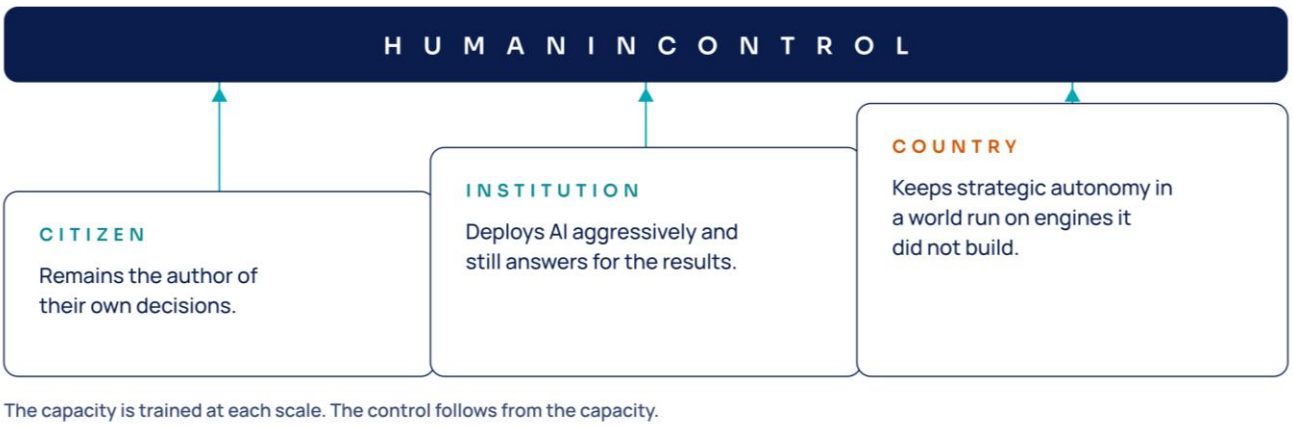

Figure 3 · Human in Control across three scales. The bar is held up by the capacity trained underneath it, at every level.

The stakes scale. A citizen who preserves the capacity remains the author of their own decisions. An institution that preserves it in its people can deploy AI aggressively and still answer for the results. A country that preserves it across citizens, institutions and law keeps strategic autonomy in a world run on engines it did not build. A country that surrenders it keeps the territory and loses the judgement. What this section asks the reader to accept: Human in Control is a capacity before it is a compliance clause, and the capacity has to be trained.

5. What makes NEPTUNE diferent

NEPTUNE works to build the right order with people and organisations: diagnosis of cognitive dependency, a brain-first protocol for acquiring skills before delegating them, critical evaluation of machine outputs, team-level delegation policies, and an institutional plan for preserving competence as AI adoption deepens. The concept is the foundation of the practice, and the practice is measurable: what a person or an organisation can still do when the tool is absent.

Non-Delegable Cognition™ · *Think first. Delegate after.*

Bibliography

[1] Kosmyna, N. et al. (2025). Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task. MIT Media Lab. Preprint, arXiv:2506.08872.

[2] Stanković et al. (2026). Comment on: Your Brain on ChatGPT: Accumulation of Cognitive Debt When Using an AI Assistant for Essay Writing Tasks. Preprint, arXiv:2601.00856.

[3] Risko, E. F. and Gilbert, S. J. (2016). Cognitive Offloading. Trends in Cognitive Sciences, 20(9), 676-688.

[4] Clark, A. and Chalmers, D. (1998). The Extended Mind. Analysis, 58(1), 7-19.

[5] Shaw, S. D. and Nave, G. (2026). Thinking Fast, Slow, and Artificial: How AI is Reshaping Human Reasoning and the Rise of Cognitive Surrender. Working paper, The Wharton School, University of Pennsylvania. SSRN 6097646.

[6] Santoni de Sio, F. and van den Hoven, J. (2018). Meaningful Human Control over Autonomous Systems: A Philosophical Account. Frontiers in Robotics and AI, 5:15.

[7] Regulation (EU) 2024/1689 of the European Parliament and of the Council (Artificial Intelligence Act), Article 14: Human oversight. Official Journal of the European Union, July 2024.

[8] High-Level Expert Group on Artificial Intelligence (2019). Ethics Guidelines for Trustworthy AI. European Commission.