SANRA

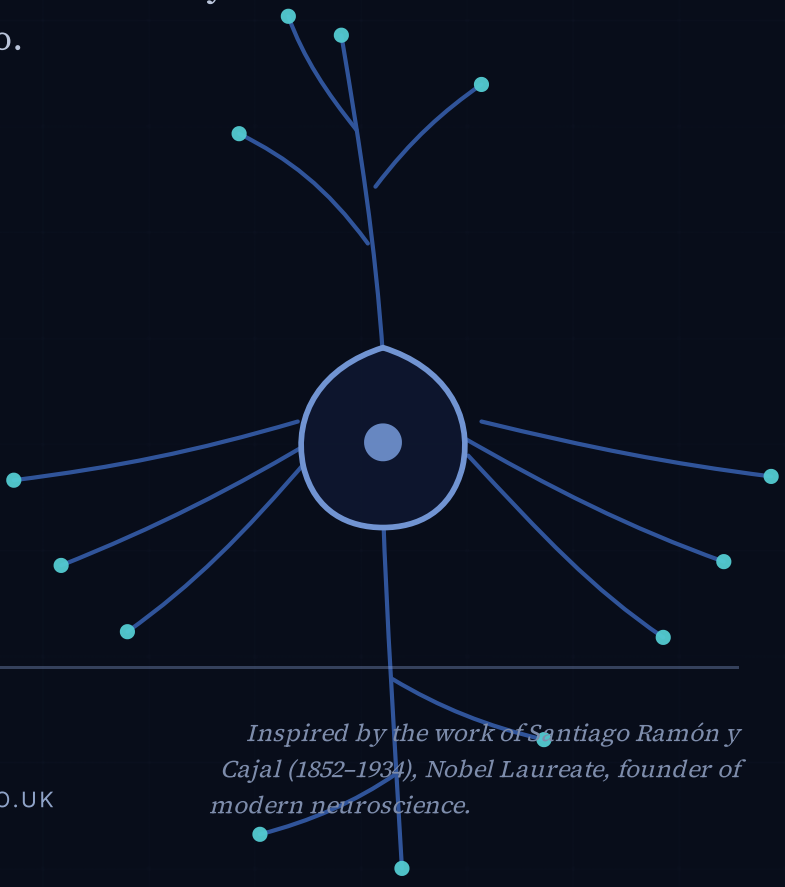
HUMAN FACTORS FOR THE AI AGE

WHITEPAPER WP-01
THOUGHT LEADERSHIP SERIES
JULY 2026

AI-BASED SECURITY

The Neuron Doctrine of Security

Why the founding father of neuroscience holds the blueprint for defending the AI age — and why the industry's perimeter model was built on a theory the brain itself disproved a century ago.



Inspired by the work of Santiago Ramón y Cajal (1852–1934), Nobel Laureate, founder of modern neuroscience.

SANRA LTD · UNITED KINGDOM
WWW.SANRA.CO.UK · RESEARCH@SANRA.CO.UK

ABOUT THIS PAPER

This is a thought-leadership paper, not a product specification. It argues a single thesis: that the most sophisticated security system ever evolved — the human brain — defends itself without anything resembling a perimeter, and that the principles Santiago Ramón y Cajal established a century ago describe, with uncanny precision, the architecture that AI-based security must now adopt. Everything that follows is a working-out of that claim.

ABOUT SANRA

SanRa is a human factors company for the AI age. Our name honours Santiago Ramón y Cajal — the Spanish histologist who first saw the nervous system clearly, and who taught science that seeing clearly is a discipline, not a gift. We build training, doctrine and tooling for the people that automation depends on: the analysts, engineers and leaders who must think with machines without ceasing to think for themselves.

A NOTE ON METHOD

Neuroscience is used here as a source of architectural analogy, not as mysticism. Where we claim a fact about the brain, it is a settled fact of the literature and is referenced. Where we map that fact onto security practice, the mapping is ours, and we argue for it on its merits. Analogy is a compass, not a proof — but some compasses point true.

DOCUMENT DETAILS

Whitepaper WP-01 · Thought Leadership Series · Version 1.0 · July 2026
© 2026 SanRa Ltd. This paper may be shared freely in unmodified form.
www.sanra.co.uk · research@sanra.co.uk

Contents

Executive Summary

01 The Reticular Fallacy

02 What Cajal Actually Proved

03 The Neuron Doctrine of Security

04 The Third Element: Security's Hidden Immune System

05 The Predictive Brain: Defence as Inference

06 The Human Neuron: Keeping the Analyst in the Circuit

07 The SanRa Architecture

08 Ten Theses for AI-Based Security

09 From Doctrine to Practice: The First Ninety Days

References & Further Reading

Executive Summary

Security has spent thirty years building walls around networks, and the walls have lost. Perimeters dissolve on contact with cloud, remote work, supply chains and now AI agents that act on our behalf. Yet the industry's reflex remains unchanged: another wall, another gate, another checkpoint. We keep rebuilding a fortress in a world that has stopped having edges.

A century ago, neuroscience went through the same error — and corrected it. The prevailing theory of the nervous system, the *reticular theory* championed by Camillo Golgi, held that the brain was one continuous, fused web: a network without internal boundaries. Santiago Ramón y Cajal disproved it. Working alone with a microscope and a staining technique he did not invent but perfected, Cajal showed that the nervous system is made of **discrete, individual cells** — neurons — that never fuse, and that all communication between them crosses controlled junctions: synapses. The brain is not a web. It is a population.[1][2]

That single correction — from continuous network to discrete, self-verifying units — is the correction security must now make. This paper develops it into a working doctrine for AI-based defence:

- **Section 01–02** trace the parallel between Golgi's reticulum and the security perimeter, and distil Cajal's findings into four architectural principles.
- **Section 03** maps those principles onto a modern security architecture: discrete verifiable units, the synapse as the universal control point, directed signalling, and plasticity.
- **Section 04** tells the story security has never heard: how Cajal's school discovered microglia — the brain's resident immune cells — and why your most important defenders are the ones you cannot see.
- **Section 05** argues that defence must become predictive inference, and that prediction without verification is hallucination — in brains and in models alike.
- **Section 06** locates the human in the circuit: the analyst as cortical neuron, and the design of dissent into machine-speed defence.
- **Sections 07–09** make it concrete: the five-layer SanRa Architecture, ten theses, and a ninety-day path from doctrine to practice.

THE CENTRAL CLAIM

The brain has no firewall. It defends three pounds of the most valuable tissue in the known universe with discrete cells, verified junctions, resident sentinels and a memory that learns. So can we.

01 The Reticular Fallacy

Every era of security inherits a theory of the network. Ours inherited the wrong one.

In 1888, the dominant model of the nervous system was beautiful, intuitive and wrong. The reticular theory, developed by Joseph von Gerlach and championed by Camillo Golgi — the finest anatomist of his generation — held that nerve cells fused into a single continuous mesh: one reticulum, one unbroken network, with signals flowing freely through it like water through a sponge. It was a comforting picture. The brain as one thing. Whole. Unified. Borderless inside.[1]

Golgi was no amateur. He invented the staining method — *la reazione nera*, the black reaction — that made individual nerve cells visible at all, and he shared the 1906 Nobel Prize for it. His theory was wrong anyway. The evidence was assembled, slide by slide, by a provincial Spanish doctor working largely alone: Santiago Ramón y Cajal. Using Golgi's own stain, refined to an art, Cajal showed that neurons never fuse. Each is a discrete cell, complete in itself, separated from its neighbours by a gap. Communication happens across that gap — later named the synapse — and it is controlled, directional and selective. The nervous system is not a mesh. It is a society of cells that touch but never merge.[2]

The two men shared the Nobel Prize in 1906 and used their lectures to contradict each other. History sided with Cajal.

The perimeter is our reticulum

Security built its Golgi-era model and called it the perimeter. Inside the firewall: the trusted, continuous network — flat, fused, unsegmented, where any machine can reach any machine and a credential opens every door. Outside: the untrusted world. The theory was beautiful, intuitive and wrong in exactly the same way as Golgi's. It mistook the network for one continuous thing, when the truth — in brains and in enterprises — is that **nothing inside is continuous, and nothing inside should be trusted by default.**

The evidence has been mounting for two decades, breach report by breach report. The flat internal network is the attacker's greatest ally: compromise one cell, and the reticulum carries you everywhere. Ransomware does not breach perimeters so much as inherit them. The industry's own doctrine has formally conceded the point — zero trust architecture, codified in NIST SP 800-207, is the explicit rejection of implicit trust based on network

location.[6] But concession is not reconstruction. Most “zero trust” programmes re-label the perimeter rather than re-think the organism.

> AXIOM I
>
> **The network is not a web. It is a population. Defend the cells, not the mesh.**

Now AI has finished what cloud started. When software agents read your email, query your databases, write your code and call external tools at machine speed, the notion of an inside and an outside stops being merely porous and becomes meaningless. An agent is simultaneously inside the perimeter and operated by whoever can reach its instruction channel — including, through indirect prompt injection, an attacker who has never touched your network at all. The reticular enterprise cannot survive the agentic era, because the agentic era turns every connected document, ticket and web page into a potential carrier of hostile instructions straight into the fused interior.

We keep rebuilding a fortress in a world that has stopped having edges.

The lesson from 1906 is not that Golgi was foolish — he was the best-equipped scientist of his age, and he was wrong. The lesson is that **a wrong theory of the network produces a wrong theory of defence**, and no quantity of tooling stacked on top of the wrong theory will fix it. We do not need a better wall. We need a better histology.

02 What Cajal Actually Proved

Four principles the brain settled a century ago — and security has not yet learned.

Cajal's achievement is usually compressed into a single phrase — "the neuron doctrine" — and then left behind in a history chapter. That is a waste. His doctrine, together with what his school discovered in the decades after, contains at least four precise architectural principles. They were established with a microscope, a stain, and the discipline to draw only what he actually saw. Each one maps onto security with uncomfortable exactness.

Principle 1 — Discreteness: the unit is the cell

Neurons do not fuse. Every neuron is an island: it maintains its own membrane, its own internal state, its own integrity. Damage one and the others do not collapse; the system degrades gracefully, cell by cell. The brain's robustness comes from the fact that its units are **separable** — there is no shared interior to poison.[1]

In security terms: workloads, identities, datasets and agents should be discrete, individually verifiable units with their own integrity boundaries — not processes swimming in a trusted interior. This is the real content of zero trust, and the reason micro-segmentation, workload identity and per-service credentials matter more than any gateway. The question is never "is it inside?" — inside does not exist. The question is always "which cell is this, and is it healthy?"

Principle 2 — The synapse: every exchange crosses a controlled junction

Between neurons there is no open channel — there is a synapse, and the synapse is a decision point. Whether a signal propagates depends on chemistry: the right receptors, the right thresholds, the right state. Synapses are selective, saturating, and above all **regulated**. The cell does not merely receive; it adjudicates.[2]

In security terms: every exchange between units — every API call, every agent tool-use, every data read — should cross an explicit junction where identity, authority and context are verified. Not a firewall at the edge of the world, but trillions of tiny adjudications at the points of contact. Policy enforcement points are our synapses; the mature architecture makes them as cheap, local and ubiquitous as the brain makes its own.

Principle 3 — Directionality: signals flow one way

Cajal's law of dynamic polarization: information flows into a neuron through its dendrites and out through its axon — not the reverse. The cell has an anatomy of direction, and the anatomy is what makes coherent signalling possible.[1]

In security terms: data flows should be as directional as axons. Egress is not ingress in reverse; it is a separate anatomy requiring separate control. Most catastrophic breaches are, in the end, directionality failures — data flowed out along paths that were built for flow in. AI agents make this urgent: an agent with read access to everything and write access to one external channel is an axon pointed at the attacker.

Principle 4 — Plasticity and pruning: the system rewires itself

The brain is not fixed wiring. Synapses strengthen with use and weaken with neglect; entire connections are pruned away — the brain deletes more synapses in development than it keeps, and prunes throughout life. Plasticity is not a feature of the brain; it is the mechanism by which the brain stays matched to its world.[5]

In security terms: a defence that cannot rewire itself is already obsolete. Detection logic, models and controls must adapt as the environment and the adversary change — and, just as important, the architecture must **prune**: decommissioned services, stale credentials, orphaned agents and forgotten data stores are unpruned synapses, and they are where attacks actually live. Attack-surface reduction is synaptic pruning. Almost nobody budgets for it.

> AXIOM II
>
> **Discreteness contains the blast radius. Synapses adjudicate every exchange. Direction controls the flow. Plasticity keeps the defence alive. Pruning keeps it honest.**

Four principles, all a century old, all demonstrated in three pounds of tissue, and all describing precisely the properties that AI-era security lacks. But Cajal's story has a second act — and it is the act that matters most for AI-based defence. Because for all his genius, there was one part of the brain Cajal could not see.

03 The Neuron Doctrine of Security

From histology to architecture: what the four principles demand of an AI-age security stack.

Doctrine is only useful when it bites. This section states the mapping plainly — what each Cajal principle obliges us to build, buy, and stop doing. The summary table is the one to pin on the wall.

Table 1. The Neuron Doctrine mapped to security practice

CAJAL PRINCIPLE	NEUROSCIENCE FACT	SECURITY IMPERATIVE	WHAT IT KILLS
Discreteness	Neurons never fuse; each maintains its own membrane and integrity	Discrete, individually verifiable units: workload identity, micro-segmentation, per-service credentials, agent identities	Lateral movement; "trusted interior"
The Synapse	Every exchange crosses a regulated junction that adjudicates before propagating	Verification at every point of contact: per-request authZ, context-aware policy, tool-call gating for agents	Implicit trust; standing access
Directionality	Dynamic polarization: in through dendrites, out through the axon	Separate anatomies for ingress and egress; egress allow-lists; agents with minimal write paths	Exfiltration; injection-driven leaks
Plasticity	Synapses strengthen and weaken with use; the system rewires continuously	Learning defences: adaptive detections, model refresh loops, adversary-informed tuning	Brittle, static rules
Pruning	The brain deletes more connections than it keeps, throughout life	Continuous decommissioning: stale identities, orphaned agents, shadow data, dead rules	Attack-surface drift

Why AI makes this doctrine mandatory, not optional

Three properties of AI systems convert these principles from good practice into survival conditions.

First, agents collapse the distance between instruction and action. A classic intruder must move, escalate and persist — each step a chance to be caught. An agent with tool access needs one poisoned input: a ticket, an email, a web page containing hostile instructions that the model cannot reliably distinguish from the content it was asked to read. Indirect prompt injection is now a first-class entry in the OWASP Top 10 for LLM applications, and it is precisely a synaptic failure: content crossed a junction without being adjudicated as instructions.[9]

Second, models act on inference, not permission. A language model does not "know" what it may not do; it predicts what a helpful system would do next. Unless the architecture wraps every consequential action in an explicit, non-model control — the synapse — the model's helpfulness becomes the attacker's capability.

Third, AI multiplies the population. Every deployment spawns new units: copilots, retrieval stores, vector databases, fine-tunes, MCP servers, browser agents. Each is a new cell in the organism, most without membranes, many invisible to the security team. You cannot defend a population you have not stained. Cajal could only fight Golgi because Golgi's stain let him *see* the units. Visibility is the precondition of doctrine — which is why the first SanRa principle is unglamorous: inventory before architecture.

> *An agent with read access to everything and write access to one external channel is an axon pointed at the attacker.*

The doctrine, then, in one sentence: **treat every workload, identity, dataset and agent as a discrete cell; force every exchange through an adjudicating synapse; give data a direction; let the defence learn; and prune relentlessly.** What this still lacks is an immune system. For that, we return to Madrid, 1918.

04 The Third Element: Security's Hidden Immune System

Cajal mapped every cell in the brain except one kind. It turned out to be the one that defends it.

For all his mastery, there was a class of cells Cajal's stains could not reveal. He found astrocytes in 1913 with his gold-sublimate method, but a third population — small cells, few or no visible processes — refused to appear. He called them, with the honesty of a man who draws only what he sees, the *third element*: neurons were the first, neuroglia the second, and these — whatever they were — the third.[10]

The resolution came from within his own school, and it was not comfortable. Pío del Río-Hortega, working first in the orbit of Cajal's Madrid laboratory and then, after a bitter estrangement, in his own lab at the Residencia de Estudiantes, developed a new stain — ammoniacal silver carbonate — and between 1918 and 1921 showed the world what the third element actually was: two cell types, the oligodendrocytes that insulate axons, and the **microglia** — the brain's resident immune cells.[3] Microglia are macrophages: they patrol tissue constantly, distinguish self from non-self, engulf pathogens and debris, coordinate inflammation, and remember. Cajal eventually conceded the discovery, and Río-Hortega — twice nominated for the Nobel Prize, exiled by the Spanish Civil War to Paris, Oxford and finally Buenos Aires — gave neuroscience its hidden defenders.[4]

The brain does not wait for the perimeter to be breached

Consider what microglia are architecturally. They are not a wall. They are not stationed at borders — the brain, sealed inside the skull behind the blood–brain barrier, *still* maintains resident sentinels in every region, because the designers of that system understood something our industry resists: **threats arrive from within, and the interior must defend itself.** Microglia are distributed, autonomous, always sampling, and they act locally first. They do not file a ticket to a central SOC and wait. They engulf.[10]

This is the precise architectural role that AI agents should play in security: resident sentinels distributed through the tissue of the enterprise — in the identity plane, the endpoint fleet, the CI/CD pipeline, the data stores, the model gateways — each continuously sampling its local tissue, each able to act locally on what is clearly pathological, and each

escalating what is ambiguous. Not one god-model in the cloud, but a population of small, local, well-bounded defenders.

Table 2. Microglial function as a design spec for AI sentinels

MICROGLIA	FUNCTION	AI SENTINEL EQUIVALENT
Constant surveillance	Resting microglia continuously extend and retract processes, sampling their entire territory every few hours	Always-on behavioural baselining of users, workloads and agents — not scheduled scans
Self / non-self discrimination	Recognise the body's own cells; attack what is foreign	Models of normal per unit; deviation triggers scrutiny, not blanket rules
Phagocytosis	Engulf and digest pathogens and debris on the spot	Local, reversible containment: isolate the workload, revoke the token, freeze the session
Inflammation signalling	Recruit and coordinate other defenders when the threat exceeds local capacity	Graduated alerting with context — escalation as a signal, not a dump of raw logs
Immune memory	Past insults shape future responses	Threat intelligence and incident history wired back into detection — institutional memory that survives staff turnover

The autoimmune warning

Immune systems fail in two directions, and the second is the one security prefers to forget. When discrimination fails, the immune system attacks the self — autoimmune disease. The security equivalent is the stack that eats the business: false positives that train analysts to ignore alerts, controls so frictionful that staff route around them into shadow AI, agents that block legitimate work at machine speed. A sentinel population that cannot tell self from non-self is worse than no sentinels, because it teaches the entire organism to stop listening. [11]

Calibration, then, is an immune property. Every sentinel action must be reversible, logged and rated for consequence: engulf the clearly pathological, flag the ambiguous, never amputate. This is also where Río-Hortega's story carries a human lesson security should sit with. The establishment — even a genius like Cajal — initially could not see the defenders, and then resisted the man who could. Every SOC contains a third element: capabilities and people the organisation has not yet learned to see. Doctrine means nothing if the culture expels its Río-Hortegas.

AXIOM III

Defence must be resident, not perimetered: small, local, always-sampling sentinels that engulf the pathological, escalate the ambiguous, and never attack the self.

05 The Predictive Brain: Defence as Inference

The brain does not react to the world. It predicts it — and survives on the errors. So should security.

Modern neuroscience's deepest idea is that the brain is fundamentally an inference engine. On the predictive-coding account, developed in its fullest form by Karl Friston, the brain maintains a generative model of its world and continuously predicts the sensory input that model implies. What travels up the hierarchy is not the world itself but **prediction error** — the difference between expectation and reality. Perception is controlled hallucination, corrected by the senses. The brain does not wait to be told; it guesses, and survives on the deltas.[5]

This is exactly the architecture that effective detection already gestures at, and that AI finally makes buildable at scale. A mature AI-based defence maintains, for every significant unit in the enterprise, a living model of normal: how this service account behaves, how this workload talks, what this agent's tool-use looks like, when this person works and from where. The signal is not the event; the signal is the **surprise** — the authenticated user whose behaviour violates the prediction, the agent whose tool calls drift from its established pattern, the data flow that has no model-consistent explanation.

Prediction-first defence changes the economics of the SOC. Signature matching asks "have we seen this exact evil before?" — a question attackers retire daily. Prediction asks "is this explicable by what we know of ourselves?" — a question that catches the novel, the slow, and the insider. It is the difference between guarding against last year's adversary and defending against the structure of abnormality itself.

Prediction without verification is hallucination

But the predictive brain carries a warning, and it is carved into the same theory. If perception is controlled hallucination, then the control is everything. Brains that stop correcting their models against reality do not become more efficient; they become psychotic. The model imprisons the perceiver.[5]

Anyone who has worked a large language model has met the same failure with a different name: hallucination — fluent, confident output unconstrained by fact. An AI security stack built on prediction alone will hallucinate threats and hallucinate all-clears with equal confidence. This is why SanRa's first law of AI-based security is **verification before output**:

no prediction — from a model or an analyst — becomes an action or a conclusion without crossing a synapse. Machine inference proposes; verified judgement disposes. The brain does this with sensory correction loops. We do it with provenance checks, second-channel confirmation, and structured human review calibrated to consequence.[11]

> AXIOM IV
>
> **Predict the normal, police the surprise, verify the conclusion. A defence that cannot hallucinate has no imagination; a defence that cannot verify has no judgement.**

And this brings us to the cell the whole doctrine exists to protect — the one running this page. The human.

06 The Human Neuron: Keeping the Analyst in the Circuit

Human factors is not the soft part of AI-based security. It is the part everything else depends on.

There is a detail in the Cajal story that should embarrass our industry. Río-Hortega made the discovery of a generation, and the establishment's first response was to expel him from the laboratory. The tissue defended itself against its own new knowledge. Organisations do this constantly: the analyst who challenges the confident model output, the engineer who flags the inconvenient risk — the system's antibodies treat them as the pathogen.[4]

AI-based security raises the stakes of this failure by an order of magnitude, because of a phenomenon human factors research documented decades ago: **automation bias**. Parasuraman and Riley's classic taxonomy — use, misuse, disuse, abuse — describes how operators of automated systems come to trust them too much (misuse), reject them too broadly (disuse), and delegate design choices to them blindly (abuse). Endsley and Kiris showed the *out-of-the-loop performance problem*: the better the automation, the worse the human's ability to take over when it fails.[11][12] In an AI-assisted SOC, the analyst supervises a system that is right often enough to destroy vigilance and wrong often enough to matter.

The neural analogy is exact. A cortical neuron does not passively relay its inputs; it integrates thousands of afferents, applies its own threshold, and fires or withholds. The analyst in the loop must be the same: not a rubber stamp on the model's output, but a cell with a threshold — integrating the machine's inference against context the machine does not have, and authorised — *required* — to withhold fire.

Designing the loop

- **Calibrated trust is a trained skill, not an attitude.** Analysts need explicit verification protocols — provenance checks, second-channel confirmation, structured challenge — exercised under realistic time pressure until they hold on the worst day, not the best.
- **Dissent must be architectural.** If the fastest path through the workflow is always "accept," the loop is decorative. Build the friction where it matters: consequential actions require a named human who has actually looked.

- **Skills must be practised against the machine.** The aviation lesson is settled: pilots who never hand-fly lose the ability to hand-fly. Analysts who never investigate without AI lose the ability to investigate. Deliberate manual-first practice — "AI-shadow" drills — is skill maintenance, not nostalgia.
- **The immune system must not attack its diagnosticians.** The analyst who catches the model being confidently wrong is your Río-Hortega. Promote them.

> *Machines signal. Humans decide. Any architecture that reverses this is not AI-based security; it is security outsourced to a prediction.*

This is SanRa's founding conviction, and the reason we exist as a human factors company rather than another tooling vendor: in AI-based security, the human is not the weakest link. The human is the only cell capable of knowing when the whole system is lying. Everything in the architecture that follows is designed to keep that cell healthy, sharp and in the circuit.

07 The SanRa Architecture

Five layers, one loop, one human in the circuit — the doctrine made buildable.

The SanRa Architecture is the reference design we use with clients to turn the Neuron Doctrine into an operating security capability. It has five layers and two permanent residents: the human in the verification loop, and governance over the whole organism.

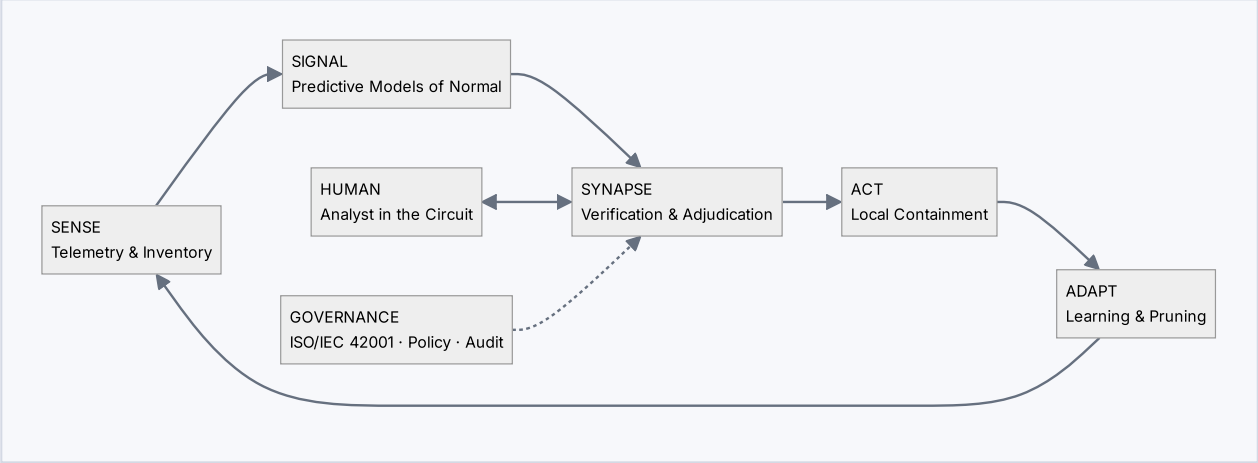


Figure 1. The SanRa Architecture: five defence layers with the human resident in the verification loop

Table 3. The five layers

LAYER	NEURAL ANALOGUE	WHAT IT DOES	KEY QUESTION IT ANSWERS
SENSE	Sensory epithelia; Cajal's stain	Complete, continuously refreshed inventory and telemetry: identities, workloads, data stores, agents, models, tools	What cells do we actually have?
SIGNAL	Predictive cortex	Living models of normal per unit; surprise detection rather than signature matching	What cannot be explained by ourselves?
SYNAPSE	The synapse	Adjudication of every consequential exchange: identity, authority, context, injection screening, human review gates	Should this signal propagate?
ACT	Microglial phagocytosis	Local, reversible, consequence-rated containment executed by resident sentinels	What do we engulf right now?

ADAPT	Plasticity and pruning	Detection tuning from outcomes, threat-intel feedback, decommissioning of stale units and rules	What did we learn — and what do we delete?

Governance is not a layer; it is the skull. The architecture operates inside an AI management system aligned to ISO/IEC 42001 and the NIST AI Risk Management Framework: every sentinel action logged and auditable, every model's role and limits documented, every consequential decision traceable to a named accountable human.[7][8] This is what makes AI-based defence defensible — to boards, to clients, to regulators, and to the analyst who has to stand behind the decision six months later.

08 Ten Theses for AI-Based Security

The doctrine, condensed to what can be nailed to a door.

1. **The network is not a web; it is a population.** Stop defending the mesh. Defend the cells.

2. **There is no inside.** Location confers no trust. Identity, health and context — verified per exchange — confer limited, revocable trust.

3. **Every synapse is a policy decision.** Any exchange that cannot be adjudicated should not occur. Especially the ones your AI agents make.

4. **Data flows one way, on purpose.** Directionality is anatomy, not configuration drift. Egress is a separate system with separate controls.

5. **Your defenders must live in the tissue.** Resident sentinels beat perimeter guards. Small, local, always sampling, reversible in action.

6. **Beware the autoimmune stack.** A defence that attacks the business teaches the business to disable the defence. Calibration is survival.

7. **Predict normal; police surprise.** Signatures chase yesterday's adversary. Models of self catch tomorrow's.

8. **Verification before output.** Machine inference proposes; verified judgement disposes. In models and in analysts alike.

9. **The human is the cell that knows when the system is lying.** Train calibrated trust, design in dissent, and practise against the machine — or lose the ability to work without it.

10. **Prune or be breached.** Stale identities, orphaned agents and shadow data are unpruned synapses. Deletion is a security control.

09 From Doctrine to Practice: The First Ninety Days

Doctrine that does not change Monday morning is literature. Here is the path.

Table 4. A ninety-day adoption path for the Neuron Doctrine

PHASE	FOCUS	KEY ACTIONS	EVIDENCE OF PROGRESS
Days 1–30 Stain the tissue	SENSE	Full inventory of identities, workloads, data stores and — critically — every AI tool, agent, copilot and retrieval store in actual use, sanctioned or not. Shadow-AI discovery included.	A living register of the cell population; shadow AI surfaced and triaged
Days 31–60 Build the synapses	SYNAPSE + ACT	Wrap AI agents in non-model controls: tool-call gating, egress allow-lists, injection screening on retrieved content, consequence-rated human approval for irreversible actions.	No agent can act externally without crossing an adjudicated junction
Days 61–90 Teach the immune system	SIGNAL + ADAPT + HUMAN	Baselines of normal for priority units; verification protocols trained with the analyst team; first pruning pass on stale identities, agents and data; governance pack started under ISO/IEC 42001.	Surprise-based detections live; pruning as a scheduled control; auditable decision trail

SanRa runs this path with clients as a structured engagement, and trains the human layer through our CREST-aligned courses — SAN-101 (*Thinking with AI*) for calibrated trust and skill maintenance, and SAN-102 (*Data at Risk*) for the data-risk surface of AI adoption. The architecture fails without the analyst; the analyst fails without the architecture. We build both because neither survives alone.

Golgi and Cajal shared a Nobel Prize while disagreeing about the fundamental structure of the nervous system. That is worth remembering when vendors tell you the structure of AI-based security is settled. It is not settled. It is being decided now — by whoever chooses a theory of the network and builds on it. Choose the population. Choose the synapse. Keep the human in the circuit.

References & Further Reading

1. Ramón y Cajal, S. (1906). The structure and connexions of neurons. *Nobel Lecture, December 12, 1906*. In *Nobel Lectures, Physiology or Medicine 1901–1921*. Amsterdam: Elsevier, 1967.

2. Golgi, C. (1906). The neuron doctrine — theory and facts. *Nobel Lecture, December 11, 1906*. In *Nobel Lectures, Physiology or Medicine 1901–1921*. Amsterdam: Elsevier, 1967.

3. Río-Hortega, P. del (1919). El "tercer elemento" de los centros nerviosos. I. La microglía en estado normal. *Boletín de la Sociedad Española de Biología*, 9, 68–82.

4. Río-Hortega, P. del (1932). Microglia. In W. Penfield (Ed.), *Cytology and Cellular Pathology of the Nervous System* (pp. 480–534). New York: Hoeber.

5. Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127–138.

6. Rose, S., Borchert, O., Mitchell, S., & Connelly, S. (2020). *Zero Trust Architecture* (NIST Special Publication 800-207). Gaithersburg, MD: National Institute of Standards and Technology.

7. International Organization for Standardization. (2023). *ISO/IEC 42001:2023 — Information technology — Artificial intelligence — Management system*. Geneva: ISO/IEC.

8. National Institute of Standards and Technology. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)* (NIST AI 100-1). Gaithersburg, MD: NIST.

9. OWASP Foundation. *OWASP Top 10 for Large Language Model Applications*. owasp.org/www-project-top-10-for-large-language-model-applications.

10. Sierra, A., de Castro, F., del Río-Hortega, J., Iglesias-Rozas, J. R., Garrosa, M., & Kettenmann, H. (2016). The "Big-Bang" for modern glial biology: Translation and comments on Pío del Río-Hortega 1919 series of papers on microglia. *Glia*, 64(11), 1801–1840.

11. Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39(2), 230–253.

12. Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors*, 37(2), 381–394.

Historical passages in this paper follow the accounts in references 1–4 and 10, including the Golgi–Cajal reticular controversy and Río-Hortega's discovery of microglia and oligodendroglia between 1918 and 1921. Interpretations for security practice are SanRa's own.

SANRA

HUMAN FACTORS FOR THE AI AGE

The brain has no firewall. It has something better.

Three pounds of tissue defends itself with discrete cells, adjudicated junctions, resident sentinels, predictive models and a memory that learns — and it has done so for half a billion years of adversarial testing. The blueprint was published in 1906. We are simply building it.

Start the conversation: research@sanra.co.uk — or train your team through SAN-101 and SAN-102 at www.sanra.co.uk.

SANRA LTD
UNITED KINGDOM
WWW.SANRA.CO.UK

WHITEPAPER WP-01 · V1.0
© 2026 SANRA LTD
SHARE FREELY, UNMODIFIED