MICKAI™

MICKAI EBOOK SERIES · No. 19

From Foundation Models to Sovereign Models.

The next layer of the stack: fine-tuned and specialised open foundations today, fully native sovereign weights as the funded roadmap.

AUTHOR

Micky Irons

Founder and named inventor, Mickai LTD.

19 June 2026 · v1 · mickai.co.uk

Mickai LTD · Companies House 17166618 · press@mickai.co.uk · mickai.co.uk
UK IPO register, named inventor Mickarle Wagstaff-Irons · Trade mark UK00004373277

TABLE OF CONTENTS

Contents

Foreword

A note from the author

Part I · The Problem

1. The Tenant and the Landlord
2. What You Actually Give Up
3. Sovereignty Is Not a Slogan

Part II · The Method

4. Owning the Weights
5. Training Your Own Models Now
6. The Sealed Corpus

Part III · The Architecture

7. Fifty Brains, Not One
8. The Substrate and the Hardware
9. Sealing Every Action

Part IV · The Path

10. The Gradient of Sovereignty
11. The Funded Roadmap to Native Weights
12. From Borrowed to Sovereign

Appendix

About the author

FOREWORD

A note from the author

I did not set out to write a book about model weights. I set out to build a system an operator could trust with the most sensitive work of their life, and I kept hitting the same wall. Every serious capability I wanted to ship depended, in the end, on a model I did not own, served from a building I would never see, governed by terms I could not change. You can wrap that in as much engineering as you like. The dependency is still there, and the people who hold it know it.

This book is the honest account of how we climbed out of that trap, and how far up the road we have actually come. I will not pretend we have arrived. Today the fifty brains of the Sovereign Intelligence Operating System run on fine-tuned and specialised open foundations: weights we have taken, retrained on our own sealed corpus, and made answer to us on the operator's own hardware. At the same time we are training our own models now, building the data, the recipes and the discipline that fully native sovereign weights will demand. Funding scales that work. It does not start it.

I am precise about that distinction because the field is not. There is a great deal of theatre around sovereignty at present, a great many flags planted on rented ground. I have no interest in any of it. What I care about is the difference between intelligence you borrow and intelligence you own, and the concrete, unglamorous, expensive steps that carry you from one to the other. This book is about those steps.

I write in the first person because the decisions were mine and so were the mistakes. I am the founder and chief executive of Mickai. If you are an operator, an investor, a policymaker, or simply someone who has felt uneasy handing your hardest problems to a machine you cannot inspect, the argument here is for you. It is not a comfortable argument. It is, I believe, the correct one.

Micky Irons

Founder and named inventor, Mickai LTD · 19 June 2026

PART I · THE PROBLEM

Borrowed intelligence is a dependency, not a foundation.

1. The Tenant and the Landlord

Most organisations using artificial intelligence today are tenants. They do not own the intelligence they rely on. They rent access to it, by the token, through a pipe, from a handful of providers who hold the weights, the infrastructure and the right to change the terms. This is rarely put so plainly, because the language of the industry is built to obscure it. We talk about platforms, partners and ecosystems. What we mean is that someone else owns the thing, and you are paying to stand near it.

A tenant accepts certain facts about their position. The landlord can raise the rent. The landlord can change the rules of the building. The landlord can decide, for reasons that have nothing to do with you, that your kind of business is no longer welcome. And the landlord can, if compelled by a court, a government or a commercial calculation, look through your windows. None of this is hypothetical in the world of frontier models. Pricing has shifted under whole companies. Access has been withdrawn. Usage has been inspected. The terms of service are not a contract between equals, they are a notice posted on the door.

> **You can wrap a borrowed model in as much engineering as you like. The dependency is still there, and the people who hold it know it.**

For a great many uses this tenancy is perfectly acceptable. If you are summarising public documents or drafting marketing copy, the landlord relationship costs you little. But the moment your work becomes consequential, the moment it touches a patient record, a legal strategy, a national security assessment, a defence supply chain, a family's private affairs, the tenancy becomes the single largest risk you carry. You have built your most sensitive operations on rented ground, and you hold no deed.

The Sovereign Intelligence Operating System exists because I refused that bargain for the work that matters most. I wanted a system where the operator holds the deed. The ambition sounds simple. Everything difficult about the last few years of my life has lived in the gap between wanting it and building it.

2. What You Actually Give Up

When you depend on a model you do not own, you surrender four things, and most people notice only the first. You surrender cost control. The price per token is set elsewhere and can move against you with little warning. For a pilot this is a line item. For a system processing millions of operations a day it is an existential exposure, because your margin becomes a derivative of someone else's pricing

committee.

The second thing you surrender is availability. A rented model can be deprecated, rate limited, geofenced or simply switched off. Models that organisations built real products on have been retired with a few months' notice, forcing expensive and risky migrations. If your business cannot function when the pipe closes, you do not have a business, you have a dependency with a logo on it.

The two you do not see

The third thing you surrender is confidentiality. Every prompt you send is data that leaves your control. Even under the best contractual assurances, your most sensitive inputs travel to infrastructure you do not own, to be processed by software you cannot inspect, under jurisdictions you did not choose. Assurances are not architecture. A promise not to look is not the same as an inability to look. Sovereignty is about the second, not the first.

The fourth thing you surrender is the right to specialise. A rented frontier model is a general instrument tuned for the average of all its users. You cannot open it. You cannot retrain its core on your own corpus. You can prompt it and you can bolt retrieval onto its side, but the intelligence itself stays generic, shaped by data and objectives that were never yours. The deepest competitive advantage in applied artificial intelligence, a model that genuinely knows your domain in its weights, is precisely the thing tenancy denies you.

Add those four together and the real shape of the problem appears. Borrowed intelligence is not a foundation. A foundation is something you build upon and own. Borrowed intelligence is a dependency you service forever, and the better it works the deeper the dependency runs.

The Mickai pantheon.

3. Sovereignty Is Not a Slogan

The word sovereignty has been thoroughly abused. It is stamped on press releases for systems that call a foreign frontier model over an encrypted connection. It is claimed by services that host an open model in a data centre on the right continent and pass the geography off as a guarantee. I want to be exact about what I mean, because the imprecision is not innocent. It lets people sell the comfort of the word without paying for the substance.

Sovereignty, as I use it, is the operator's ability to run, inspect, retrain and govern their intelligence without depending on any external party for permission, availability or secrecy. It is a property of the architecture, not the marketing. The honest test is simple. Cut the wire. Disconnect the system from every network it does not control. Does the intelligence still work, in full, on the operator's own hardware. If yes, you have sovereignty. If no, you have a dependency wearing the word as a costume.

> **Cut the wire. If the intelligence still works in full on your own hardware, you have sovereignty. If not, you have a dependency wearing the word as a costume.**

By that test the fifty specialised brains of the Sovereign Intelligence Operating System are sovereign today. They run on the operator's own hardware. They are fully offline-capable. Every consequential action they take is sealed into a post-quantum Open Audit Record under the FIPS 204 ML-DSA-65 standard, so that governance does not depend on a vendor's logs. The weights are ours to inspect and retrain. The wire can be cut and the system keeps its mind.

I am equally clear about where the road still runs ahead of us. Those brains are built today on fine-tuned and specialised open foundations, not yet on fully native weights trained end to end by us. That is the next layer of the stack, and the subject of this book. Sovereignty is not a binary you declare on launch day. It is a gradient you climb deliberately, and the climb has stages. I would rather tell you exactly which stage we are on than sell you a flag.

PART II · THE METHOD

You earn ownership of intelligence by specialising and training, not by buying.

4. Owning the Weights

To own intelligence you must own the weights. The weights are the model. Everything else, the prompts, the retrieval, the orchestration, the interface, is scaffolding around a core of numbers that encodes what the system knows and how it reasons. If you do not hold those numbers, in a form you can load, inspect, modify and run, you do not own the intelligence, however clever your scaffolding.

This is why open foundation models matter so much to the sovereign project. Open weights are the raw material of ownership. When a capable foundation model is released under terms that let you download, run and retrain it, a door opens that no amount of prompting against a closed model can replicate. You can take those weights onto your own hardware. You can study them. You can change them. You can make them yours in a way a rented endpoint will never allow.

Open foundation, sovereign specialisation

Today the Sovereign Intelligence Operating System builds on open foundations in the Llama and Qwen families, and others, taken as a starting point and then specialised hard. The larger models on the workstation reach up to the seventy to seventy-two billion parameter class through hybrid processing that splits the work across the graphics processor and the central processor, so the system is not confined to small models. I am honest about the mechanics: that largest class runs through offload, it is not served live on a single twenty-four gigabyte graphics card. We size each brain to the hardware it will actually run on, and the architecture scales up the lineup as the hardware does.

Starting from open weights is not the same as dressing up a borrowed model. The licences are real and we honour them, with private third-party notices and no fabricated provenance, ever. What changes is the relationship. We are not tenants of these weights. We hold them, we retrain them, we run them on our own metal, and we shape them into something that did not exist before we started. Ownership of intelligence begins with ownership of the weights, and open foundations are how an operator who is not a trillion-dollar laboratory can begin to hold them at all.

The Mickai pantheon.

5. Training Your Own Models Now

There is a comfortable lie in the sovereign AI conversation, and the word is later. Organisations say they will train their own models later, when they are funded, when they are bigger, when the time is right. Later never comes, because the capability to train is not a switch you flip with money. It is a discipline you build with practice, and you cannot buy years of practice on the day the cheque clears.

So we train now. We are actively training our own models today, fine-tuning and specialising the open foundations on a sealed corpus we built ourselves, developing the data pipelines, the evaluation harnesses, the recipes and the institutional muscle that fully native weights will demand. This is not a promise deferred to a funding round. It is work in progress in the building right now, on the hardware we have, at the scale we can currently reach.

> **The capability to train is not a switch you flip with money. It is a discipline you build with practice, and you cannot buy years of practice on the day the cheque clears.**

What does funding actually change, then. It scales the work, it does not initiate it. More compute lets us train larger models, on more data, more often, and ultimately carry the specialisation all the way down from the foundation into fully native sovereign weights. The funded roadmap is a question of scale and completeness, not of whether we know how. The knowing is being earned now, every week, in the gap between a borrowed foundation and a model that genuinely answers to us.

I labour this point because investors and operators deserve precision, and the field rewards vagueness. When I say we are training our own models, I do not mean we have already produced a frontier-scale sovereign model from scratch. I mean the training capability is live, the corpus is real, the

specialisation is shipping in the fifty brains today, and the path from here to native weights is one of scale rather than discovery. That is a very different claim from later, and I will not let it be confused with one.

6. The Sealed Corpus

A model is what it eats. The single most important asset in training a sovereign model is not the compute and it is not the architecture, it is the corpus, the body of data the model learns from. Whoever controls the corpus controls what the model knows, how it reasons, what it refuses and what it values. For a system meant to handle an operator's most sensitive work, the corpus cannot be a scrape of the open internet of unknown provenance. It must be sealed, curated and owned.

By a sealed corpus I mean a controlled body of training data whose provenance is known, whose contents are governed, and whose boundary is defended. We know what is in it and what is not. We know where each part came from and under what right we hold it. It does not leak the operator's private inputs back into a shared pool, because there is no shared pool. The corpus sits inside the sovereign boundary, it is not a pipe out of it.

Why provenance is sovereignty

Provenance is not a compliance nicety, it is the heart of the matter. A model trained on data of unknown origin is a liability you cannot audit. You cannot say where a behaviour came from, you cannot defend the rights you relied on, and you cannot reason about what the model may have absorbed that it should not have. A sealed corpus with known provenance turns the model from a black box of borrowed text into an asset you can stand behind. When a regulator or an operator asks what this model learned and from where, the sovereign answer is a manifest, not a shrug.

This discipline is also what makes specialisation honest. The Mickai canon, the facts about our own system, our filed patents, our hardware and our architecture, lives in a governed knowledge collection authored from our own primary documents and ingested so the brains answer with current facts rather than stale guesses. The corpus is not just training fuel, it is the institutional memory of the system, and keeping it sealed and current is a permanent operational duty, not a one-time event.

The Mickai pantheon.

PART III · THE ARCHITECTURE

Fifty specialised brains, sealed and governed, on the operator's own hardware.

7. Fifty Brains, Not One

The dominant pattern in artificial intelligence today is the single giant generalist, one enormous model asked to do everything from poetry to protein folding. It is an extraordinary engineering achievement and, for a sovereign operator, the wrong shape. One model that does everything is one model you must trust with everything, one boundary that holds every secret, one failure that touches every function. We chose a different architecture deliberately.

The Sovereign Intelligence Operating System is built around fifty specialised brains. Each is a Mickai model in its own right, a foundation specialised for a defined purpose, sized for the hardware it runs on. The fifty divide into twenty-five domain brains covering areas such as intelligence and defence, governance and strategy, health and humanity, science and engineering, and identity, and twenty-five operational brains that run the system itself. Specialisation is not a limitation here, it is the source of both quality and containment.

> **One model that does everything is one boundary that holds every secret, and one failure that touches every function.**

The operational tier is where the system governs its own behaviour. It holds the kernel functions that route and arbitrate work between brains, custodian brains that keep knowledge fresh and repair the system, and a tier of specialists. The substrate they all run on, the silicon and serving layer, is its own component and not a brain itself. This is an operating system in the full sense, not a chatbot with extra prompts. The intelligence is organised the way a serious institution is organised, by function, with separation of duties and lines of accountability.

Specialisation also makes sovereignty achievable at the hardware an operator can actually own. A single frontier generalist of the largest scale is out of reach for almost everyone outside a handful of laboratories. Fifty focused brains, each sized to its task, can run on hardware an operator holds in their own building, and the most demanding of them scale up the lineup as the operator's hardware scales up. The architecture is what turns sovereignty from a slogan into something that fits in a room you control.

8. The Substrate and the Hardware

Sovereign intelligence is not only a software claim, it is a hardware claim. If the model runs on someone else's computer it is not sovereign, however open the weights. So the architecture treats hardware as

a first-class concern. The fifty brains run on a serving substrate on the operator's own machines, and the system detects the hardware it is on and scales its behaviour to match, rather than assuming a fixed deployment.

We build every capability into the system even when the current hardware cannot run it at full strength. A feature that exceeds the present machine is not omitted, it is shown as unavailable on the current hardware with clear guidance on what would be required, never quietly dropped. This is a deliberate design rule. It keeps the system honest about its own limits and ready for the hardware lineup ahead, from a single workstation today to the flagship edge servers of the funded roadmap, without a rewrite.

Right-sizing the intelligence

On a current workstation the system serves models up to the seventy to seventy-two billion parameter class through hybrid offload across graphics and central processors. I state the mechanism plainly because the field is full of inflated claims: that largest class runs through offload, it is not served live on a single twenty-four gigabyte graphics card, and I will not pretend otherwise. Smaller and mid-sized brains run comfortably and quickly, and the architecture assigns each brain a footprint appropriate to both its task and the metal available.

All this hardware discipline exists to serve the cut-the-wire test from earlier in the book. Because the brains run on the operator's own machines, the network can be severed and the intelligence persists. Because consequential actions are sealed locally into post-quantum audit records, governance does not depend on a remote service either. The hardware is not an implementation detail beneath the sovereignty claim, it is the sovereignty claim made physical.

The Mickai pantheon.

9. Sealing Every Action

Ownership of intelligence is necessary but not sufficient. A sovereign system that cannot prove what it did is a powerful instrument with no accountability, and that is its own kind of danger. So the architecture pairs ownership with evidence. Every consequential action the system takes is sealed into an Open Audit Record, a tamper-evident entry that records what happened in a form anyone can verify later without trusting the system that produced it.

These records are signed using post-quantum cryptography, specifically the FIPS 204 ML-DSA-65 standard for digital signatures. The choice of a post-quantum scheme is deliberate and forward-looking. Audit records are meant to hold their integrity for years, and a signature a future quantum computer could forge is no signature at all for evidence that must outlive today's cryptography. We sign for the threat model of the decade ahead, not only the threat model of this morning.

Verification without trust

In practice the operator asks the system to act, the action is sealed, and the seal can be independently verified. Trust does not rest on the operator's faith in the vendor or in the running software. It rests on cryptographic evidence that anyone holding the public verification material can check. This is the difference between a system that asks to be trusted and a system that can be checked, and for sovereign work that handles the most sensitive matters, only the second is acceptable.

This sealing layer is also what lets sovereignty coexist with governance, which sceptics often assume must trade off. A sovereign system is not an unaccountable one. Because every consequential action carries a verifiable record, the operator keeps full control and full evidence at the same time. Sovereignty without accountability is just opacity with better marketing. The Open Audit Record is how we refuse that trade.

PART IV · THE PATH

From borrowed foundations today to fully native sovereign weights ahead.

10. The Gradient of Sovereignty

Sovereignty is not a line you cross once. It is a gradient you climb, and pretending otherwise is how people end up either overclaiming or never starting. I find it useful to name the stages plainly, because then an operator can locate themselves honestly and an investor can see exactly what funding moves and what it does not.

The first stage is tenancy, the borrowed model behind an interface, which we have already examined. The second stage is local hosting, running an open model on your own hardware without changing it, which buys you availability and confidentiality but not depth. The third stage, the one the Sovereign Intelligence Operating System occupies today, is sovereign specialisation: open foundations taken onto your own hardware, retrained on your own sealed corpus, organised into fifty governed brains, and sealed with post-quantum audit. The fourth and final stage is fully native sovereign weights, models trained end to end under the operator's own control.

> **Sovereignty is a gradient you climb deliberately. I would rather tell you exactly which stage we are on than sell you a flag.**

Each stage delivers real value, and each is a foundation for the next. This matters because the gradient is also a defence against the two failure modes of the field. Overclaiming plants the flag of stage four on the ground of stage one. Paralysis refuses to move until stage four is fully reached, and so never leaves stage one at all. The honest path is to occupy the highest stage you have genuinely earned, deliver from it now, and climb deliberately. That is exactly what we are doing.

So let me state our position without ornament. We are at stage three today, in full, in production architecture: specialised, sealed, sovereign on the operator's hardware. We are actively doing the training work that stage four requires. The funded roadmap carries us up the final part of the gradient. There is no flag planted on rented ground in that sentence, and that is the point.

The Mickai pantheon.

11. The Funded Roadmap to Native Weights

What stands between sovereign specialisation today and fully native sovereign weights tomorrow is not a mystery of method, it is a question of scale. Training a model end to end at competitive quality demands compute, data and time in quantities that grow steeply with capability. The recipes, the corpus discipline and the evaluation harnesses are being built now. What the funded roadmap adds is the scale to carry that work all the way down to weights that owe nothing to a borrowed starting point.

I am careful here because this is exactly the kind of claim the field routinely inflates. Funding does not buy us the ability to train, we already have it and we are using it. Funding lets us train bigger, more often, on more data, and ultimately replace the open foundation at the core of each brain with weights we trained ourselves. The progression runs from specialising someone else's foundation, to training native weights for the brains where it matters most, to a fully native lineup across the fifty. It is a staged climb funded in steps, not a single leap bought with a single round.

Hardware-gated, honestly built

The hardware roadmap mirrors the model roadmap. The architecture already names the lineup it is built for, from current workstations through to the flagship sovereign edge servers the most demanding native training will require. Capabilities that exceed the present hardware are built and gated, shown as requiring an upgrade rather than hidden, so the system is ready for the metal as it arrives. This is why I insist on building everything now even when today's hardware cannot run it at full strength. The roadmap is not a wish list, it is a system already shaped to grow into it.

And beneath all of it sits the economic question of sovereignty, which native weights sharpen rather than create. The more of the intelligence you own outright, the less of your future you owe to anyone else's pricing, availability or permission. Stage four is not an engineering vanity, it is the completion of the deed. When the weights are fully ours, the last thread of tenancy is cut, and the operator holds the

intelligence the way they hold the building it runs in.

12. From Borrowed to Sovereign

I began this book with the image of a tenant and a landlord, and I want to end by closing that circle. The whole argument has been a single movement: from intelligence you borrow to intelligence you own, traced through the concrete steps that make the movement real rather than rhetorical. Owning the weights. Training your own models now. Sealing the corpus. Organising fifty governed brains on your own hardware. Sealing every action into post-quantum evidence. Climbing the gradient deliberately toward fully native weights.

None of these steps is free, and none is finished by declaring it. That is precisely why so few systems actually walk this path and so many simply claim the destination. The work is expensive, unglamorous and slow, and it does not photograph well. But it is the only work that produces the thing operators in serious domains actually need, which is intelligence they can run, inspect, retrain and govern without asking anyone's permission, and prove they did so.

The Sovereign Intelligence Operating System is my attempt to do that work honestly and completely. Today it stands as fifty specialised brains, built on fine-tuned and specialised open foundations, running fully offline-capable on the operator's own hardware, every consequential action sealed under post-quantum signatures, with our own sovereign Bitcoin-anchored Layer 1, Pantheon, anchoring the economic and trust substrate beneath it. That is a real position, earned and verifiable, not a flag on rented ground.

Tomorrow it becomes more sovereign still, as the training we are doing now scales up the gradient toward fully native weights and the funded roadmap completes the climb. I will not tell you we have arrived, because we have not, and because the operators who depend on this system deserve the truth about where it stands. What I will tell you is that the direction is set, the method is proven, the work is live, and the deed is being written in our names. That is the difference between borrowed intelligence and sovereign intelligence, and it is the difference this entire system exists to make.

The Mickai pantheon.

APPENDIX · ABOUT THE AUTHOR

Micky Irons

Founder and chief executive of Mickai LTD (Companies House 17166618, registered office 20 Wenlock Road, London, N1 7GU) and named inventor on the Mickai SIOS patent corpus: 101 filed UK patent applications, around 2,234 claims. Trade mark Mickai registered at UK00004373277.

Profiles

mickai.co.uk

crunchbase.com/person/micky-irons

linkedin.com/in/mickyirons

References and further reading

· National Institute of Standards and Technology, FIPS 204: Module-Lattice-Based Digital Signature Standard (ML-DSA), 2024.

· Bommasani, R. et al., On the Opportunities and Risks of Foundation Models, Stanford Center for Research on Foundation Models, 2021.

· Touvron, H. et al., Llama 2 and Llama 3: Open Foundation and Fine-Tuned Chat Models, Meta AI, 2023 to 2024.

· Bai, J. et al., Qwen Technical Report, Alibaba Cloud, 2023 to 2024.

· Longpre, S. et al., The Data Provenance Initiative: A Large-Scale Audit of Dataset Licensing and Attribution in AI, 2023.

· Bender, E. M., Gebru, T. et al., On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?, Proceedings of FAccT, 2021.