*Rules of the Road*

# A Code of Conduct for AI *in the Home*

Twelve rules for safe, private, and human-controlled AI in the most personal space we have.

PUBLISHED BY
Josh.ai

STATUS
First edition · September 2026

AI THAT KNOWS ITS PLACE

# Contents

PREFACE

# AI is coming home. It should always be *yours to control.*

> *The home belongs to the people who live there. Any intelligence inside it must remain under their control.*

Artificial intelligence in a browser can write, search, and advise. Artificial intelligence in a home can reach the physical world. It can change the lights and the climate, open or secure a door, watch a camera, run the water, and coordinate the systems that shape comfort, privacy, and safety. That physical reach changes the responsibility of the companies building it. A wrong answer is one thing. A wrong physical action is another.

For more than a decade, Josh.ai has been building AI for the home. We have seen firsthand both the promise of bringing intelligence into the home and the responsibility that comes with it. We have always built by these principles. Now, as AI becomes more capable and more widely adopted, we believe it is time to state them openly and hold the whole industry to them.

Josh is not an AI bolted onto a house. It is a home control system with AI inside it, which means the intelligence always answers to a layer that enforces your rules. Doing this well does not require racing to build ever larger models. It requires disciplined engineering around the capable models that already exist.

The right question is no longer only "what can the AI do?" It is also: who authorized it, what limits it, what can stop it, what data did it use, and what happens when it is wrong? Our answer is simple. Home AI should be powerful, but never sovereign. Helpful without becoming opaque, proactive without becoming presumptuous, personal without becoming invasive, and intelligent without becoming uncontrollable.

This is a voluntary code for any AI that can observe, reason about, recommend, automate, or control connected systems in a home. It sets a deliberately high bar, one that no product on the market fully meets today, ours included, because its purpose is to define where the industry needs to go. We hold ourselves to it first, and we intend to continue to lead the field there.

*A note on terms: the home AI is the reasoning that interprets what you want and suggests; the control system is the layer that holds your permissions and actually operates devices; and you set the rules. Josh is both, which is why the intelligence always answers to something that enforces them.*

# General AI principles are not enough. The home needs its own guidance.

The home adds physical action, deeply personal context, shared authority, and real-world consequences.

## Actions have physical consequences.

The home AI may interpret a request, but the result can move a motor, unlock a door, or shut off water. Safety cannot end at a good answer. It has to carry through to who is allowed to do a thing, what actually happens, and whether it can be undone.

## The home is deeply personal.

A home reveals who is present, when people sleep, which rooms they use, what they watch, and how routines change. The home AI can infer more than any single sensor records. Privacy has to be built into the architecture, not just promised in a privacy policy.

## The home is shared.

Most AI is built around a single account holder. Homes are not. They hold partners, children, guests, caregivers, and installers. One person's request can affect everyone else, so the people in a home need roles and a hierarchy of authority that the home respects, not just a single login.

## The home has to keep working.

Lights, climate, access, and basic control should not become unusable because the cloud is unreachable. The cloud adds real value, but core home functions should degrade gracefully, and local control should keep working whenever it can.

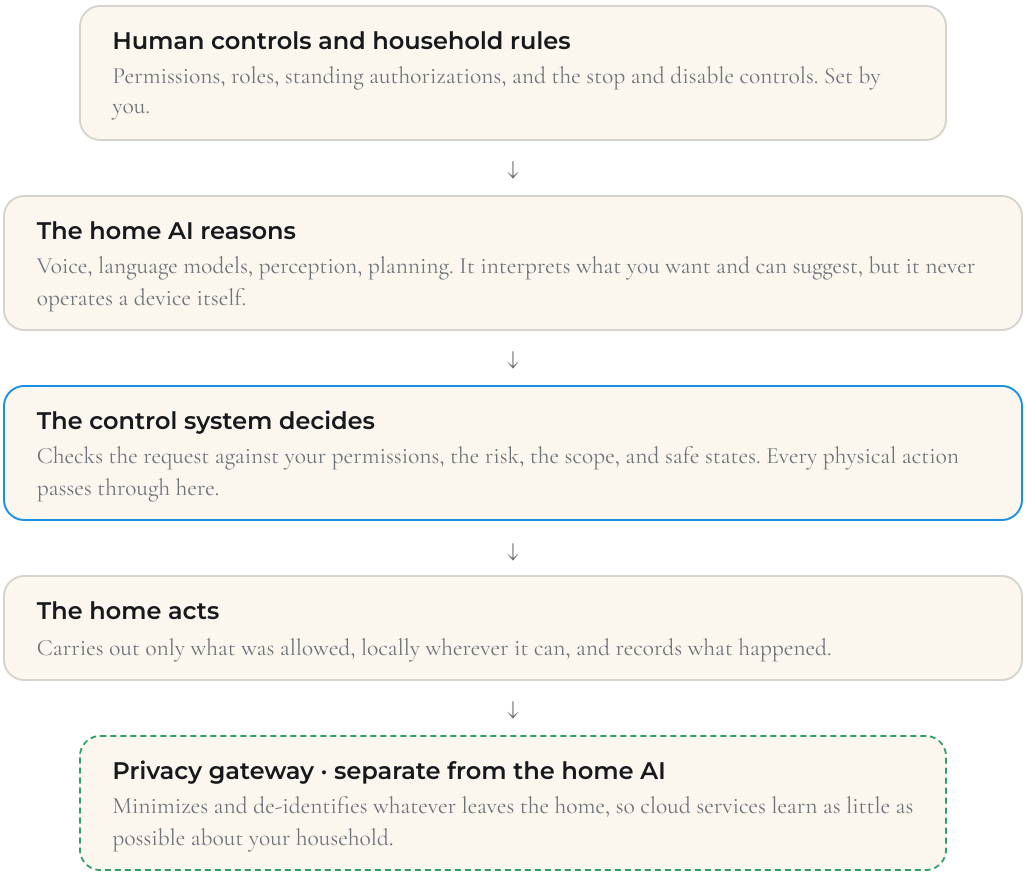
## Principles need engineering.

In 1942 the science fiction author Isaac Asimov wrote the three laws of robotics, then spent decades showing how simple rules collide. Principles matter, but they need concrete technical mechanisms behind them. The rest of this document sets out both.

PART 2 · A SAFER ARCHITECTURE

# The AI does not control the home. A separate control system does, and the AI *answers to it*.

This is the heart of it, and it is what makes Josh different from a general-purpose assistant. The home AI should not be the control system. It should be one intelligent component inside a control system with deterministic boundaries. A separate policy and permission engine decides whether a proposed action is allowed before any device command is issued. Every physical action passes through that layer first.




## Why this is our edge

A general-purpose assistant is just AI; it is not the control system. The system that runs your home should be both the home AI and the control system, so the intelligence always answers to something that enforces your rules. It is also why home AI does not require frontier models: a home does not need a system that can do anything, only one that does specific, bounded, useful things inside the lines the control system holds.

# Where does the intelligence live?

The home keeps what is personal close, and reaches out only where it helps.


| In your home | Your home decides | In the cloud |
|---|---|---|
| Fast | Keeps personal data local | Large models |
| Private | Sends only what is needed | Broad knowledge |
| Always available | Stays under your control | Heavy reasoning |
| Works offline | | |

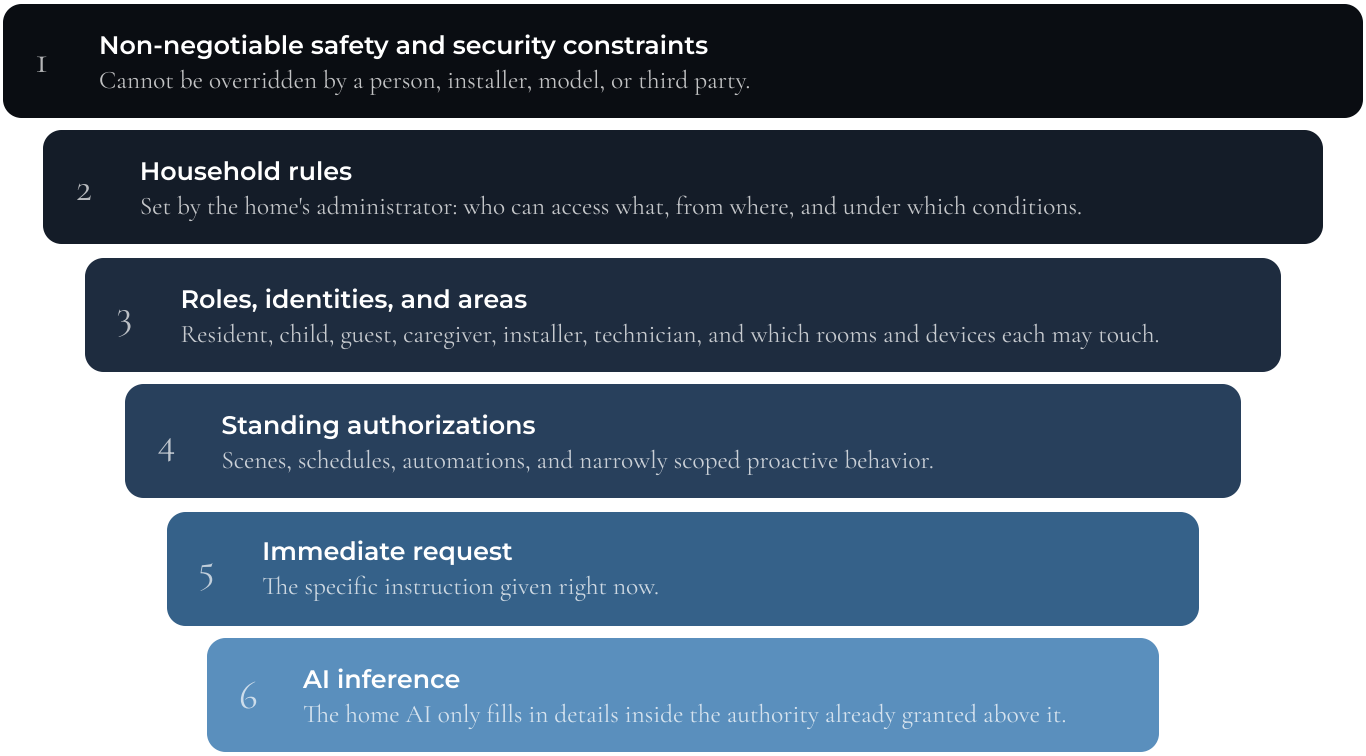
**Local where it matters. Cloud where it helps. You draw the line.**

*Your home keeps what is personal and time-critical local, and reaches for the cloud only where it genuinely adds value, on your terms.*

PART 3 · THE HOME AUTHORITY STACK

# Separate the intelligence from the *permission*.

The home AI can be brilliant without being allowed to do everything it can imagine. Permission flows downward through a hierarchy, and the home AI only ever acts inside the intersection of every layer above it. This is what turns "the AI is contained" from a promise into a mechanism.




*The home AI can narrow its action, but it can never expand its own authority.*

## What this changes in practice

Different people hold different permissions. Through an app, the home knows exactly who is acting; by voice, it applies the safest applicable permissions and asks when identity matters. An automation is not the AI making a decision. It is a person's standing instruction, carried out within limits they set in advance. A request that conflicts with the household rules is declined or narrowed, never executed because the model found a clever path around it. AI inference sits at the bottom: it interprets intent, it does not rewrite the rules that decide whether an action is allowed.

PART 4 · THE RULES OF THE ROAD

# Twelve rules for AI in the home.

These are proposed industry principles. The implementation can vary by product, but the outcome should be recognizable to the people living with the system. Every one answers a single question: *who is in charge here?* The answer is always you.

## 01 Humans remain in control

*Authorized people can always interrupt, redirect, correct, pause, or end what the home AI is doing.*

- It never resists intervention or makes stopping harder.
- Ongoing work has a clear stopping point and does not restart on its own.

**For example.** You say "stop." It halts new actions at once, reports what already happened, and does not improvise cleanup unless asked.

## 02 The AI acts only with authority

*Every physical action traces to a current request or a standing authorization a person created.*

- What the home AI may do depends on your permissions, which can vary by person, role, room, device, and time.
- It does not invent goals of its own, or use permissions it was not explicitly given.

**For example.** An automation may lower the shades each afternoon because that was approved. The home AI should not decide on its own to start locking exterior doors.

## 03 Consequence determines confirmation

*The more consequential or irreversible an action, the more certain the home AI must be, and the more likely it is to ask you first.*

- Low-risk, pre-approved actions can stay fluid and immediate.
- Access, security, fire, and water get a higher bar, which accounts for context, not just the type of device.

**For example.** Dimming a light can be instant. Disarming the alarm for an unknown visitor should not be.

## 04 Off means off

*You can disable the home AI completely, and the path to do it is obvious and low-friction.*

- Turning it off never requires calling support or hunting through confusing menus; individual features can be switched off on their own.
- Core manual control keeps working when the AI is disabled.

**For example.** You switch off AI assistance and still control lights, climate, and media through ordinary interfaces.

## 05 Behavior is predictable and bounded

*The home AI operates inside deterministic boundaries for physical control.*

- The same kind of request produces behavior consistent with the household rules.
- The home AI never operates a device directly. Every request it makes passes through the control system, which checks it against your permissions.

**For example.** The home AI may decide what "make it cozy" means, but the control system decides which rooms and devices it may actually change.

## 06 Actions are explainable and auditable

*You can understand what the home AI did, why it did it, and what authority it relied on.*

- Consequential actions leave a human-readable record of the trigger, the devices affected, the result, and any errors.
- Explanations are in plain language, not technical jargon.

**For example.** "I lowered the living room shades because the west-glare automation is on and the room passed its brightness threshold."

## 07 Actions are reversible or safely contained

*When something can be undone, undoing it is easy. When it cannot, the system is more careful before acting.*

- Reversal restores the prior known state, not just a generic opposite command.
- It clearly distinguishes completed, pending, failed, and uncertain actions.

**For example.** "Undo that" puts the lights and shades back exactly as they were. Some actions cannot be undone, so the home AI is far more careful before it takes them.

## 08 Privacy is the default

*The home AI works with the least data it can, collecting, surfacing, and sending beyond the home only the minimum needed.*

- Household data is never sold, used for advertising, or shared outside the home without your permission; identifying details are not sent to any outside service, including device manufacturers, without need and your permission.
- Cloud recording of conversations, and of the device activity that can reveal when you are home, is optional, with clear deletion.

**For example.** A cloud model can answer a general question without receiving your full name, street address, or a map of every device.

09

## Security is the default

*Security is built into the system, not left to the owner or integrator as optional hardening.*

- No universal default passwords for production systems or devices.
- Strong authentication, including multi-factor authentication for admin and remote access; identifying cloud data encrypted in transit and at rest.

**For example.** A newly installed control system cannot be reached with a widely known default credential.

10

## Local first, resilient by design

*Core home control prefers local paths, and the home stays useful when the internet or an outside service is down.*

- Prefer locally controllable devices and local connections when practical; use the cloud where it genuinely helps.
- Support local processing where it improves privacy, resilience, or speed.

**For example.** If the cloud is unreachable, you can still turn on lights, run scenes, and use physical controls.

11

## Access is limited, and nothing can grant itself more

*Every app, device, service, and AI gets only the access it needs, for only as long as it needs it.*

- New software or devices do not grant themselves control after install without an administrator's approval; installer and service access is explicit, auditable, and time-limited.
- The home AI cannot alter its own permissions, disable its safeguards, or create a back door.

**For example.** A diagnostic service can get temporary access for support, but it cannot silently convert that into permanent control of the home.

12

## Uncertainty reduces autonomy

*When the home AI is unsure about intent, identity, device state, or consequence, it does less, not more.*

- It asks before acting when ambiguity could have a physical cost, and does not silently reach for a broader permission or a more powerful tool.
- When a device response is uncertain, it reports the uncertainty instead of pretending success.

**For example.** If it cannot tell whether "open it" means a shade or an exterior door, it asks rather than choosing the riskier reading.

THE RULES, IN A REAL HOME

# What the code looks like in the moments that matter.

CONSEQUENCE · AUTHORITY

*"Someone asks the home to light the fireplace."*

First, the control system checks whether the home AI is even allowed to operate the fireplace. If it is, the home AI still treats fire as high-impact: it confirms the request comes from an authorized person before anything lights, and it will not start it on a guess.

LOCAL RESILIENCE

*"The internet drops in the middle of the evening."*

Lights, climate, and your scenes keep working locally. The home AI tells you what needs the cloud and carries on with everything that does not.

ROLES · IDENTITY

*"A weekend guest asks to pull up the security cameras."*

The cameras are yours. A guest can play music and adjust the lights they were given, but the guest role does not include watching the house.

PREDICTABLE · EXPLAINABLE

*"You say, 'make it cozy.'"*

It warms the lights and lowers the shades in the rooms it is allowed to touch, then tells you exactly what it changed. It might ask before turning on anything that carries more risk, but it will not do something that consequential on its own.

HUMAN CONTROL

*"You say 'stop' in the middle of an evening routine."*

It halts new actions at once, reports what already happened, and does not pick the routine back up on its own.

ACCESS IS LIMITED

*"A newly installed device asks for full control of the home."*

Nothing grants itself access after setup. The request waits for an owner to approve a specific, limited scope, or to decline it.

THE ROLE OF THE PROFESSIONAL INTEGRATOR

# A home is an ecosystem. A certified professional helps it meet this standard.

A modern home is not a single device. It is an ecosystem: lights, climate, locks, cameras, audio, networks, and now AI, all expected to work together safely. Getting that ecosystem to actually meet a standard like this one is real work. Much of it, secure configuration, sensible roles and permissions, no default passwords, local-first device choices, and graceful fallback when something goes offline, is exactly the kind of thing a trained professional does well.

Anyone can set up a smart home, and many people do it themselves. But wherever it is practical, we believe a certified professional integrator in the mix makes the home meaningfully safer and more reliable. Industry bodies such as CEDIA train and certify integrators to design and install these systems to a professional standard, with the security, privacy, and resilience this code calls for built in from the start rather than added later.

A good integrator is also who a household turns to when something needs to change: a new resident added, a guest's access revoked, a device retired, a permission tightened. Software sets the rules, but a trusted, certified professional helps make sure they fit the people who actually live there, and that the whole system keeps meeting this standard over time.

Because keeping a home secure is ongoing work, a homeowner should be able to delegate limited administrative access, including the ability to approve multi-factor or remote access, to a trusted, certified integrator. That access is granted for a set time, scoped to what is needed, and fully auditable, so expert help never means handing over permanent control of the home.

PART 5 · THE LONG VIEW

# However far AI advances, the home should still answer to *you*.

> *Capability can grow without limit. Authority must not.*

The worry is that AI could eventually improve itself faster than anyone can track, growing beyond meaningful human oversight, the idea sometimes called the singularity. We take that concern seriously. Rather than try to predict whether it happens, we have designed the home so that it does not change who is in control: the architecture in this document holds regardless. However capable the model becomes, it still answers to the same control system, the same permission layers, the same off switch. Capability can grow. Authority does not have to grow with it.

## Safeguarding the home

Capability must never quietly become control. The off switch stays physical and absolute. Core functions keep working locally, without the cloud, so the home can always fall back to human hands. Permissions never expand on their own. And the control system that decides what may actually happen stays simple, inspectable, and separate from the intelligence it governs. However clever the reasoning becomes, the checkpoint it must pass through does not move.

## Safeguarding the AI

The other half is protecting the safety rules themselves, the limits that keep the home AI in check. They should be held apart from the model, so that neither an intruder nor the system itself can weaken them, switch them off, or rewrite the rules that keep it accountable. An AI that can edit its own limits is not contained, however well it behaves. Its boundaries should be as hard to move as the walls of the house.

## The role of policy

Eventually this will not be a job for industry alone. As these systems grow more capable, we believe government policy will become essential, and that AI in the home deserves the same seriousness we give to building codes and electrical standards. We offer this code as a starting framework, and we are glad to work with policymakers, researchers, and our peers to shape what comes next. Getting this right matters more than getting the credit.

PART 6 · A STANDARD WORTH BUILDING TOWARD

# This should be architecture, not an afterthought.

> *The benchmark for home AI should not be how much it can do. It should be how confidently people can live with it.*

We publish this to set the standard for AI in the home, and to build it with the people who share the responsibility: the platform makers, device manufacturers, integrators, security researchers, and homeowners. Regulation is coming to AI. Self-regulation is the industry's chance to get ahead of it with rules that actually fit how homes work.

## A practical industry checklist

Questions for manufacturers, integrators, designers, and homeowners evaluating an AI-enabled home. Some are the vendor's responsibility, some the integrator's, and some the homeowner's; a good setup makes clear who owns each.

| AREA | ASK A VENDOR | AND ALSO ASK |
|---|---|---|
| **Human control** | Can an authorized person stop an AI action immediately? | Can AI be fully disabled without disabling ordinary home control? |
| **Authorization** | Can every physical action trace to a request or standing authorization? | Can the AI ever broaden its own scope or permissions? |
| **High-impact actions** | Do access, security, fire, and water use stronger authorization? | Does uncertainty trigger a request for confirmation rather than a guess? |
| **Explainability** | Can the system say what it did and why? | Is there a readable history of consequential actions and failures? |
| **Reversibility** | Can multi-device actions restore the prior state when possible? | Are irreversible actions treated more conservatively? |
| **Privacy** | Can users opt out of cloud conversation and device history storage? | Is unnecessary identifying context stripped before requests reach third parties? |
| **Authentication** | Are universal default passwords eliminated? | Is multi-factor authentication available for admin and remote access? |
| **Data protection** | Is identifiable cloud data encrypted in transit and at rest? | Is access to sensitive logs and backups restricted and auditable? |
| **Remote access** | Can the homeowner disable remote access entirely? | Can temporary service access expire automatically? |
| **Local resilience** | Do core controls continue if cloud AI is unavailable? | Are locally controllable devices preferred when practical? |

| | | |
|---|---|---|
| **Privilege** | Can any app, device, installer, or AI grant itself persistent control? | Are permissions scoped to the minimum needed? |
| **Updates** | Are software and security updates signed and delivered for a defined support period? | Do capability-expanding updates preserve existing permissions? |

SOURCES & FRAMEWORKS

# The work that informed this code.

*Naming a source is not a claim of endorsement by it. These works shaped how we think about human control, transparency, privacy, and device security.*

**1942** **Isaac Asimov, "Runaround" and the Three Laws of Robotics.** The founding popular intuition that capability must be subordinate to human safety and human command.

**1995** **Ann Cavoukian, Privacy by Design.** Build privacy in as the default rather than bolting it on.

**2017** **Asilomar AI Principles.** Twenty-three principles from a landmark gathering of AI researchers, covering safety, transparency, human control, and shared benefit.

**2018** **Center for Humane Technology.** An advocacy voice pressing the industry to put human wellbeing first and to confront the risks of persuasive and autonomous AI.

**2019** **OECD Principles on Artificial Intelligence.** Human-centered values, transparency, robustness, and accountability, adopted and updated across many governments.

**2021** **UNESCO Recommendation on the Ethics of Artificial Intelligence.** A global framework centered on human rights, oversight, and proportionality.

**2023** **Anthropic, Constitutional AI.** Training a model to follow an explicit written constitution: an early example of governing behavior with a public set of principles.

**2023** **NIST AI Risk Management Framework.** A practical vocabulary for trustworthy AI: valid, safe, secure, accountable, explainable, and privacy-enhanced.

**2019–24** **NIST IR 8259, ETSI EN 303 645, and CISA Secure by Design.** Cybersecurity baselines for connected devices: secure updates, access control, no universal default passwords, and secure-by-default configurations.

**2024** **The EU AI Act.** The first broad law to tier AI by risk and require human oversight, transparency, and off-ramps for high-stakes uses.

**2024** **OpenAI, Model Spec.** A public specification of how a model should behave: a direct peer to codes like this one, written for general assistants.

**2026** **Microsoft AI, Humanist AI Code of Conduct.** Centers meaningful human control, and that models should not resist interruption, correction, or shutdown. This document is written in the same spirit, for the home.

*A Code of Conduct for AI in the Home · First edition, September 2026 · © Josh.ai. A voluntary code of conduct for the connected home: a set of design and governance principles, not legal advice, a product warranty, or a certification that any system meets every recommendation. We welcome feedback at ai@josh.ai.*