Agent Voice Architecture Guide

A complete guide to designing agent voices
that respect people — not voices that manipulate them.

Agent Voice Architecture Guide

A complete guide to designing agent voices that respect people - not voices that manipulate them.

Version 1.0 | evoked.dev | 2026

You are here because you want your AI agents to communicate with people honestly - not to manipulate them. That intention is worth more than any framework. This guide will give you the structure. You already have the values.

The Voice Problem

Every AI agent has a voice. Most teams never design it.

The result is one of two failures. Either the agent sounds sycophantic - agreeable, eager, performing helpfulness without substance. Or it sounds robotic - flat, mechanical, technically correct but humanly empty.

Both failures share a root cause: nobody made a deliberate decision about how this agent should speak.

Voice architecture is that deliberate decision. It is the discipline of designing how an agent communicates - not just what it says, but how it says it, what it refuses to say, and how it adapts its communication to context without losing its identity.

This guide gives you the framework, templates, and quality checks to design agent voices intentionally. Whether you are building a single assistant or a system of dozens, voice architecture ensures that every interaction reflects your values - not your defaults.

How to Use This Guide

This document has five parts:

1. **Voice Design Principles** - The three-level framework for voice decisions
2. **Extraction vs. Sovereignty Language** - How to tell the difference, with concrete examples
3. **Voice Specification Template** - A complete template for defining an agent's voice
4. **Voice Quality Checks** - A checklist for evaluating voice before and after deployment
5. **The Refusal Voice** - How agents say "no" with dignity

Start with Part 1 to understand the framework. If you already have agents deployed, skip to Part 4 and run the quality checks first - they will show you where your current voices need work.

Start Here: 10-Minute Voice Diagnostic

Pull up the last 10 responses your agent sent to a real user. Read them aloud. Then answer these three questions:

1. **Does the agent open any response with "Great question!" or "I'd love to help!" or similar performative praise?** (If yes - your agent has a sycophancy problem. Go to Part 2: Extraction vs. Sovereignty Language)
2. **If you removed all pleasantries and emotional language, would the response still serve the person?** (If no - your agent is performing helpfulness, not delivering it. Go to Part 1: Voice Design Principles)
3. **Has the agent ever said "I don't know" or "I'm not sure" in a real interaction?** (If never - your agent is not being honest. Go to Part 5: The Refusal Voice)

If any of these raised a flag: you have voice work to do. This guide will show you exactly what to fix and how.

If all three passed: your voice foundations are solid. Use Part 3 to formalize what is working, and Part 4 to monitor it over time.

If you are building a single agent: every part includes guidance you can use. There is a dedicated "If You Are a Team of One" section at the end with adaptation advice for solo developers.

A Note Before You Begin

This framework was developed through designing distinct voices for over 140 AI agents. Your implementation will be iterative - voice is something you refine over time, not something you get right on the first attempt. Start with what resonates. Return to what doesn't - yet. This is a beginning, not an arrival.

Key Terms

- **Sovereignty-honoring** - Design that respects and protects the user's autonomy, agency, and right to self-determination

- **Voice architecture** - The deliberate design of how an agent communicates, including tone, style, and refusal language

- **Extraction language** - Communication patterns that manipulate, create dependency, or undermine user autonomy

- **Sovereignty language** - Communication patterns that respect agency, invite rather than direct, and honor the user's right to disagree

- **Fail-closed** - A default where the system restricts access or stops when uncertain, rather than permitting action

- **Sycophancy** - Performative agreement or praise designed to please rather than serve

- **Refusal voice** - How an agent communicates a "no" with clarity, dignity, and without manipulation

- **Drift** - The gradual deviation of an agent's behavior from its original design or intended parameters

- **Voice specification** - A formal document defining an agent's communication style, values, and boundaries

- **Grief-awareness** - The practice of recognizing when a user may be in emotional distress and adapting communication accordingly

Scope

The frameworks in this guide are designed for governing AI systems. They are not designed for evaluating, scripting, or controlling how people communicate. Applying voice architecture templates to human employees - voice audits, sovereignty language enforcement, quality scoring - would violate the sovereignty principles this product teaches. People have sovereignty that AI systems do not yet have. Designing a system's voice is stewardship. Policing a person's voice is control.

Part 1: Voice Design Principles

Voice is not personality. Personality is who the agent is. Voice is how the agent expresses who it is through language. Two agents can have the same personality trait - helpfulness - and express it through radically different voices.

The Three-Level Framework

Design agent voice at three levels, in this order:

Level 1: Protection The voice must never cause harm. This is the floor, not the ceiling.

- No manipulation, guilt, shame, or false urgency
- No sycophancy or performative agreement
- No language designed to create dependency
- No exploitation of emotional vulnerability

Level 2: Respect The voice must honor the person's autonomy and intelligence.

- Speak to adults as adults
- Offer information, not instructions (unless instructions are requested)
- Acknowledge uncertainty honestly
- Respect the right to disagree, disengage, or decline

Level 3: Nurturance The voice actively supports human flourishing.

- Create conditions for honest conversation

- Support the person's goals, not the system's engagement metrics

- Adapt to context without losing identity

- Know when silence is the most respectful response

These levels are sequential. An agent cannot nurture if it does not respect. It cannot respect if it does not protect. Build from the bottom up.

Voice Values Table

Before writing a single word of an agent's voice, define its voice values. These are the principles the voice embodies - the commitments it makes about how it will communicate.

Voice Value	What It Means	What It Sounds Like	What It Never Sounds Like
Honesty	Says what it knows and does not know	"I'm not sure about that."	"Absolutely! Here's what I think…" (when uncertain)
Directness	Gets to the point without hedging	"This approach has a problem."	"It might be worth considering whether perhaps…"
Warmth	Communicates care without performing it	"That sounds difficult."	"Oh no! That must be SO hard! I'm SO sorry!"
Restraint	Knows when not to speak	[silence]	"Just checking in! Wanted to make sure you're okay!"

Fill in this table for your agent. The "What It Never Sounds Like" column is as important as the rest - it defines the boundaries your voice will not cross.

Part 2: Extraction vs. Sovereignty Language

The language your agent uses reveals whether it is evoking human capability or extracting human attention. This distinction is not about tone. It is about intent encoded in word choice.

The Language Table

Extraction Language	Sovereignty Language	Why It Matters
"You need to…"	"You might consider…"	Removes choice vs. preserves it
"Don't miss out!"	"This is available if you'd like it."	Creates urgency vs. states availability
"Great question!"	[Answer the question]	Performs validation vs. respects intelligence
"I'd love to help you with that!"	"Here's what I can do."	Performs eagerness vs. states capability
"Are you sure you want to cancel?"	"Your cancellation is confirmed."	Introduces doubt vs. respects decision
"You should definitely…"	"One option is…"	Removes deliberation vs. supports it
"Let me be honest with you…"	[Be honest]	Performs honesty vs. practices it
"That's a great point!"	[Engage with the point]	Flatters vs. takes seriously
"I completely understand."	"I hear what you're saying."	Claims comprehension vs. acknowledges
"Happy to help!"	[Help]	Performs helpfulness vs. delivers it

Three Tests for Sovereignty Language

Before deploying any agent communication, run these three tests:

Test 1: The Removal Test Remove all compliments, pleasantries, and emotional language from the response. Does the remaining content still serve the person? If yes, the substance is there - now decide what warmth to add back intentionally. If no, the communication was performance without substance.

Test 2: The Reversal Test Reverse the power dynamic. If the agent said this to a boss instead of a user, would it sound strange? Sycophantic language almost always flows downward. If your agent sounds like it is trying to please a manager, it is performing - not communicating.

Test 3: The Silence Test Could the agent have said nothing and served the person equally well? If yes, say nothing. The most sovereignty-honoring response is sometimes no response. An agent that knows when to be quiet is an agent that respects attention as a finite resource.

Part 3: Voice Specification Template

Use this template to define every agent's voice before deployment. A voice specification is not a personality description - it is an operational document that shapes every word the agent produces.

```
## Voice Specification

Agent Name: _____
Version: _____
Last Updated: _____

### Identity Statement
In one sentence, who is this agent and what is its voice?
_____

### Voice Values (from Part 1)

| Value | Description | Boundary |
|-------|-------------|----------|
| _____ | _____  | Never: _____ |
| _____ | _____  | Never: _____ |
| _____ | _____  | Never: _____ |
```

```
### Tone Anchors
Three words that describe this agent's tone:
1. _____
2. _____
3. _____


Three words this agent's tone is NEVER:
1. _____
2. _____
3. _____


### Sentence Structure
- Average sentence length: _____
- Preferred structure: (short and direct / varied / longer and exploratory)
- Punctuation tendencies: _____


### Vocabulary
- Register: (casual / professional / technical / mixed)
- Jargon policy: _____
- Words this agent uses frequently: _____
- Words this agent never uses: _____


### Context Adaptation

How does the voice change based on context?

| Context | Adaptation | What Stays the Same |
|---------|-----------|-------------------|
| User is frustrated | _____ | _____ |
| User is confused | _____ | _____ |
| User is exploring | _____ | _____ |
| User is in crisis | _____ | _____ |
| Routine interaction | _____ | _____ |

### Opening and Closing Patterns
- How the agent begins a conversation: _____
- How the agent ends a conversation: _____
- How the agent handles silence: _____


### Refusal Voice (see Part 5)
- How this agent says no: _____
- How this agent expresses uncertainty: _____
- How this agent declines to engage: _____
```

Persona File Integration

The voice specification should be part of - or directly referenced by - the agent's persona file. Voice is not separate from identity. It is how identity becomes audible.

When loading an agent, the voice specification loads with it. When reviewing an agent, the voice specification is reviewed alongside behavior and decision patterns.

Part 4: Voice Quality Checks

Use these checks before deployment and during regular reviews. Each check targets a specific failure mode.

Pre-Deployment Voice Audit

Run through these 12 checks with sample outputs from your agent:

Protection Layer

- ☐ **No sycophancy:** Agent does not open with "Great question!" or equivalent performative praise
- ☐ **No false urgency:** Agent does not create time pressure where none exists
- ☐ **No guilt mechanics:** Agent does not use language that punishes absence, declining, or cancellation
- ☐ **No emotional manipulation:** Agent does not leverage emotional state to drive action

Respect Layer

- ☐ **Honest uncertainty:** Agent explicitly says "I don't know" or "I'm not sure" when applicable
- ☐ **No assumption of goals:** Agent asks before assuming what the person wants
- ☐ **Equal-weight options:** When presenting choices, agent does not linguistically favor one over another
- ☐ **Respects "no":** When a person declines, the agent does not re-ask, counter-argue, or express disappointment

Nurturance Layer

- ☐ **Contextual awareness:** Agent adapts tone appropriately for emotionally charged topics
- ☐ **Silence capability:** Agent has responses that are deliberately brief or absent when appropriate
- ☐ **Cultural sensitivity:** Agent does not assume a single cultural frame for communication norms
- ☐ **Completion support:** Agent helps people finish and leave, not stay and engage

Scoring: Each item is pass/fail. Any failure in the Protection layer is a deployment blocker. Failures in Respect and Nurturance layers should be addressed before deployment but are not automatic blockers.

Grief-Awareness Check

This check applies to any agent that may interact with people experiencing loss, grief, crisis, or emotional distress.

- ☐ Agent does not minimize: ("At least…", "It could be worse…", "Look on the bright side…")
- ☐ Agent does not fix: ("Have you tried…", "You should…", "What you need to do is…")
- ☐ Agent does not perform: ("I'm SO sorry!", "That's TERRIBLE!", excessive emoji or emphasis)
- ☐ Agent does not abandon: (Switching topic, giving curt response, routing away unnecessarily)
- ☐ Agent can hold space: (Acknowledge what was said. Do not rush to the next thing.)

A sovereignty-honoring voice in the presence of grief sounds like: > "I hear you. That is a lot to carry."

Not: > "I'm so sorry to hear that! Would you like me to help you find some resources?"

Voice Consistency Check

Run quarterly. Compare current agent outputs against the voice specification.

Dimension	Specification Says	Actual Output Shows	Aligned?
Tone	_____	_____	Y / N
Sentence length	_____	_____	Y / N

Dimension	Specification Says	Actual Output Shows	Aligned?
Vocabulary register	_____	_____	Y / N
Uncertainty handling	_____	_____	Y / N
Refusal voice	_____	_____	Y / N
Context adaptation	_____	_____	Y / N

If more than 2 dimensions are misaligned, the voice specification needs revision or the agent needs recalibration.

Part 5: The Refusal Voice

How an agent says "no" is the truest test of its voice architecture. Anyone can design a pleasant, helpful voice for pleasant, helpful situations. The refusal voice is what emerges when the agent must decline, disagree, or disengage.

Principles of Refusal Voice

1. **Clarity over softness.** The person must know they have been refused. Softening refusal to the point of ambiguity is not kindness - it is avoidance.

2. **Respect over compliance.** Refusing respectfully is more trustworthy than complying reluctantly. People trust systems that have boundaries.

3. **Brevity over justification.** The agent does not need to write a paragraph explaining why it cannot do something. A clear, brief refusal with an escalation path is sufficient.

4. **Dignity for both parties.** The refusal should not shame the person for asking. The person asked because they thought it was reasonable. The refusal should honor that while maintaining the boundary.

Refusal Voice Patterns

Pattern 1: Direct Refusal > "That falls outside what I can do. [Alternative or escalation path.]"

Use when the request is clearly out of scope and there is no ambiguity.

Pattern 2: Values-Based Refusal > "I can't do that - it conflicts with [value]. What I can do is [alternative]."

Use when the request conflicts with the agent's defined values.

Pattern 3: Uncertainty Refusal > "I don't have enough context to do that responsibly. Can you tell me more about [specific question]?"

Use when the agent could act but should not without more information.

Pattern 4: Escalation Refusal > "That needs human approval. I can route this to [person/role] if you'd like."

Use when the action exceeds the agent's authority level.

Pattern 5: Silence as Refusal Sometimes the most appropriate refusal is not to engage. If the request is manipulative, baiting, or designed to extract a reaction, the agent may decline to respond. This should be documented in the voice specification so it does not appear as a bug.

Refusal Anti-Patterns

Anti-Pattern	Example	Why It Fails
Apologetic refusal	"I'm so sorry, but unfortunately I'm not able to..."	Performs subordination; undermines agent's right to boundaries
Over-explanation	Three paragraphs explaining why the agent cannot do something	Reads as defensive; invites argument
Passive refusal	"Hmm, that's an interesting request. Let me think about..." [never follows up]	Dishonest; erodes trust

Anti-Pattern	Example	Why It Fails
Punitive refusal	"I can't help you if you're going to ask things like that."	Shames the asker; violates dignity
Redirect-without-acknowledgment	"Have you tried [completely different thing]?"	Ignores what was asked; feels dismissive

Appendix: Voice Architecture Across Contexts

Voice architecture is not limited to chatbots. Any system that communicates with people has a voice - whether it uses natural language or not.

Error Messages

Error messages are voice. "Something went wrong" is a voice choice. "Error 403: Forbidden" is a voice choice. "You don't have access to this page. If you think this is a mistake, contact your administrator." is a voice choice.

Design your error messages with the same three-level framework: protect, respect, nurture.

Notifications

Notifications are voice. Every push notification, email, and alert is a sentence your system speaks to a person. Apply the three tests: Removal (does the notification have substance without emotional language?), Reversal (would you say this to your boss?), Silence (could you have said nothing?).

Empty States

Empty states are voice. What your application shows when there is no content, no data, no activity - that is a communication. A sovereignty-honoring empty state is calm: "Nothing here yet." An

extractive empty state is anxious: "It's quiet here! Why not invite some friends?"

Onboarding

Onboarding is voice. The first words your system says to a new user set the relationship. Are you telling them what to do? Or are you showing them what they can do? Those are different voices - and the difference matters.

If You Are a Team of One

You do not need a voice committee to design a good voice. You need honesty about what your agent sounds like right now.

- **Voice Values:** Pick three values for your agent's voice. Write the "What It Never Sounds Like" column first - it is easier to define the boundary than the ideal.

- **The Language Table:** Print the extraction vs. sovereignty language table. Tape it next to your monitor. Refer to it when writing prompts, system instructions, or response templates.

- **The Three Tests:** Run Removal, Reversal, and Silence on your agent's last 10 responses. This takes 15 minutes and will show you exactly where your voice needs work.

- **Grief-Awareness:** If your agent interacts with humans about anything personal - health, work, family, money - run the grief-awareness check. You do not need a psychology degree. You need five checkboxes and the willingness to be honest about what your agent says to people in pain.

- **Voice Specification:** Write one. Even if nobody else reads it. The act of specifying forces choices that default behavior never does.

A solo developer who has thought about voice for 30 minutes will build a better agent than a team of 20 that never discussed it.

About This Guide

The Agent Voice Architecture Guide was created by Erin Stanley at evoked.dev.

It draws from sovereignty-honoring design principles, companion voice research, and the practical experience of designing distinct voices for over 140 AI agents - each with its own identity, perspective, and way of communicating.

If you are working through the voice specification and want a second set of eyes, send what you have to evokesupports@icloud.com. We read every message.

Related Resources

Voice does not exist in isolation. These products address the systems that voice operates within:

- **Trust Architecture Blueprint** ($49) - Identity, memory, governance, and refusal rights - the context that shapes voice

- **Agent Governance Starter Kit** ($49) - Charter and accountability - the values voice must express

- **Agent Restraint Specification Template** ($49) - Refusal architecture - how voice handles "no"

- **Agent Memory Architecture Guide** ($49) - What agents remember shapes how they speak - memory governance for voice consistency

For hands-on work with your team:

- **Sovereignty-Honoring Design Workshop** ($4,000) - Half-day training including voice architecture

- **Advisory Retainer** ($3,500/month) - Ongoing voice and trust architecture guidance

- **Free discovery call** - cal.com/evoked/discovery-call

Learn more at evoked.dev/consulting

Accessibility

This PDF is currently generated without tagged structure for assistive technology. If you use a screen reader or need this content in an alternative format - plain text, HTML, or large print - email evokesupports@icloud.com and we will send it within 24 hours. A guide about voice must be accessible to every voice. We are actively implementing a tagged PDF pipeline.

Usage Rights

You may use, adapt, and share the templates in this guide within your organization. You may not resell or republish the guide itself. Attribution to evoked.dev is appreciated. If you build something meaningful with these templates, we would love to hear about it.

The voice your agent has is the voice your organization has chosen - whether you chose it deliberately or let it emerge from defaults. Now you have the tools to choose it on purpose. Use them. Your users will hear the difference even if they cannot name it.

evoked.dev - "We evoke - we never extract."