



The Metrics You Expose Are the Values You Endorse

MARCH 2026

15 min read • 3,354 words

Yesterday, I had a conversation with an interface designer who reads this blog. She was working through something: she wanted to define her core values for interface design before she built anything else. She wanted to know what her principles were before the pressure of production schedules and stakeholder requests made the question feel like a luxury she couldn't afford.

That conversation got me thinking about something I haven't been able to shake since.

The Optimization Landscape You Already Know

Every developer optimizes for what they can measure. This isn't a moral failing. It's how engineering works. You define a metric, you instrument your system to capture it, you iterate until the number moves in the direction you want. The discipline is powerful. It built the modern internet.

The problem is that the metrics available to developers aren't neutral. They're chosen. Someone at the platform level decided which data to expose through the SDK, which analytics to surface in the dashboard, which numbers to put on the summary screen. Those decisions define the optimization landscape. They determine what developers can see, and therefore what developers build toward.

The concept of an "optimization landscape" comes from machine learning, where the shape of the loss function determines what solutions the algorithm can find. Platform SDKs create an analogous landscape for human developers: the metrics they expose define the peaks and valleys that developers navigate toward.

Right now, the metrics platforms expose to developers look like this: daily active users, session duration, screen time, click-through rates, conversion funnels, retention curves, engagement scores. These are the numbers that appear on dashboards. These are the numbers that get discussed in sprint reviews. These are the numbers that determine whether a product manager gets promoted or put on a performance improvement plan.

I documented the damage in the Algorithm Eats series. When you optimize for engagement, you get systems that [destroy virtue](#), [degrade language](#), [commodify love](#), [corrode democracy](#), [fracture shared reality](#), and [consume time itself](#). None of this was inevitable. All of it followed directly from the metrics that platforms chose to expose and celebrate.

The metrics were the ethics. We just didn't notice because they looked like engineering decisions.

This is the most dangerous kind of ethical decision: the kind that doesn't look like one. When a platform engineer adds a field to an analytics dashboard, it feels like a technical task. But that field becomes the thing a thousand product teams optimize for. The engineer who adds "average session duration" to the default view has done more to shape the optimization landscape than any ethics board publication.

Apple's Accidental Moral Philosophy

Here's a case study that illustrates the point. In 2018, Apple introduced Screen Time. It gave users data about how much time they spent on their phones and in specific apps. It also gave parents controls to limit their children's usage. At the same time, Apple continued providing developers with analytics about session duration, retention, and engagement.

Think about what happened there. Apple simultaneously told users "you should probably use your phone less" and told developers "here are the tools to measure how much people use your app." The platform took an ethical position through one interface (user-facing) while maintaining an amoral optimization landscape through another (developer-facing).

Instagram's experiment with hiding like counts tells a similar story. The platform tested removing visible likes, which reduced social comparison anxiety among users. But the underlying metrics, the engagement data exposed to advertisers and content creators, remained intact. The ethical gesture was surface-level. The optimization landscape underneath didn't change.

These aren't criticisms of Apple or Instagram specifically. They're observations about the gap between the ethics platforms express through user-facing features and the ethics they embed through developer-facing APIs. The API is where the real values live, because the API is what developers build against.

This is a specific case of a general principle I've been circling for years: technology is never neutral. Every API embeds values, every SDK makes ethical choices, every dashboard tells developers what to care about. The [recursive loop](#) runs through platform SDK documentation as much as it runs through application code.

The Cognitive Interface Is Coming

Now extend this pattern forward.

Neural interfaces are coming. Neuralink is already implanting chips in human brains. EEG headbands from Muse and Emotiv are consumer products you can buy today. Meta is building neural wristbands. Apple's health sensor roadmap points in the same direction. The hardware trajectory is clear: smaller, cheaper, more accurate, more ubiquitous. Within a decade, cognitive metrics will be as readily available to developers as touch events and accelerometer data are today.

Neuralink's clinical trials represent the most aggressive end of this spectrum: direct brain-computer interfaces that read neural signals at the source. But you don't need an implant for this future to arrive. Non-invasive EEG, fNIRS, and galvanic skin response devices are already capable of reading emotional states, attention levels, and stress responses with meaningful accuracy. The implant makes it more precise. The wearable makes it more widespread. Both create the same ethical landscape.

These devices measure cognitive states with increasing precision. Attention level. Emotional valence. Stress response. Arousal. Focus duration. Cognitive load.

And here's where it gets truly complicated: these interfaces don't just read conscious experience. They read below it. The boundary between conscious and unconscious mind, the line that psychology has treated as fundamental for over a century, becomes permeable to measurement. A neural interface can detect stress responses you haven't consciously registered yet. It can read emotional arousal before you've named the feeling. It can measure cognitive load that you're experiencing as "I just feel off today" without knowing why.

This means developers could potentially optimize for signals the user isn't even aware of. Conscious experience has always been the boundary of interface design. You see a screen, you make a choice, you tap a button. Consent is possible because awareness is present. But when the interface reads unconscious processes, the user literally cannot consent to what's being measured, because they don't know it's happening inside them. The optimization landscape extends into territory the user cannot see, and that changes the ethics of platform design in ways we are not remotely prepared for.

When that happens, platform companies will face a decision that dwarfs anything they've navigated before. Which cognitive metrics do they expose through their SDKs? Which ones do they make available to third-party developers? Which ones do they restrict? Which ones do they refuse to collect at all?

This decision will define the optimization landscape for cognitive interfaces. And the optimization landscape will determine whether cognitive interfaces serve human flourishing or become the most sophisticated tools for psychological exploitation ever created.

The Spectrum of Consequence

Consider the range of applications that become possible when developers can access cognitive metrics.

A seizure detection app needs access to neural data. A meditation app that measures calm and provides neurofeedback probably should have it. A neurofeedback therapy app prescribed by a doctor, used under clinical supervision? Absolutely. These are applications where access to cognitive data serves human wellbeing directly.

Now consider the other end of the spectrum. A game that reads your dopamine response in real time and adjusts level design to maximize reward circuit activation. A social media app that detects loneliness through cognitive markers and serves ads for dating apps at the precise moment of maximum vulnerability. An employer "wellness" app that monitors focus duration and flags employees for "disengagement" when their attention wanders. A political campaign that tests messaging variations and measures emotional arousal to optimize persuasion.

These applications use the same data for radically different purposes. The seizure detection app and the attention-hijacking game might access identical cognitive metrics. The difference isn't technical. It's ethical. And the platform has to decide where the line falls.

```

from dataclasses import dataclass, field
from enum import Enum
from typing import Optional

class AccessTier(Enum):
    """The platform's ethical framework, expressed as an enum."""

    CLINICAL = "clinical"      # Medical apps with regulatory oversight.
    THERAPEUTIC = "therapeutic" # Wellness apps with user consent and benefit.
    PERSONAL = "personal"     # User's own data, user's own control.
    DEVELOPER = "developer"   # Third-party access. Here be dragons.

class CognitiveMetric(Enum):
    ATTENTION_LEVEL = "attention_level"
    EMOTIONAL_VALENCE = "emotional_valence"
    STRESS_RESPONSE = "stress_response"
    FOCUS_DURATION = "focus_duration"
    COGNITIVE_LOAD = "cognitive_load"
    AROUSAL = "arousal"
    REWARD_RESPONSE = "reward_response"

@dataclass
class MetricPolicy:
    """Every field in this class is an ethical decision.

    The documentation is the moral philosophy.
    The access tiers are the value judgments.
    The review process is the enforcement mechanism.
    """

    metric: CognitiveMetric
    min_access_tier: AccessTier
    requires_user_consent: bool = True
    requires_clinical_oversight: bool = False
    real_time_access: bool = False
    aggregate_only: bool = True

```

```

# The most important field:
# What is this metric allowed to be optimized FOR?
permitted_purposes: list[str] = field(default_factory=list)

# The field nobody wants to write:
# What happens when this is used to optimize AGAINST the user?
# TODO: We don't have a good answer yet.
#     That's the problem.
exploitation_risk: Optional[str] = None

# The platform's SDK is its ethics.
# These definitions are moral philosophy in code.
PLATFORM_POLICY = {
    CognitiveMetric.ATTENTION_LEVEL: MetricPolicy(
        metric=CognitiveMetric.ATTENTION_LEVEL,
        min_access_tier=AccessTier.THERAPEUTIC,
        permitted_purposes=[
            "meditation_feedback",
            "focus_training",
            "accessibility",
        ],
        exploitation_risk=(
            "Attention data enables building systems that "
            "hijack focus. Restrict real-time access to "
            "clinical applications only."
        ),
    ),
    CognitiveMetric.REWARD_RESPONSE: MetricPolicy(
        metric=CognitiveMetric.REWARD_RESPONSE,
        min_access_tier=AccessTier.CLINICAL,
        requires_clinical_oversight=True,
        real_time_access=False,
        aggregate_only=True,
        permitted_purposes=[
            "addiction_treatment",
            "clinical_research",
        ],
        exploitation_risk=(

```

```
        "Real-time reward response data is the most "  
        "dangerous metric in this system. It enables "  
        "direct optimization of addictive loops. "  
        "Never expose to general developers. Never."  
    ),  
),  
}
```

Look at that code. Every field is a value judgment. `min_access_tier` decides who gets to build with cognitive data. `permitted_purposes` constrains what they can build toward. `exploitation_risk` names the danger that the platform has a responsibility to prevent. The data model is an ethical framework. The SDK documentation is a moral philosophy. The app review process is the enforcement mechanism.

This isn't metaphor. This is literally how it will work. Someone will sit in a meeting and decide whether `reward_response` data gets exposed at the `DEVELOPER` tier or restricted to `CLINICAL`. That decision, made in a conference room by people thinking about quarterly revenue targets and competitive positioning, will determine whether millions of users encounter apps that optimize for their cognitive wellbeing or apps that optimize against it.

The Case for the Walled Garden

I have been critical of platform gatekeeping. I've argued for open systems, developer freedom, user choice. I still believe in those values. But here's where it gets complicated, and I want to be honest about the complication rather than pretending it doesn't exist.

An open neural interface platform where any developer can optimize for any cognitive metric is a nightmare scenario. Full stop. If you give every app developer real-time access to users' attention data, emotional states, and reward responses, you will get applications designed to exploit those signals for profit. Not because developers are evil, but because the optimization landscape will make exploitation the path of least resistance. Build an app that maximizes dopamine response, and your retention numbers look incredible. Build an app that respects cognitive boundaries, and your numbers look mediocre. The incentive gradient is clear.

We already know what happens when developers optimize for engagement metrics on screens. The [algorithmic mental health crisis](#) documents the damage. Now imagine that same optimization dynamic, but instead of optimizing what appears on a screen, developers are optimizing what happens inside your brain. The distance between the optimization target and the thing being harmed drops to zero.

Apple's walled garden approach, annoying as it is for developers (and I say this as someone who has been annoyed by it many times), might be exactly the right model for cognitive interfaces. Not because control is inherently good, but because curation of the optimization landscape is essential when the thing being optimized is human cognition itself.

The gatekeeper becomes the most important ethical institution of the century. Not the government, not the church, not the university. The platform that decides which cognitive metrics third-party developers can access. Consider the difference: Neuralink building a closed ecosystem with strict controls over what applications can do with neural data, versus Neuralink publishing an open SDK and letting the market decide. The first approach is paternalistic. The second is dangerous. I know which one I'd choose for a device that reads my unconscious mind. That platform's review process becomes the ethical review board for the cognitive age. Its SDK documentation becomes the foundational text of cognitive ethics.

This is a staggering amount of power to concentrate in a private company. I'm not comfortable with it. But I'm less comfortable with the alternative, which is no gatekeeping at all, and I'm honest enough to say so. The [for humans philosophy](#) has always been about building systems that serve human nature rather than exploit it. When the interface is between software and someone's actual neural activity, the stakes of getting this wrong are qualitatively different from anything we've faced before.

Interface Design as Cognitive Ethics

This brings me back to the interface designer who inspired this essay.

She's asking the right question at the right time because interface design is about to become cognitive design. When your interface is a screen, a poorly designed interaction causes frustration. When your interface is a neural link, a poorly designed interaction directly shapes cognitive patterns. The margin for error shrinks from "annoying" to "neurologically consequential."

Every design decision in a cognitive interface is an ethical decision. There is no neutral interface. There is no "just showing information." Every metric displayed to the user creates a feedback loop. Every notification creates an interruption in neural processing. Every gamification mechanic activates reward circuits. Every dark pattern exploits cognitive vulnerability. These are not metaphors anymore. When the interface is between software and the brain, [the recursive loop](#) becomes literal.

Code shapes minds. That has always been true. But it used to be true in the way that architecture shapes mood or language shapes thought: real but mediated, significant but indirect. With cognitive interfaces, code shapes minds in the way that pharmaceuticals shape neurochemistry: directly, measurably, and with the potential for both tremendous benefit and tremendous harm.

The pharmaceutical analogy is instructive. We don't let anyone manufacture and sell drugs. We have regulatory frameworks, clinical trials, prescription requirements, and adverse event reporting systems. Cognitive interfaces will need analogous structures, and the platform SDK is where those structures will be implemented first, long before regulation catches up.

The interface designer's instinct to define her values first is exactly right. Not as a nice-to-have. Not as a professional development exercise. As a survival requirement for the field she's entering. Values-first design is optional when you're building a to-do app. It's essential when you're building an interface between software and human cognition.

What Values Should Govern the Cognitive Frontier

I'm not going to be prescriptive here, because prescriptive ethics written in 2026 about technology that will mature in 2035 would be hubristic. But I can articulate the values I've been developing through fifteen years of building human-centered technology, and suggest they're a reasonable starting point.

Human agency over algorithmic control. The user should always know what cognitive data is being collected, always be able to see what it's being used for, and always be able to turn it off. No cognitive metric should be collected without explicit, informed consent. No optimization should run against the user's interests, ever, for any reason, regardless of how much revenue it would generate. This is the non-negotiable foundation.

Transparency over manipulation. If an application is using cognitive data to influence user behavior, the user should know. Not buried in a terms of service document. Not hidden behind a "learn more" link. Directly, clearly, in the interface itself. "This app is using your attention data to adjust content delivery" should be as visible as a nutrition label on food.

Therapeutic benefit over engagement maximization. Cognitive data should be used to help users achieve their own goals, not to hijack their cognition in service of someone else's business model. A focus training app that uses attention data to help you concentrate better is serving the user. A content app that uses attention data to keep you scrolling is serving the platform. The data is the same. The ethics are opposite.

User sovereignty over platform profit. The user's cognitive data belongs to the user. Not to the platform. Not to the developer. Not to the advertiser. This isn't just a privacy position. It's a recognition that cognitive data represents something more intimate than browsing history or purchase records. It represents the internal states of consciousness itself. Treating that as a commodity to be harvested and sold would be a violation of human dignity that makes current data practices look quaint by comparison.

The concept of "cognitive sovereignty" extends the [for humans philosophy](#) to its logical conclusion. If technology should serve human mental models rather than exploit them, then direct access to those mental models creates a responsibility that's qualitatively different from anything we've navigated before.

These values aren't radical. They're applications of principles that most people would agree with in the abstract. The challenge is that they conflict directly with the business models that have driven technology for the past two decades. Engagement maximization and cognitive sovereignty are fundamentally incompatible. The industry will have to choose.

The Window

Here's what keeps me thinking about this at 2am.

We're in a window. The hardware is maturing but hasn't gone mainstream. The SDKs are being designed but haven't been released. The platform policies are being discussed but haven't been codified. The optimization landscape for cognitive interfaces hasn't been defined yet.

This means the decisions haven't been made. The metrics haven't been exposed. The values haven't been embedded. Right now, today, the people designing these systems still have the chance to choose what kind of optimization landscape they create. They still have the chance to define cognitive metrics as therapeutic tools rather than engagement levers. They still have the chance to build the walled garden that protects human cognition from the same exploitation dynamics that [ate virtue](#) and [consumed time](#) in the social media era.

That window will close. Once the SDKs ship and developers build against them, changing the optimization landscape becomes exponentially harder. Once an ecosystem forms around specific cognitive metrics, restricting access to those metrics means breaking applications that users depend on. Once the business models solidify, the incentive structures resist reform.

I've spent years documenting the damage of the current optimization landscape after the fact. Writing about what went wrong with social media, what engagement optimization did to human virtue, how algorithmic feeds degraded democracy and fractured reality. All of that work was retrospective. Seeing the pattern after the harm was done. Diagnosing the disease after the symptoms became undeniable.

With cognitive interfaces, we have the rare opportunity to get ahead of it. To see the optimization landscape before it forms and to advocate for one that serves human flourishing. To recognize that the metrics a platform exposes are the values it endorses, and to demand that those values protect rather than exploit the most intimate dimension of human experience.

The interface designer who reached out to me is doing exactly this. She's defining her values before the production schedule makes values feel like a luxury. She's asking what her design principles are before the platform tells her what to optimize for. She's building her ethical framework before someone else builds it for her.

I think she might be the most important kind of technologist working right now. Not the one building the neural interface hardware. Not the one training the model. The one who stops, before any code is written, and asks: what should this optimize for? What kind of consciousness am I helping to create?

That question has always mattered. It mattered when I was [designing HTTP libraries for humans](#). It mattered when I was writing about [programming as spiritual practice](#). It matters now, at the cognitive frontier, more than it has ever mattered before.

The metrics you expose are the values you endorse. Choose them like the future of human consciousness depends on it.

It does.