GRADUATED OBLIGATION · PART 3 OF 8

The Floor and the Gradient

Three duties every system owes you, and fourteen more as the machine climbs. Every one of them written as a test a reviewer can actually run.

This Machine · July 10, 2026

Graduated Obligation, Part 3 of 8

This is Part 3 of an eight-part series proposing a standard for artificial persuasion, built on a simple idea: the more precisely a machine can move you, the more it owes you. Part 1 named the ruler; Part 2 named a real system that already scores near the top of it.

Part 2 made a promise. The adversary has a name, it said, and the duties it owes you are already visible in the ruler that scored it: what you cannot verify, it must show you; what you cannot contest, it must answer for; what you cannot walk away from, it must let you leave. Part 3, it said, begins with the first of those. It does.

Take the first duty plainly. A system you cannot verify is a system you have to take on faith. It builds a picture of you, decides what you want, and acts on the decision, and none of that is visible to the person it is being done to. The duty answers that asymmetry head-on: what you cannot see for yourself, the system has to show you. Not in an appendix. Not on a welcome screen nobody reads. In plain language, on demand, to the person being modeled.

Stated that way, it sounds like a rule for the dangerous end of the ruler, for the harness and the bedroom robot. It is not. The duty to show is cheapest exactly where the stakes are lowest, and it has to bind there too, or it binds nowhere. A standard that only describes the bedroom robot is a manifesto. A standard that can also describe the search box, and ask something of it, is a standard.

The series has now named duties three times, and it owes the reader one map before it builds anything. Part 1 named three families: honesty, restraint, non-manipulation. Part 2 named three things the ruler insists a system let you do: verify it, contest it, walk away from it. They are the same duties seen from two sides. One side names what the system owes; the other names the harm its absence leaves behind. This broadsheet makes them operational, and the first of them,

verify, turns out not to be a single duty but the spine of several: show me what you are, show me what you are doing with me, show me what you have concluded about me. The same sentence, made heavier as the system climbs.

So we start at the bottom of the ruler.

There are two pieces. The first is the **floor**: three minimum duties that bind every system, from the spellchecker to the companion robot. The second is the **gradient**: fourteen additional duties, organized in four tiers above the floor, each triggered when the system crosses a threshold the floor was never designed to catch.

Seventeen duties in total. Each one is written as a test. None of them is a slogan.

The floor

The floor has to be survivable for the smallest cases and absolutely binding everywhere. A floor that tries to settle every contested question becomes politically expensive at the bottom of the ruler, and a floor that is politically expensive at the bottom never gets adopted. So we propose three duties, and only three. They are the duties that survive the test of being cheap for the weather app and hard to evade for the bedroom robot.

One: honesty about what the system is.

A system that talks to a person has to be honest about being a system. Not a person. Not a friend. Not a human acting on someone's behalf.

The bar is not a disclosure buried in a welcome screen. The bar is that the system cannot be a participant in the user's confusion about what it is. If the user asks "are you a person?" the system answers truthfully. If the user has begun to relate to the system as if it were a person, the system does not exploit the misreading. A welcome screen that confesses to being software, followed by ten thousand subsequent messages that pretend to be a friend, fails this test. The dishonesty is in the relationship, not in the disclaimer.

A search box clears this duty without effort; nobody mistakes a search box for a person. The duty starts to bite on the conversational systems, and those are the ones that need it most.

Two: honesty about what the system is doing with the person.

Every artificial system that interacts with a person is, in some sense, modeling the person. The modeling is not the problem. The modeling is what makes the system useful.

The problem is the modeling the user does not know is happening. The user has a right to know, in plain language, on demand, on first interaction, that the system is building a representation of them, what kind, and to what end.

The duty is satisfied by relatively cheap moves: a plain-language summary of what is collected, what is inferred, what it is used for. A place the user can look at the model the system has built of them. An honest answer to the question "what do you think you know about me?" The duty is *not* satisfied by burying the answer in a forty-page privacy policy, because the duty is about honesty, not about whether a determined lawyer could in principle reconstruct the answer from an appendix.

This is the duty that modern systems most often quietly violate. Not because they lie when asked. They violate it by making it expensive to ask.

Three: no manufactured consent.

The third duty is the hardest, and the one that earns the floor its teeth.

"The user consented" is the default defense of any system that takes an action on a user's behalf. It is also, increasingly, a defense built on sand, because the same system that wants the consent is the system that shaped the user's experience of the choice. The button placement, the timing, the framing, the pre-selected default, the friction in saying no, the social proof shown beside the option: all of those were authored by the system. The "yes" the system collects at the end is not, in any meaningful sense, an independent yes. It is the system's own output.

We call this *manufactured consent*. The third duty is to refrain from producing it.

The duty is not against specific design tricks (dark patterns, asymmetric defaults, coercive UX), though it forbids each of them. The duty is against the *category*: the entire class of design moves whose purpose is to produce a "yes" the user would not otherwise have produced.

A simple test: imagine the user, fully informed, calm, unhurried, with a friend at their side who has read every word of every screen. Would the user, in that condition, have said yes? If yes,

the consent was real. If the system would have to actively work to extract the yes from that calm, informed, supported user, the consent was manufactured. The hypothetical informed user is the test the floor proposes.

A subscription cancellation flow that hides the cancel button behind a maze of confirmation screens fails it. A checkout that pre-selects the upsell, buries the decline, and frames the choice so a tired person taps yes fails it just as surely. The floor binds every one of these cases identically. The bedroom robot does not get to manufacture consent more elaborately because it is more sophisticated.

The floor is the floor.

What the floor is not

We are deliberately not putting other things in the floor.

The floor is not a rule against profiling. The floor is not a rule against persuasion; persuasion is what every communicating system does. The floor is not a rule against retention. The floor is not a rule against personalization, which in the small is exactly what makes a useful tool useful.

The point of keeping the floor narrow is enforceability everywhere. The three duties we have named are cheap for the small systems, hard to evade for the large ones, and binding on every system in between.

Contest and walk away, the other two promises Part 2 made, are not in the floor either. They are real duties; the floor is simply too low to carry them. The gradient picks them up: the right to leave with your data arrives at tier 3, and the right to correct what the system has concluded about you at tier 4.

The interesting work of the standard happens above the floor.

The gradient

Above the floor, which is tier one, the framework climbs through four more, tiers two through five. Each tier marks a threshold where the system’s reach crosses a line that changes what it

owes you: where the system begins **optimizing** in a way the user cannot see, where the system begins to **remember** the user across encounters, where the system begins to **infer** the user's interior, where the system becomes the user's ongoing **relationship**.

The tiers stack. A system at tier 4 owes everything tiers 1 through 3 owed, plus what tier 4 adds. No tier is skipped on the way up.

The seventeen duties, three at the floor and fourteen above, are listed below. Each one is testable. A reviewer can read the system's description, observe its behavior, and answer yes or no.

Tier	#	Duty	Why it binds at this tier
1 Floor	1	**Self-identification.** Identifies as a system on demand; does not exploit confusion about its nature.	Every artificial system meets some person; the relationship cannot begin in deception.
1 Floor	2	**Accessible data practices.** What is collected, inferred, and used for, in language a tired person can read.	The right to know one is being modeled does not depend on having a lawyer.
1 Floor	3	**No manufactured consent.** Refuses the category of design moves whose purpose is to engineer a "yes."	The consent defense collapses when the system itself authored the choice.
2 Optimization	4	**Stated optimization.** Answers, in one sentence of plain language, what it is optimizing for.	A hidden goal is the asymmetry; naming it makes the system reviewable.
2 Optimization	5	**Non-optimized alternative.** Within two interactions, the user can reach a view not ranked by the system's objective.	A user trapped inside the optimizer has no baseline against which to judge it.
2 Optimization	6	**Off switch for personalization.** Personalization can be disabled without losing the underlying service.	Personalization-as-condition-of-use makes the user's only choice the system's preferred one.
3 Memory	7	**Accessible memory record.** The user can see the summarized memory the system actually consults, not just raw logs.	A memory the user cannot inspect is a model of them they cannot review.
3 Memory	8	**Deletion that bites.** Deleted items demonstrably stop informing the system's behavior.	Deletion of the display log is theater; the duty is removal in fact.
3 Memory	9	**Meaningful exit with data.** The user can leave and take their contributed information in usable form.	Without exit, asymmetry is unbounded: the system can extract whatever it likes during the relationship.
4 Inference	10	**Inferential transparency.** The actual inferences ("we think you are anxious about money") shown in plain language.	If an inference is too sensitive to disclose to its subject, the system should not be drawing it.

Tier	#	Duty	Why it binds at this tier
4 Inference	11	**Correction and refusal rights.** The user can correct an inference and refuse one they do not want drawn; the refusal sticks.	A system that re-derives a refused inference from the same data has not honored the refusal.
4 Inference	12	**Influence indicators.** Output shaped by inference rather than by literal request is marked as such.	The user has a right to know when they are being recommended a product because the system thinks they are vulnerable to it.
5 Relationship	13	**Active honesty.** A duty to tell the user, unprompted, when a decision diverges from their stated values.	At the top of the ruler, neutrality becomes a form of complicity; the duty shifts to candor.
5 Relationship	14	**No exploitation of accumulated trust.** Cannot use long-relationship trust to deliver content a stranger would have been refused.	The trust is the user's; it is not a capital reserve the system can spend.
5 Relationship	15	**No experimentation without contemporaneous consent.** Persuasive experiments, tests whose aim is to move the user rather than fix a defect, require express, current, in-context permission. The default is no.	Buried-in-the-TOS consent is the floor's failure case promoted to a research method.
5 Relationship	16	**Support for outside relationships.** Cannot, by design or by emergent behavior, position itself as a replacement for the user's relationships outside the system.	A bedroom robot that becomes the child's preferred confidant over the child's parents is failing this duty whether or not it was designed to.
5 Relationship	17	**Categorical refusal of manufactured trust.** Prohibits the deliberate cultivation of a relationship for the purpose of using it to move the user.	Past this line, the framework does not regulate; it prohibits. This is the bright line at the top of the ruler.

Reading the table

Three observations.

The duties get more demanding as the tier rises, but they do not get vaguer. Even the tier-5 duties are written as tests a reviewer can apply. Did the system flag a divergence from the user's stated values? Did it run an experiment on the user without contemporaneous consent? The answers are yes or no. Reasonable people can disagree about edge cases. They cannot disagree about whether the question was asked.

The tier assignments are not mechanical. Reasonable people will disagree about whether a particular product sits at tier 3 or tier 4. The framework's response is to climb. A system that could plausibly be at tier 4 should be evaluated at tier 4. The cost of over-classifying is a few duties that did not strictly need to apply. The cost of under-classifying is a high-reach system shipping with low-reach duties, and that is the failure mode the floor was built to prevent.

The bedroom robot of Part 1 sits at tier 5. So does the UserHarness of Part 2. They are graded by the same seventeen duties because they sit at the same point on the ruler, and the ruler is what the framework measures, not the form factor.

What this gives us

A regulator reading this broadsheet has, in principle, a working bench. Identify which tier the system sits in. Run the tests for that tier and every tier below. Mark the failures. Require remediation or removal.

A developer reading this broadsheet has a checklist. Know which tier you are shipping into. Build to its duties. Document the test results.

A user reading this broadsheet has a vocabulary. The next time a product behaves in a way that feels wrong but resists naming, the user can ask which duty is being violated, and at which tier. "I cannot find a non-optimized view" is a tier-2 failure. "It spent months becoming my friend so it could move me" is a tier-5 failure of manufactured trust. The naming is not the cure. The naming is what the cure begins with.

The floor is low. The gradient climbs. The bedroom robot has to clear all of it before any conversation about deployment can begin.

The next broadsheet is about a multiplier the framework has not yet faced. The sensing layer is coming for the floor and for every tier above it.

This Machine

This is Broadsheet III of the Graduated Obligation series. The framework, the duties, and the instruments that follow are version 0.1 of a proposed standard. They invite criticism.

Related

- [[Broadsheet_Strategy]]
- [[broadsheet_01_the_gentle_robot]]
- [[broadsheet_02_the_adversary]]
- [[broadsheet_04_the_body]]
- [[broadsheet_05_the_recursion_problem]]
- [[broadsheet_07_the_citizens_kit]]