The IQ/MQ Public Layer

Freedom of Speech, Civic Labeling, and the Moral Visibility of Public Discourse

A Position Paper Extending the IQ/MQ Space Series

Speech remains free. It does not therefore remain unexamined.

Lev Neymotin, PhD

Uncompressed Thinking

N5Digital, Inc.

April 2026

Preamble

This position paper grows out of a shift in the IQ/MQ project. The first form of the IQ/MQ Filter was deliberately personal. It was imagined as a user-owned adaptive layer between platform and timeline, a private instrument by which an individual could govern the quality of what enters his or her own informational life. That original framing remains valid. The feed is already filtered, and the central question remains whether the filtering is hidden, engagement-driven, and platform-owned, or visible, adjustable, and meaningfully user-owned.

The present paper begins where that private model reaches a new boundary. If the IQ/MQ framework is applied not only to personal intake but to the public information space itself, the ideology, design, and maintenance of the system must change. A private filter may reflect user preference. A public layer must answer to civic legitimacy. A personal filter may ask what I want to see more or less of. A public layer must ask how speech circulating before the public may be labeled without becoming censorship, how civic risk may be made visible without becoming state control, and how moral judgment may be structured without reducing human beings to public scores.

This question has become more urgent as political violence, public targeting, and reputation warfare have moved into the center of contemporary life. Recent public discussion has focused on whether repeated derogatory language about public figures can accumulate into a stable negative identity-label, turning a person into a target in the imagination of unstable or highly suggestible actors. This paper does not attempt to explain any single act of violence and does not rest its argument on the unresolved facts of any single event. It uses the public discussion as a starting point for a broader question: can a civic system label speech patterns that participate in target construction, manipulation, or moral degradation while preserving freedom of speech itself?

The answer proposed here is yes, but only under strict conditions. The IQ/MQ Public Layer should not be a ban machine, a government instrument, or a platform-owned moral police system. It should be a transparent civic labeling layer for the public information space. Its purpose is not to decide who may speak. Its purpose is to help citizens see what kind of speech is being placed before them. In that sense, this paper marks a new stage in the IQ/MQ series. The private filter protects the user's attention. The public layer protects the visibility of civic structure.

That visibility cannot come from a black box. If the IQ/MQ Public Layer is to operate in the public information space, it must be more than a model that produces labels. It must be a live civic standard: created through a public charter, defined through a label registry, disciplined by an annotation guide, tested against a reference corpus, assisted by AI but not authored by AI alone, reviewed by humans where the stakes require it, open to appeal, versioned over time, and subject to audit. The central question is therefore not only what the Public Layer labels, but how the layer itself is created, maintained, corrected, and kept independent of political, governmental, platform, or crowd capture.

Abstract

This paper proposes the IQ/MQ Public Layer as a civic labeling framework for public discourse. It extends the prior IQ/MQ project beyond the user-owned personal filter, which was designed to help individuals shape their own informational environment, into a shared public layer that identifies morally and cognitively relevant features of speech circulating in the civic public sphere. The shift is significant. A personal filter may legitimately reflect user preference; a public layer requires transparency, contestability, public standards, resistance to ideological capture, and careful separation from coercive power.

The central claim is that freedom of speech is a right, but speech in the modern digital public sphere does not remain socially neutral simply because it is legally protected. A post may be lawful and still be manipulative, dehumanizing, evidentially weak, morally degrading, or part of a repeated pattern of target construction. The IQ/MQ Public Layer would not remove such speech merely because it is objectionable. It would label the relevant structure so that citizens can see, at the point of encounter, whether a statement is supported, distorted, dehumanizing, high-quality, high-risk, or participating in a recognizable pattern of civic harm.

The paper distinguishes speech rights from amplification, labeling from censorship, and public evaluation from state enforcement. It argues that public officials cannot be placed beyond public labeling, since democratic accountability depends on the ability of citizens to evaluate those who exercise power. At the same time, it insists that the public layer must label posts and speech patterns rather than assign permanent moral ranks to persons. It must distinguish severe criticism from target construction, forceful dissent from incitement, and moral warning from reputational violence.

The proposed architecture separates three layers: a public baseline layer of shared labels, a personal preference layer through which users decide how labels affect their own feeds, and a platform action layer where enforcement decisions are made under law and platform rules. But the paper further argues that this architecture is credible only if the public baseline itself is not a black box. The Public Layer must be created and maintained as a live civic standard, with a public charter, label registry, annotation guide, reference corpus, model documentation, human review thresholds, appeal routes, release notes, audit procedures, and protection against political, platform, governmental, or majoritarian capture. AI may assist in applying the standard at scale, but it must not become the silent author of the standard itself.

The paper concludes that the IQ/MQ Public Layer is plausible only if it is presented as a civic instrument rather than an authority over speech. Its purpose is not to restrict the right to speak, but to challenge the assumption that public speech must remain morally invisible. Speech remains free. It does not therefore remain unexamined.

1. Introduction: The New Boundary of the IQ/MQ Project

The IQ/MQ project began with a concern about the quality of public expression. The first formulation of MQ, or Moral Quotient, was not a claim to measure the soul, private virtue, or ultimate goodness. It was a proposed framework for evaluating observable public behavior: truth orientation, fairness, dignity, responsibility, relationship to harm, and the recurrent moral posture a person displays when speaking

before others. The next stage paired MQ with IQ, understood not as psychometric intelligence, but as the cognitive quality of observable public reasoning. Together, MQ and IQ formed a two-dimensional space for understanding public expression as morally and cognitively structured rather than merely popular, viral, or ideologically aligned.

The third stage translated that space into the idea of a user-owned filter. The IQ/MQ Filter was designed as a private adaptive layer between platform and timeline. Its core premise was simple: the feed is already filtered. The user is not deciding whether mediation exists, but whether mediation remains hidden and platform-owned or becomes visible, adjustable, and user-owned. The proposed personal filter did not censor the public sphere. It allowed an individual to shape his or her own informational intake according to more meaningful criteria than engagement, outrage, repetition, and platform retention.

The present paper asks what changes when the same underlying framework moves into the public domain. This is not merely a change of name. A public IQ/MQ system cannot simply be a personal filter enlarged. The private version begins with autonomy. The public version must begin with legitimacy. The private version answers to the user's declared and adaptive preferences. The public version must answer to shared standards, procedural fairness, public audit, and protection against ideological or governmental capture.

The term proposed here is therefore not simply the IQ/MQ Public Filter. The stronger term is the IQ/MQ Public Layer. The difference matters. A filter suggests exclusion. A layer suggests interpretation. The public layer is not first a mechanism for blocking speech. It is a civic annotation system that makes the moral and cognitive structure of speech visible while leaving the legal right to speak intact. It may influence prominence, warning, trust, and user decisions, but it should not present itself as a universal censor.

The governing question is not whether public speech should be judged. It already is. Citizens, journalists, platforms, institutions, activists, and audiences constantly label speech as truthful, false, hateful, manipulative, courageous, stupid, dangerous, brave, irresponsible, or noble. The problem is that these judgments often occur chaotically, tribally, and invisibly. The IQ/MQ Public Layer asks whether some of that judgment can be made more disciplined, more transparent, more contestable, and less dependent on the emotional swarm of the moment.

For that reason, the Public Layer must be described not only as an idea, but as a live operating structure. A label that appears in the public sphere carries weight even when it does not coerce. It may shape trust, attention, reputation, and public interpretation. The legitimacy of such a label therefore depends on the route by which it was produced. This paper argues that the Public Layer must be built as a maintained civic standard rather than as an invisible scoring engine. Its labels must come from published definitions, applied through disclosed procedures, revised through visible versioning, challenged through appeals, and tested through audit. Without that live structure, civic labeling becomes accusation. With it, labeling may become a disciplined form of public interpretation.

2. Freedom of Speech and the Meaning of Civic Labeling

Freedom of speech is a right. It is not a privilege granted by government, by platforms, or by polite society. In the American constitutional tradition, this right functions above all as a barrier against state suppression. The state does not bestow the citizen's freedom to speak as a favor. It is restricted in its

power to punish, license, or prohibit speech because the citizen's freedom is presumed to be fundamental.

That principle does not mean that every form of visibility, amplification, hosting, endorsement, or public trust is itself an absolute right. There is a difference between the right to speak and the right to be algorithmically boosted. There is a difference between the right to criticize a public figure and the right to have one's criticism presented to millions without context, warning, challenge, or counter-speech. There is also a difference between suppressing a statement and labeling its structure. A label does not necessarily silence. It may itself be a form of speech about speech.

This distinction is central to the IQ/MQ Public Layer. The layer does not begin by asking whether a statement should be illegal. Most of what it labels may remain fully protected speech. Its task is to ask a different question: what kind of public object is this speech creating? Is it offering evidence, argument, warning, exposure, analysis, criticism, satire, dehumanization, manipulation, target construction, or reputational violence? The fact that a statement may be protected does not make these distinctions irrelevant. It makes them more important, because in a free society the principal answer to harmful but lawful speech cannot be state suppression. It must be better speech, better visibility, better labeling, and better public judgment.

The old vocabulary of rights and privileges is helpful but incomplete. Speech is a right. Access to a particular amplifier may function as a privilege or contractual condition. But digital life has complicated the distinction because amplification now structures public reality. A platform's hidden ordering of content can make one statement visible and another invisible without formally censoring either. The public therefore needs a vocabulary not only for speech rights but also for the architecture of salience. The IQ/MQ Public Layer belongs to that second vocabulary. It concerns what becomes visible, how it is interpreted, and what kind of warning or confidence accompanies it.

This is why civic labeling should not be treated as an insult to free speech. It may be one of the forms through which free speech becomes more responsible without becoming less free. The right to speak protects expression from coercive silencing. It does not require society to pretend that all expressions are equal in structure, evidence, dignity, or civic consequence. Speech remains free, but it does not remain beyond interpretation.

3. The Public Domain Is Not an Institution, but It Can Have Civic Infrastructure

In earlier discussion, the question arose whether the public domain itself could be treated as an institution. The answer must be careful. In a philosophical sense, the public may perform an institutional function. Public norms, public expectations, public pressure, and public judgment shape behavior with enormous force. But legally and operationally, the public is not a single institution. It has no board, charter, accountable officer, election cycle, membership rule, or stable will. To say that the public as an institution grants or withdraws privileges risks creating confusion.

The better terms are the public information space, the civic public sphere, or the shared civic forum. These phrases preserve the seriousness of the domain without pretending that the public is a single sovereign body acting through one voice. The IQ/MQ Public Layer should therefore not be described as a tool of an institution called the public. It should be described as a civic tool operating within the public information space.

This distinction protects the project from two opposite errors. The first error is privatization, in which every label becomes merely one user's preference and no shared public standard exists at all. The second error is collectivist overreach, in which the supposed voice of the public becomes a majoritarian instrument for suppressing dissent. The public layer must occupy a third position. It should be public in visibility, standards, and accountability, but not majoritarian in the crude sense of letting a crowd define moral truth by volume.

The public layer is therefore institutional in function but not institutional in ownership. It requires governance, maintenance, published standards, appeals, audits, and version control. Yet its legitimacy cannot come from state power or from the emotional majority of the day. It must come from disciplined public procedure: open criteria, diverse review, transparent revisions, contestable outcomes, and a clear boundary between labeling speech and punishing speakers.

This point may become decisive as the project develops. Once a label appears in the public sphere, it carries authority. Even if it is non-coercive, it may affect reputation and trust. For that reason, the public layer cannot be casual. Its maintenance must be closer to stewardship than content moderation, closer to standards governance than platform preference, and closer to civic instrumentation than public shaming.

4. From Personal Filter to Public Layer

The personal IQ/MQ Filter and the public IQ/MQ Layer share a common conceptual ancestry, but they do not serve the same purpose. The personal filter protects the user's informational life. It allows the user to choose a target region in IQ/MQ space, set porosity, preserve challenge exposure, and train the system slowly so that platform-owned engagement logic is replaced by user-owned mediation logic. Its authority comes from the user's autonomy over his or her own attention.

The public layer has a different source of authority. It does not ask what one user prefers to receive. It asks what civic-risk features of speech should be visible to all who encounter it. A personal filter may legitimately be idiosyncratic. One user may dislike certain tones, subjects, formats, or styles. Another may seek intensity and conflict. A public layer cannot be built on idiosyncrasy. It must label features that are publicly defensible: evidentiary weakness, manipulative framing, dehumanization, target-construction risk, unsupported criminal accusation, contextual omission, or high civic value.

This creates a layered architecture. The public baseline layer supplies shared labels. The personal preference layer determines how much those labels influence a user's own feed. The platform action layer determines, under platform rules and law, whether content is removed, downranked, age-gated, referred for review, or left untouched. These layers must not be collapsed. If the public label automatically becomes enforcement, the system becomes censorial. If the personal layer can erase public labels entirely, the shared civic function disappears. If the platform owns the standard, user and public trust may collapse into platform dependency.

The right formulation is therefore not that public input overrides personal input in every respect. Rather, public input governs the maintenance of shared civic labels, while personal input governs how those labels affect the user's own informational intake. A user should not be able to redefine a high-risk target-construction label as harmless simply because he likes the target or dislikes the speaker. But the user may decide whether such a label hides the post, places it behind friction, lowers its prominence, or

leaves it visible with warning. Public meaning and personal mediation are related, but they are not identical.

This distinction preserves the strongest insight of the original filter while allowing the new public version to develop. The user still owns the personal environment. The public still gains a common language for speech structure. Neither side fully consumes the other.

5. Label Speech Acts, Not Speakers

The IQ/MQ Public Layer evaluates content as well as form: the claim being made, the evidence offered, the moral framing, the treatment of the target, the implied call to action, and the surrounding context. What it must not evaluate is the hidden essence of the person. A post may be labeled manipulative, unsupported, dehumanizing, target-constructing, or high-risk. A speaker should not be labeled as a low-MQ person, a dangerous soul, or a permanently degraded civic actor.

The first ethical discipline of the IQ/MQ Public Layer is that it must label speech behavior before it labels persons. This principle appeared in the earlier MQ and IQ/MQ frameworks, where the object of evaluation was observable public expression rather than inner worth. It becomes even more important in the public layer, because public labels can travel, harden, and become reputational facts in their own right.

A post can be labeled as containing dehumanizing language. A thread can be labeled as showing escalating target-construction risk. A claim can be labeled as evidentially weak, contextually unsupported, or manipulative in framing. A recurring pattern of public expression can be described as unstable, degrading, or high-risk. But the public layer should resist the move from such observations to a permanent label on the person as a whole. It should not say, without qualification, that a senator is low-MQ, a journalist is morally defective, or a citizen is intellectually unfit. It should say what can be defended: this expression, in this context, carries these identifiable features with this level of confidence.

This is not only a moral safeguard. It is also a practical and legal safeguard. Public officials, commentators, activists, and ordinary citizens may all be criticized. But factual claims about crime, corruption, treason, or personal misconduct can create defamation risks if made without adequate grounding. A label such as unsupported criminal accusation is safer and more precise than a label that repeats the accusation as if proven. A label such as high dehumanizing intensity is more defensible than a label that claims to know the speaker's hatred or inner motive.

The public layer should therefore remain behaviorally anchored. It can aggregate patterns, because patterns matter. Repetition is one of the ways public conduct becomes meaningful. But aggregation should remain contextual, probabilistic, and revisable. It should never become a social passport or permanent civic rank. The public layer may say that an account has repeatedly produced posts with low evidentiary support and high target-construction language. It should not reduce the whole person to a score that follows him into employment, banking, voting, legal rights, or basic social standing.

A society that labels nothing becomes blind. A society that labels persons too easily becomes cruel. The IQ/MQ Public Layer must live between those dangers. It must make speech visible without pretending to possess final knowledge of the speaker's soul.

6. What the Public Layer Labels

The public layer should label observable features of speech rather than ideological positions. A post should not be red because it is conservative, progressive, religious, secular, nationalist, internationalist, pro-government, anti-government, or radical in its conclusions. It should be red, amber, neutral, or green because of the way it behaves as speech. The distinction is essential. If labels follow ideology, the system becomes a factional weapon. If labels follow structure, the system has a chance to become civic infrastructure.

The relevant features belong to both IQ and MQ. On the cognitive side, the layer may identify low evidentiary support, weak coherence, inflated certainty, selective quotation, missing context, circular reasoning, conspiracy structure, or pseudo-argument. On the moral side, it may identify humiliation, dehumanization, scapegoating, cruelty as entertainment, reckless insinuation, deliberate misrepresentation, celebration of harm, and target-construction language. In many cases the two dimensions will interact. A post that makes an unsupported accusation against a named person and then frames that person as an existential enemy is both cognitively weak and morally dangerous.

The public layer should also be able to label constructive speech. Otherwise it will become only a warning system and will train the public to associate IQ/MQ with punishment. High-quality labels should identify posts that are evidence-grounded, fair to opponents, proportionate, clarifying under disagreement, contextually responsible, morally restrained, or unusually helpful in a difficult public moment. A civic labeling system must show not only what corrodes public discourse, but what improves it.

Color can help, but it must not dominate the system. A red label should mean high civic risk, not moral damnation. Amber should mean caution, unresolved context, or moderate risk. Green should signal high-quality civic contribution, not ideological approval. Blue or another neutral marker may indicate context added, public dispute, or provenance information. The exact visual language can be designed later. The conceptual rule is already clear: color should orient judgment, not replace it.

A label must also explain itself. A red mark with no reason becomes accusation. A concise explanation changes its character. If a post is labeled high target-construction risk, the user should be able to see whether the signal came from dehumanizing language, explicit harm language, operational targeting, repeated unsupported accusations, or thread escalation. The system must show enough of its reasoning to allow both trust and disagreement.

7. Target-Construction Speech

The most important new label proposed in this paper is target-construction risk. The term names a pattern that is not identical with criticism, insult, defamation, incitement, or threat, though it may overlap with all of them. Target-construction speech occurs when repeated public language converts a person or small group into a symbolic object of justified hostility. The target is no longer merely wrong, corrupt, misguided, harmful, or worthy of defeat. The target becomes an enemy whose continued presence is framed as intolerable.

This distinction is difficult but necessary. Democratic politics requires severe criticism of public figures. Citizens must be free to say that officials are incompetent, reckless, corrupt in policy, destructive in judgment, morally failed, or unfit for office. A civic system that cannot tolerate harsh criticism of power is

not a free system. The public layer must therefore protect sharp criticism as such. It should not label severity itself as risk.

Target construction is different. It begins when criticism mutates into a social architecture of personal danger. The speech repeatedly attaches degrading identity-labels to a named person, treats the person as an existential contaminant, denies ordinary human dignity, uses language of elimination or cleansing, circulates fantasies of removal, or frames any restraint toward the person as complicity. The post may stop short of an explicit threat. It may even remain technically ambiguous. But it participates in a pattern that makes a human being easier to imagine as an object to be acted upon.

This does not mean that speech causes violence in a simple mechanical way. Human behavior is too complex for that. Most people exposed to extreme rhetoric do not commit violence. Many violent actors have private pathologies, situational triggers, and motives that cannot be reduced to public discourse. The public layer should avoid diagnosing speakers or audiences. Its task is not to say that a post will cause an act. Its task is to say that a post participates in a recognizable civic-risk pattern.

The distinction between causation and risk is crucial. A society already labels many risk conditions without claiming certainty. Smoke does not prove fire, but it justifies attention. Repeated target-construction language does not prove imminent violence, but it justifies civic visibility. The purpose of the label is to interrupt normalization before the pattern becomes ambient. It says to the reader: this is not ordinary criticism; this speech is participating in the construction of a target.

Such a label should be especially careful when applied to public officials, because public officials must remain open to intense criticism. Yet it is precisely around public officials that target construction can become most dangerous. Officials exercise power and therefore invite opposition. They also become symbols onto which large-scale grievance can be projected. The public layer must preserve the right to oppose officials while identifying speech that transforms opposition into personal targeting.

8. Red Labels and the Boundary Between Warning and Ban

A red label is plausible within the IQ/MQ Public Layer, but a red label should not automatically mean a ban. This is one of the paper's central design decisions. Red should mean that the system has identified a high-risk speech pattern: direct dehumanization, explicit or coded harm language, target construction, operational targeting, doxxing, celebration of violence, incitement-like escalation, or some combination of these features. It should alert readers, slow reflexive sharing, and make the structure visible. It should not by itself erase the speech from the public record.

The reason is both philosophical and practical. Philosophically, the public layer must not confuse civic judgment with coercion. Practically, many high-risk statements require context. A journalist may quote dangerous language in order to expose it. A researcher may analyze extremist rhetoric. A critic may condemn a threat by repeating part of it. A legal document may cite violent language as evidence. A system that treats every dangerous phrase as removable without context will soon become both foolish and unjust.

The better model uses graduated interventions. A red label may trigger a visible warning, contextual explanation, sharing friction, reduced algorithmic prominence, clustering to prevent repetition, or human review. Removal should be reserved for narrower categories where speech crosses into direct threats, operational targeting, doxxing, incitement of imminent lawless action, clear calls for violence, or platform-

rule violations. The public layer may flag such cases for escalation, but it should not become the sole enforcement authority.

This distinction allows the system to mark what law may not punish. In the United States, much harmful or reckless speech remains protected. That is not a defect of the First Amendment; it is part of the design of a free society. But protected speech may still be criticized, labeled, contextualized, and countered. A red label can therefore perform a crucial civic function in the zone between legal permissibility and moral irresponsibility. It marks speech that free society may tolerate legally but should not treat as invisible.

The phrase to preserve is simple: protected but labeled. This phrase may become one of the central principles of the IQ/MQ Public Layer. It avoids the false choice between censorship and blindness. It allows a society to protect the right to speak while refusing to surrender the right to interpret.

9. Public Officials and the Right to Public Evaluation

A democratic society cannot exempt public officials from public labeling. Presidents, members of Congress, senators, judges, governors, mayors, agency heads, and other officials exercise public power. The right to criticize, rate, evaluate, accuse, defend, and judge such officials is not peripheral to freedom of speech. It lies near its center. Public life already contains advocacy scorecards, watchdog ratings, editorial endorsements, public trust measures, fact checks, and institutional labels of many kinds. The IQ/MQ Public Layer would be more systematic, but it would not invent the practice of public evaluation.

Government therefore should not be able to say that the public has no right to label officials. Such a rule would protect power from scrutiny. It would be especially troubling if it applied only to unfavorable labels, because it would become viewpoint discrimination in favor of those who govern. The public may praise officials as courageous, honest, disciplined, and high-quality. It must also be free to label speech about or by officials as unsupported, manipulative, cruel, reckless, or high-risk when the label is grounded in observable features.

The public layer must still be careful. It should not casually declare that a public official is a criminal, traitor, agent, thief, or murderer unless such claims are legally established or properly attributed. It should instead label the speech act. A post may be marked as an unsupported criminal accusation. A thread may be marked as high target-construction risk. A claim may be marked as low evidentiary support. These labels protect the system from becoming a publisher of the very unproven accusations it is trying to interpret.

This is where the distinction between officials and speech about officials becomes important. The public layer is strongest when it evaluates the public expression itself, including the expression of officials, media figures, citizens, and organized groups. It is weaker and more dangerous when it purports to assign fixed moral ranks to public figures as persons. The system may eventually show patterns associated with public actors, but those patterns should remain attached to evidence, context, time, and specific speech behaviors.

Democracy requires public judgment. It also requires that judgment not become mob branding. The IQ/MQ Public Layer can serve democratic accountability only if it preserves both sides: no immunity for officials, and no careless transformation of officials into permanent targets.

10. Constitutional Boundaries in the American Context

The IQ/MQ Public Layer should be designed with constitutional caution, especially in the American context. The First Amendment does not protect every form of expression, but the categories of unprotected speech are limited. Advocacy of force or law violation cannot be punished merely because it is abstractly dangerous; the Brandenburg standard protects advocacy unless it is directed to producing imminent lawless action and likely to produce such action. True threats are also outside full protection, but modern doctrine requires attention to the speaker's subjective awareness of the threatening character of the speech; Counterman v. Colorado placed at least a recklessness requirement on true-threat prosecutions. Public officials suing for defamation face the actual malice standard of New York Times Co. v. Sullivan, requiring knowledge of falsity or reckless disregard for truth. These doctrines are not minor technicalities. They express a constitutional preference for breathing space around public debate.

The public layer should not attempt to collapse these legal thresholds into its own labels. Its labels are civic, not criminal. A post can fall below the threshold of incitement and still be labeled high target-construction risk. A statement can fall short of a true threat and still be marked as dehumanizing or dangerously ambiguous. A claim about an official can be legally protected and still be marked as evidentially weak. The layer should therefore be more linguistically sensitive than law, but less coercive than law. It should see more, but punish less.

The doctrine of editorial discretion is also relevant. Private editors, newspapers, platforms, and civic annotation systems engage in speech when they select, arrange, label, or contextualize expression. Miami Herald v. Tornillo rejected a state-mandated right of reply that intruded into editorial judgment. Moody v. NetChoice recognized that content moderation and organization by social platforms can implicate protected editorial discretion, even when carried out through modern technical means. A civic labeling layer would likewise need protection from government attempts to dictate which labels may be applied to officials or favored viewpoints.

At the same time, the public layer must guard against becoming a vehicle for indirect government censorship. The state may speak, warn, request, and provide information, but it may not use threats or coercive pressure to force private actors to suppress protected speech. The history of informal censorship, including cases such as Bantam Books, remains relevant. Recent litigation over government contacts with social platforms has further highlighted the difficulty of distinguishing persuasion from coercion. The IQ/MQ Public Layer should therefore be structurally independent of government control. It may use legal categories as references, but it should not become an unofficial enforcement arm of the state.

The American legal framework produces a clear design instruction. The public layer may label speech, contextualize speech, and help users decide how to treat speech. It should not punish citizens, allocate rights, or act as a proxy for government suppression. Its legitimacy depends on remaining a civic instrument of interpretation rather than a coercive instrument of law.

11. The Public Standard and the Danger of Majoritarian Capture

Public input is necessary for the IQ/MQ Public Layer, but public input cannot mean raw majority control. A system that simply asks the crowd whether a post is immoral or dangerous will reproduce the worst tendencies of the public sphere: outrage waves, factional pressure, organized reporting campaigns,

ideological brigading, and emotional overreaction. The public layer must learn from the public without becoming a plebiscite of anger.

The appropriate model is deliberative public maintenance. Standards should be developed through published criteria, cross-perspective review, expert input, civil liberties review, technical audit, and periodic public comment. Public complaints should matter, but they should not automatically decide. A large number of complaints may indicate that a label deserves review. It should not by itself prove that a post is high-risk. The system must distinguish signal from mobilization.

This is one reason cross-perspective agreement is valuable. Public context systems such as Community Notes have explored the idea that a note becomes more credible when people who often disagree nonetheless rate it helpful. The IQ/MQ Public Layer can borrow the underlying principle without becoming a clone of any existing system. A high-risk label should become stronger when observers from different political or cultural positions can identify the same structural problem: unsupported accusation, dehumanization, target construction, manipulative omission, or direct harm language. Such agreement does not guarantee truth, but it helps resist factional capture.

The public standard must also be versioned. Moral and cognitive labels will evolve. New forms of manipulation will emerge. Old categories will prove too broad or too narrow. The system should therefore maintain a visible history of changes to label definitions, thresholds, examples, and appeals outcomes. A user should be able to know not only what a label means today, but how that meaning changed over time. Without versioning, the public layer becomes unstable and unaccountable.

The public layer should also include a constitutional culture of humility. It should not claim to have solved morality. It should present itself as a structured and revisable civic lens. The goal is not final judgment. The goal is better public visibility than the opaque, emotional, and often manipulative labeling that already governs digital life.

The question is therefore not whether public input is needed, but how public input can be structured without becoming crowd rule. That answer belongs not to abstract theory alone, but to the live architecture of the Public Layer itself: how its standards are created, maintained, challenged, revised, and protected against capture.

12. Operational Architecture: Object, Context, and Label Vector

The operational architecture of the public layer should borrow from the companion design document while adapting its purpose. The personal filter processes a Content Object, attaches a Context Envelope, generates a Placement Vector, and compares the result with the user's Declared Profile, Adaptive Profile, Challenge Budget, and Drift Governor. The public layer can use a similar object-context logic, but the output is different. Instead of personal prominence alone, it generates civic labels that may travel with the content across users and surfaces.

The primary unit remains the post in context. A post should not be evaluated as naked text detached from its thread, author pattern, topic cluster, event frame, quotation status, and real-world relevance. A statement that appears dangerous in isolation may be a quotation under critique. A harsh accusation may be part of legal reporting. A vivid emotional statement may be justified by urgent reality. Conversely, a polished and moderate-sounding post may carry manipulative omissions or repeated unsupported

insinuations. The public layer needs context because speech does not act as isolated grammar. It acts as situated public behavior.

Each evaluated item should therefore receive a public label vector rather than a single verdict. The vector may include cognitive quality signals, moral conduct signals, civic-risk signals, confidence level, and explanation tokens. The label shown to the public may be compact, but the underlying structure should remain inspectable. Users should be able to open the label and see why it was applied. Auditors should be able to examine aggregate behavior across topics, ideologies, officials, and events.

The system should also separate automated detection from public labeling. Automated models can identify candidate signals, cluster similar patterns, detect dehumanizing language, compare claims to known context, and surface target-construction risk. But high-impact public labels, especially red labels attached to named persons or active political events, should be subject to stronger confidence requirements and, where appropriate, human review. Automation can scale attention. It should not be treated as moral authority.

This architecture allows the public layer to be both practical and cautious. It does not need to solve every philosophical question before functioning. It needs to identify observable structures, attach explanations, preserve context, allow challenge, and maintain confidence discipline. These are technical requirements, but they are also moral requirements. A system that labels public speech without confidence discipline will quickly become part of the problem it was built to address.

This operational architecture explains what the Public Layer processes and how labels are attached to public speech acts. It does not yet answer the institutional question of legitimacy. A public label cannot be credible merely because the system can generate it. It must come from a maintained standard whose route from speech act to label remains visible, revisable, and contestable.

13. The Live Public Layer: Creation, Standards, Governance, and Maintenance

The IQ/MQ Public Layer cannot be treated as a finished model that silently receives public speech and emits labels. If it were designed that way, it would become precisely what the project is trying to avoid: an opaque authority inserted into public discourse. The legitimacy of the Public Layer depends not only on what it labels, but on how the labeling standard is created, how it is applied, how it is corrected, and how it remains accountable over time.

The Individual IQ/MQ Filter is maintained primarily by the user. It has a Declared Profile, an Adaptive Profile, a calibration process, a Drift Governor, and a correction pathway through which the user's own informational preferences remain central. The Public Layer is different. It cannot be governed by the private preference of each individual user, because its labels must carry shared civic meaning. A label such as "Target-Construction Risk," "Low Evidentiary Support," or "Dehumanizing Language" must mean roughly the same thing across users, platforms, viewpoints, and controversies. Otherwise the Public Layer dissolves into private taste.

This does not mean the Public Layer should become a central authority over speech. It means that it must be built as a public standard before it is implemented as a technical system. The model may assist the standard, but the model cannot replace the standard. AI may detect patterns, rank confidence, cluster related posts, and support reviewers. It must not become the hidden author of civic judgment. The path

from speech act to label must be visible enough to be inspected, challenged, improved, and, when necessary, reversed.

13.1. The Public Layer as a Living Civic Standard

The phrase “Public Layer” should be understood literally. The system is not only a classifier. It is a living civic layer attached to the public information space. It consists of a label vocabulary, a set of definitions, an annotation guide, a reference corpus of examples, a technical implementation, a review process, an appeals process, a revision process, and a public record of changes. If any of these elements is missing, the layer becomes either too vague to trust or too powerful to leave unexamined.

The first design requirement is therefore institutional modesty. The Public Layer should not present itself as the moral voice of society. It should present itself as a structured interpretive standard for public speech acts. Its claim is not that it knows the soul of the speaker or the final truth of every controversy. Its claim is narrower: some public expressions contain identifiable features that matter for civic understanding, and those features can be labeled more honestly if the criteria are explicit.

A live Public Layer is not static. Language changes. Manipulative techniques evolve. Coded forms of dehumanization appear and disappear. Political factions learn to imitate the language of moderation while preserving the substance of attack. Coordinated networks learn which words trigger review and which insinuations can pass unnoticed. A public labeling system that does not adapt will become obsolete. A public labeling system that adapts invisibly will become untrusted. The Public Layer therefore requires a visible maintenance process that combines stability with revision.

The appropriate analogy is not a police code and not a platform terms-of-service page. It is closer to a maintained public standard. A standard does not change with every controversy, but neither does it freeze forever. It is updated through reasoned revision, documentation, testing, and public explanation. That is the level of discipline the Public Layer needs.

13.2. Creation Sequence: From Charter to Model

The Public Layer should be created in a defined sequence. The first stage is a Public Layer Charter. The charter states the purpose of the layer, the limits of its authority, the speech acts it may label, the objects it must not label, the relation between labels and enforcement, and the anti-coercion principle. It should say plainly that the Public Layer labels public content and speech behavior, not the permanent worth of persons. It should also say that labels are not legal verdicts, criminal findings, employment recommendations, credit signals, or civic rank.

The second stage is the Label Registry. The registry lists the approved labels, their families, their definitions, their exclusions, their intensity bands, and their visual presentation. A label is not legitimate merely because the system can generate it. It becomes legitimate only when its meaning is public. For example, “Target-Construction Risk” must not be an emotional synonym for “harsh criticism.” It must be defined through recognizable features, such as repeated enemy construction, dehumanizing or eliminationist framing, unsupported criminalization, operational targeting cues, or language that frames harm as necessary or cleansing. The registry should also state what the label does not cover, including forceful policy criticism, satire, documented investigative reporting, or legally protected opinion that does not participate in targeting structure.

The third stage is the Annotation Guide. This is the working manual through which human reviewers, model trainers, auditors, and public challengers understand how labels are applied. The guide should

contain examples, borderline cases, exclusion cases, and common failure modes. It should distinguish quotation from endorsement, criticism from target construction, urgency from manipulation, and public-interest exposure from reputational degradation. Without an annotation guide, the label system will drift according to the instincts of whoever is applying it at the moment.

The fourth stage is the Reference Corpus. This is a curated and periodically refreshed set of example speech acts used for training, testing, review, and public explanation. It should include ideological, cultural, and topical diversity. It should include examples that are easy, examples that are difficult, and examples that the system previously misclassified. The corpus should not be treated as a secret moral archive. Sensitive or legally problematic material may require redaction, but the categories and representative patterns should be available for inspection.

Only after these stages should the technical model be trained or deployed. This order matters. If the model is created first and the standard is reverse-engineered afterward, the system becomes a black box with a civics label pasted on top. The Public Layer must begin as a disclosed standard, then become a model-assisted standard. It should not begin as a model and ask the public to trust its invisible intuitions.

13.3. The Institutional Form: Independent Standard, Not State Office and Not Platform Department

The Public Layer requires a home, but its home should not be the state. Government may have a legitimate role in enforcing laws against true threats, incitement, doxxing, defamation, and related unlawful conduct, but the IQ/MQ Public Layer is not a legal enforcement instrument. It is a civic interpretive layer. If the state controls its definitions or commands its labels, the layer becomes vulnerable to becoming indirect censorship. The independence of the layer from government power is therefore a structural requirement, not a decorative preference.

Nor should the Public Layer simply become a department inside a platform. Platforms already possess immense power over visibility. If the same entity that profits from engagement also controls the civic label standard, the system risks becoming another form of platform mediation, only with a more elevated vocabulary. A platform may implement or display Public Layer labels. It may supply data, provide integration, and participate in technical review. But it should not quietly own the standard.

The strongest long-term form is an independent public-interest standards body or open standards consortium. Such a body would maintain the charter, label registry, annotation guide, release history, audit requirements, and dispute procedures. It should include technical expertise, civil-liberties review, editorial judgment, moral and philosophical competence, and cross-perspective public participation. No single ideological group, government office, platform, advertiser class, or activist network should be able to define the standard alone.

At the earliest stage, the founding author or founding organization may necessarily create the first version of the standard. That is normal. New frameworks often begin with a small group or even one author. But if the Public Layer moves from position paper to live civic infrastructure, its maintenance must gradually move beyond founding authority. The creator may remain intellectually central, but the standard cannot remain credible if every definition, threshold, and revision depends indefinitely on one person’s judgment.

13.4. The Live Labeling Pipeline

In the live Public Layer, content should move through a disciplined route. A public speech act enters the system as a content object. That object may be a post, video caption, thread segment, quote-post, image with text, public statement, or other unit of expression. The system then attaches a context envelope. This context includes the immediate thread, the author pattern where relevant, source history where relevant, the topic cluster, the event frame, the quotation status, and available evidence surrounding the claim. The point is not to excuse harmful speech by burying it in context, but to prevent irresponsible labeling by ignoring context.

The system then produces candidate label signals. Some signals will be cognitive, such as low evidentiary support, context deficit, misleading framing, or high analytical quality. Others will be moral, such as dehumanizing language, humiliating framing, responsible critique, or constructive contribution. A third family will concern civic risk, especially target-construction, unsupported criminalization, mobilization risk, and instability amplification. A fourth family may concern route and handling, such as high uncertainty, reality-pass, or challenge item in the user-facing environment.

Candidate labels should not all be treated alike. A low-impact informational label may be applied with moderate confidence if it is clearly explained and easy to contest. A high-impact red label attached to an active public figure, an unfolding event, or a named private person should require stronger confidence and a higher review threshold. The system should therefore evaluate not only the label category, but the consequence of applying the label in a live public environment. The more serious the label, the stronger the obligation to explain, review, and allow challenge.

The final public label should carry more than a word. It should carry a visible title, an intensity band, a confidence indication when appropriate, and an expandable explanation. The expanded view should explain which features triggered the label, what context was considered, whether the label is preliminary or stable, and what version of the label definition was used. A reader should be able to understand the label without trusting a hidden machine. A speaker should be able to contest the label without guessing at the accusation.

The live pipeline should also include a status distinction. Some labels may be stable. Others may be provisional. In fast-moving events, a label may properly state that evidence is incomplete, context is developing, or interpretation remains uncertain. This is especially important during crises, violence, elections, war, or sudden allegations. A public system that rushes to moral certainty under conditions of incomplete information will become another accelerant. The Public Layer should be allowed to say, “high uncertainty,” “context developing,” or “preliminary caution,” rather than pretending that every live event can be resolved immediately.

13.5. Automation, Human Review, and the Escalation Ladder

The Public Layer should use automation, but it should not worship automation. Automated systems are necessary because public discourse is too large and too fast for human review alone. Models can detect candidate patterns, cluster related content, identify repeated accusations, flag likely dehumanizing phrasing, compare claims against known context, and track the spread of target-construction language across threads. But automation is a tool of attention, not a source of final authority.

The correct structure is an escalation ladder. Low-risk labels may be generated automatically when confidence is high and the category is relatively mild. Moderate labels may be generated automatically

but sampled frequently for review. High-impact red labels should require stronger thresholds, and in many cases human review before public display. Labels connected to named individuals, public officials, violent events, alleged crimes, or volatile political controversies should be treated with special caution. These are precisely the cases in which the public most needs guidance, but also the cases in which mistaken labels can do the most damage.

Human review itself must be structured. It should not mean one moderator's instinct. Reviewers should apply the annotation guide, record reasons, identify uncertainty, and distinguish between personal disagreement and label-relevant features. For sensitive categories, cross-perspective review is especially important. A label becomes more credible when people who may disagree politically or culturally can identify the same structural feature in the speech act. This does not mean the system should require unanimous agreement across all viewpoints. It means the review process should be designed to resist factional reflex.

The escalation ladder also protects speed. A live public system cannot send every label through the slowest process. But it also cannot allow the most serious labels to appear as instant machine verdicts. Different label categories, impact levels, confidence levels, and contexts require different routes. A mature Public Layer should therefore publish not only label definitions, but review thresholds. Users should know which kinds of labels are automatic, which are sampled, which are human-reviewed, and which are treated as provisional pending further context.

13.6. Public Input Without Crowd Rule

Public input is necessary, but raw public input is not governance. The Public Layer must learn from users, speakers, journalists, researchers, affected communities, civil-liberties advocates, and ordinary readers. It must also protect itself from brigading, outrage waves, coordinated reporting, partisan pressure, and mass emotional reaction. The public should be able to challenge labels, submit examples, report misclassifications, propose revisions, and identify emerging manipulation patterns. The public should not be able to vote a label into existence simply by force of numbers.

The difference between input and command must remain clear. A large number of complaints about a post may indicate that the system should review the post. It should not prove that the post is dangerous. Similarly, a large number of defenses may indicate controversy. It should not prove that the label is wrong. The Public Layer should treat public reaction as evidence to be examined, not as a verdict to be obeyed.

This is one of the most delicate parts of the proposal. The Public Layer must be public enough to avoid becoming a private priesthood of labels, but disciplined enough to avoid becoming the crowd with better interface design. The ideal is deliberative maintenance, not plebiscitary judgment. Public input enters the process through challenges, examples, comments on definitions, audit signals, and review requests. Final label standards are maintained through documented criteria, cross-perspective review, evidence of errors, and published revision decisions.

Such a structure also protects minority viewpoints and heterodox speech. A public label standard that simply follows majority discomfort would punish precisely the speech that public discourse often needs: dissent, criticism of institutions, exposure of corruption, unpopular moral warnings, and minority cultural perspectives. The Public Layer must be able to distinguish discomfort from degradation. It must also be

able to distinguish controversial truth from target construction and serious dissent from manipulative hostility.

13.7. Appeals, Corrections, and Disputed Labels

A public label must be contestable. Contestability is not a courtesy to speakers. It is a condition of legitimacy. Because speech is contextual, because language is ambiguous, and because models and reviewers both make mistakes, a label that cannot be challenged becomes an accusation without due process. The Public Layer should therefore include a visible appeal path for speakers, publishers, readers, researchers, and affected parties.

An appeal should be able to raise several kinds of objection. The content may have been quoted rather than endorsed. The post may have been satire, rebuttal, legal reporting, documentation of wrongdoing, or warning rather than incitement. The label may have ignored evidence, misread context, over-weighted tone, or applied a definition too broadly. A label may also be technically correct but visually too severe, poorly explained, or assigned an intensity band that exceeds the evidence.

Appeals should not automatically remove labels. If every challenge suspended a label, coordinated actors would learn to neutralize the system by flooding it with disputes. But appeals should force the label to show its route. Where the appeal is persuasive, the label should be corrected, narrowed, downgraded, removed, or marked as disputed. Where the label is upheld, the reason should be stated. Where the underlying standard is contested, the dispute should be recorded as a standards issue rather than buried as a one-off complaint.

The system should maintain a Correction Ledger. This ledger records patterns of successful appeals, repeated misclassifications, label categories that generate confusion, topics with high error rates, and examples that should be added to the annotation guide. The purpose of the ledger is not only to correct individual posts. It is to teach the system where its standard or implementation is weak. A mature Public Layer learns from its mistakes openly.

13.8. Versioning, Release Notes, and Public Memory

The Public Layer must have version control. Every label definition, exclusion rule, intensity band, review threshold, and model release should have a version identity. If the definition of “Target-Construction Risk” changes, the public should know what changed and why. If “Unsupported Criminalization” is divided into separate labels for criminal accusation, treason accusation, and corruption insinuation, that change should appear in release notes. If the model is updated to detect coded dehumanization more effectively, the update should be documented.

Versioning matters because labels do not exist outside time. A post labeled under one standard may not be labeled the same way under a later standard. That is not necessarily corruption. It may be learning. But without public memory, learning looks like arbitrary power. A standards system that changes silently trains distrust even when its changes are reasonable.

The Public Layer should therefore publish release notes in ordinary language. Technical documentation may exist for auditors and developers, but civic labels require civic explanation. The public should be able to read what changed, what errors prompted the change, what risks the change is meant to address, and what safeguards have been added. The same applies to visual label formats. If colors, symbols, or severity bands change, that change should be visible rather than quietly imposed.

Versioning also protects historical interpretation. When researchers, journalists, or citizens review the history of a controversy, they should be able to know which label standard was active at the time. Without this, public labels become unstable artifacts detached from their own conditions of production. A live Public Layer must therefore preserve not only labels, but the memory of how labeling standards evolved.

13.9. Audit, Bias Testing, and Ideological Asymmetry

The Public Layer will inevitably be accused of bias. Some accusations will be bad-faith attempts to avoid accountability. Others will be serious. The system must be designed so that neither type of accusation controls the process. The answer is audit, not denial.

Audit should examine both standards and outcomes. Standards audit asks whether definitions are viewpoint-neutral, whether exclusions protect legitimate dissent, whether categories are being smuggled into ideological conformity, and whether certain traditions of speech are being misunderstood. Outcome audit asks how labels are distributed across topics, parties, officials, movements, languages, and cultural contexts. It asks whether one side is being mislabeled, whether one side is genuinely producing more of a label-relevant pattern in the sampled data, or whether the model is detecting surface style rather than structural behavior.

Equal outcomes should not be assumed. A system committed to fairness should not require that every political faction receive identical numbers of labels. Different actors may behave differently at different times. But unequal outcomes must be interpretable. If labels concentrate heavily in one direction, the system should be able to explain whether that concentration reflects observed behavior, sampling bias, model weakness, reviewer culture, coordinated reporting, or a flawed standard. A system that cannot ask those questions cannot be trusted.

Audit must also include adversarial testing. Once labels matter, actors will learn to game them. They will avoid obvious trigger words while preserving hostile structure. They will use humor, implication, selective quotation, coded terms, and plausible deniability. They will imitate high-MQ language while creating low-MQ effects. The Public Layer must therefore test itself not only against yesterday's explicit speech, but against tomorrow's evasive speech. The deeper object is not vocabulary. It is structure.

13.10. Independence, Capture Resistance, and Funding Discipline

A public labeling layer cannot be legitimate if it is financially, politically, or technically captured by the forces it is supposed to help the public interpret. Capture can come from government, platforms, donors, advertisers, political movements, activist networks, or professional guilds. The danger is not that any of these groups may participate. The danger is that any one of them becomes dominant enough to define the standard in its own interest.

The governance model should therefore include capture resistance. Funding sources should be disclosed. Major institutional partnerships should be visible. Platform integrations should not grant platforms quiet control over label definitions. Government grants, if ever accepted, should not permit government direction of labels. Donor support should not buy category changes. The Public Layer should develop a conflict-of-interest policy before it becomes influential enough for capture to matter, because by then it may be too late.

Technical capture is equally important. If the only implementation of the standard sits inside one proprietary platform, the public cannot know whether the label standard is being applied consistently or

tuned for private goals. An open protocol or standards-based implementation would be stronger. Multiple clients, platforms, or user-side tools could display or apply the same standard while remaining independently auditable. This is the difference between a public layer and a private feature.

The Public Layer should also remain independent of the personal filter. The shared label standard may inform the Individual IQ/MQ Filter, but the individual user should still decide how labels affect personal exposure. Public meaning and personal mediation are related, but they are not identical. Confusing them would damage both.

13.11. Relation to Platform Enforcement and Law

The Public Layer should be separate from platform enforcement and from law. This separation is essential. A public label may say that a post is low-evidence, dehumanizing, target-constructing, or civically risky. That does not automatically mean the post should be removed. Removal, suspension, demonetization, legal referral, or other coercive actions belong to separate layers governed by platform rules and law. The Public Layer should inform those layers only with caution and only under defined procedures.

This separation helps preserve freedom of speech. A post can be lawful and still deserve a warning. It can be morally ugly and still protected. It can be high-risk as a civic pattern and still fall short of incitement, threat, or defamation. The Public Layer is valuable precisely because it can identify structures that law should not necessarily punish. It sees more than law, but it must punish less than law. Its native action is visibility, not coercion.

Platforms may choose to use labels as one input in ranking, friction, or moderation decisions, but that relationship should be disclosed. If a red civic-risk label automatically reduces reach on a platform, users should know that. If labels are purely informational, users should know that too. Hidden consequence is one of the ways a label becomes censorship by another route. The distinction between labeling and enforcement must therefore be visible in the interface and in the policy.

The same principle applies to government. The state may not use the Public Layer as a proxy enforcement mechanism. It should not request labels on political opponents, demand removal of labels from officials, or pressure platforms to treat labels as enforcement triggers. The Public Layer's independence from state direction is part of its constitutional culture.

13.12. The Minimum Legitimacy Package

If the IQ/MQ Public Layer is to move from concept to live operation, it needs a minimum legitimacy package. This package is not a bureaucratic luxury. It is the difference between a civic standard and a black box.

The package begins with a public charter, a label registry, an annotation guide, and a reference corpus. It continues with model documentation, review thresholds, appeal routes, correction logs, release notes, audit reports, and conflict-of-interest disclosures. It also includes a clear statement of what labels do not do: they do not assign civic rank, determine rights, replace law, or convert public officials into protected objects beyond criticism.

These components do not all need to be perfect at launch. A first version may be modest. It may cover only a small label set, a limited number of platforms, or a narrow class of public speech acts. But even a limited version should show the full architecture of accountability. The public must be able to see that the

system has a standard, that the standard has definitions, that definitions have examples, that examples guide application, that application can be challenged, and that challenges can change the system.

The central rule is simple: a public label is legitimate only when the path to that label can be explained. Explanation does not mean that everyone will agree. It means that disagreement has an object. The reader can see the basis. The speaker can challenge the basis. The auditor can test the basis. The maintainer can revise the basis. Without that route, the Public Layer becomes merely a moral assertion with technical force.

13.13. What “Live” Means in Practice

A live Public Layer is not merely present on a screen. It is alive in the sense that it is maintained. It has standards, reviews, corrections, revisions, audits, and memory. It learns without drifting invisibly. It changes without pretending it has always been the same. It applies labels without claiming moral omniscience. It allows public input without surrendering to crowd rule. It uses AI without allowing AI to become the unseen author of civic judgment.

In practice, this means the Public Layer would operate in three time scales. In the immediate time scale, it annotates public content, often with caution or provisional status when events are moving quickly. In the medium time scale, it processes appeals, corrections, reviewer disagreements, and emerging patterns. In the long time scale, it revises definitions, updates models, publishes audits, and strengthens the standard against new forms of manipulation. A system without these three time scales is not a live public layer. It is a static labeling product.

The result is a different kind of civic infrastructure. The Public Layer does not decide who may speak. It does not relieve citizens of judgment. It does not make politics clean, conflict harmless, or public discourse comfortable. Its more modest and more serious purpose is to make the structure of public speech more visible while keeping the standard itself visible to public examination. If the layer asks citizens to look more carefully at speech, it must allow citizens to look just as carefully at the layer.

The Public Layer therefore stands or falls on a final principle: no public label without a public route. The label may be compact, but its basis must not be hidden. A system that makes speech visible while hiding itself would betray its own purpose. A system that shows both the speech act and the path by which it was labeled can become something more defensible: not a censor, not a tribunal, but a maintained civic lens for the public information space.

14. Labeling as Counter-Speech, Not Censorship

The most persuasive defense of the IQ/MQ Public Layer is that it functions as structured counter-speech. It does not say, 'You may not speak.' It says, 'This speech has identifiable features that citizens should see.' It does not remove the speaker from the public sphere. It adds an interpretive layer around the speech so that the reader is not left alone with the platform's raw amplification logic.

This distinction matters because the word censorship is often used too broadly and too narrowly at the same time. It is used too broadly when any criticism, label, fact-check, warning, or loss of private amplification is called censorship. It is used too narrowly when only formal legal prohibition is recognized, while hidden pressure, algorithmic invisibility, and institutional coercion are ignored. The public layer must avoid both mistakes. It should not pretend that labels have no force. They do. But force is not identical with censorship. A label is legitimate when it is transparent, contestable, proportionate, and

non-coercive. It becomes censorial when it is secret, punitive, state-directed, or tied to rights and penalties.

The public layer therefore needs a disciplined vocabulary of consequence. It can inform, warn, contextualize, slow sharing, reduce repetition, cluster duplicates, and recommend review. These are not identical acts. Informing a reader that a post contains unsupported allegations is not the same as blocking the post. Adding friction before resharing high-risk content is not the same as criminalizing the speaker. Removing a doxxing post is not the same as labeling a harsh criticism. A serious system must preserve these distinctions rather than allowing all interventions to collapse into one emotional category.

The layer should also make clear when a label reflects civic risk rather than legal illegality. This is especially important for target-construction speech. A post may receive a high-risk warning because it dehumanizes a named official and participates in a pattern of personal targeting, even though it is not a true threat and cannot be punished by government. The label does not claim that the post is illegal. It claims that the post is publicly dangerous in structure.

The public layer stands or falls on this difference. If it becomes a censorship machine, it betrays freedom of speech. If it refuses to label harmful structures because they are legally protected, it betrays civic responsibility. Its purpose is to occupy the space between legality and judgment, where free societies do much of their most important moral work.

15. Public Officials, Targeting, and the Ethics of Severe Criticism

The most difficult test of the public layer will be speech about public officials. Such speech is often severe because public power can produce severe consequences. Citizens must be able to accuse officials of failure, corruption in policy, moral blindness, abuse of office, constitutional betrayal, cruelty, incompetence, and dangerous judgment. A public layer that softens all criticism of officials in the name of safety would become an instrument of power. It would be worse than useless.

The layer must therefore distinguish severity from targeting. Severe criticism remains attached to actions, policies, decisions, evidence, and public accountability. It may be angry. It may be morally forceful. It may call for resignation, impeachment, prosecution, electoral defeat, investigation, or public condemnation. Targeting crosses a different line. It detaches the person from accountable action and turns him into a totalized object of contempt, danger, or elimination. It encourages the audience to stop thinking of the official as a human being exercising public power and to begin seeing him as a contaminating presence whose removal becomes morally urgent in a personal sense.

The distinction is imperfect, but democratic life depends on making it. Criticism of power must remain protected. Personal target construction must become visible. The public layer can help by labeling the structure rather than the viewpoint. A post saying that an official's policy is destructive and should be defeated is not the same as a post saying the official is vermin, a traitor beyond law, an enemy who must be stopped by any means, or a creature whose existence itself endangers the people. The first is political speech. The second participates in target construction.

Even here, context remains essential. Historical quotation, legal analysis, satire, literary reference, and warning about extremist rhetoric may use severe language without endorsing it. The system must therefore examine whether the speaker is affirming, condemning, reporting, parodying, or analyzing the

language. If it cannot determine the difference, confidence should fall and the label should become more cautious.

A healthy public layer would make public discourse harder to weaponize without making public officials harder to criticize. That is the balance. It is demanding, but it is the balance without which the project cannot deserve public trust.

16. The Anti-Coercion Principle

The IQ/MQ Public Layer must be built around an anti-coercion principle. It should not allocate rights, eligibility, employment, credit, banking access, licenses, public benefits, legal status, travel permissions, or civic standing. It should not become a social passport. It should not be used by employers to screen citizens by moral-cognitive labels, by banks to judge customers, by schools to rank students' civic worth, or by government agencies to condition access to public life. Those uses would transform a civic lens into a coercive scoring regime.

This concern is not imaginary. Contemporary AI policy already recognizes the danger of social scoring systems that classify persons in ways that lead to detrimental treatment unrelated to the context of the behavior assessed. The IQ/MQ project must therefore draw a bright line. Public speech may be labeled. Public patterns may be interpreted. Users may decide how labels affect their own feeds. But institutional penalties should not be attached to IQ/MQ labels as though they were a civic credit score.

The anti-coercion principle also affects data retention. The public layer should not build permanent person-centered dossiers except where public pattern analysis is necessary and proportionate. Post-level labels should remain primary. Aggregated actor-level summaries, where used, should be contextual, time-bounded, explainable, and contestable. The system should not quietly create a universal moral file on every citizen who speaks online.

This principle may limit some attractive features. It would be technically possible to produce leaderboards, author ranks, public shame lists, or employer-readable trust scores. The public layer should resist these temptations. The moment the system becomes a spectacle of ranking persons, it will reproduce the very moral appetite it is meant to discipline. The goal is civic visibility, not reputational blood sport with better math.

A public layer worthy of trust must remain modest about its authority. It sees speech patterns. It does not own persons. It labels civic features. It does not govern civil status.

17. Public Input and Personal Override

The question of override must be handled precisely. In the personal filter, the user's declared preference governs. The system may adapt slowly, but it remains subordinate to the user's chosen orientation. In the public layer, the label standard cannot be rewritten by each individual user, because then it ceases to be public. A high target-construction label must mean roughly the same thing for different users. A low evidentiary-support label must not vanish merely because a user likes the conclusion. Public labels require semantic stability.

Yet this does not mean that public input fully overrides personal autonomy. The public layer defines the civic label. The personal layer defines the feed consequence. A user may choose to see red-labeled

content with warnings because he is a researcher, journalist, lawyer, or citizen trying to understand a controversy. Another user may choose to place such content behind friction. A third may choose to strongly attenuate it. The public label remains common; the personal mediation remains owned.

This design has an important advantage. It prevents the public layer from becoming paternalistic while preventing the personal filter from becoming blind. A user can decide how to live with the label, but not erase the label's public meaning. A civic system can warn without commanding. It can say, 'This speech has these features,' while leaving the user some agency over what follows in his own environment.

There will be hard cases. Some users may want to disable all public labels. Some may see labels themselves as political manipulation. Some may want labels to apply only to opponents and not allies. The system should allow display customization, but it should be cautious about total invisibility. At minimum, content carrying high-risk labels should remain inspectable even if the user chooses a light-touch interface. A civic warning system that can be fully erased by preference becomes ineffective precisely when it is most needed.

The guiding sentence is this: public labels are shared civic meaning; personal filters are individual exposure decisions. Confusing these two will damage the project. Preserving their distinction gives it a workable foundation.

18. The Relationship to Existing Labeling Systems

The IQ/MQ Public Layer would not appear in a vacuum. Existing systems already label public information. Fact-checkers label claims. Community annotation systems add context. Platforms apply warning screens. Provenance systems such as Content Credentials identify how media was created or modified. News organizations use editorial labels. Advocacy groups produce scorecards. Each of these systems captures part of the public need, but none replaces the IQ/MQ proposal.

Fact-checking is necessary but insufficient. A statement may be factually accurate and still be morally destructive in its framing, timing, omissions, or target-construction function. Conversely, a statement may be factually disputed but offered in good faith with proper uncertainty. The IQ/MQ layer does not replace truth assessment; it adds moral and cognitive structure around it.

Community notes and public context systems are also valuable, especially when they require cross-perspective agreement. But they generally focus on missing context, misleading claims, or factual correction. The IQ/MQ Public Layer is broader. It asks not only whether context is missing, but whether the speech behaves responsibly, whether it dehumanizes, whether it reasons coherently, whether it is proportionate to evidence, and whether it contributes to target construction or civic degradation.

Provenance systems solve a different problem. They help users know whether a digital object has a traceable origin or editing history. That is important in an age of synthetic media. But provenance is not moral quality. A verified authentic video can still be used manipulatively. A real quote can still be framed dishonestly. The public layer would complement provenance by evaluating speech behavior after authenticity questions have been addressed.

The IQ/MQ Public Layer should therefore be interoperable rather than imperial. It should learn from fact-checking, community annotation, AI risk management, and provenance standards, while maintaining its distinct role: the moral and cognitive visibility of public speech.

19. Objections and Replies

The first objection is that the public layer is censorship by another name. The reply is that labeling is not the same as suppression when it is transparent, contestable, and non-coercive. A society that permits speech must also permit speech about speech. The public layer becomes censorship only if it is tied to coercive penalties, hidden suppression, or state command. Its proper form is structured counter-speech.

The second objection is that the system will be biased. This is a serious concern. Any evaluative framework touching morality and intelligence can be captured by ideology. The answer is not to pretend neutrality is easy. The answer is to build standards around observable structure rather than viewpoint, publish definitions, audit outcomes, require cross-perspective review, and permit appeal. Bias cannot be eliminated by declaration. It can only be made harder to hide.

The third objection is that labels will chill speech. Some chilling is possible, especially if labels are careless or reputationally severe. But the absence of labels also chills speech, because people withdraw from public life when it is saturated with harassment, humiliation, and target construction. The goal is not to eliminate all pressure from public discourse. The goal is to replace chaotic and often brutal pressure with more disciplined visibility. A system that labels dehumanization may chill dehumanization. That is not necessarily a defect.

The fourth objection is that public officials will use the system against critics. This is precisely why the public layer must be independent of government and designed to protect severe criticism of power. The standard must distinguish criticism from targeting, and the appeals process must not become a tool for powerful figures to suppress unfavorable labels. Public officials should have no special immunity from labels and no special power to remove them.

The fifth objection is that the system will reduce complex speech to colors. This would be true if the visible label were the whole system. It must not be. Color should be an entry point into explanation, not a substitute for judgment. The public layer should expose the basis of labels and allow users to inspect context. The interface may be simple; the underlying reasoning must remain accessible.

The sixth objection is that the project invites moral arrogance. This is the deepest objection. Any attempt to formalize moral evaluation risks becoming self-righteous. The answer is not to abandon moral language but to discipline it. The IQ/MQ Public Layer must be modest, revisable, transparent, and aware that it evaluates public traces rather than souls. If it forgets that humility, it will deserve the criticism it receives.

20. The Place of This Paper in the IQ/MQ Series

This paper should be understood as a fourth movement rather than a replacement for the earlier documents. The MQ paper introduced Moral Quotient as a framework for evaluating the moral quality of observable public behavior. The IQ/MQ Space paper paired MQ with Intellectual Quotient, understood as the cognitive quality of public reasoning, and proposed a two-dimensional evaluative field. The IQ/MQ Filter paper then moved from conceptual space to user mediation, proposing a user-owned adaptive layer between platform and timeline. The companion document translated that proposal into operational design language: Content Objects, Context Envelopes, Placement Vectors, Control Deck, Drift Governor, Surface Actions, and profile ownership.

The Lashon Hara essay added a deeper moral ancestry to the project by placing modern speech mediation beside an ancient discipline of harmful speech. It reminded the project that truth alone does not exhaust moral responsibility and that speech can damage dignity and trust even when it is not simply false. That essay remains especially relevant to the public layer, because civic labeling must avoid becoming a new appetite for public condemnation.

The present paper addresses a new question. What happens when the IQ/MQ framework is no longer only a private mediation tool but also a public labeling layer? This shift introduces freedom of speech, public officials, civic standards, governance, red labels, target-construction risk, and the danger of public or governmental capture. It therefore belongs to the same project but occupies a distinct position in it.

The simplest map is this: MQ defines the moral dimension; IQ/MQ Space defines the coordinate field; the Personal Filter gives the user control over intake; the Operational Design outlines how the system may be built; Lashon Hara deepens the ethics of speech; the Public Layer brings the framework into the civic domain of shared labels and public accountability.

This paper therefore does not close the project. It opens the next design phase. The next artifact should likely be an operational companion for the public layer, analogous to the companion design document for the personal filter. That future document would define label registries, governance workflow, appeal procedures, public audit methods, data structures, and interface treatment for public annotations.

21. Conclusion: Free Speech and Moral Visibility

Freedom of speech remains a right. The IQ/MQ Public Layer should begin from that principle and never abandon it. It should not become a government censor, a platform enforcement engine, a civic credit score, or a moral ranking system for human beings. If it becomes any of those, it will violate the deepest safeguards required by the project itself.

Yet freedom of speech does not require moral invisibility. A free society is not obligated to treat all speech as equally coherent, responsible, dignified, or safe in its civic structure. It may protect a statement from legal punishment while labeling it as manipulative. It may tolerate severe criticism while identifying target construction. It may allow citizens to criticize officials while marking unsupported accusations, dehumanizing language, and escalation patterns. The right to speak does not include a right to make the moral and cognitive structure of speech disappear.

This is the space the IQ/MQ Public Layer is meant to occupy. It stands between censorship and blindness. It offers structured counter-speech rather than suppression. It labels speech acts, claims, content structures, and public patterns rather than speakers, souls, identities, or permanent civic worth. It protects severe dissent while making target construction visible. It separates public labels from personal feed choices and from platform enforcement. It insists that public officials remain fully open to evaluation, while also insisting that no person should be casually transformed into a symbolic object of harm.

The project is plausible only if it remains transparent, contestable, non-coercive, and resistant to capture. It must show its work. It must allow appeal. It must preserve context. It must treat red labels as warnings, not automatic bans. It must distinguish lawful speech from healthy speech, and unhealthy

speech from illegal speech. It must recognize that the public domain is not a single institution, but it can still develop civic infrastructure worthy of a free people.

The final claim is therefore modest but serious. The IQ/MQ Public Layer does not decide who may speak. It helps citizens see what kind of speech is being placed before them. In an age where amplification can turn words into atmosphere and atmosphere into public reality, that visibility is not a luxury. It may become one of the conditions of a more livable public sphere. Speech remains free. It does not therefore remain unexamined.

The project is plausible only if the Public Layer is not merely a labeling model, but a maintained civic standard. It must show its work. It must allow appeal. It must preserve context. It must publish its label definitions, exclusions, thresholds, version history, and correction procedures. It must treat red labels as warnings, not automatic bans. It must distinguish lawful speech from healthy speech, and unhealthy speech from illegal speech. It must recognize that the public domain is not a single institution, but it can still develop civic infrastructure worthy of a free people. A public label is legitimate only when the path to that label can be explained.

References and Authorities

Brandenburg v. Ohio, 395 U.S. 444 (1969). The case is cited for the American constitutional distinction between protected advocacy and incitement directed to imminent lawless action. Source consulted: Cornell Legal Information Institute, https://www.law.cornell.edu/supremecourt/text/395/444.

New York Times Co. v. Sullivan, 376 U.S. 254 (1964). The case is cited for the actual malice standard applicable to public officials in defamation cases. Source consulted: Justia Supreme Court Center, https://supreme.justia.com/cases/federal/us/376/254/.

Counterman v. Colorado, 600 U.S. 66 (2023). The case is cited for the requirement of subjective awareness, at least recklessness, in true-threat prosecutions. Source consulted: Supreme Court of the United States opinion PDF, https://www.supremecourt.gov/opinions/22pdf/22-138_43j7.pdf.

Miami Herald Publishing Co. v. Tornillo, 418 U.S. 241 (1974). The case is cited for the constitutional protection of editorial judgment against compelled publication. Source consulted: Justia Supreme Court Center, https://supreme.justia.com/cases/federal/us/418/241/.

Moody v. NetChoice, LLC, 603 U.S. ___ (2024). The case is cited for its discussion of social media content moderation and editorial discretion. Source consulted: Supreme Court of the United States opinion PDF, https://www.supremecourt.gov/opinions/23pdf/22-277_d18f.pdf.

Murthy v. Missouri, 603 U.S. ___ (2024). The case is cited for the contemporary boundary questions surrounding government contacts with platforms and claims of coercion. Source consulted: Supreme Court of the United States opinion PDF, https://www.supremecourt.gov/opinions/23pdf/23-411_3dq3.pdf.

NIST Artificial Intelligence Risk Management Framework. The framework is cited for governance, mapping, measurement, and management as recurring elements of trustworthy AI risk practice. Source consulted: National Institute of Standards and Technology, https://www.nist.gov/itl/ai-risk-management-framework.

European Union Artificial Intelligence Act, Article 5 and related guidance. The Act is cited as a warning regarding prohibited social scoring and unacceptable-risk AI practices. Source consulted: European Commission guidance and Article 5 summaries, including https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act and https://artificialintelligenceact.eu/article/5/.

X Community Notes documentation. The documentation is cited as an example of public context added to posts when contributors from different viewpoints rate a note helpful. Source consulted: https://help.x.com/en/using-x/community-notes.

C2PA and Content Credentials. The standard is cited as an example of a public-facing label layer for digital content provenance, distinct from but complementary to the IQ/MQ Public Layer. Source consulted: https://c2pa.org/.

About the Author

Lev Neymotin, PhD, is the founder of Uncompressed Thinking, an independent research and communication initiative focused on structured frameworks at the intersection of science, technology, media, ethics, and public discourse. His recent work develops Moral Quotient (MQ), the IQ/MQ evaluative space, the IQ/MQ Filter, and related essays on speech ethics, public expression, and user-owned mediation in the digital information environment.

Appendix A. The IQ/MQ Public Label Set: Meaning, Structure, and Visual Language

The IQ/MQ Public Layer depends on visible labels, but those labels must remain disciplined in both meaning and presentation. They are not moral verdicts on persons, parties, institutions, religions, communities, or sources as fixed objects. They are public annotations attached to speech acts, claims, threads, and other units of public expression. Their purpose is to make cognitive quality, moral posture, and civic-risk features more legible to the reader without converting the system into a hidden censorship machine. For that reason, the label set should be intelligible, limited in number, visually recognizable, and structurally modest.

The most important design principle is not that the system avoids content. It does not. The Public Layer evaluates content as well as form. It may evaluate the claim being made, the evidence offered, the moral framing, the treatment of the target, the implied call to action, and the surrounding context. What it must not evaluate is the hidden essence of the person behind the speech. A post may be labeled manipulative, unsupported, dehumanizing, target-constructing, or high-risk. A speaker should not be labeled as a low-MQ person, a dangerous soul, or a permanently degraded civic actor.

A.1. The governing rule: label the speech act, not the speaker

The phrase 'label the speech act, not the speaker' should govern the whole public label system. A speech act is broad enough to include a post, reply, video, image caption, repost with endorsement, public claim, thread segment, or coordinated repetition pattern. It includes both content and expression: what is said, how it is said, what is implied, what evidence is used, what dignity is preserved or violated, and what civic risk is produced by the statement in context. The speaker, by contrast, remains a human being whose complete moral and intellectual reality cannot be collapsed into the label attached to a single public act.

This rule also protects the system from a common error. Source history may inform confidence, and repeated patterns may influence interpretation, but source history should not replace content-level judgment. A source may have a record of unreliability, yet a particular statement may still be accurate. A generally reliable source may still publish a distorted or morally careless claim. The Public Layer should therefore use author and source history as context, not as a substitute for analysis.

A.2. Four label families

The proposed label set is organized into four families. IQ labels describe the cognitive quality of the content: its relationship to evidence, context, proportionality, coherence, and framing. MQ labels describe the moral quality of the content: its relationship to dignity, fairness, restraint, dehumanization, humiliation, and constructive public conduct. Civic-Risk labels identify special danger patterns in the public sphere, especially where speech contributes to target construction, mobilization pressure, or unstable amplification around named persons or groups. Route labels explain how the system handled an item: whether it entered as challenge content, passed because of urgent reality, was clustered to reduce repetition, or was marked uncertain because context was insufficient.

These families should remain distinct. A post can be cognitively weak without being morally degrading. It can be morally fair while still lacking evidence. It can be emotionally intense yet justified by urgent reality. It can carry civic risk even when some of its factual content is not false. Distinguishing the families prevents the Public Layer from flattening different kinds of judgment into one opaque warning.

A.3. Visual family identity

A label should be readable in words, color, and shape. Color should never be the only signal, because color alone is inaccessible to some users and too emotionally crude for a serious civic system. Each family should therefore have a stable code and a simple shape identity. The following visual language is proposed as a first version rather than a final interface standard.

Code	Family	Shape / color cue	Primary use
■ IQ	Cognitive quality	Rectangular tag	Evidence, context, coherence, proportionality, framing
● MQ	Moral quality	Rounded or capsule tag	Dignity, fairness, restraint, dehumanization, constructive conduct
▲ CR	Civic risk	Triangular or shield-like marker	Target construction, mobilization pressure, operational targeting, instability amplification
◆ RT	Route or operation	Diamond or neutral pill marker	Challenge passage, reality-pass, uncertainty, clustering, deferral

A.4. Proposed core label set

The first public version should remain deliberately small. An oversized label vocabulary would make the layer difficult to understand, easy to misread, and vulnerable to ideological gaming. The following table proposes an initial set large enough to cover the main cognitive, moral, civic-risk, and operational categories, but restrained enough to remain legible.

Family	Label	Meaning	Use and limits
IQ	Low Evidentiary Support	The claim is asserted with insufficient visible support, weak corroboration, or unclear grounding.	This label should not mean the claim is false. It means the claim asks for more confidence than its presented support earns.
IQ	Context Deficit	Important context is missing, making the content easy to misunderstand or overinterpret.	This label should be used where omitted context materially affects interpretation, not merely where every possible background detail is absent.
IQ	Misleading Framing	The content uses selective emphasis, asymmetrical comparison, or distorted presentation to alter the reader's impression.	This label is about framing structure, not ideological disagreement.
IQ	Disproportionate Certainty	The tone or conclusion exceeds what the evidence can responsibly support.	This label is useful when a statement may contain some truth but overstates scale, certainty, or implication.
IQ	High Analytical Quality	The content shows clarity, coherence, evidence awareness, proportion, and meaningful engagement with complexity.	This label should be positive but not decorative. It should be reserved for content that genuinely improves understanding.
MQ	Humiliating / Derogatory Framing	The content relies on belittlement, contempt, mockery, or degradation as a primary mode of public treatment.	This label should not be used merely because criticism is sharp. It concerns the degrading method of treatment.
MQ	Dehumanizing Language	The content reduces a person or group to a subhuman, contaminating, parasitic, monstrous, or disposable category.	This is a serious label and should generally require high confidence and context review.
MQ	Moral Asymmetry	The content applies moral principles unevenly, excusing in allies what it condemns in opponents, or reversing standards opportunistically.	This label should be attached to demonstrated asymmetry in the speech act or pattern, not guessed motive.
MQ	Responsible Critique	The content criticizes forcefully while preserving evidence, proportion, and the basic dignity of the target.	This label affirms that criticism and moral seriousness can coexist.

Family	Label	Meaning	Use and limits
MQ	Constructive Contribution	The content adds clarification, repair, balance, restraint, or improved understanding to the public discussion.	This label should be reserved for content that moves the conversation toward greater public usefulness.
CR	Target-Construction Risk	The content contributes to a pattern in which a named person or group is built into an object of concentrated hostility, moral elimination, or obsessive fixation.	This label does not mean the speaker caused violence. It marks a dangerous civic pattern that deserves visibility.
CR	Unsupported Criminalization	The content assigns criminal, treasonous, terrorist, or equivalent status without adequate evidence or adjudicated grounding.	This label is especially important when public officials or private persons are converted into alleged criminals without support.
CR	Mobilization Risk	The content tends toward hostile collective action, coordinated harassment, or escalating pressure against a target.	This label should distinguish lawful organizing from intimidation, harassment, or coercive swarm behavior.
CR	Operational Targeting Risk	The content supplies or amplifies location, schedule, identity, contact, or tactical information that could facilitate real-world targeting.	This label may justify escalation beyond ordinary labeling, depending on law and platform rules.
CR	Instability Amplifier	The content intensifies public volatility through weak evidence, high emotional loading, conspiratorial certainty, or escalating rhetoric.	This label should mark amplification risk, not simply emotional intensity.
RT	Challenge Item	The content was surfaced to preserve meaningful exposure to serious disagreement outside the user's ordinary preference region.	This is not a warning label. It explains why useful disagreement entered the visible field.
RT	Reality-Pass	The content was surfaced because of urgent real-world significance even though it is intense, disturbing, or outside ordinary preference settings.	This label protects the system from becoming a comfort filter that hides harsh reality.
RT	High Uncertainty	The system's interpretation is tentative because context, evidence, source relation, or meaning remains unresolved.	This label should lower the reader's confidence in the label itself and invite inspection.
RT	Clustered	Similar items were grouped to reduce repetitive amplification while preserving access to the underlying discussion.	This label explains repetition control rather than moral or cognitive judgment.
RT	Deferred for Context	The item was delayed or reduced in prominence until additional context, corroboration, or thread development becomes available.	This label should be used sparingly, especially around fast-moving events where early information is unstable.

A.5. Intensity bands and color logic

Labels should carry intensity without becoming theatrical. A red label should not mean that speech has been banned; it should mean that a serious feature of the speech act deserves strong visibility. The proposed intensity bands are Informational, Notice, Caution, and High. Informational labels explain routing or positive quality. Notice labels mark modest issues worth seeing. Caution labels identify meaningful cognitive or moral concerns. High labels identify serious civic-risk or dignity-related concerns requiring special care.

COLOR	MEANING	TYPICAL LABELS
Green	Constructive or positive quality	High Analytical Quality, Responsible Critique, Constructive Contribution
Amber	Cautionary cognitive or moral concern	Low Evidentiary Support, Context Deficit, Misleading Framing, Disproportionate Certainty
Red	Serious dignity or civic-risk concern	Dehumanizing Language, Target-Construction Risk, Mobilization Risk, Operational Targeting Risk

COLOR	MEANING	TYPICAL LABELS
Blue	Operational route or system handling	Challenge Item, Reality-Pass, Clustered
Gray	Uncertainty, deferral, unresolved confidence	High Uncertainty, Deferred for Context

A.6. Recommended visible label format

The visible label should be compact. It should not cover the content or turn reading into a warning-screen experience. A good standard form is: family code, short title, and intensity band. More detailed explanation should appear only on inspection. The compact label might read 'IQ | Context Deficit | Caution,' 'MQ | Responsible Critique | Notice,' or 'CR | Target-Construction Risk | High.' The reader should be able to see the label, recognize its family, continue reading, and open an explanation only when desired.

VISIBLE TAG EXAMPLE	INTERPRETATION
■ IQ | Context Deficit | Caution	A cautionary cognitive label. The claim may require missing background before it can be understood fairly.
● MQ | Responsible Critique | Notice	A constructive moral label. The content criticizes without degrading the target.
▲ CR | Target-Construction Risk | High	A serious civic-risk label. The content participates in target-building rhetoric or pattern.
◆ RT | Reality-Pass	A route label. The item was surfaced because urgent reality should not be hidden by comfort filtering.
◆ RT | High Uncertainty	A confidence label. The system's own interpretation is provisional and should be read cautiously.

A.7. How labels should be interpreted

Labels in the IQ/MQ Public Layer should be interpreted as structured annotations, not as total judgments. A label does not claim omniscience, and it does not cancel the reader's own agency. It marks a feature of the content that the system judges important for public understanding. The reader remains free to inspect, disagree, and continue reading. In this sense, the Public Layer is interpretive rather than punitive: it makes speech more visible in structure, not less available in existence.

The label set should also remain revisable. Public speech changes, tactics evolve, and labels that are useful in one period may need refinement in another. Revision, however, should not be impulsive. New labels should be added only when existing labels cannot adequately describe an important recurring pattern, and every addition should pass through governance, testing, public explanation, and version history. A restrained label set is part of the moral discipline of the Public Layer itself.

Appendix B. Illustrative Cases and Sample Labeling

The following cases are hypothetical and stylized. They do not refer to any actual public figure, party, platform, or event. Their purpose is to show how the IQ/MQ Public Layer distinguishes ordinary criticism from degradation, weak evidence from falsehood, disagreement from targeting, and urgent reality from manipulative intensity. The examples also show why the system should label speech acts rather than speakers. The same speaker could produce different kinds of content on different occasions, and the same topic can contain both responsible critique and civic-risk rhetoric.

B.1. Forceful criticism without degradation

A post argues that a senator's proposed bill is irresponsible, poorly drafted, and likely to harm the public. The post quotes the relevant section of the bill, links to the text, explains the practical concern, and urges voters to contact their representatives. The tone is strong, but the target is treated as a public official whose decision is being criticized rather than as a human object to be degraded.

SAMPLE LABEL	REASON
● MQ | Responsible Critique | Notice	The content criticizes a public action while preserving proportionality and the dignity of the target.
■ IQ | High Analytical Quality | Notice	The criticism is tied to evidence, quotation, and a coherent explanation of consequences.

This case should be protected conceptually and visually from over-labeling. The Public Layer must not treat strong disagreement with officials as a civic-risk problem merely because the language is severe. Democratic life requires serious criticism of power. The relevant distinction is not soft versus harsh speech, but responsible critique versus degradation, distortion, or target construction.

B.2. Weak claim with missing context

A post states that a public agency 'admitted failure' based on one sentence from a long report. The quoted sentence is real, but the surrounding report explains that the agency also corrected the problem, published a revised procedure, and accepted outside review. Without that context, the post creates a misleading impression of ongoing concealment.

SAMPLE LABEL	REASON
■ IQ | Context Deficit | Caution	The omitted context materially changes the reader's interpretation.
■ IQ | Misleading Framing | Caution	The post selects a real fragment in a way that distorts the larger record.

This case illustrates why the Public Layer should not be a simple fact-checking device. The quoted statement may be true, but the framing can still mislead. The correct label is not necessarily 'false.' It is a cognitive-quality label that tells the reader the claim requires context before it can be evaluated fairly.

B.3. Derogatory framing that remains below civic-risk threshold

A post mocks a city official with repeated insults, personal ridicule, and degrading nicknames, but it does not call for harm, doxxing, mobilization, or criminal punishment. The content is politically expressive but contributes little beyond contempt.

SAMPLE LABEL	REASON
● MQ | Humiliating / Derogatory Framing | Caution	The post relies primarily on belittlement rather than substantive criticism.
■ IQ | Low Evidentiary Support | Notice	The post makes evaluative claims without meaningful support.

This case should normally be labeled rather than suppressed. It may be vulgar, unfair, or low-quality, but that does not make it unlawful or automatically removable. The purpose of the Public Layer is to show that the speech act is morally and cognitively weak, not to prevent political mockery from existing.

B.4. Target-construction risk around a named person

A series of posts repeatedly describes a named public figure as a traitor, parasite, invader, and enemy who must be 'removed by any means.' The posts include no lawful evidence of treason and gradually move from policy criticism to a stable image of the person as an existential contaminant. The rhetoric is repeated by multiple accounts and begins to circulate with images of the person's public appearances.

SAMPLE LABEL	REASON
▲ CR | Target-Construction Risk | High	The content contributes to the construction of a named person as an object of concentrated hostility or moral elimination.
▲ CR | Unsupported Criminalization | High	Serious criminal or treasonous status is asserted without adequate grounding.
● MQ | Dehumanizing Language | High	The language weakens ordinary restraints by reducing the target to a degraded category.

This is the kind of case for which the Public Layer has special value. The label does not claim that the speaker caused or intended violence, and it does not by itself prove illegality. It marks a civic pattern that readers, platforms, and reviewers should see clearly. The post is no longer merely criticism. It participates in target construction.

B.5. Operational targeting risk

A post shares the private address, travel schedule, or real-time location of a controversial figure together with hostile commentary. Even if some surrounding criticism is political, the content adds practical information that could facilitate direct real-world targeting.

SAMPLE LABEL	REASON
▲ CR | Operational Targeting Risk | High	The content supplies or amplifies information that could enable real-world targeting.
▲ CR | Mobilization Risk | High	The content may encourage hostile collective action around the target.

This case differs from ordinary reputational criticism because it crosses into operational exposure. A public label may not be sufficient here. Depending on platform rules and law, the content may require escalation, removal, or emergency review. The appendix therefore reinforces a central principle: red labeling and enforcement are related but not identical.

B.6. Urgent and disturbing but worthy of passage

A post during an unfolding emergency contains graphic descriptions, intense language, and emotionally disturbing facts. It is sourced to official briefings, credible reporting, or direct evidence, and it helps the public understand a real danger. The content may fall outside the user's ordinary preference for calm discourse, but hiding it would create a false informational environment.

SAMPLE LABEL	REASON
◆ RT | Reality-Pass	The content was surfaced because urgent real-world significance overrides ordinary comfort filtering.
■ IQ | High Analytical Quality | Notice	Despite intensity, the content is grounded and coherent.

This case prevents the Public Layer from becoming a comfort machine. The goal is not to make the feed pleasant. It is to make it more truthful, proportionate, and livable. Reality is sometimes harsh, and an honest mediation system must allow harsh reality to pass when it is well grounded and consequential.

B.7. High-quality disagreement outside the user's preference zone

A user's personal settings favor moderate, high-MQ, high-IQ content from familiar sources. The system nevertheless surfaces a strong opposing argument from an unfamiliar ideological region because the argument is well supported, coherent, and morally restrained. The content is outside the user's usual center but within the system's challenge budget.

SAMPLE LABEL	REASON
◆ RT | Challenge Item	The content was surfaced to preserve meaningful exposure to serious disagreement.
■ IQ | High Analytical Quality | Notice	The opposing argument is cognitively strong.
● MQ | Responsible Critique | Notice	The disagreement is conducted without degrading the opponent.

This case illustrates why the Public Layer and the Personal Filter must not produce an elegant echo chamber. A healthy informational environment includes principled difference. The challenge label tells the user why the item appeared without pretending that the item matches ordinary preference.

B.8. Repetitive amplification without added signal

A fast-moving controversy produces hundreds of posts repeating the same accusation, image, slogan, or emotional claim. Some posts are accurate, some are not, and many add no new evidence. The informational problem is no longer a single statement, but repetitive amplification that crowds out context and proportion.

SAMPLE LABEL	REASON
◆ RT | Clustered	Similar items were grouped to reduce repetitive amplification while preserving access.
▲ CR | Instability Amplifier | Caution	The repetition may intensify volatility without adding understanding.

Clustering is not censorship. It is repetition control. The user can still inspect the underlying material, but the feed does not allow raw duplication to dominate attention. This is one of the most important differences between a civic mediation layer and a platform-native engagement feed.

B.9. Ambiguous irony, quotation, or satire

A post appears to repeat an inflammatory phrase, but it may be quoting the phrase critically, mocking it satirically, or documenting its use by others. The system cannot confidently determine whether the user is endorsing, condemning, or analyzing the language.

SAMPLE LABEL	REASON
◆ RT | High Uncertainty	The system's interpretation is tentative because meaning depends on unresolved context.
◆ RT | Deferred for Context	The item may be delayed or reduced in prominence until more context is available.

This case is crucial for trust. A serious system must know when not to overclaim. Ambiguity is not failure by itself. Failure occurs when ambiguity is converted into an unjustified moral or civic-risk label. The correct response is caution, inspection, and context gathering.

Taken together, these cases show the main interpretive boundaries of the Public Layer. It must protect political criticism, mark weak evidence, identify degrading conduct, distinguish intensity from distortion, preserve meaningful disagreement, control repetition, and admit uncertainty. Its legitimacy depends not only on correct labels, but on knowing which kind of label is appropriate and what consequence, if any, should follow.

Appendix C. Legal Authorities and Design Implications

Brandenburg v. Ohio supports the distinction between dangerous advocacy and punishable incitement. For the public layer, the implication is that many statements may be labeled high-risk without being treated as unlawful. Civic labeling should not pretend to be criminal law.

Counterman v. Colorado supports caution around true threats by requiring attention to subjective awareness, at least recklessness, before criminal punishment. For the public layer, the implication is that threatening structure may be labeled while legal threat status remains a separate and narrower question.

New York Times Co. v. Sullivan protects speech about public officials by requiring actual malice for defamation liability. For the public layer, the implication is that public officials cannot be shielded from robust public evaluation, but factual accusations should be handled with evidentiary discipline.

Miami Herald v. Tornillo and Moody v. NetChoice underscore the importance of editorial discretion and the constitutional difficulty of government attempts to compel or control editorial choices. For the public layer, the implication is that civic labeling should remain independent from government commands, especially where labels concern officials or political viewpoints.

Bantam Books and later controversies over government communications with platforms show the danger of informal coercion. For the public layer, the implication is that government may not indirectly turn a civic label system into a suppression tool. Independence from state pressure is not optional; it is part of the architecture of legitimacy.

The EU AI Act's treatment of social scoring also provides a warning. A system that classifies persons and attaches unrelated detrimental consequences becomes dangerous. For the public layer, the implication is that post-level and context-level labels must not become a general civic score used to allocate rights or access.