Building the IQ/MQ Public Layer

Three Methods for Creating a Publicly Owned, Living, and Transparent Grading and Labeling System

A Strategy Paper for the IQ/MQ Public Layer

Lev Neymotin, PhD

Uncompressed Thinking | N5Digital, Inc. | May 2026

A public label is legitimate only when the route to that label can be explained.

Abstract

The IQ/MQ Public Layer was proposed as a civic labeling layer for public discourse: a way to make the moral and cognitive structure of speech visible without converting visibility into censorship. The Public Layer extends the earlier user-owned IQ/MQ Filter into a shared civic domain. That shift creates a new problem. A personal filter can be maintained by an individual user through declared preferences, adaptive calibration, correction, and drift control. A public layer cannot be maintained that way. Its labels must carry shared public meaning, and therefore its grading and labeling system must be created, maintained, corrected, and governed through procedures that are themselves visible.

This strategy paper proposes three methods for creating such a Public Layer. The first is a Civic Standards Charter Method, in which the Public Layer begins as a public standard: a charter, label registry, annotation guide, reference corpus, review process, and revision process. The second is a Blind Multi-Rater Calibration Method, in which author-identifying information is removed where possible and diverse raters evaluate speech acts on the IQ/MQ plane and through label categories, producing a public calibration corpus and empirical thresholds. The third is a Living Case-Law and Appeal Method, in which labels evolve through public challenges, reviewed decisions, release notes, audits, and precedent-like examples that guide future labeling. Each method solves part of the problem and creates its own risks. The first provides conceptual discipline, the second provides empirical grounding, and the third provides continuity and adaptation.

The recommended strategy is not to choose one method exclusively. A credible Public Layer should begin with the standards discipline of Method One, use the calibration strength of Method Two to test and refine the label system, and rely on the living correction structure of Method Three to maintain legitimacy over time. The result is not a black box and not a crowd vote. It is a public civic standard assisted by AI, informed by public judgment, constrained by transparent definitions, and continuously corrected through appeal and audit.

1. Introduction: From the Personal Filter to a Public Standard

The earlier IQ/MQ Filter was personal by design. It was imagined as a user-owned adaptive layer between platform and timeline, allowing an individual to shape the moral and cognitive character of his own informational intake. Its authority came from the user. The user could choose a target region in IQ/MQ space, adjust porosity, preserve challenge exposure, correct the system, and allow slow adaptation under drift control. This made sense because the personal filter was fundamentally an instrument of informational self-governance.

The IQ/MQ Public Layer changes the problem. It is not merely a larger personal filter. It is a public labeling system, and labels in the public sphere are not private preferences. A label such as Low Evidentiary Support, Dehumanizing Language, Target-Construction Risk, or Responsible Critique must mean approximately the same thing across users, platforms, controversies, and political positions. If each user privately defines the label, the Public Layer loses civic meaning. If one platform or government defines it alone, the Public Layer risks capture. If a crowd defines it by volume, it becomes majoritarian pressure wearing the costume of public legitimacy.

The present question is therefore foundational. How can the grading and labeling system of the IQ/MQ Public Layer be created and maintained so that it is publicly owned, living, transparent, and resistant to abuse? This paper does not yet address product implementation, platform integration, business structure, machine-learning architecture, or legal deployment in detail. The focus here is earlier and more important: the creation of the labeling standard itself.

The working premise is that the Public Layer must label speech acts, not speakers. It 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. But it does not declare the hidden essence, moral worth,

civic rank, or permanent intellectual status of the person behind the speech. This distinction is central to the whole proposal. A system that refuses to label content becomes blind. A system that labels persons too easily becomes cruel. The Public Layer must live between those failures.

A second premise is that anonymous public labels can be legitimate only if the process behind them is not anonymous. The visible label may not need to name the individual raters or reviewers who participated in its production. Indeed, exposing individual reviewers could invite harassment, ideological retaliation, or strategic pressure. But the standard itself must not be hidden. The public must be able to inspect label definitions, examples, thresholds, exclusions, review procedures, appeal routes, and revision history. The label can be anonymous at the point of display; the route to the label cannot be anonymous as a civic process.

The three methods proposed below are not software modules. They are governance and standard-creation strategies. Each one answers a different part of the legitimacy problem. The first method asks how to define labels before applying them. The second asks how public judgment can be gathered without surrendering to tribe or source bias. The third asks how the layer remains alive after launch, when new controversies, new rhetoric, and new errors inevitably appear.

2. What the Public Layer Must Produce

Before considering methods, it is useful to define the output. The Public Layer is not trying merely to produce a single score. Nor is it simply trying to decide whether a post is good or bad. Its task is to produce a visible and inspectable label structure. That structure may include IQ signals, MQ signals, civic-risk signals, intensity bands, confidence information, and a short explanation of the route by which the label was reached.

The grading system must therefore produce several linked objects. The first is a label vocabulary. This is the set of permissible labels, such as Low Evidentiary Support, Context Deficit, Misleading Framing, Disproportionate Certainty, High Analytical Quality, Humiliating or Derogatory Framing, Dehumanizing Language, Moral Asymmetry, Responsible Critique, Constructive Contribution, Target-Construction Risk, Unsupported Criminalization, Mobilization Risk, Operational Targeting Risk, and Instability Amplifier. The exact list may evolve, but the principle is stable: a label must name a feature of the speech act, not merely express approval or disapproval.

The second object is a definition for each label. A label without a definition is an accusation. A definition should say what the label means, what features justify it, what evidence is required, what intensity levels are possible, and what kinds of cases are excluded. The exclusions are as important as the inclusions. Target-Construction Risk, for example, must not mean harsh criticism. Low Evidentiary Support must not mean ideological disagreement. Dehumanizing Language must not be stretched so widely that ordinary criticism becomes morally suspect.

The third object is an annotation guide. The guide translates definitions into working judgment. It gives examples, borderline cases, and failure cases. It explains how to handle quotation, satire, legal reporting, public-interest exposure, urgent warnings, moral outrage, and developing events. Without such a guide, different reviewers or models will use the same label vocabulary while silently applying different standards.

The fourth object is a reference corpus. This is a carefully selected body of examples used for training, calibration, testing, audit, and public explanation. It should include obvious cases, difficult cases, historical cases, current cases, and cases where earlier labeling was wrong. It should include ideological and cultural diversity. It should include ordinary social media posts as well as mass-media headlines, broadcast excerpts, podcast claims, public statements, and article passages, because the Public Layer may eventually extend beyond social media into the wider public domain.

The fifth object is a confidence and explanation structure. A public label should not merely appear as a color or symbol. It should carry enough explanation to be inspected. A reader should be able to see whether the label came from missing context, unsupported evidence, dehumanizing metaphor, repeated enemy construction, operational targeting cues, or reviewer uncertainty. A speaker should be able to challenge the label by pointing to missing context, satire, quotation, correction, or stronger evidence.

The methods proposed in this document should be judged by how well they produce these objects. A method that produces fast labels but no standard is not enough. A method that produces beautiful definitions but no empirical testing is not enough. A method that allows public participation but no protection against mob pressure is not enough. The Public Layer needs all three: standards, public calibration, and living correction.

3. Criteria for Evaluating the Three Methods

A method for creating the IQ/MQ Public Layer should be judged by several criteria. The first is legitimacy. The public must be able to understand why the method deserves trust. Legitimacy does not require unanimity; no public labeling system will receive that. But it does require that the method not be reducible to one founder, one platform, one party, one profession, one ideology, or one emotional crowd.

The second criterion is transparency. The method should produce public documents, visible definitions, and inspectable procedures. Transparency does not mean that every internal note or every reviewer identity must be exposed. It means that the standard and the route to labels can be examined by those affected by them and by independent observers.

The third criterion is resistance to capture. A Public Layer may be captured by government, by platforms, by donors, by activist networks, by professional elites, by media institutions, by ideological factions, or by organized user brigades. A good method should make capture harder, more visible, and more reversible.

The fourth criterion is contextual intelligence. Speech does not act as isolated grammar. It acts as situated public behavior. A method that strips away all context will mislabel quotation, reporting, satire, and rebuttal. A method that lets context overwhelm content will excuse too much. The Public Layer needs disciplined context rather than contextlessness or infinite context.

The fifth criterion is public ownership. Public ownership does not mean state ownership. It means that the standard is maintained as a public civic object: documented, versioned, contestable, open to review, and not privately locked inside a platform. Public ownership is achieved less by slogans than by governance structure.

The sixth criterion is adaptability. Language changes, manipulative tactics evolve, and public controversies create new edge cases. A dead standard will fail. A constantly shifting standard will also fail. The right method must combine stability with revision.

The final criterion is operational readiness. A method may be philosophically attractive but impossible to use at scale. Another may scale easily but produce shallow judgments. The Public Layer needs methods that can begin modestly and grow into a practical public labeling standard without pretending to solve every case from the start.

4. Method One: The Civic Standards Charter Method

The first method begins not with ratings, not with AI, and not with live labeling, but with a public standard. It treats the IQ/MQ Public Layer as a civic standard before it becomes a technical layer. Its central premise is that public labels must be defined before they are applied. The method therefore creates a Public Layer Charter, a Label Registry, an Annotation Guide, and a Reference Corpus before broad deployment.

The Public Layer Charter would state the purpose and limits of the system. It would say that the Public Layer labels public content and speech behavior, not the hidden worth of persons. It would state that labels are not legal verdicts, employment recommendations, credit signals, school admissions tools, or civic rank. It would define the anti-coercion principle and separate civic labeling from platform enforcement and law. This charter is essential because it prevents the system from drifting into a social-scoring regime.

The Label Registry would list the approved labels, their families, definitions, exclusions, intensity bands, and visual presentation. The Annotation Guide would explain how to apply the labels. The Reference Corpus would provide representative examples. Together, these materials would form the first public object of the Public Layer. They would not yet produce labels at scale. They would produce the grammar by which labels can later become meaningful.

The governing body for this method should not be the state and should not be a single platform. The strongest form would be an independent public-interest standards body or open standards consortium. At an early stage, the founding author or founding organization may create the first draft. That is normal. But if the Public Layer becomes a live civic instrument, its stewardship must gradually move beyond founder authority. The creator may remain conceptually central, but the standard cannot remain credible if every future definition and threshold depends on one person's judgment.

This method also allows expert and public input without confusing public input with crowd rule. Experts in constitutional law, journalism, AI governance, ethics, social psychology, political communication, platform design, and civil liberties can help define labels and exclusions. Ordinary citizens and public users can comment on whether definitions make sense in real public discourse. But final adoption should occur through a documented standards process, not through raw popularity.

4.1. Strengths of Method One

The greatest strength of the Civic Standards Charter Method is conceptual discipline. It creates a common language before judgment begins. This matters because labels such as Target-Construction Risk or Dehumanizing Language are too serious to be improvised. They need definitions, examples, and exclusions. Without this, the Public Layer becomes a moral weather system: labels appear according to mood, pressure, or institutional instinct.

A second strength is stability. Public discourse is fast and reactive. A standards-first method slows the system down at the right point. It does not slow every future label, but it slows the creation of the categories by which labels will be assigned. This is especially important for high-risk labels that may affect reputation or public interpretation during volatile events.

A third strength is legitimacy. A published charter and registry allow critics to engage the standard itself rather than guessing at hidden motives. A person may disagree with the definition of Target-Construction Risk, but the disagreement has an object. That is a major improvement over a black box.

A fourth strength is interoperability. If the Public Layer is defined as an open civic standard, it can be implemented by more than one platform, client, media tool, or user-side system. This prevents the standard from becoming merely a private feature inside one company.

4.2. Weaknesses of Method One

The weakness of this method is that it can become too institutional, too slow, or too expert-driven. A standards body may speak with procedural authority while drifting away from ordinary public understanding. It may also become vulnerable to professional capture. If journalists, academics, civil-liberties lawyers, technologists, or policy experts dominate too completely, the system may acquire an elite accent that weakens public trust.

Another weakness is abstraction. A label registry can look coherent on paper and fail in practice. Public speech is messy. It contains irony, coded language, genuine anger, legal reporting, bad-faith evasion, emotional testimony, and rhetorical traditions that differ across communities. A standards document alone cannot anticipate all of this.

This method also requires careful funding discipline. If a standards body is funded by one platform, one donor class, one ideological network, or one government program, critics will reasonably ask whether the standard is independent. Even if the work is honest, the appearance of capture can damage legitimacy.

For these reasons, Method One is necessary but insufficient. It should create the initial public grammar of the layer. It should not be the only source of legitimacy. It needs empirical public calibration and live correction.

4.3. Best Use of Method One

Method One is best used at the founding stage and at every major revision stage. It should define the charter, label registry, annotation guide, intensity bands, confidence rules, exclusions, and review thresholds. It should be especially strong around red labels and civic-risk categories, because those labels carry the greatest danger of abuse. The method provides the constitutional structure of the Public Layer. Other methods can then test, calibrate, and revise that structure.

5. Method Two: The Blind Multi-Rater Calibration Method

The second method begins from a different intuition. If the Public Layer is to be publicly owned, the public must help teach it what the labels mean in practice. One possible route is to strip posts, articles, headlines, comments, or public statements of author-identifying information where possible and ask multiple raters to evaluate them on the IQ/MQ plane and through candidate labels. The results are then aggregated, tested for cross-perspective agreement, examined for variance, and used to build a calibration corpus for the Public Layer.

This method resembles the user's initial thought but should be expanded and disciplined. It should not simply ask random people to rate posts and then average the scores. Averaging can conceal disagreement, polarization, misunderstanding, or group bias. A high-quality calibration method should preserve distribution as well as central tendency. It should record whether raters converge across political, cultural, educational, and professional lines. It should distinguish consensus from factional clustering. It should also track confidence and uncertainty.

The blind element is important. Removing author, source, party, platform, and follower-count information can reduce some forms of bias. A post from a disliked public figure may be judged more harshly if the author is visible. A post from a trusted source may be excused too easily. Anonymization helps the system see the structure of the speech act before the reputation of the speaker intervenes.

But blind rating must not be naive. Some context is necessary for interpretation. A post about a breaking event may be unintelligible without knowing the event frame. A quotation may be misread as endorsement if the surrounding thread is removed. A legal or journalistic citation may appear reckless if the source context is hidden. Therefore this method should use staged anonymization rather than total stripping. In Stage A, raters see the content with minimal context and no author identity. In Stage B, raters see the thread or event context, still without author identity where feasible. In Stage C, trained reviewers or auditors examine source-aware context for cases where source history, pattern, or public role matters.

The rater pool should also be structured. It should include ordinary citizens, trained reviewers, subject-matter readers, and cross-perspective panels. Raters should not simply be selected for demographic diversity, although that matters. They should also represent different cognitive styles, media diets, political orientations, moral temperaments, and familiarity with public discourse. The purpose is not to manufacture

false balance. It is to prevent the Public Layer from learning only one social group's instinctive sense of speech quality.

The output of Method Two would be a Public Calibration Corpus. Each item in the corpus would carry ratings on IQ, MQ, label applicability, intensity, confidence, and disagreement pattern. Some examples would become anchor cases. Others would become ambiguous cases. Still others would become exclusion examples: content that looks risky but should not be labeled because it is satire, quotation, responsible reporting, or severe but legitimate criticism.

5.1. Strengths of Method Two

The greatest strength of the Blind Multi-Rater Calibration Method is empirical grounding. It tests whether the labels defined in Method One can actually be recognized by people. A label that seems clear in a charter may be confusing in live examples. A label that experts apply easily may be unstable among ordinary readers. Calibration exposes these weaknesses early.

A second strength is public participation. The Public Layer cannot be publicly owned only in name. A rater-based calibration process gives the public a real role in shaping the system. It also helps demonstrate that the standard is not simply a founder's moral preference, a platform's policy preference, or an expert committee's private vocabulary.

A third strength is bias reduction. Blind rating does not eliminate bias, but it reduces certain obvious forms of source bias. It helps reveal whether a speech act is being judged because of its structure or because of the author. It also allows comparison between blinded and source-aware judgments, which can be highly informative. If a post receives one label when anonymous and a very different label when attributed, the system learns where reputation is influencing interpretation.

A fourth strength is statistical visibility. This method can reveal disagreement rather than hiding it. A post may receive strong consensus that it is low-evidence. Another may divide raters sharply by political orientation. A third may produce high uncertainty across all groups. These patterns are themselves useful. They can inform confidence levels, provisional labels, or the need for human review.

5.2. Weaknesses of Method Two

The first weakness is context loss. Public speech is situated. Blind rating can protect against author bias, but it can also remove important meaning. A severe sentence may be a quotation under critique. A claim may be a reference to a documented investigation. A sarcastic line may be misread if detached from surrounding conversation. The method must therefore avoid the false purity of total decontextualization.

The second weakness is public bias in another form. Removing the author's name does not remove ideology from the content. Raters may still identify political direction, cultural context, or target identity. They may still judge according to group loyalty, moral temperament, or media habit. Blind rating helps, but it is not magic.

The third weakness is aggregation danger. A simple average can flatten meaningful disagreement. If half the raters see target construction and half see legitimate criticism, the average may produce an amber label when the real result is unresolved conflict. The system should therefore report variance, cross-perspective agreement, and uncertainty instead of relying only on mean scores.

The fourth weakness is gaming and labor quality. Rater pools can be infiltrated, coordinated, fatigued, or poorly trained. Public rating at scale can become mechanical. Participants may learn the expected answer or respond according to ideology. The calibration process therefore requires quality controls, attention checks, rater reliability metrics, and periodic refreshment of the rater pool.

A fifth weakness is legitimacy under controversy. If the public learns that a controversial label was based on a pool of raters, critics will ask who those raters were, how they were selected, how they were trained, what they were shown, and whether they reflected the relevant public. These questions must be answered

without exposing individual identities. The process must be transparent even if individual raters remain protected.

5.3. Best Use of Method Two

Method Two is best used to calibrate the standard, not to govern the standard alone. It should test definitions, identify ambiguous zones, produce anchor examples, evaluate intensity thresholds, and reveal where labels are too broad or too narrow. It is especially useful for building the reference corpus and for training AI models under publicly understandable criteria.

It should not be the sole method for high-impact live labeling. A blind multi-rater average should not automatically place a red civic-risk label on a named public figure during a volatile event. In such cases, the calibration corpus may inform the model and review process, but the live label should still pass through confidence discipline, context review, and appealability. Method Two teaches the system; it should not become the system's only judge.

6. Method Three: The Living Case-Law and Appeal Method

The third method treats the Public Layer as something that must learn from real cases over time. It is modeled less on polling and more on a common-law style of public reasoning. In this method, the standard begins with a charter and label registry, and it is calibrated through public rating, but it is maintained through reviewed cases, appeals, disputes, corrections, and published precedent-like explanations.

This method begins from the fact that no initial label system will be sufficient. New forms of manipulation will emerge. Political groups will learn to avoid obvious trigger words while preserving hostile structure. Media organizations will produce borderline framing. Public figures will weaponize both victimhood and accusation. Ordinary citizens will make mistakes, use satire, quote harmful language, or participate unknowingly in escalating patterns. A static label registry cannot handle this world. The Public Layer must have memory.

In the Living Case-Law Method, labels applied to public speech acts can be challenged. A challenge might come from the speaker, a reader, a journalist, an affected person, a researcher, a civil-liberties group, or an auditor. The challenge may argue that the label ignored context, misread satire, confused quotation with endorsement, exaggerated intensity, missed evidence, applied the wrong label family, or failed to distinguish severe criticism from target construction.

The appeal process should not simply erase labels whenever challenged. That would allow organized actors to neutralize the system. Instead, appeals should force the label to show its route. Where the challenge is persuasive, the label may be removed, downgraded, revised, or marked as disputed. Where the challenge is rejected, the reason should be stated. Where the challenge reveals a weakness in the standard itself, the case should become part of a standards revision docket.

Over time, this method produces a public body of examples: not merely labels, but explained labeling decisions. These examples become a kind of civic case law. They show why a post was labeled as Target-Construction Risk rather than Responsible Critique, why a headline was labeled Misleading Framing rather than Low Evidentiary Support, why a post was marked High Uncertainty rather than red, and why a public-interest warning was not treated as mobilization risk. This is not legal precedent in the formal sense. It is interpretive precedent for a public standard.

The method also requires release notes and revision history. When a label definition changes, the change should be explained. When a new exclusion is added, the public should know why. When an audit finds over-labeling in one category or under-labeling in another, the correction should be visible. This is how the layer remains living without becoming arbitrary.

6.1. Strengths of Method Three

The greatest strength of the Living Case-Law and Appeal Method is adaptability with memory. It allows the Public Layer to change in response to real errors and new forms of speech while preserving a public record of how and why it changed. This prevents two opposite failures: rigidity and invisible drift. A rigid system becomes obsolete. An invisibly drifting system becomes untrusted. A case-law system can learn without pretending it has always been the same.

A second strength is contestability. A label that cannot be challenged becomes an accusation. Appeals make the system answerable. They also improve the standard because they expose weak definitions, missing exclusions, ambiguous label boundaries, and model failures. A mature Public Layer should not fear successful appeals. Successful appeals are one of the ways the system becomes more credible.

A third strength is public education. Explained cases teach users how to interpret labels. They also teach speakers how to avoid destructive structures without becoming silent. Over time, the public learns the difference between severe criticism and target construction, between urgency and manipulation, between low evidence and honest uncertainty, between dehumanization and moral condemnation of conduct.

A fourth strength is protection against black-box authority. If labels are accompanied by public cases and decision history, the public can inspect not only the definitions but the application of definitions. This is more meaningful than abstract transparency alone.

6.2. Weaknesses of Method Three

The first weakness is administrative burden. Appeals, case decisions, release notes, and audits require time and institutional discipline. A public system operating at large scale could be overwhelmed if every label becomes a contested proceeding. The method therefore needs triage. Not every label should receive the same appeal route. High-impact labels deserve stronger process than low-impact informational labels.

The second weakness is politicized challenge. Organized groups may use appeals strategically, not to correct errors but to exhaust the system, create doubt, or delegitimize labels. The Public Layer must distinguish good-faith challenge from coordinated abuse without suppressing legitimate dissent. This is difficult.

The third weakness is precedent capture. If early decisions are made by a narrow group, the resulting case memory may encode narrow assumptions. Later reviewers may follow precedent too mechanically. The system must therefore allow precedent to be revised, not merely accumulated.

The fourth weakness is public misunderstanding. A case-law style process may appear too complex for ordinary users. Most people will not read full decisions. The system therefore needs layered explanation: compact labels for the feed, short rationales for inspection, longer case notes for disputed or precedent-setting decisions, and technical documentation for auditors.

6.3. Best Use of Method Three

Method Three is best used after the initial standard and calibration corpus exist. It is the maintenance engine of the Public Layer. It should handle disputed labels, emerging manipulation patterns, new exclusions, confidence failures, red-label appeals, and periodic revisions. It should also feed back into Methods One and Two. Cases that repeatedly produce confusion should trigger registry revision. Cases that split reviewers or raters should be added to the calibration corpus. Cases that reveal new manipulative tactics should become new examples in the annotation guide.

7. How the Three Methods Compare

The three methods should not be understood as rivals in the ordinary sense. They are different answers to different questions. Method One asks, what should the labels mean? Method Two asks, can the labels be recognized across people and perspectives? Method Three asks, how does the system correct itself when real public speech exceeds the first design?

Method One is strongest at founding legitimacy. It prevents the Public Layer from being born as a model with a hidden moral intuition. Its weakness is that it can become too abstract or too elite. Method Two is strongest at public grounding. It tests the labels against blinded and cross-perspective judgment. Its weakness is that rater judgment can lose context or become another form of bias. Method Three is strongest at maintenance. It lets the layer learn from dispute and error. Its weakness is complexity and administrative burden.

A Public Layer built only through Method One would be principled but possibly sterile. A Public Layer built only through Method Two would be public but unstable, vulnerable to aggregation errors and crowd instincts. A Public Layer built only through Method Three would be adaptive but may lack sufficient initial structure. The strongest strategy is to combine the three in sequence and then allow them to reinforce one another.

8. Recommended Composite Strategy

The most defensible path is a composite strategy. The Public Layer should begin with Method One, be tested and calibrated through Method Two, and then be maintained through Method Three. This is not a compromise in the weak sense. It is a layered answer to the legitimacy problem.

The first phase should be a Founding Standards Phase. This phase produces the Public Layer Charter, the Label Registry, the Annotation Guide, and a first Reference Corpus. The purpose is not to perfect the system but to create a visible standard with clear limits. The charter should state that the Public Layer labels speech acts, not speakers; that labels are civic annotations, not legal judgments; that red labels are warnings, not automatic bans; and that public officials remain open to criticism while target-construction structures may still be labeled.

The second phase should be a Blind Calibration Phase. A carefully selected body of public speech acts should be anonymized where possible and presented to multiple rater groups. Raters should place examples in IQ/MQ space, apply candidate labels, identify intensity, and mark uncertainty. Some examples should then be reviewed with additional context to measure how context changes judgment. The output should not be a simple average score. It should be a calibration map showing consensus, disagreement, variance, and cross-perspective convergence.

The third phase should be a Standards Revision Phase. The findings from calibration should be used to revise definitions, exclusions, label names, and intensity thresholds. If raters repeatedly confuse severe criticism with target construction, the Target-Construction Risk definition must be sharpened. If they consistently miss manipulative framing that experts identify, the guide needs better examples. If a positive label such as Responsible Critique is too rarely used, the system may need stronger examples of constructive speech.

The fourth phase should be a Limited Public Pilot. The Public Layer should not begin with universal deployment. It should begin with a limited label set and a limited domain, perhaps public political posts, mass-media headlines, or public-figure discourse. The pilot should display labels with confidence bands and explanation routes. It should also make clear which labels are automatic, which are human-reviewed, which are provisional, and which are under appeal.

The fifth phase should be a Living Maintenance Phase. Once the Public Layer is operating, Method Three becomes central. Appeals, corrections, audits, release notes, and case decisions become the routine life of

the standard. Periodic recalibration through Method Two should continue, especially when new labels are introduced or old labels show high dispute rates. Method One returns whenever major revisions are needed.

This composite strategy also answers the question of public ownership. The public owns the layer not because every label is crowd-voted, but because the standard is visible, challengeable, versioned, and maintained in public. Public ownership means the public can see the rules, contest the labels, inspect the changes, and participate in calibration. It does not mean the public emotional majority governs every case.

9. Design Implications for the Grading and Labeling System

Several design implications follow from the three methods. The first is that the IQ/MQ plane should be used as a conceptual foundation but not forced into a single public number at the start. It may be tempting to assign a precise IQ score and MQ score to every speech act. But early public deployment should probably use bands, labels, and explanations before precise numerical scores. Numbers can come later, after definitions and calibration mature. A banded system is less falsely precise and easier to explain.

The second implication is that positive labels are essential. A public layer that only warns will feel punitive. The label set should also identify High Analytical Quality, Responsible Critique, and Constructive Contribution. These labels remind readers that the purpose is not moral policing but public visibility. The system should show what improves discourse, not only what degrades it.

The third implication is that source history should be contextual, not decisive. The system may use source history to assess confidence and pattern, but the primary object must remain the speech act in context. A generally unreliable account may make a supported claim. A prestigious outlet may publish a misleading headline. A public figure with a history of inflammatory language may still make a responsible critique in a particular case.

The fourth implication is that label explanations should be layered. The feed-facing label should be compact, perhaps a family code, label title, and intensity band. The first expansion should give a plain-language reason. A deeper view should show the evidence path, context considered, applicable definition version, and appeal status. This preserves readability while avoiding black-box authority.

The fifth implication is that the system should preserve uncertainty. Not every case should be forced into red, amber, green, or clean. High Uncertainty, Context Developing, and Deferred for Context are important labels because public life contains incomplete information. A civic system that cannot say 'we do not yet know' will become another engine of premature judgment.

The sixth implication is that the public labels should be available across media forms. The first application may be social media posts because they are easy to segment as speech acts. But the same logic can apply to headlines, article excerpts, broadcast transcripts, podcast claims, campaign materials, institutional statements, and public official communications. The unit of judgment may change, but the principle remains: label the public speech act in context.

The seventh implication is that anonymity of reviewers must be balanced with transparency of process. Reviewers and raters may need protection, especially in politically heated cases. But definitions, examples, routes, and aggregate decision logic should be public. The user need not know the name of every reviewer. The user must know what standard the reviewer applied.

10. Risks Shared by All Three Methods

All three methods face common risks. The first is moral overreach. The Public Layer must not come to believe that it possesses moral finality. It operates on public traces, not souls. It labels content and expression, not the whole human being. If the system forgets this, even transparent procedure will not save it.

The second risk is ideological asymmetry. Equal outcomes should not be required, because different actors may produce different levels of problematic speech at different times. But unequal outcomes must be explainable. If one political side receives far more labels, the system must ask whether that reflects behavior, sampling, rater bias, definition bias, model bias, source selection, or organized reporting. The Public Layer cannot promise perfect neutrality, but it must make bias harder to hide.

The third risk is social scoring by drift. Even if the Public Layer begins as post-level labeling, external actors may try to aggregate labels into person scores, institutional rankings, employment screens, or reputation lists. The standard must resist this. It should explicitly discourage coercive or administrative uses and avoid producing convenient public shame leaderboards.

The fourth risk is performative compliance. Once creators know the label system, some will imitate the surface signs of high IQ/MQ content while preserving manipulation underneath. This is unavoidable. The answer is not secrecy but deeper structure: context, pattern, human review, adversarial testing, and correction through case law.

The fifth risk is exhaustion. A living public standard takes work. If maintenance becomes too heavy, the system may either collapse or simplify itself into crude automation. The design must therefore be modest at first. A small, well-governed label set is better than an ambitious but unmanageable taxonomy.

The sixth risk is public misunderstanding. Many users will interpret any warning label as censorship or any positive label as endorsement. The interface and public education around the system must repeat the core distinction: labels are structured annotations; they are not bans, legal verdicts, or moral absolution.

11. Conclusion: A Public Layer Must Be Built Like a Public Standard

The IQ/MQ Public Layer is plausible only if its grading and labeling system is not a black box. The central question is not whether AI can classify speech. AI can classify at scale, but classification is not legitimacy. The central question is whether a public standard can be created, calibrated, maintained, challenged, and revised in a way that citizens can understand and inspect.

Three methods can contribute to that answer. The Civic Standards Charter Method creates the grammar of the layer. It defines the labels, limits, exclusions, and anti-coercion principles. The Blind Multi-Rater Calibration Method grounds the standard in structured public judgment and builds the calibration corpus needed for threshold-setting and model training. The Living Case-Law and Appeal Method keeps the layer alive through challenges, corrections, precedents, audits, and release notes. Each method is incomplete alone. Together they form a serious path.

The recommended strategy is therefore sequential and integrated. Start with a charter and registry. Test them through blinded and contextual public calibration. Revise them. Pilot a small label set. Then maintain the system through appeals, audits, and versioned public memory. This gives the Public Layer both a spine and a nervous system: stable enough to be trusted, alive enough to learn.

The deeper principle is simple. A public label is legitimate only when the route to that label can be explained. The reader should know what the label means. The speaker should be able to challenge it. The auditor should be able to test it. The public should be able to see how the standard changes. Without that route, civic labeling becomes accusation. With that route, it may become a disciplined form of public interpretation.

The Public Layer should not decide who may speak. It should help citizens see what kind of speech is being placed before them. That task is difficult, and it is dangerous if done carelessly. But the alternative is not neutrality. The alternative is the continued rule of hidden ranking, emotional amplification, tribal labeling, and opaque platform mediation. A free society need not choose between censorship and blindness. It can build civic visibility.

Source Documents Consulted

This strategy paper was drafted as a continuation of the IQ/MQ project and draws on four prior documents supplied for this discussion: The IQ/MQ Public Layer: Freedom of Speech, Civic Labeling, and the Moral Visibility of Public Discourse; MQ: Toward a Moral Quotient; A Two-Dimensional Evaluative Framework for Public Expression: Moral Quotient (MQ) and Intellectual Quotient (IQ); and The IQ/MQ Filter: Toward a User-Owned Adaptive Layer Between Platform and Timeline. These documents provide the conceptual basis for MQ as observable moral behavior, the IQ/MQ two-dimensional evaluative space, the personal user-owned filter, and the shift from a private filter to a shared public layer.

Appendix. Compact Comparison of the Three Methods

The following summary is included only as a quick reference. The main analysis above should remain primary, because the differences between the methods are better understood narratively than as a checklist.

Method	Primary Function	Main Strength	Main Risk
Civic Standards Charter Method	Creates the charter, label registry, annotation guide, reference corpus, and governance spine.	Strong conceptual legitimacy and clear definitions before deployment.	Can become too expert-driven, slow, abstract, or institutionally captured.
Blind Multi-Rater Calibration Method	Uses anonymized and contextualized public judgment to calibrate labels and thresholds.	Empirically grounds the standard and reveals consensus, disagreement, and bias patterns.	Can lose context, hide polarization in averages, or become another form of crowd bias.
Living Case-Law and Appeal Method	Maintains the layer through challenges, corrections, reviewed decisions, audits, and release history.	Keeps the system adaptive, contestable, and publicly accountable over time.	Can become administratively heavy or vulnerable to politicized challenge campaigns.