EDYANT

Responsible AI. Consciously Designed.

edyant.com | consensus@edyant.com | +91 620 391 9445

v1.1.1, 2026

DOI: 10.5281/zenodo.19219685

GRACE : Global Repository of AI Consensus at Edyant

GRACE is a living consensus document. It is built by synthesising diverse inputs—academic research, institutional submissions, practitioner reports, open letters, and civil-society perspectives—into a single, authoritative reference for those who develop, deploy, govern, or are affected by artificial intelligence.

No single submission dominates GRACE. Each input is weighed against the accumulated body of consensus: strong individual opinions inform the document; they do not override it. Where submissions agree, existing language is deepened. Where submissions introduce new ground, new guidance is added. Where genuine tension exists between positions, GRACE reflects that tension honestly rather than papering over it.

GRACE speaks in one voice. It does not attribute sections to contributors, and it does not reproduce the perspectives of any single source verbatim. It distils. Every principle, guideline, guardrail, and best practice recorded here can be traced to at least one submitted input document; nothing has been inferred or imported from outside the submitted record.

This document is structured as follows: Foundational Principles establish the ethical bedrock. Policies and Guidelines translate those principles into operational direction. Guardrails define hard limits. Best Practices provide actionable, implementable instructions organised by theme. Domain-Specific Considerations address sectors where particular guidance has emerged. Governance and Accountability addresses how AI systems and their developers should be held responsible. The Citations section records every source document that has contributed to this version.

Foundational Principles

The following principles represent the ethical foundations on which all further GRACE guidance rests. They are not exhaustive; they will deepen as further inputs are submitted.

1. Societal Responsibility Extends Beyond Individual Harm

AI research and development can cause significant harm to human society even when no individual human subject is directly harmed in the conventional sense. Computing researchers and organisations bear responsibility for taking the ethical and social dimensions of their work into account in design and implementation—not only when their systems are tested, but from the earliest stages of conception.

Existing mechanisms for research oversight—such as Institutional Review Boards—are designed to evaluate risks to human subjects and do not, under most current interpretations, cover risks to human society at large. This gap means that a substantial proportion of AI research proceeds without any structured ethical review. GRACE affirms that this gap must be addressed through deliberate institutional action.

2. Ethical Reflection Must Begin Early

Ethical and societal reflection is most effective when it occurs at the beginning of a project's lifecycle, before decisions about stakeholders, models, data, and evaluation strategy become entrenched. End-of-pipeline requirements—such as ethics statements appended to paper submissions—have value but are insufficient on their own. Organisations and institutions should build structures that inject ethical consideration early and continuously, not only at the point of publication or deployment.

3. High-Level Principles Are Insufficient Without Institutional Structure

The existence of ethical principles and guidelines—however numerous and well-crafted—does not by itself produce ethical behaviour. Voluntary processes reach only those who self-select. Structural and institutional mechanisms that apply to all participants, not only those already disposed toward ethical reflection, are necessary to translate principles into consistent practice. Incentive structures, including funding conditions, are legitimate and effective levers for achieving this.

4. Harms Are Not Distributed Equally

The risks and harms arising from AI systems are not distributed evenly across society. Marginalised groups—including but not limited to racial minorities, people of low socioeconomic status, LGBTQ+ individuals, and communities in the Global South—are frequently more vulnerable to AI-related harms and less represented in the data, design, and deployment processes that shape those systems. Any ethical framework for AI must attend specifically to subgroup harms and must not treat aggregate societal impact as a proxy for equitable impact.

5. Dual Use Is a Systemic Risk

AI systems and research outputs can be co-opted for purposes that differ from, or are directly contrary to, their intended use. Authoritarian surveillance, discriminatory decision-making, and the amplification of disinformation are among the documented risks. Acknowledging the possibility of malicious use is necessary but insufficient; researchers and developers must translate that acknowledgement into concrete design decisions and mitigation strategies.

6. Compliance Is Not Ethics

Organisational compliance with regulation or policy requirements does not guarantee ethical behaviour. Compliance efforts can become performative, producing the appearance of ethical conduct—sometimes described as "ethics washing"—without substantive change. GRACE affirms that genuine ethical practice requires cultural and structural change within organisations, not only procedural adherence to external requirements.

7. The Ethical Status of Public Data Is Contested

Common practices in AI research—including the mining and use of publicly available data—remain ethically contested. The fact that data is technically public does not imply that those who generated it have consented to its use in research. Anonymisation can be compromised. Participants may be unaware that their data

is being used or may object to such use. These questions of consent, publicness, and anonymity must be actively considered, not assumed away.

8. Interdisciplinary Oversight Strengthens Ethical Review

Ethical review of AI research is strengthened when it draws on a diversity of disciplinary perspectives—including but not limited to the humanities, social sciences, engineering, medicine, philosophy, and law. No single discipline holds a complete view of the societal implications of AI. Panels and committees responsible for ethical oversight should be constituted to reflect this breadth.

Policies and Guidelines

Ethics Review as a Condition of Funding

Institutions and funding bodies should consider making the completion of a structured ethics and society review a condition for the release of research funding for AI projects. Funding represents a rare moment of institutional leverage: it reaches all funded researchers, not only those who voluntarily engage with ethics processes. Tying ethics review to funding access transforms ethical reflection from an optional extra into a standard requirement of research practice.

Where such a requirement is adopted, the review process should be designed to engage researchers early—prior to the finalisation of research design—so that ethical considerations can influence decisions about stakeholders, data, models, and evaluation strategy before those decisions become difficult to reverse.

Scope of Review: Society, Subgroups, and the World

Ethics review of AI research should address risks across three levels:

- **Society as a whole**: risks to the social, political, or economic fabric of the communities targeted by or affected by the research.
- **Groups within society**: risks that fall disproportionately on marginalised or vulnerable subgroups, recognising that harms are not distributed equally.
- **Global and cross-border impacts**: risks to societies beyond those directly targeted by the research, including potential for misuse in other geopolitical or cultural contexts.

Iterative Process Over Binary Gatekeeping

The primary purpose of ethics review is not to filter out research but to improve it. Review processes should be structured to enable iterative dialogue between reviewers and researchers, with the goal of helping researchers identify and implement appropriate mitigation strategies. Outright rejection should be a last resort, reached only after sustained iteration has failed to produce a feasible path forward.

At the same time, review bodies should retain the authority to recommend against funding where iteration has not resolved fundamental ethical concerns. The basis for such a recommendation should be transparent, documented, and proportionate. The threshold for rejection should be high; where protected characteristics or groups are being directly and materially harmed, that threshold may be met.

Distinguishing Ethics Review from Regulatory Compliance

Ethics review bodies focused on societal impact are distinct from—and complementary to—existing regulatory mechanisms such as Institutional Review Boards. IRBs address risks to human subjects in research; ethics and society review addresses risks to human society. Both are necessary. Organisations and institutions should be attentive to the difference and should not treat the existence of one as a substitute for the other.

Regulatory compliance alone does not constitute ethical review. Compliance frameworks can create ethical window dressing rather than genuine ethical practice. Ethics review processes should be designed to encourage authentic reflection, not procedural box-ticking.

Annual Review and Ongoing Oversight

Ethical obligations do not end at the point of initial approval. Research projects—and deployed AI systems—evolve, and changes in design, deployment context, or use may introduce new risks not identified at the outset. Funding bodies and institutions should build mechanisms for ongoing review, including annual updates against the original ethics statement, to ensure that ethical commitments remain current throughout a project's lifecycle.

Transparency of Review Principles

Ethics review bodies should publish the principles and guidelines they use. Transparency in review criteria helps researchers understand expectations, builds legitimacy for the review process, and reduces the risk that decisions are perceived as arbitrary. It also mitigates concerns about academic freedom by making clear that the review is focused on societal impact, not on the scientific merit or direction of research.

Guardrails

The following represent areas where GRACE affirms hard limits or non-negotiables, based on the submitted record to date.

1. **Ethics review must not be purely voluntary.** Voluntary processes reach only those already disposed toward ethical reflection. Structural requirements—tied to funding, publication, or deployment approval—are necessary to ensure consistent engagement across an entire research or practitioner community.
2. **Ethical review must not be reduced to individual human subject risk.** Restricting ethics review to direct risks to human research participants leaves the broader societal harms of AI systems unaddressed. Any ethics framework that does not account for risks to society, to subgroups, and to the world is incomplete.
3. **Acknowledging risks without mitigating them is insufficient.** A researcher or developer who names a potential harm but takes no steps to address it has not met the ethical standard. Risk identification must be paired with articulated principles for mitigation and concrete instantiation of those principles in design and practice.
4. **Performative ethics is not ethics.** Processes that produce documentation of ethical consideration without substantive change in research or product design—sometimes described as ethics washing—

undermine the purpose of ethical governance. Organisations should design review processes that create genuine accountability, not reputational cover.

5. **Representation in data is not optional.** AI systems trained on unrepresentative data risk encoding and amplifying existing inequalities. Ensuring adequate representation across demographic groups in training and evaluation data is a baseline requirement, not an enhancement.

Best Practices

All best practices in this section are directly traceable to submitted input documents. Each is written as an implementable instruction. Aspirational language without actionable content is not included.

Transparency

- **Document societal risks explicitly and in advance.** Before a research project or AI system proceeds to funding, deployment, or publication, produce a written statement that identifies the most significant ethical challenges and societal risks. The statement should address risks to society as a whole, to specific subgroups, and to populations beyond the primary target context. This statement should be the starting point of a conversation, not a final declaration.
- **Articulate mitigation principles and their instantiation.** For each identified risk, state the general principle that should govern its mitigation, and then describe specifically how that principle is implemented in the research or system design. Naming a risk without describing a concrete response to it does not constitute adequate ethical disclosure.
- **Commit to publishing privacy-preserving design decisions.** Where a system involves risks to privacy or surveillance, researchers and developers should commit to explaining the privacy-preserving aspects of their design in all resulting publications and communications, contributing to the development of field-wide norms.
- **Make review principles public.** Institutions operating ethics review processes should publish the criteria, principles, and guidelines their panels use, so that researchers understand expectations and the review process is perceived as legitimate rather than arbitrary.

Accountability

- **Use funding conditions as a lever for ethical review.** Organisations and institutions that distribute research or development funding should require completion of a structured ethics and society review as a condition for fund release. This ensures that all funded projects engage with ethical review, not only those whose principal investigators choose to do so voluntarily.
- **Assign clear responsibility for the review process.** Establish who within the organisation or institution is responsible for convening, conducting, and following up on ethics review. Diffuse responsibility produces inconsistent outcomes. The review function should have named ownership and clear authority.
- **Retain review authority to recommend against funding or deployment where necessary.** Ethics review bodies should have the authority to recommend against funding or deployment of a project that has not produced a feasible mitigation plan after sustained iteration. This authority should be

exercised as a last resort, but its existence gives the review process credibility and prevents it from becoming a purely advisory formality.

- **Build annual review into grant and project cycles.** Require researchers and developers to submit brief annual updates on their projects as they relate to their original ethics statements. Where the project has changed in ways that introduce new risks, this triggers renewed review and dialogue.

Fairness

- **Audit AI systems for differential performance across demographic groups.** Algorithm audits—controlled, replicable assessments of system behaviour across different groups—should be a standard practice for organisations and government bodies that develop or deploy AI. Audits should specifically test for disparities affecting racial minorities, gender groups, low-income populations, and other groups at elevated risk of harm.
- **Ensure representation in training and test data.** Assess and document whether the training and evaluation data for an AI system adequately represents the populations it will affect, including groups that have historically been under-represented. Where gaps are identified, take concrete steps to address them before deployment.
- **Guard optimisation criteria against subgroup bias.** When designing systems that optimise for long-term outcomes (e.g., engagement, retention, or efficiency), explicitly assess whether the optimisation criterion creates incentives to deprioritise groups that are harder to serve or already disadvantaged. Commit to testing for and mitigating such effects.
- **Consider whether an internally fair system can still cause external harm.** Meeting formal fairness criteria does not guarantee equitable outcomes. Assess whether a system, even if technically unbiased, could be embedded in or used to support unjust social structures or deployed by actors with harmful intent.

Human Oversight

- **Ensure human control at all levels of automation.** Design AI systems so that human oversight and control remain available regardless of the degree of automation involved. Where AI systems are intended to support or augment human decision-makers, design them explicitly to function as tools for human empowerment rather than replacements for human judgement.
- **Convene interdisciplinary review panels.** Ethics review panels should include members from across the humanities, social sciences, engineering, medicine, and other relevant fields. No single discipline holds sufficient perspective to evaluate the full range of societal implications of AI research. Panels should also include, where possible, representation from communities likely to be affected by the systems under review.
- **Orient review panels as coaches, not gatekeepers.** Train and orient ethics reviewers to function primarily as collaborators helping researchers improve their work, rather than as enforcers seeking to identify and penalise failings. The goal of iterative feedback is to produce better research, not to produce compliance documentation.

Stakeholder Inclusion

- **Include diverse stakeholders in research and product design.** Actively involve representatives of communities that will be affected by an AI system in its design and deployment planning. This includes marginalised or disadvantaged communities who may be at heightened risk of harm. Participation should be substantive, not tokenistic.
- **Identify relevant collaborators and expertise gaps.** Where a research or development team lacks the expertise or perspectives needed to evaluate the societal implications of their work—for example, expertise in inclusive design, educational equity, or the experiences of affected communities—identify and bring in collaborators with that expertise before the research design is finalised.
- **Consider the interests of communities not directly targeted by the research.** Assess whether an AI system or research output could harm societies or communities beyond those it is primarily designed to serve, including communities in other countries or regions where the system might be deployed or replicated.

Data Governance

- **Do not assume that public data implies consent for research use.** Before using publicly available data in AI research or development, assess whether those who generated the data would reasonably have consented to such use. The technical availability of data does not settle the ethical question of its appropriate use.
- **Document dataset composition and limitations.** Produce explicit documentation of who is and is not represented in training and evaluation datasets, including known gaps in demographic coverage. Make this documentation available to those who build on or deploy the resulting systems.
- **Assess anonymisation robustness.** Do not rely on anonymisation as a complete guarantee of privacy protection. Assess whether anonymised data could be re-identified, and design data handling practices accordingly.

Dual Use and Malicious Deployment

- **Name dual-use risks and respond to them concretely.** Where a research output or AI system could be co-opted for harmful purposes—including surveillance, discrimination, or the amplification of disinformation—name those risks explicitly and describe concrete design decisions or constraints that reduce the likelihood or severity of such misuse. Naming the risk without taking steps to address it does not discharge the ethical obligation.
- **Consider risks specific to authoritarian or adversarial contexts.** Assess whether a system could be used by authoritarian governments or other actors with harmful intent to harm individuals or groups, including through mass surveillance or the targeting of dissidents. Where such risks are identified, consider what design constraints, usage limitations, or technical countermeasures are appropriate.

Safety

- **Conduct ethics review early, not only at publication.** Do not defer ethical reflection to the point of publication or product launch. Ethical considerations should inform decisions made at the earliest

stages of a project—before data is collected, before models are designed, and before deployment contexts are fixed. Retrospective review cannot undo choices that have become structurally embedded.

- **Anticipate downstream consequences beyond the immediate application.** Assess what happens when someone else builds on the research output, or when an organisation outside the researcher's control decides to replicate it. Design with the downstream use case in mind, not only the immediate research context.

Domain-Specific Considerations

This section records domain- or sector-specific guidance where sufficient consensus has emerged from submitted inputs. New subsections are added as submissions support them.

AI Research and Academic Institutions

Academic institutions occupy a distinctive position in the AI ecosystem. They generate foundational research that is subsequently built upon by industry, government, and civil society, often in contexts the original researchers did not anticipate. This downstream reach creates obligations that extend beyond the immediate research context.

Unlike professions such as law and medicine, computing does not currently have widely applied legal and professional accountability mechanisms or formal fiduciary relationships. This absence places greater responsibility on institutional structures—funding bodies, research programmes, and professional societies—to create the frameworks that individual ethical commitments alone cannot sustain.

The following considerations apply specifically to AI research conducted in academic settings:

- **Funding bodies at universities should require ethics review as a condition of grant release.** This ensures engagement from all funded researchers, not only those who self-select into voluntary processes.
- **Ethics review in academic research should be structurally distinct from, and complementary to, IRB review.** IRBs focus on risks to research participants. An ethics and society review focuses on risks to society. Both are needed, and each should be clear about its scope so that researchers understand the difference and neither process becomes a substitute for the other.
- **Professional societies in computing should consider developing institutional ethics processes, not only codes of ethics.** A code of ethics sets norms; institutional processes—tied to conference submission, funding, or membership—create structures that apply those norms consistently across a community.
- **Ethics education should be integrated throughout computing curricula, not confined to standalone courses.** Ethics and societal considerations should be present across the curriculum, including in technical machine learning courses, where they currently arise infrequently.
- **Research teams should document and share the ethical issues raised in their projects.** Anonymised collections of issues identified during ethics review, together with the principles and design decisions used to address them, provide valuable scaffolding for researchers encountering similar challenges for the first time.

Governance and Accountability

Effective governance of AI requires more than voluntary commitment. It requires institutional structures that create consistent incentives and apply across entire communities of practice, not only to those already disposed to ethical engagement.

The Role of Institutional Structures

No technology directly changes organisational practice or culture. Engineers, product managers, researchers, and others working with AI must navigate accountability gaps and unclear responsibility within their organisations. High-level ethical principles consistently fail to translate into on-the-ground decisions without institutional structures that close this gap. Rules, incentives, and processes that apply to all—not only to those who choose to engage—are the instruments through which broad behavioural change is achieved.

Policy and regulatory action can complement organisational measures by enforcing standards across all actors, not only those who opt in. However, the typical organisational response to regulation is compliance, and compliance alone does not produce ethical culture. Institutional governance structures should therefore be designed to foster genuine reflection, not only documentary compliance.

Ethics Review Bodies: Design Principles

Where organisations or institutions establish bodies responsible for ethics review of AI research or development, the following design principles reflect current evidence on what makes such bodies effective:

- **Link review to resource allocation.** Ethics review that is tied to access to funding, compute, headcount, or other resources reaches a wider and more consistent population of researchers and developers than voluntary review. Identifying the relevant resource lever within a given institution is a prerequisite for effective design.
- **Compose panels with disciplinary breadth.** Panels should include perspectives from technology, social science, humanities, medicine, ethics, and other relevant fields. Many other disciplines and identities should be included as the process matures. Diverse intellectual composition improves the quality of feedback and the legitimacy of decisions.
- **Assign proposals to reviewers with relevant and complementary expertise.** Each proposal should be reviewed by at least two panellists: one familiar with the broad field of the proposal and one providing a complementary perspective. This reduces blind spots and improves the quality of engagement.
- **Prioritise iteration over rejection.** The first response to an identified concern should be to engage the researcher or developer in dialogue aimed at producing a better mitigation plan. Rejection should follow only if sustained iteration fails to produce a feasible path forward. This orientation keeps the process collaborative rather than adversarial.
- **Establish and publish clear norms and criteria.** Review bodies should document and share the principles and guidelines they use, provide example cases and archetypes for good practice, and offer scaffolding to help researchers think through common categories of ethical concern. Vague prompts produce inconsistent reflection; structured guidance produces more rigorous engagement.
- **Extend review over the project lifecycle.** Initial review at the funding stage should be complemented by mechanisms for ongoing oversight. Annual reporting requirements tied to grant renewals are one

practical mechanism. Where a project changes in ways that introduce new risks, renewed review should be triggered.

- **Acknowledge the limits of review.** Ethics review reduces the probability that harms will be unforeseen; it does not eliminate it. Review bodies should be transparent about this limitation, and should establish processes for acknowledging and learning from cases where approved projects later produce unanticipated harms.

Accountability for Unforeseen Harms

When a project that has received ethics approval is later found to produce harmful outcomes not identified in the review process, the review body should acknowledge the issue publicly, reflect on what the review process failed to anticipate, and adapt its processes accordingly. This ongoing accountability is necessary to maintain the legitimacy of the review function and to ensure that it improves over time.

Citations

Every source document that has contributed to this version of GRACE is listed below. This list is cumulative and will never be trimmed.

[Academic Paper] | "ESR: Ethics and Society Review of Artificial Intelligence Research" | Michael S. Bernstein, Margaret Levi, David Magnus, Betsy Rajala, Debra Satz, Charla Waeiss (Stanford University) | 9 July 2021 | Academic / United States

How to cite GRACE

For citation guidelines, see `https://edyant.com/resources/about_grace`.