AGAINST THE CODE — Manifesto

Safety is not optional. It’s table stakes.

Systems must ship with rigorous risk assessments, red-team coverage, and clear rollback procedures.

Human agency > automated optimization.

Humans retain override authority at critical decision points, with audit trails preserved.

Transparency over mystery boxes.

Disclose risks, data provenance, and limits. Hidden behaviors are liabilities, not features.

Accountability for deployed systems, not just labs.

Operators and integrators share responsibility for downstream harms and timely remediation.

Design for fail-safes, not only capabilities.

Degrade gracefully, sandbox risky actions, and verify before execution in the real world.

© Against The Code 2025 — https://against-the-code.com