

Building AI Systems That Earn Trust

AI goes wrong when implementation is careless, ungoverned, or unchecked

7 Principles of Responsible AI

 **Fairness & Inclusion**
Diverse representation, bias-free models

 **Accuracy & Decision Integrity**
Validated, precise insights

 **Transparency & Interpretability**
Explainable decisions

 **Accountability & Governance**
Clear ownership and oversight

 **Privacy & Consent**
User control over data

 **Purpose & Human Intent**
Empower humans, prevent harm

 **Reliability & Safety**
Consistent performance, even in edge cases

Build Trust Through Action

Don't blindly trust AI
Build transparent, human-centered systems



How Explorance MLY Embodies the 7 Principles of Responsible AI



Fairness & Inclusion

MLY is trained using **supervised, expert-verified annotations** with cultural nuance across diverse educational and workplace contexts.

It avoids internet-scale bias by using **purpose-curated, representative data** and domain-specific language models.



Transparency & Interpretability

Every insight in MLY is **traceable to source comments** and presented in human-readable summaries.

Users can view themes, sentiment, emotion, and alerts and understand why the system flagged or recommended something.



Accuracy & Decision Integrity

MLY insights are **solely based on human-validated data**. Every learning signal is reviewed by **multiple trained annotators**, ensuring accuracy in tone, emotion, and theme detection.

We invested **6 years of R&D** before launching to ensure integrity at every layer.



Accountability & Governance

MLY includes **role-based access**, human override controls, and audit trails to ensure transparency and oversight.

Only authorized individuals can act on sensitive alerts or insights, ensuring decisions are **ethically owned and documented**.



Privacy & Consent

Explorance follows **strict privacy commitments**: customer data is never **used for training** unless clients **explicitly opt-in**.



Purpose & Human Intent

MLY is designed to **empower human judgment**, not replace it. Alerts, summaries, and recommendations guide reflection and care, never punitive action.



Reliability & Safety

MLY models are **tested across multilingual expression, sarcasm, emotion types, and rare phrasing**.

Continuous model tuning ensures reliability even in high-stakes or low-signal scenarios. Safeguards include false-positive filtering, human escalation, and system monitoring.

