AI PLAYBOOK 15

Three Guardrails to Manage AI Risks

July 2025

CANADIAN MARKETING ASSOCIATION CMA

theCMA.ca

Content partially generated by artificial intelligence, refined by human expertise.

This Playbook is part of the CMA's AI Mastery Series, empowering marketers to Implement AI in ways that earn regulatory confidence, maintain strong brand reputation, and foster consumer trust

Guardrails are critical

Research shows that companies implementing generative AI with appropriate guardrails are 27% more likely to achieve higher revenue performance than their peers. To leverage AI effectively while managing associated risks, there are three guardrails to establish that provide specific, measurable parameters marketing teams can implement immediately and refine over time.

Each guardrail serves a distinct purpose within the governance framework, creating a comprehensive protection system that enables confident usage and innovation.

- Brand risk thresholds protect organizational reputation and customer trust by establishing clear parameters for acceptable risk exposure.
- Decision authority limits establish clear human oversight parameters, defining when AI systems can operate autonomously versus when human judgment is required.
- Ethical red lines define non-negotiable standards for responsible AI use, reflecting organizational values and compliance requirements.

These guardrails should be documented, quantified where possible, and regularly reviewed as both AI capabilities and market expectations evolve.

Effective implementation combines objective metrics such as complaint volumes, sentiment analysis scores, and engagement rates with qualitative indicators including media coverage sentiment and customer feedback themes.

Clear guardrails enable innovation.

CANADIAN MARKETING ASSOCIATION CMA

Guardrail 1: Brand risk thresholds for AI

Brand risk thresholds establish clear parameters for acceptable risk exposure related to AI implementation in marketing activities. Organizations must establish baseline measurements before AI implementation, then continuously monitor against defined risk thresholds with clear escalation protocols if those thresholds are breached.

Key risk areas that require defined risk thresholds include:

- Customer complaint volume increases, measured as increase month-over-month (recommended threshold: >15% increase triggers escalation);
- Negative sentiment shifts in social media monitoring against established baseline (recommended threshold: >20% above baseline triggers review; and
- Regulatory body attention measured by contact monitoring (recommended threshold: any regulatory agency contact triggering crisis protocols).

Measurement examples include tracking monthly complaint volume via CRM systems, social listening tools and maintaining a log tracking of regulatory agency communications.

Guardrail 2: AI decision authority limits

Managing AI risks requires setting clear parameters regarding when AI systems can operate autonomously versus when human judgment and oversight are required. These limits should be documented in AI-decision making authority matrices that specify the level of human involvement required for different marketing activities and decisions. This practice can ensure appropriate control while leveraging AI's analytical capabilities.

For example, the following categorizes AI decision-making limits into three risk-based tiers:

- High-risk AI decisions requiring human approval, include budget allocation decisions beyond defined monetary thresholds (e.g., over $5,000) where AI algorithms automatically adjust spending based on performance data, customer data usage decisions like AI systems determining data retention periods, and brand messaging approval before publication;
- Medium-risk AI decisions leveraging AI recommendations with human review, include content creation, audience targeting, and campaign optimization; and
- Low-risk AI decisions where AI can operate autonomously with limited human monitoring, involve scheduling of pre-approved content, basic personalization (e.g., adding a first name to a marketing email), and routine performance reporting.

Implementing these guardrails requires organizations to define clear roles and responsibilities, establish oversight mechanisms, maintain audit trails of decision processes, and regularly review and update authority limits as AI capabilities mature.

By clearly delineating human and AI decision domains, organizations achieve an optimal balance that maximizes both efficiency and safety.

CANADIAN MARKETING ASSOCIATION CMA

Guardrail 3: Ethical red lines

Ethical red lines establish absolute prohibitions and conditional restrictions that reflect organizational values and compliance requirements. These clearly documented standards should be embedded in all AI marketing processes and reinforced through regular training and compliance monitoring to ensure alignment with evolving regulations and stakeholder expectations.

Absolute prohibitions or "no-go zones" for your organization should include:

- No deceptive or manipulative AI practices;
- No processing personal, sensitive, or confidential data using AI without consent;
- No unfair or discriminatory target marketing using AI; and
- No AI-generated content that misrepresents products or services.

Conditional restrictions might include enhanced oversight requirements for sensitive decisions using AI, such as AI systems determining customer access to services or offers or AI-created content for communications related to product recalls.

From principles to practice

Implementing these guardrails for your organization's use of AI requires a systematic approach with distinct phases and clear responsibilities.
Begin by establishing AI governance frameworks through creating oversight structures, such as cross-functional AI committees for larger organizations, or designated AI accountability roles for smaller organizations.

Next, map your organization's current AI-driven marketing use cases and evaluate them against our recommended guardrails. Following this assessment, develop and document your organization-specific guardrails, including a formal AI policy, decision authority matrix, risk threshold definitions, and ethical standards tailored to Canadian and other applicable regulatory requirements. This should be reinforced with communication and training to ensure all team members understand your organization has set AI boundaries and how they apply to their daily work activities.

Regardless of your organization's size or level of AI maturity, it is important to integrate these guardrails into existing workflows by adding risk threshold checkpoints to project templates, including reminders in AI tool training, quick decision trees for common situations, and don't forget to celebrate examples of good AI guardrail setting and "ethical wins".

Success metrics and monitoring

Measuring guardrail effectiveness requires metrics across compliance, performance, risk, and innovation dimensions. Some examples of KRIs and KPIs include:

Key risk metrics (KRIs)	Key performance indicators (KPIs)
• Zero regulatory violations • Issue resolution within 24 hours • Maintained audit trail integrity • Proactive issue identification rates • Incident prevention vs. reaction ratios • 100% team training completion • AI model bias detection alerts	• Successful AI project deployment rates • Stakeholder trust maintenance scores • Maintained or improved customer satisfaction scores • Return on AI investment • Marketing efficiency gains exceeding 25% • Reduced time-to-market for campaigns

Deployment should follow a phased approach, starting with lower-risk marketing functions, such as SEO content optimization, social media content scheduling before expanding to more sensitive areas of concern, such as customer complaint resolution, legal notice communications and compliance messaging. Organizations should conduct ongoing reviews to revisit and refine risk thresholds, authority limits, and red lines and "no-go zones" based on performance data and evolving best practices. This systematic measurement approach provides objective evidence of guardrail effectiveness while identifying opportunities for continuous improvement to ensure the framework evolves with organizational needs and rapidly evolving technological capabilities.

CANADIAN MARKETING ASSOCIATION CMA

AI guardrails – Detailed recommendations

The following build on the prior metrics providing more detailed examples and serve as a starting point for your own development. Organizations must tailor these to their specific business, risk tolerance, regulatory obligations and operational context.

Brand risk thresholds - When to stop and escalate

- Customer complaints increase >15% in one month
- Negative social media sentiment spikes >20% above normal
- Any media inquiry about AI practices
- Regulatory agency contact of any kind

→ Escalate to: Marketing Director within 4 hours

Decision authority - Who decides what

- Human required: Budget >$5,000, data policy changes, brand messaging
- Human review: Content creation, targeting, campaign changes
- AI autonomous: Scheduling, basic personalization, reporting

→ Questions? Ask: AI Governance Lead

Ethical red lines - Never cross

- No deceptive AI practices
- Consent required for personal information
- No unauthorized data sharing
- No discriminatory targeting
- Always disclose AI-generated content

→ Concerns? Contact: Compliance Officer immediately

CANADIAN MARKETING ASSOCIATION
CMA

Recommended reading and references

CMA Resources:

- CMA Guide on AI for Marketers
- Setting the Stage on Artificial Intelligence: A CMA Primer on AI for Marketers
- CMA Accountability Checklists for AI In Marketing
- CMA Mastery Series: AI Playbooks

External References

- The Office of the Privacy Commissioner of Canada: Principles for responsible, trustworthy and privacy protective generative AI technologies
- Innovation, Science and Economic Development Canada: Voluntary Code of Conduct on the Reponsible Development and Management of Advanced Generative AI Systems
- California Management Review: On the ROI of AI Ethics and Governance Investment – From Loss Aversion to Value Generation
- IBM Report: Escalating Data Breach Disruption Pushes Cybersecurity Costs in Canada
- Stanford University HAI: 2025 AI Index Report
- National Institute of Standards and Technology: AI Risk Management Framework
- McKinsey & Company: The State of AI – How organizations are rewiring to capture value

The CMA

As the voice of Canadian marketing, the CMA champions our profession's powerful impact. We are the catalyst to help Canada's marketers thrive today, while building the marketing mindset and environment of tomorrow.

We provide opportunities for our members from coast to coast to develop professionally, to contribute to marketing thought leadership, to build strong networks, and to strengthen the regulatory climate for business success. Our Chartered Marketer (CM) designation signifies that recipients are highly qualified and up to date with best practices, as reflected in the CMA's Canadian Marketing Code of Ethics and Standards.

We represent virtually all of Canada's major business sectors, and all marketing disciplines, channels and technologies. Our Consumer Centre helps Canadians better understand their rights and obligations.

AI Mastery Series: This playbook is part of the CMA's comprehensive AI initiative designed to empower Canadian marketers with the knowledge, skills, and ethical frameworks needed to implement AI responsibly and effectively.

For more information, visit thecma.ca.

CANADIAN MARKETING ASSOCIATION CMA

The best guardrails are the ones you deploy.

CANADIAN
MARKETING
ASSOCIATION
CMA