Presentation Script

Introduction

Good [morning/afternoon], everyone. Thank you for being here today. I will be presenting our framework for ethical standards in artificial intelligence, focusing on three foundational topics: fairness, safety, and responsibility.

These principles are critical not only for compliance and regulation, but also for long-term trust, sustainability, and business success.

Today, I will begin by outlining the ethical challenges AI presents, then describe our proposed standards, provide normative arguments for adopting them, and finally explain their practical implications.

Ethical Challenge

Artificial intelligence is now embedded in decision-making across business sectors: hiring, finance, healthcare, and customer service. But this raises three core challenges:

- Fairness: Algorithms can unintentionally amplify social inequalities. For example, the COMPAS criminal justice case revealed how predictive models may reflect racial bias, raising fairness concerns.

- Safety: AI generates risks ranging from misinformation and deepfakes to privacy violations and manipulative targeting. These harms can be both direct and indirect.

- Responsibility: When harm occurs, accountability is often unclear. Should we hold designers, companies, institutions, or end-users responsible? The absence of clear standards weakens public trust.

Together, these challenges show why we need actionable ethical standards.

Proposed Standards

We propose three interlocking standards:

1. Fairness Standards: Companies should implement parity-based audits such as demographic parity, predictive parity, and equality of opportunity. Tools like adversarial de-biasing and counterfactual dashboards ensure decisions are equitable and explainable.

2. Safety Standards: AI must be benchmarked against harm taxonomies and safety frameworks like ALERT and SALAD-Bench. This ensures coverage of direct harms (e.g., violence, privacy violations) and indirect harms (e.g., deception, manipulation).

3. Responsibility Standards: Adopt a shared enterprise liability model where responsibility is distributed across designers, practitioners, and institutions. Human-in-the-loop oversight must be maintained for meaningful accountability.

Without these standards, companies expose themselves to reputational damage, lawsuits, and regulatory fines.

Normative Arguments

So, why should companies adopt these standards? The justification is strong across multiple ethical perspectives:

- Deontological: These standards protect rights and prevent discrimination, ensuring equal opportunity.
-
- Consequentialist: They reduce foreseeable harms while strengthening long-term trust in AI products and services.
-
- Agent-Based: They embody corporate responsibility, demonstrating due care and ethical integrity.

No matter which ethical lens we apply, these standards are preferable to weaker or permissive policies.

Practical Implications

Since they are not abstract ideals, they can be operationalized through:

1. Quarterly fairness audits disaggregated by demographics.
2. Deployment of counterfactual dashboards for transparency.
3. Mandatory safety benchmarks before releasing models.
4. Clear liability templates to assign responsibility in case of harm.
5. Training stakeholders to understand ethical tradeoffs.

These practices create a culture of accountability and trust, satisfying compliance while ensuring sustainable adoption.

Conclusion

Briefly wrapping this literature review up, I hope we can all understand that fairness, safety, and responsibility are not optional. They are cornerstones of trustworthy AI in business.

Adopting these standards protects companies legally, preserves reputations, and creates long-term value. By committing to them, businesses will not only meet compliance requirements, but also lead in building a future where AI serves society responsibly.

Thank you, and I welcome your questions.