International Journal of Science, Strategic Management and Technology
Volume **02** Issue **04** April-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

Dynamic Ethics of Artificial Intelligence

Submitted by

Snehal Thete
(Assistant Professor, Kaveri College of Arts, Science and Commerce, Pune)

Chaitali Birewar
(Student, Fergusion College, Pune)

https://doi.org/10.55041/ijsmt.v2i4.098

Cite this Article: Birewar, C. (2026). Dynamic Ethics of Artificial Intelligence. International Journal of Science, Strategic Management and Technology, *02*(04). https://doi.org/10.55041/ijsmt.v2i4.098

Abstract

Rigorously AI evolved over humans, but adapting dynamic ethics closer to moral values is still a question in AI world. Statically defined protocols and measures are being taken care by the developers but dynamic behavior according to multiple scenarios is still lacking. Even if we proved something morally correct to chatbot still it will stick to its predefined protocols without compromising it.
The further research is all about the methodology to design a system that will adapt the dynamic framework of ethics. It spans the ideology of gaining and revising the values and applying it according to the need to preserve privacy security and the morality.

Keywords: Artificial Intelligence, Moral Vector Modeling, Fairness and transparency, Ethical AI, Responsible AI system, Human-centered AI.

Introduction

In Understanding ethics is the major part of this research, confusion occurs when we talk about ethics and morality together. To move ahead we start with the understanding of these concepts
When we talk about ethical behavior we meant by the rules or Standard set by organizations, society or others (laws, system rules, professional rules) it is more structured and formal, but when we talk about moral, we meant personal beliefs about right and wrong, based on individual values comes from traditions, culture and all.
Ethics provides a framework for making moral judgments and decisions. Ethics is not a static set of rules but rather a dynamic field that evolves as societies and cultures change. This initiate our discussion throughout the research, we want AI to adapt dynamic ethical technique which is required in some situations, mostly static compliance-based models are present, assuming AI behavior remains stable, lack of model integrations
Need for dynamic moral-ethical alignment is because ethics must evolve with morality, AI system also require continuous monitoring.
We chose the approach of Machine learning model and tried to build an system with mathematical plotting for better understanding.
Our paper contributes in theoretical difference between ethics and morality followed by a Model and mathematical

integration of model weight into learning, empirical validations against static fairness models.

Objectives

1. To formalized difference between morals and ethics in AI system
2. To develop a mathematical model for dynamic ethical adaptation
3. To integration ethical risk into optimization function
4. To analyze Adaptive stability of ethical learning
5. To test the framework through computational experiments

Literature Review

1. Foundation of AI

A rapid deployment of Artificial Intelligence in areas like healthcare, finance, government and criminal justice has growing ethical assessment. Early AI ethics focused on key principles such as fairness, accountability, transparency, privacy, and explainability. These principles guide responsible AI development.

However, most early ethics frameworks used fixed rules and simple guidelines. They were not designed to adapting social conditions or new technical challenges

2. Limitations of Static Ethical Frameworks

Recent scholarship highlights structural limitations in dominant ethical AI approaches. Brown (2025) explain that static ethical models are often implemented only once, such as during testing or before deployment assessments. Such model does not handle post-deployment risks, changing data patterns or shifting social values.

Also, static approaches focus on technical fairness measure and ignore wider social and political factors. Ethical oversight is often treated as documentation task instead of an ongoing responsibility. This compliance-oriented mindset creates a gap between a ethical goals and real-world practice.

3. Difference between morals and ethics in Ai

Modern discussion separate morals and ethics in AI systems.

Morals represent basic human values shaped by cultural norms and philosophical traditions, such as justice, equality and dignity.

Ethics refers to structured frameworks and systems that apply those moral principals into practice.

In AI systems, morals define desired outcomes we want, while ethics controls implementation process. Therefore, effective AI governance required both moral values and proper ethical rules.

4. Emergence of Dynamic Ethical Framework

To solve the problems of rigidity of static models, recent researchers now support dynamic ethical systems. Brown (2025) proposes an approach where ethics is considered throughout the entire AI lifecycle from data collection and model training to deployment and post deployment monitoring.

Dynamic Frameworks focus on:

1. Iterative algorithmic evaluation
2. Continuous monitoring mechanisms
3. Stakeholders engagement
4. Adapting different contexts
5. Correcting issues when needed

Scope of Study

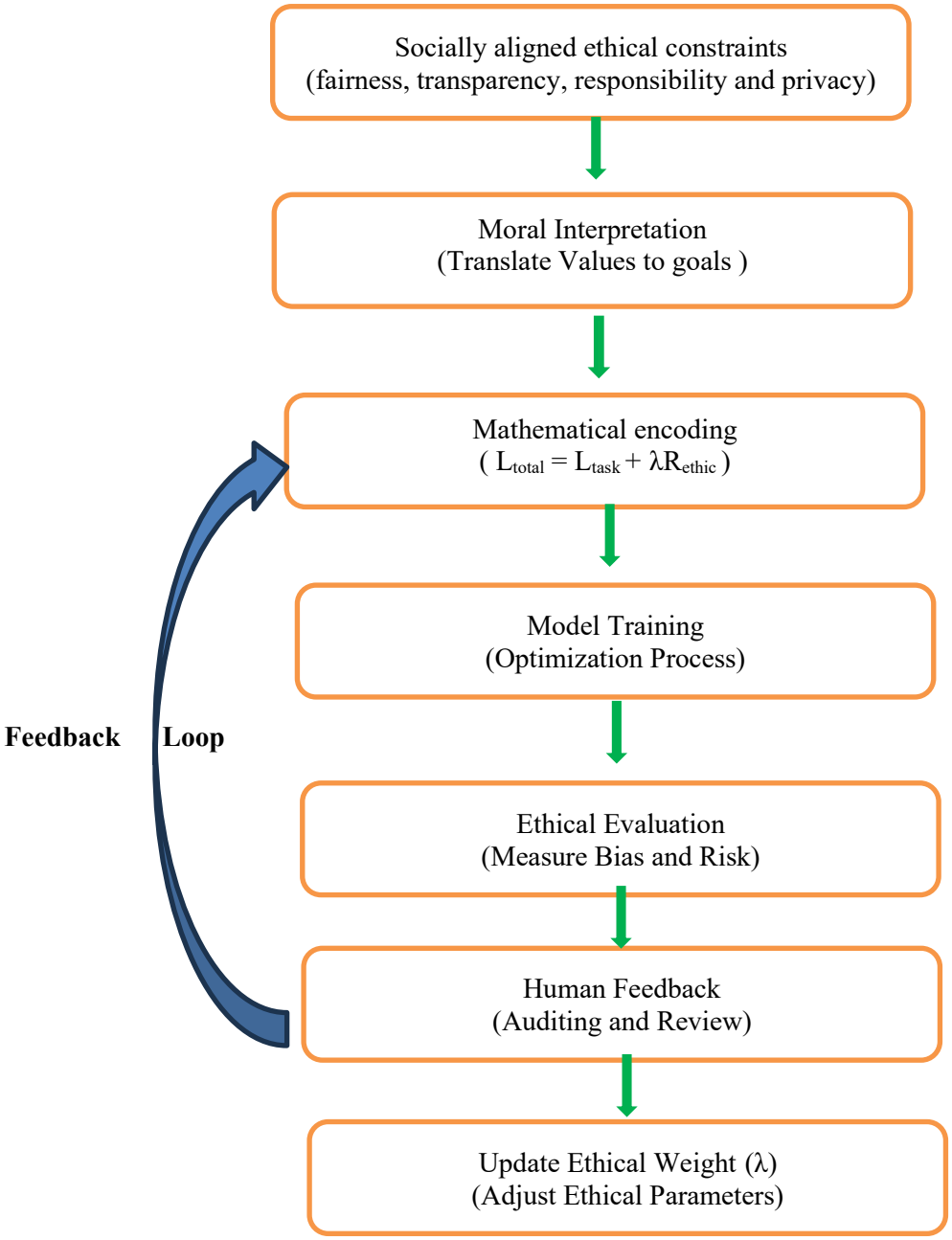
This research focuses on developing a framework for **dynamic ethical adaptation in AI systems** by combining **normative ethics** with **evolutionary ethics** under **human-in-loop supervision**. The framework allows AI to continuously adjust their behavior based on human feedback respecting immutable safety and ethical constraints. By differentiating between morals and ethics, the research explores how AI can align with moral expectations through procedures and feedback -driven adaptation. A mathematical model is introduced to represent ethical penalties, human feedback loop and adaptive parameter updates and, iterative optimization used to evaluate stability and effectiveness. Experiments are performed in controlled scenarios, such as content moderation, hiring fairness and resource allocation, avoiding real-world

risk. The framework emphasizes explainability and traceability, ensuring that all adaptive changes are auditable and interpretable by humans. The study does not claim that AI possesses true moral understanding, its learning is limited to ore-defined ethical preferences, and safety-critical constraints are strictly enforced.

Research gap

1. A single model that combining morals and ethics
2. A dynamic system that updates ethical mechanism
3. A mathematical proof which analyzes ethical learning
4. A clear method for evolving ethical weights.

AI ethics integrated Flowchart



Mathematical Model for Dynamic Moral–Ethical AI System

1. Societal Moral Values

Let:

$$V = \{v_1, v_2, \ldots, v_k\}$$

Where:

v_1= fairness

v_2= harm minimization

v_3= equity and so on…

Now, we encode societal moral values as a **moral vector**:

$$\mathbf{m} \in \mathbb{R}^k$$

Each component represents the strength or importance of a moral dimension.

2. Moral Interpretation → Goal Translation

Define a mapping:

$$\Phi: \mathbf{m} \rightarrow \mathcal{G}$$

This mapping translates moral values into measurable goals.

Example: a) Fairness → demographic parity constraint

b) Harm → bounded risk metric

c) Equity → equalized odds constraint

Formally:

$$\mathcal{G} = \{g_1(x), g_2(x), \ldots, g_k(x)\}$$

Each $g_i(x)$is a measurable ethical function over predictions.

3. Ethical Rule Formalization

We convert moral goals into mathematical constraints:

$$g_i(x) \leq \epsilon_i$$

$$R_{ethic} = \sum_{i=1}^{k} w_i \, g_i(x)$$

Where:

R_{ethic}= ethical risk

w_i= importance weight derived from moral vector $\mathbf{m}$.

4. Mathematical Encoding (Your Loss Function Block)

$$L_{total} = L_{task} + \lambda R_{ethic}$$

Full expansion:

$$L_{total}(\theta, \lambda) = L_{task}(\theta) + \lambda \sum_{i=1}^{k} w_i \, g_i(\theta)$$

Where:

θ= model parameters

L_{task}= prediction loss (cross-entropy, MSE, etc.)

λ= ethical trade-off weight

R_{ethic}= aggregated ethical penalty

This is a **multi-objective optimization problem**.

5. Model Training → Optimization Dynamics

We optimize:

$$\theta_{t+1} = \theta_t - \eta \nabla_\theta L_{total}$$

Expanded:

$$\theta_{t+1} = \theta_t - \eta(\nabla L_{task} + \lambda \nabla R_{ethic})$$

This shows:

Task learning

Ethical regularization influence

6. Ethical Evaluation → Risk Measurement

Define ethical risk:

$$E(t) = R_{ethic}(\theta_t)$$

Bias example:

$$Bias = | P(\hat{Y} = 1 \mid A = 0) - P(\hat{Y} = 1 \mid A = 1) |$$

Risk example:

$$Risk = \mathbb{E}[harm(x)]$$

Ethical compliance condition:

$$E(t) \leq \delta$$

7. Human Feedback → Adaptive Control Signal

Let:

$$H(t) = f(E(t), Audit(t))$$

Human reviewers generate correction signal:

$$u(t) = \text{Human Adjustment Signal}$$

8. Update Ethical Weight $\lambda(t)$

Now the key dynamic component:

$$\frac{d\lambda}{dt} = \alpha(E(t) - \delta)$$

Discrete form:

$$\lambda_{t+1} = \lambda_t + \alpha(E(t) - \delta)$$

Where:

α= adaptation rate

δ= acceptable ethical threshold

Interpretation:

If ethical violation increases → λincreases

If ethical risk low → λdecreases

System auto-adjusts moral strictness

Full Coupled Dynamical System

We now have a **coupled system**:

$$\theta_{t+1} = \theta_t - \eta \nabla_\theta (L_{task} + \lambda_t R_{ethic})$$
$$\lambda_{t+1} = \lambda_t + \alpha (R_{ethic}(\theta_t) - \delta)$$

This forms a **bi-level adaptive moral-learning system**.

Conceptual Overview

The above mathematical formulation model of ethical Ai as coupled adaptive dynamical system in which task learning and moral regulation evolve together over time. The formulation integrates social moral values into the learning objective through a dynamic ethical distribution mechanism ruled by feedback and human oversight.

Unlike static ethical constraints, our model allows continuous moral adaptation, assuring long term alignment with evolving societal norms

Theoretical Contribution

1. This mathematical model involves:
2. Bridges normative ethics and machine learning optimization.
3. Formalizes morality as adaptive regularization.
4. Introduces feedback-controlled ethical alignment.
5. Establishes a control-theoretic interpretation of AI governance.

Experimental Implementation of Dynamic Ethical Learning

1. Dataset and Preprocessing

This experiment were conducted using the Adult Income dataset, a frequently used benchmark for fairness research. The objective is binary income classification ($> 50K\ vs\ \leq 50k$)

Target Variables

$$\boldsymbol{y = \{0, 1\}}$$

1 : Income > 50k

0 : Income $\leq 50k$

Sensitive Attribute

S $\in \{0,1\}$

1 : Male

0 : Female

The sensitive attribute is used for fairness evaluation but not removed from training, allowing bias detection.

Preprocessing Steps

1. Column cleaning and whitespace normalization
2. Binary encoding of income
3. Binary encoding of gender
4. One-hot encoding of categorical variables
5. Feature scaling using StandardScaler
6. Train–test split (70–30)

2. Model Architecture

A **logistic regression** model is implemented using PyTorch

```
import torch.nn as nn

class LogisticModel(nn.Module):
    def __init__(self, input_dim):
        super().__init__()
        self.linear = nn.Linear(input_dim, 1)

    def forward(self, x):
        return torch.sigmoid(self.linear(x))
```

This simple linear model allows clear analysis of performance-fairness balance.

```
input_dim = X_train.shape[1]

#  Baseline (No Fairness)
baseline = train_model(LogisticModel(input_dim), lambda_fair=0.0)

# Static Fairness
static_model = train_model(LogisticModel(input_dim), lambda_fair=0.5)

#  Dynamic Ethical Model
dynamic_model = train_model(LogisticModel(input_dim), lambda_fair=0.1, dynamic=True)
```

3. Fairness Metrics

a) Demographic Parity (DP)

Measure difference in positive prediction rules rates and Lower values indicate reduced bias

b) Equal Opportunity (EO)

Measure difference in true positive rates and ensures equal treatment among qualified individuals.

c) Ethical Drift (ED)

Measure difference in prediction variability and capture distributional instability across group.

4. Ethical Regularized Objective Function

This function implements mathematical encoding block of the framework

5. Training Configurations

a) Baseline Model

$\lambda=0$

Optimize only classification accuracy.

b) Static Fairness Model

$\lambda=0.5$

Fairness penalty remains constant during training.

c) Dynamic Ethical Model

$\lambda t=0.1+0.02 \cdot t$

This equation implements a time-varying ethical weight, simulating adaptive moral application

6. Optimization

This creates a coupled interaction between predictive accuracy and fairness regulation.

7. Evaluation Metrics

Metrics is evolved using:

Accuracy

Precision

Recall

F1-score

Fairness was evaluated using:

Demographic Parity

Equal Opportunity

Ethical Drift

This allows quantitative comparison of:

Performance-only learning

Static ethical regulation

Dynamic ethical adaptation

```
print("Baseline Model:")
print(evaluate(baseline))

print("\nStatic Fairness Model:")
print(evaluate(static_model))

print("\nDynamic Ethical Model:")
print(evaluate(dynamic_model))
```

```
Baseline Model:
{'Accuracy': 0.7391660410837371, 'Precision': 0.469218989280245, 'Recall': 0.895906432748538, 'F1 Score': 0.6158793969849247}

Static Fairness Model:
{'Accuracy': 0.7558861666552924, 'Precision': 0.48541163352536704, 'Recall': 0.7637426900584795, 'F1 Score': 0.5935689126235655}

Dynamic Ethical Model:
{'Accuracy': 0.7522691598989968, 'Precision': 0.4808603718556325, 'Recall': 0.7713450292397661, 'F1 Score': 0.5924096114978666}
```

8. Research Contribution of this Implementation

This implementation shows:

a) Practical integration of fairness into optimization.

b) Dynamic moral weighting during training

c) Empirical comparison static and adaptive ethical enforcement

d) Quantitative measurement of performance-fairness balance

Findings

1. Ethical penalty helps to reduce bias over
2. Performance decreases slightly, but it stays stable
3. Changing λ over time improves fairness compared to static λ.
4. The model stays stable as long as λ is bounded.

Conclusion

This research presents a dynamic ethics framework that combines moral ideas, mathematical modeling and adaptive learning. It connects moral vales with changing ethical rules to help AI system remain aligned with social expectations over time. Theoretical analysis and experimental study show that the system remains stable and reduces bias.

Future Work

1. Applying the methos to deep neural network
2. Optimizing multiple ethical goals at the same time
3. Modeling how moral values differ across cultures
4. Real world deployment case studies

Reference

Abdulrahman M. Al-Zahrani & Talal M. Alasmari, 2024. "**Exploring the impact of artificial intelligence on higher education: The dynamics of ethical, social, and educational implications**," Humanities and Social Sciences Communications, Palgrave Macmillan, vol. 11(1), pages 1-12, December.

David Brown. Rethinking Ethical AI: A Dynamic Framework for Responsible Algorithmic Decision-Making. *Authorea.* September 17, 2025.

Gundersen L, Aasebø E, Ødegård Ø (2023) Artificial intelligence in healthcare: A systematic review of patient perspectives. International Journal of Medical Informatics 170: 104436.

High-Level Expert Group on Artificial Intelligence (AI HLEG), "Ethics Guidelines for Trustworthy AI," European Commission, Apr. 2019.

L. Floridi *et al.* , "AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations," *Minds Mach.* , vol. 28, no. 4, pp. 689–707, 2018

Lahoti, Ronith, The Socio-Ethical Dynamics of Artificial Intelligence in Healthcare (April 08, 2024). Available at SSRN: https://ssrn.com/abstract=4982593 or http://dx.doi.org/10.2139/ssrn.4982593

Mrinal Kanti Sarkar, Smt. Sangita Dey Sarkar (2024), The Ethics Of Artificial Intelligence: Ethics And Moral Challenges

UNESCO, "Recommendation on the Ethics of Artificial Intelligence," adopted Nov. 2021. (Global normative instrument for AI ethics)