STRIDE: A Novel Kernel-Based Framework for Explainable AI

Submitted to IEEE TNNLS (Under Review) & Patent-Pending

Chaeyun Ko | M.S. in Mathematics
chaeyuniris@gmail.com | linkedin.com/in/chaeyunko | github.com/chaeyuniris

Abstract

This document provides a technical overview of STRIDE (Subset-free Kernel-Based Decomposition for Explainable AI). Born from a fundamental inquiry into the mathematical principles governing AI, STRIDE is a novel framework designed to overcome the limitations of traditional XAI methods. Grounded in rigorous mathematics, it utilizes orthogonal decomposition in Reproducing Kernel Hilbert Space (RKHS) to offer a transparent and efficient interpretation of complex AI models. We present its core philosophy, key performance benchmarks, and its ultimate vision to expand the frontiers of human knowledge through trustworthy AI.

1 Vision & The Problem: A Question-Driven Approach

My journey into AI began not with code, but with a question: What are the fundamental mathematical truths that govern this artificial "intelligence"? **For me, a powerful question is the key that unlocks intuition.** This philosophy, bridging my background in mathematics, economics, and philosophy, led me to challenge the existing paradigms of Explainable AI.

While foundational, conventional methods like SHAP presented a paradox: they sought to explain "black boxes" while their own computational complexity created new kinds of opaqueness. This inspired an ambitious goal: to develop a new framework from first principles. This is the "why" behind STRIDE—a personal quest to transform a **fundamental question into a real-world impact.**

2 Core Idea: The "What"

STRIDE introduces a new paradigm for model interpretation based on a simple yet powerful philosophy: **any complex function can be decomposed into a sum of orthogonal, simpler components.**

Instead of relying on combinatorial subset sampling, STRIDE leverages the mathematical elegance of Reproducing Kernel Hilbert Space (RKHS). The core mechanism is an orthogonal decomposition algorithm that analytically separates a model's output into main effects (single features) and high-order interaction effects.

This approach is embodied in the STRIDE framework, a modular system designed for comprehensive model interpretation. As detailed in our patent filing, the framework consists of six core modules that work in concert: (1) a Subset-free Kernel Operation module, (2) a Data Transform module, (3) a Domain-Graph integration module, (4) an Intrinsic Component-based decomposition module, (5) a Differential analysis and distribution module, and (6) an Explanation generation module.

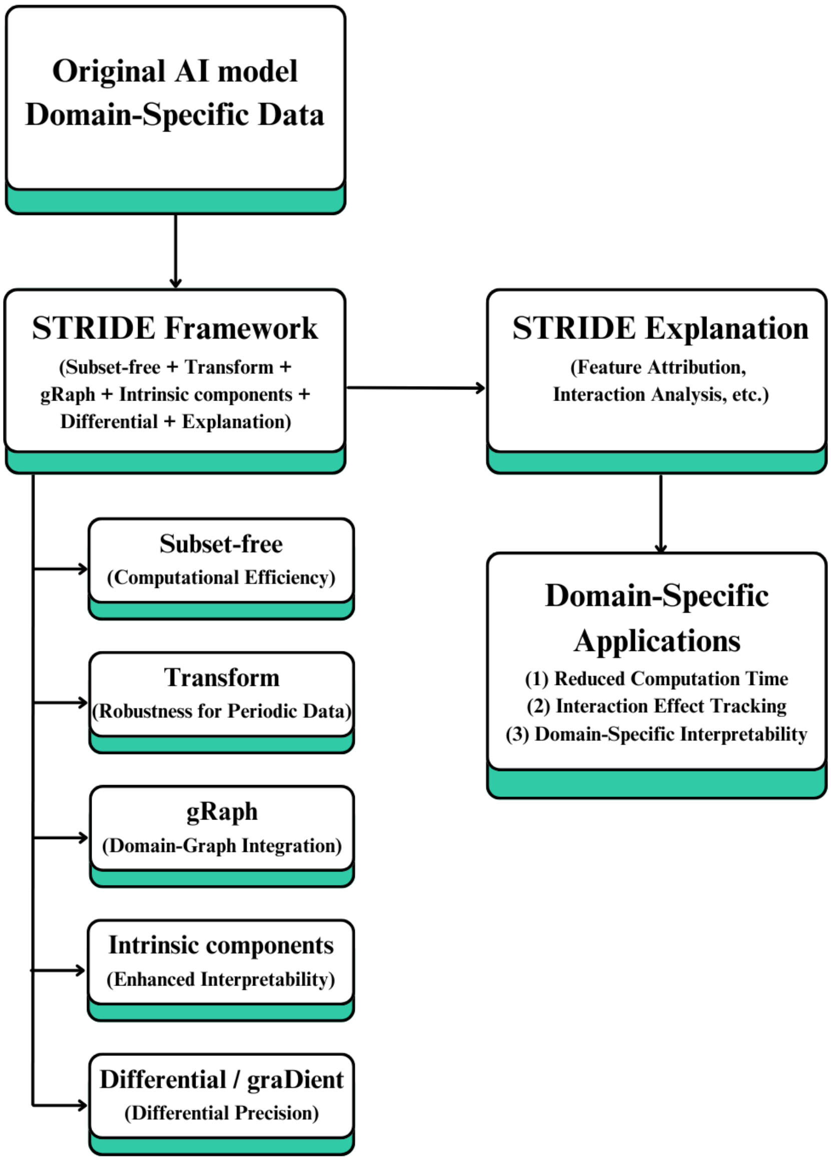

Original AI model
Domain-Specific Data
STRIDE Framework
(Subset-free + Transform +
gRaph + Intrinsic components +
Differential + Explanation)
STRIDE Explanation
(Feature Attribution,
Interaction Analysis, etc.)
Subset-free
(Computational Efficiency)
Transform
(Robustness for Periodic Data)
gRaph
(Domain-Graph Integration)
Intrinsic components
(Enhanced Interpretability)
Differential / graDient
(Differential Precision)
Domain-Specific
Applications
(1) Reduced Computation Time
(2) Interaction Effect Tracking
(3) Domain-Specific Interpretability

Figure 1. System Architecture of the STRIDE Framework

3 Key Results: Empirical Validation

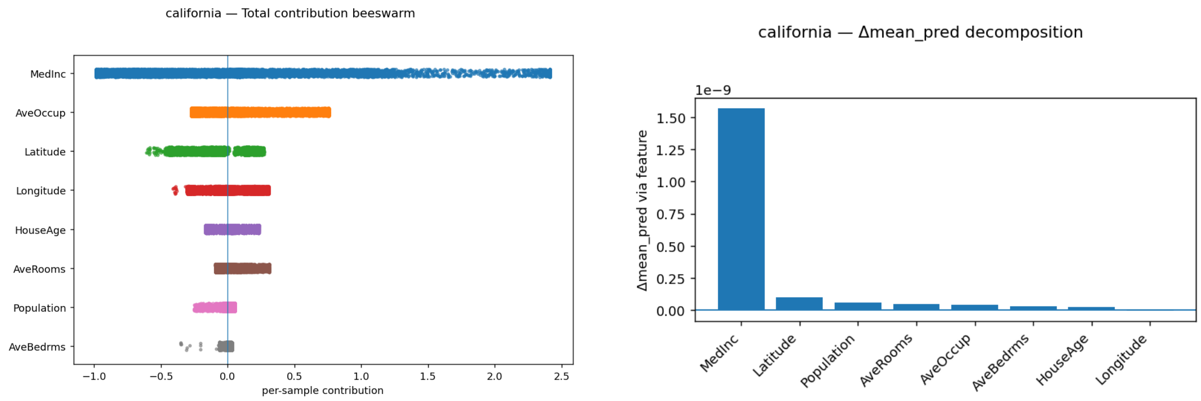
The efficacy of STRIDE is validated through its superior performance benchmarks and comprehensive explanation capabilities. The framework consistently demonstrates high fidelity, significant computational speedup against industry standards, and offers unique analytical functions.

Table 1. Benchmark: STRIDE vs. TreeSHAP (Averaged over 3 seeds)

Dataset	Metric	STRIDE	TreeSHAP	Result
California Housing	Time (s)	**0.55s** ± 0.01s	5.21s ± 0.02s	**9.5x Speedup**
(Regression, d=8)	Fidelity (R^2)	0.931 ± 0.003	1.000	High Fidelity
Bank Marketing	Time (s)	**2.70s** ± 0.03s	14.65s ± 0.02s	**5.4x Speedup**
(Classification, d=62)	Fidelity (R^2)	0.990 ± 0.001	1.000	High Fidelity
Credit Default	Time (s)	**1.50s** ± 0.01s	11.30s ± 0.06s	**7.6x Speedup**
(Classification, d=23)	Fidelity (R^2)	0.988 ± 0.001	1.000	High Fidelity

Beyond these quantitative benchmarks, STRIDE provides a rich suite of visual explanations that illustrate its comprehensive capabilities, reliability, and unique functions.

Comprehensive Explanations (Global & Local)



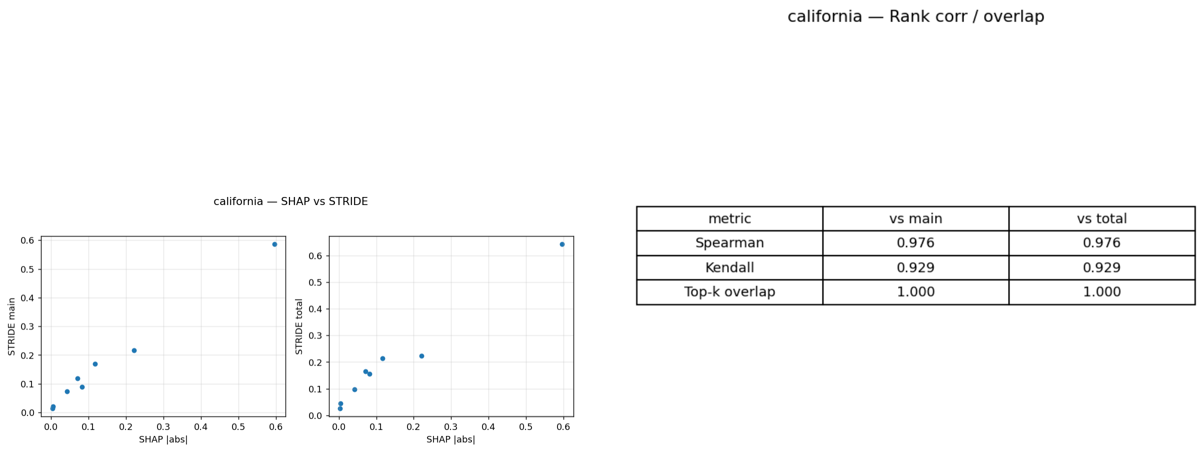
(a) Global Explanation (Beeswarm Plot)

(b) Local Explanation (Delta Breakdown)

Figure 2. STRIDE delivers a full explanatory suite, from a global summary of feature impacts across the entire dataset (a) to a detailed delta breakdown for a single prediction (b).

Interpretation: *These visualizations confirm that STRIDE masters the fundamentals of XAI, providing the foundational explanations necessary for any practical application.*

Reliability & Trustworthiness



metric	vs main	vs total
Spearman	0.976	0.976
Kendall	0.929	0.929
Top-k overlap	1.000	1.000

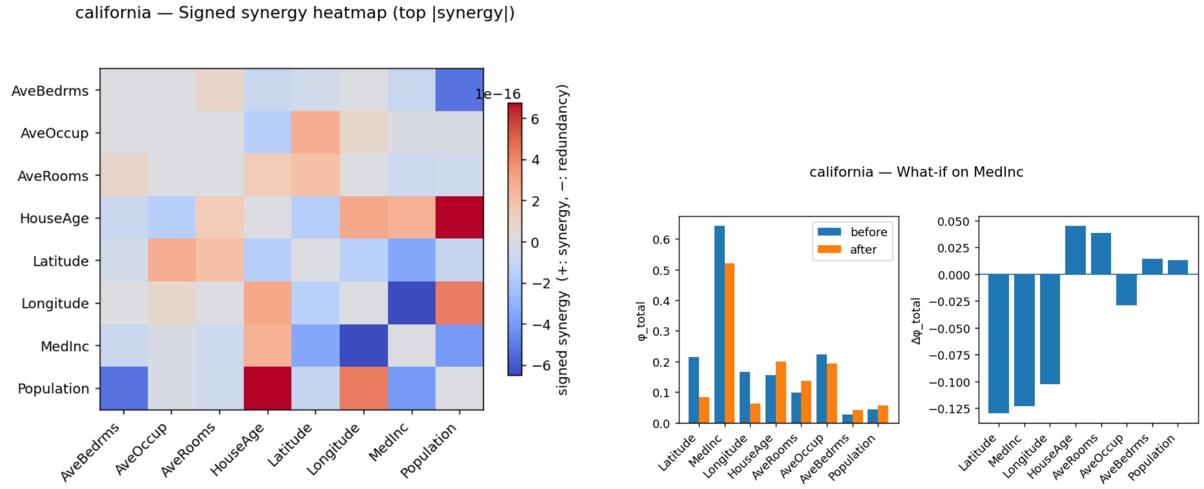
(a) High Value Correlation with TreeSHAP

(b) High Rank Correlation with TreeSHAP

Figure 3. Reliability validation. (a) A direct comparison shows a high correlation between STRIDE's and TreeSHAP's attribution values. (b) The high rank correlation scores further confirm this consistency.

Interpretation: *This demonstrates that the significant speedup shown in Table 1 is achieved without compromising the accuracy and integrity of the explanations, validating STRIDE as a trustworthy solution.*

Unique Analytical Functions Beyond Attribution



(a) Interaction Synergy Heatmap

(b) What-if Analysis Simulation

Figure 4. Advanced analytical capabilities. (a) The synergy heatmap quantifies complex high-order interactions. (b) The "What-if" analysis provides a simulation of how changing a feature's value impacts other features' contributions.

Interpretation: *These functions showcase that STRIDE is more than just a faster SHAP. It represents a new frontier in XAI, enabling deeper insights and providing a dynamic, simulation-based tool for proactive decision-making.*

4 Impact & Future Work: Expanding the Frontiers of Knowledge

STRIDE is more than an algorithm; it's a step towards a future where AI acts not as an opaque oracle, but as a trustworthy partner in discovery. By providing reliable explanations, it can unlock new insights in science, finance, and medicine, ultimately contributing to the expansion of human knowledge.

My vision for this research extends to answering even deeper questions:

- **From "What" to "Why":** Integrating causal inference to move beyond correlation-based explanations.
- **Beyond Tabular Data:** Applying the core principles to interpret the complex, multi-modal world of graphs, images, and language.

I believe this work serves as a foundational step towards building a truly collaborative human-AI future, where technology helps us answer questions we haven't even thought to ask yet.