# AI/ML Soot Blowing Optimization Project

## Current Status Documentation

### DEVELOPMENT CONTEXT & CHALLENGES

#### Generative AI for Coding (First-Time Implementation)

* Challenge: Principal investigator had no prior experience using generative AI for software development
* Approach: Systematic evaluation of multiple AI coding assistants and development workflows
* Outcome: Successfully integrated AI-assisted development for complex engineering simulation code
* Impact: Accelerated development timeline while maintaining code quality and physics accuracy

#### Advanced Thermodynamics Knowledge Acquisition

* Challenge: Simulation complexity required deeper thermodynamic understanding beyond operations experience
* Learning Areas:
  + Advanced heat transfer coefficient correlations
  + Multi-phase flow and steam property calculations
  + Complex fouling mechanisms and deposition physics
  + Combustion gas composition and stack gas analysis
* Approach: Self-directed study to bridge practical operations knowledge with simulation requirements
* Outcome: Successfully guided AI coding efforts with enhanced technical understanding

#### AI Model Evaluation Process

* Evaluated Models: Multiple generative AI platforms for coding assistance
* Selection Criteria:
  + Understanding of complex engineering problems
  + Ability to generate thermodynamically sound code
  + Quality of technical explanations and documentation
  + Reliability for iterative development cycles
* Result: Optimized AI-assisted development workflow for engineering simulation projects

#### Development Methodology Impact

* Knowledge-Guided AI Development: Combined domain expertise with AI assistance
* Iterative Learning: Enhanced technical knowledge in parallel with code development
* Quality Assurance: Used engineering judgment to validate AI-generated thermodynamic calculations
* Efficiency Gains: AI assistance significantly accelerated coding while maintaining technical rigor

### Project Overview

**Project: AI/ML-based soot blowing optimization for pulverized coal-fired boilers Principal Investigator: Principal Data Scientist (with boiler operations background) Goal: Commercial demo to win client contracts Timeline: 2 months (8 weeks) to working demo**

#### Current Development Status

✅ COMPLETED COMPONENTS

1. Comprehensive Boiler Simulation Framework

* thermodynamic\_properties.py: ✅ COMPLETE
  + Thermo library integration for accurate steam/water/gas properties
  + Temperature range: 32°F to 1500°F
  + Pressure range: 1-5000 psia
  + Includes safety fallbacks for property calculations
* heat\_transfer\_calculations.py: ✅ COMPLETE
  + Enhanced heat transfer coefficient calculations
  + Individual segment-level analysis
  + Nusselt number correlations for various geometries
  + Overall U-value calculations with fouling effects
* fouling\_and\_soot\_blowing.py: ✅ COMPLETE
  + Realistic fouling progression models
  + Individual segment soot blowing simulation
  + Fouling gradient calculations based on temperature/position
  + Cleaning effectiveness modeling

2. Coal Combustion Integration

* coal\_combustion\_models.py: ✅ COMPLETE
  + Coal ultimate analysis processing
  + NOx formation modeling (thermal and fuel NOx)
  + Combustion efficiency calculations
  + Soot production modeling with deposition tendencies
  + Realistic fouling distribution based on gas temperatures

3. Complete Boiler System Model

* boiler\_system.py: ✅ COMPLETE
  + 7-section boiler model (furnace → generating bank → superheaters → economizers → air heater)
  + Heat transfer cascading through sections
  + Steam/water flow modeling
  + System performance calculations

4. Annual Operation Simulation

* annual\_boiler\_simulator.py: ✅ COMPLETE
  + Full year simulation capability (8,760 hourly data points)
  + Massachusetts weather patterns integration
  + 4 different coal quality profiles
  + Variable load operation (45-100% capacity)
  + Realistic soot blowing schedules (8-168 hour cycles)

5. Data Analysis and Visualization Tools

* analysis\_and\_visualization.py: ✅ COMPLETE
  + Comprehensive system analysis tools
  + Performance trending and comparison
  + Fouling progression analysis
  + Economic impact calculations
* data\_analysis\_tools.py: ✅ COMPLETE
  + Annual operation data analysis
  + Seasonal performance patterns
  + Coal quality impact analysis
  + Optimization opportunity identification

6. ML Dataset Generation Framework

* ml\_dataset\_generator.py: ✅ COMPLETE
  + Comprehensive dataset generation for ML training
  + Multiple cleaning scenarios and strategies
  + Economic optimization features
  + Target variable creation for supervised learning

7. Demonstration and Integration

* main\_demonstration.py: ✅ COMPLETE
  + Complete system integration demonstration
  + Multiple test scenarios
  + Performance validation
  + Economic analysis
* run\_annual\_simulation.py: ✅ COMPLETE
  + End-to-end annual simulation runner
  + Automated data generation and analysis
  + Report generation capabilities

#### 🔧 CURRENT ISSUES (Week 1 Priority)

Primary Issue: Stack Temperature Realism

* Problem: Stack temperature is completely static at 250°F with no variation
* Root Cause: Heat transfer calculations not properly responding to changing operating conditions
* Impact: Static temperature makes simulation output unbelievable for commercial demo
* Expected Behavior: Should vary 250-350°F based on load, fouling, coal quality, ambient conditions
* Estimated Fix Time: 1 week (aligns with your assessment)

Secondary Issues:

* Temperature Responsiveness: Stack temperature should respond to:
  + Load variations (45-100% capacity)
  + Fouling buildup over time
  + Coal quality changes
  + Ambient temperature variations
  + Soot blowing effectiveness
* Convergence Stability: Some operating conditions cause solver instability
* Heat Balance: Energy conservation needs refinement across all sections

#### 📊 COMPLETED DATASETS

Extensive Dataset Generation Experience

* Generated Datasets: 6 complete annual datasets (8,760 hours each)
* Total Data Points: 52,560 hours of simulated operation across multiple scenarios
* Exploration Status: All 6 datasets analyzed for patterns, trends, and simulation behavior
* Learning Outcome: Deep understanding of simulation parameter effects and data quality requirements

Target Dataset for Demo

* Goal: Generate 12-month dataset (8,760 hourly data points) with realistic temperature variation
* Requirements:
  + Stack temperature varying 250-350°F based on operating conditions
  + Responsive to load changes, fouling progression, coal quality
  + Realistic seasonal patterns and maintenance cycles

Previous Dataset Analysis Results

* Variables: 142 total columns consistently generated across all datasets
* Operational Coverage: Full range of loads, coal types, weather conditions
* Issue Identified: Stack temperature non-responsiveness discovered through extensive data exploration
* Data Architecture: Proven framework ready for corrected simulation

#### 🏗️ ARCHITECTURE STATUS

Technology Stack

* Language: Python 3.x
* Core Libraries:
  + ✅ thermo (thermodynamic properties)
  + ✅ numpy, pandas (data processing)
  + ✅ matplotlib, seaborn (visualization)
  + ✅ dataclasses, typing (code structure)

Code Organization

* Modular Design: ✅ Well-structured, individual modules for each component
* Documentation: ✅ Comprehensive docstrings and comments
* Version Control: Assumed Git repository structure
* Testing: Limited unit tests (typical for research code)

Integration Status

* Module Dependencies: ✅ All modules integrate successfully
* Data Flow: ✅ Clean data pipeline from simulation → analysis → ML dataset
* Error Handling: Basic error handling present

#### 🎯 WHAT'S WORKING WELL

##### Technical Strengths

1. Comprehensive Physics: Realistic boiler modeling with proper thermodynamics
2. Domain Expertise: Clear understanding of industrial boiler operations
3. Data Richness: 142-variable dataset covers all relevant operational aspects
4. Modular Architecture: Easy to modify and extend individual components
5. Industry Realism: Proper coal types, weather patterns, operating schedules

##### Commercial Demo Readiness

1. Professional Code Quality: Well-documented, maintainable codebase
2. Compelling Data: Rich operational scenarios for demonstration
3. Economic Framework: Built-in ROI and cost analysis capabilities
4. Visualization Tools: Professional plotting and analysis capabilities

#### ⚠️ RISK ASSESSMENT

Critical Risks (Week 1)

* Stack Temperature Fix: If this takes >1 week, timeline compression needed
* Fundamental Physics: If heat transfer model needs complete redesign

Medium Risks (Weeks 2-4)

* ML Model Performance: Models may not show dramatic improvement
* Data Quality: Simulation fixes may require dataset regeneration

Low Risks (Weeks 5-8)

* Demo Interface: Streamlit/Dash development is straightforward
* Presentation Prep: Investigator has strong technical communication skills

#### 📈 CURRENT READINESS FOR COMMERCIAL DEMO

**Current Readiness: 70-75%**

* ✅ Simulation Framework: Complete and sophisticated
* ✅ Data Generation: Proven capability - 6 annual datasets already generated and analyzed
* ✅ Analysis Tools: Professional-grade visualization and reporting
* ✅ Dataset Architecture: 142-variable structure validated across multiple generation cycles
* 🔧 Temperature Responsiveness: Critical issue affecting simulation believability
* ⏳ ML Models: Not yet developed (Weeks 2-4)
* ⏳ Demo Interface: Not yet built (Weeks 5-6)

**Path to 100% Readiness**

1. Week 1: Fix stack temperature responsiveness → 85% ready
2. Week 4: Working ML models on 12-month dataset → 90% ready
3. Week 6: Professional demo interface → 95% ready
4. Week 8: Polished presentation materials → 100% ready

## Commercial Demo Remaining Timeline

### AI/ML Soot Blowing Optimization - Sales-Focused POC

**Executive Summary**

**Context**: Building a commercial demo

**Timeline**: 6-8 weeks optimal from current status (70-75%) to completion

**Goal**: Compelling sales demo that showcases technical capability and business value

#### Optimal 6-8 Week Commercial Timeline From Current Project Status

**Week 1-2: Simulation & Baseline Data**

*Get "good enough" simulation working*

**Week 1: Fix Core Simulation Issues**

* [ ] Achieve realistic stack temperatures (don't perfect them)
* [ ] Stable convergence for demo scenarios
* [ ] Test with 3-4 coal types and load ranges

**Week 2: Generate Demo Dataset**

* [ ] 6-12 months of operational data
* [ ] Include clear patterns clients can understand
* [ ] Document baseline cyclical cleaning performance
* [ ] Create 3-4 compelling scenarios for demo

**Outcome**: Simulation that produces believable results for sales purposes

**Week 3-4: ML Development & Value Demonstration**

*Build models that show clear improvement over scheduled soot blowing approach*

**Week 3: Predictive Models**

* [ ] Random Forest for fouling prediction (focus on economizers)
* [ ] Simple optimization algorithm (rules + ML hybrid)
* [ ] Target 20-30% improvement over baseline (achievable goal)

**Week 4: Economic Impact Calculation**

* [ ] ROI calculator with realistic assumptions
* [ ] Fuel savings projections
* [ ] Maintenance cost optimization
* [ ] Payback period analysis (target <18 months)

**Outcome**: Models that demonstrate clear, quantifiable business value

**Week 5-6: Professional Demo Interface**

*Create impressive visual presentation*

**Week 5: Dashboard Development**

* [ ] Professional-looking Streamlit or Dash interface
* [ ] Real-time fouling status visualization
* [ ] Before/after efficiency comparisons
* [ ] Cost savings calculator

**Week 6: Demo Scenarios & Story**

* [ ] 3-4 compelling use cases
* [ ] "Emergency fouling" response scenario
* [ ] "Seasonal optimization" scenario
* [ ] "Coal quality change" adaptation
* [ ] Clear narrative arc for each demo

**Outcome**: Impressive visual demo that tells a compelling story