Lessons from AI Success for Financial Institution Model Development

I. Executive Summary

This "Complexity Simplified" podcast series focuses on how financial institutions (FIs) can leverage Artificial Intelligence (AI) and Machine Learning (ML) best practices to enhance their model development processes. The central theme revolves around bridging the gap between traditional econometric modeling and the rapidly evolving AI/ML landscape. Key takeaways include the urgent need for FIs to conduct AI readiness assessments, strategically adapt their practices to be more open, collaborative, standardized, and usage-centric, and proactively address the impending transformation of roles and tasks within modeling organizations. While the technology is rapidly maturing, the primary challenge for FIs lies in fostering the necessary process and cultural shifts to enable deep adoption of AI/ML.

II. The Shifting Landscape: Traditional vs. AI/ML Modeling

The podcast series initiates by defining the core distinctions between traditional econometric models and AI/ML models:

- **Traditional Econometric Models:**
  - **Specification:** Guided by economic or theoretical concepts, focusing on "what theory says would make sense" for variable selection and functional form. Model coefficients are "established upfront by fitting the model to the data, and the model changes only through discrete model revisions." (Episode 1)
  - **Estimation:** Utilizes "Classical Statistical methods, using formal hypothesis testing to guide the model specification search, and using a coefficient estimator which has desirable statistical properties." (Episode 1)
  - **Monitoring:** Regular monitoring of predictive accuracy, often through backtesting, with attention to "representativeness of the development data for future periods and to identify limitations such as extrapolation risk." (Episode 1)
  - **Limitations:** Can struggle with "high-dimensional data" and "multi-collinearity problems." (Episode 1)
- **AI/Machine Learning Models:**
  - **Specification:** Machine Learning algorithms "automatically detect complex, non-linear relationships and interactions among variables," allowing for "general" model architectures and less "manual feature engineering." Modelers can "cast a wide net in searching for which data variables to use." (Episode 1)
  - **Estimation (Machine Learning):** An "algorithmic" training process where the estimator takes "distinct steps to better fitting the data until it reaches a point where it is satisfactory to stop." (Episode 1) "ML models cast a wider net in the specification search, nesting many possible specifications in the estimation stage itself." (Episode 1)
  - **Artificial Intelligence:** Broader than ML, AI is "the ability of the Machine to mimic human capabilities we call intelligence, doing things like understanding what we see, the meaning of words seen or heard, planning future actions, and using reasoning." (Episode 1) AI systems leverage ML *and* other learning types, including Transfer Learning.
  - **Advantages:** Often achieve "higher predictive accuracy," adapt "more easily to new patterns as data evolves," and are "more robust in dynamic environments." (Episode 1)
  - **Challenges:** Require "interpretability techniques to meet regulatory and consumer expectations," which traditional models inherently satisfy due to simplicity. (Episode 1)

Key Concepts:

• **Transfer Learning:** "The ability to use knowledge learned earlier, often from a different domain and source of data, to build and apply knowledge in a context different from that of the initial model training." (Episode 1)
• **Foundation Models:** "General" AI models (like Large Language Models) that "can do a wide range of things" and "need to be built upon to do better at specific tasks." (Episode 2)
• **Agentic AI:** AI systems that "autonomously plan, reason, and take actions toward achieving goals." (Episode 1) They are poised to transform roles by automating repetitive decision-making and task execution.

III. Current State of Model Development at Financial Institutions
Financial institutions operate within a "highly regulated" environment due to "responsibility to do things right" and being the location "where the money is." (Episode 1). This regulatory intensity drives specific practices:
• **Formalized Processes:** "Regulations require ways-of-work to be written down formally in policies, standards, and procedures." (Episode 1)
• **Three Lines of Defense:** A structured risk management framework:
  ◦ **First Line:** Model developers and users (business units).
  ◦ **Second Line:** Model validators and centralized model governance.
  ◦ **Third Line:** Internal Audit for independent evaluation.
• **Limited Model Inventories:** Typically "at most a few internal models per model use or a single vendor-provided solution," partly due to regulatory incentives and cost. (Episode 2)
• **Discrete Lifecycle Stages:** A sequential process including "Development, Implementation, and Use," with "explicit hand-off from one stage to the next and gatekeepers." (Episode 2)
• **Focus on Explanation:** Due to regulatory and consumer protection needs (e.g., explaining credit denials), models "need to be explained." (Episode 1)
Current AI Readiness Assessment (Industry Level):
The series assesses current AI readiness in FIs across several dimensions, using credit modeling and financial instrument valuation as examples:
• **Openness:**
  ◦ **Data Availability:** * **Credit Modeling:** "About 3 or so" on a scale of 1-10. Primarily relies on proprietary data, limiting benchmarking. "Most of that data [publicly available] is aggregated, grouped in some way, not at the individual loan and borrower level." (Episode 3) * **Financial Instrument Valuation:** "About a 5." More open due to transactional market data being published by exchanges or readily available from vendors.
  ◦ **Model Results Distribution:** * **Credit Models:** "Limited." Few research papers from companies or open code. Purchased models (e.g., FICO) are "pre-engineered feature[s]" not full predictive models. (Episode 3) * **Financial Instrument Valuation Models:** "Generally are a bit more open," with "open code for any of the major interest rate models." (Episode 3)
  ◦ **Software Languages:** * **Credit Modeling:** Historically proprietary (SAS), moving "towards R and now more Python, both of which are open-source." (Episode 3) * **Financial Instrument Valuation:** Often developed in "C++, which is open source." (Episode 3)
• **Collaboration & Standardization:**
  ◦ **General Lack of Standardization:** "Generally, institutions do not standardize their model development ways-of-working enough." (Episode 3) Standardization often exists only within "silos" (e.g., marketing models or specific statistical methods).
  ◦ **IT Shortfall:** Lack of "end-to-end dev-ops engineering to encapsulate the results of the model development activities." (Episode 3)
  ◦ **Workforce Education Gap:** Differences in staff education and training, reflecting shifts in university curricula (from econometrics to data science/ML engineering), hinder standardization.

• **Benchmarking:** Limited without openness and standardization. Current practices often involve "champion/challenger" models rather than "task-specific" leaderboards. (Episode 2)
• **Usage-centricity:** While model registration and user engagement exist, FIs often lack the "input from Users at the run-time of a model" seen in AI/ML (e.g., prompt engineering, collecting user feedback). (Episode 3)

IV. AI/ML Best Practices and Lessons for Financial Institutions
The podcast identifies several key AI/ML best practices that FIs can adopt:
1. **Open and Collaborative Development Communities:**
◦ **AI/ML Practice:** Characterized by "Openness and sharing," where "Foundation models are widely used and adapted," and researchers "share its work at various conferences" and "publish research papers, the details of the models themselves... and the code." (Episode 1, 2)
◦ **Lesson for FIs:** Move "beyond strict ownership roles to encourage broader peer input and collective problem-solving." (Episode 2) This means fostering internal sharing of datasets and code, documenting processes clearly, and using open-source languages. (Episode 3)
2. **Usage-Based Leaderboards (Benchmarking):**
◦ **AI/ML Practice:** Performance is tracked "across shared tasks, allowing multiple models to compete in a more dynamic, usage-driven environment." (Episode 2) Models are evaluated based on "State-of-the-Art performance levels, as measured through benchmarking" on "task-specific metrics." (Episode 2)
◦ **Lesson for FIs:** Shift evaluation toward "task-specific performance, not just model-versus-model comparisons." (Episode 2) This requires building out "a rich task and sub-task taxonomy" and providing guidance on "how the quality of performance of each can be measured." (Episode 3)
3. **Reinforcement Learning Through Human Feedback (RLHF) and User-Provided Inputs:**
◦ **AI/ML Practice:** "Quality of outputs can be estimated by model or judged by humans." Feedback on model outputs can be "explicitly included in the objective function used in training the model," a technique called "Reinforcement Learning from Human Feedback." (Episode 2) Generative AI systems also allow users to "tailor the model outputs to their specific usages" at runtime via "prompt engineering, providing information sources and context." (Episode 2)
◦ **Lesson for FIs:** Incorporate "structured, iterative feedback from users and domain experts into model refinement." (Episode 2) FIs can start by capturing information whenever "Assumptions are changed or Adjustments made" by model users, providing valuable feedback for future model improvements. (Episode 3)
4. **Flexible Model Architecture Design:**
◦ **AI/ML Practice:** Emphasis on "model architecture as a design discipline," considering "how different modules or components interact, pass information, and improve performance when combined." (Episode 2)
◦ **Lesson for FIs:** Adopt this mindset, even for non-neural network models, to design "a flexible architecture that allows you to iterate, re-use components, and experiment with different structures." (Episode 2) This includes thinking about "systems-of-equations design questions." (Episode 2)
5. **Thoughtful Feature Engineering:**
◦ **AI/ML Practice:** A data pre-processing step where "raw data is transformed into some form of standardized input which makes it easier for the AI-ML model to complete the training stage well," often involving identifying "bite-size" pieces of information. (Episode 2)
◦ **Lesson for FIs:** Pay "more attention to this too," while being cautious of "spurious correlation" or "data mining" in contexts with limited historical variation. (Episode 2)

V. Strategic Recommendations for Executive Leaders
Given the uncertainty surrounding AI's evolution, executive leaders in FI model development groups should focus on proactive positioning:

1. **Conduct an AI Readiness Assessment:** "Use the results of this to update their Strategy and Strategic and Operational Plans." (Episode 3) This assessment should judge "how far a group is from being ready for that end-state" where AI helps with much of the model development work. (Episode 3)
    ◦ **Key Characteristics of the End State:** Open, Collaborative, Standardized, Benchmarked, Task-specific, Usage-centric, Doing Transfer Learning, Improving Performance-in-Use via User-provided Inputs. (Episode 3)
2. **Develop Roadmaps and Clear Obstacles:** "It's going to be a transformational change which is going to require a lot of planning and internal buy-in." Leaders need to "clear some of the readily identifiable obstacles, so that when the organization is ready, they can move quickly and decisively." (Episode 3)
3. **Prioritize Process Transformation:** While "human capital that can be redeployed or trained" and "Technology itself is at least close," the real risk is "shallow adoption, which give the appearance that the organization has been transformed but really leaves a lot of work to be done." (Episode 3)
4. **Promote Internal Openness:** Leaders have "a lot of autonomy to make internal changes." They can "Share datasets... widely circulate... code... Write clear and concise detailed white papers... Use open-source software languages." (Episode 3)
5. **Foster Standardization with Autonomy:** While traditional "Procedural documents" are an option, leaders should consider approaches that grant "more autonomy" to innovative AI/ML staff. (Episode 3) Examples from tech firms (Stripe's Shepherd, Flyte) show how "procedural discipline" can be embedded through "flexible, developer-centric infrastructure." (Episode 3)
6. **Embrace Task-Specific Performance Measurement:** Even before full benchmarking, FIs can introduce "task-specific performance measurement" and develop a "task and sub-task taxonomy." (Episode 3) This can be enhanced by introducing "more humans and machines into the task output evaluation process." (Episode 3)
7. **Integrate AI Thoughtfully via Transfer Learning:** Recognize that "some capabilities—such as learning how to learn, how to research, how to plan, how to take specific actions such as authoring software code or executing it—are generic and transferable to model development." (Episode 3) Focus on "augmenting key cross-cutting capabilities like decision-making, forecasting, and risk assessment" rather than full automation. (Episode 3)
    ◦ **Human-Machine Interaction Models:** Consider AI Advisor and Human Performer, AI Performer and Human Assistant, or AI Performer and Human Validator configurations. (Episode 3)
    ◦ **Leverage Generative AI for Readiness Assessment:** Use Generative AI "to do some of the re-engineering work itself," such as documenting current processes or extracting tasks from documentation. (Episode 3)
8. **Ensure AI-Compatible Technology Stack:** Align with broader organizational technology choices, but ensure access to "one of the big Cloud Providers" for scalability and access to "Foundational Models and fuller Generative AI systems." (Episode 3) Also, consider specialized "tools to conduct reasoning and evaluate the soundness of individual reasoning steps." (Episode 3)

VI. Impact on Jobs and Future Outlook
The AI journey is not a question of "whether" but "when" it will transform ways-of-working in model development. (Episode 3)

• **Job Transformation:** "All jobs are on the table for being re-engineered or eliminated and replaced by machines." (Episode 1)
    ◦ Job listings show a significant shift, with "more than 10 times as many job openings for Machine Learning roles than traditional modeling roles, and jobs emphasizing AI are an even higher multiple, more than 50 times as many." (Episode 1)

    ◦ The trend suggests "we are rapidly approaching an era in which models build and use other models." (Episode 1)

- **Human-AI Collaboration:** While AI models are becoming sophisticated enough to perform "very advanced tasks," human roles will shift "toward higher-level oversight, strategic thinking, and creativity." (Episode 1)
- **Three Lines of Defense Evolution:** Agentic AI "has the potential to fundamentally change the way each line of defense works by shifting their engagement process from being sequential and reactive to one of integrated, proactive collaboration." (Episode 1)

The series concludes with an optimistic yet practical outlook, emphasizing that leaders have a "real opportunity for model risk and development leaders to shape how their institutions adapt" by rethinking architecture, embracing iterative workflows, and balancing innovation with control. (Episode 3)