76th Harry E. Salzberg Memorial Lecture  
Academic Day Honoring Georgia Perakis

November 14, 2024

9:00 A.M. – 4:00 P.M.

Whitman School of Management, Syracuse University  
Official speaker order and talk titles/abstracts coming soon

**Georgia Perakis**

John C Head III Dean (Interim), MIT Sloan School of Management

William F. Pounds Professor of Management

Professor, Operations Management

Professor, Operations Research and Statistics

<https://mitsloan.mit.edu/faculty/directory/georgia-perakis>

**Joline Uichanco**

Associate Professor of Technology Management and Innovation  
NYU Tandon School of Engineering  
<https://engineering.nyu.edu/faculty/joline-uichanco>

**Wei Sun**

Senior Research Scientist at IBM Research  
Technical Assistant to David Cox, VP of AI Models at IBM Research and IBM Director of the MIT-IBM Watson AI Lab.   
<https://research.ibm.com/people/wei-sun>

**Gonzalo Romero**

Associate Professor  
Operations Management and Statistics  
Rotman School of Management  
University of Toronto  
<https://www.rotman.utoronto.ca/the-rotman-experience/our-community/people/romero-gonzalo/>

**Ye Liu**

Assistant Professor Supply Chain Management   
Whitman School of Management  
Syracuse University  
<https://whitman.syracuse.edu/faculty-and-research/faculty-staff-directory/details/yliu623>

**Leann Thayaparan**

Assistant Professor  
Carey Business School  
Johns Hopkins  
<https://carey.jhu.edu/faculty/leann-thayaparan-phd-mba>

**Pavithra Harsha**

Principal Research Scientist  
AI division of IMB T.J. Watson Research Center  
<https://mitsloan.mit.edu/staff/directory/pavithra-harsha>

**Joline Uichanco**

Title: Understanding Consumer Choice for Smarter Pricing and Assortment Decisions

Abstract

Consumer choice models provide a foundation for predicting demand and guiding decisions in retail, from pricing to assortment design. Classical models such as the multinomial logit have been widely used because of their tractability, but they impose restrictive assumptions on substitution patterns. In this talk, I will present insights from my research on decision making under consumer choice models, including extensions that capture sequential search and product framing effects. I will also share findings on the marginal distribution model, which relaxes the need to know the full joint distribution of choices. By focusing only on marginal distributions, this framework allows for more flexible and data-compatible representations of substitution behavior. Together, these examples illustrate how modeling consumer choice leads to sharper demand predictions and more effective operational decisions.

**Ye Liu**

**Title:** Make Your Chatbot Price Like You Do: Identify and Customize Distributional Preferences of Large Language Models in Supply Chains

**Abstract:**

Companies increasingly use large language model (LLM) agents for decision-making in various business processes, including supply chain procurement. How an LLM agent sets the price in a supply chain transaction, however, depends on its hidden traits, such as its self-interest level and preferences between achieving high efficiency or high fairness. Our paper examines these embedded distributional preferences of GPT-4o, Llama-4 Maverick, and the reasoning model GPT-o3 in a supply chain. We show that the type of decisions an LLM makes plays a significant role in shaping its distributional preferences. While a simple persona change from human to supplier or buyer reduces all LLMs' self-interest level, making pricing decisions as a firm increases it. Descriptions of supply chain relationships can customize all LLMs' distributional preferences, but to different extents. GPT-4o and Llama-4 have a strong "personality", with GPT-4o making the most balanced decisions and Llama-4 the most selfish GPT-o3, on the other hand, is the most customizable and may resemble GPT-4o or Llama-4. We find that while concrete descriptions lead to similar and intuitive changes in all LLMs, generic descriptions such as "good relationship" lead to diverging results for GPT and Llama LLMs, making the former more fairness-minded and the latter more efficiency-driven. For a firm, the same relationship description applied to the same LLM may lead to opposite profit share outcomes depending on whether it is more or less financially constrained than its counterparty. Our research provides guidance on the characterization and design of LLM tools in supply chains.

**Leann Thayaparan**

**Title:** The Role of Driver Behavior and Interpretability in Distributed Renewable Energy Storage

**Abstract:**

As we move towards more renewable resources, the need for energy storage increases. Electric Vehicles (EVs) are a way of providing a distributed energy storage resource to the electric grid. However, before EVs can be used to support energy inventory, through optimal EV charging and discharging, highly complex, non-linear driver behavior must be accounted for. In this work, we propose a driver-centric dynamic programming (DP) model for the charging and discharging of EVs. We show that the policies this model gives rise to are easy to communicate, allowing for easier adoption. We determine regimes under which optimal polices have the required structural properties and can be solved in closed form. Finally, through a close collaboration with a large American EV manufacturer, we apply the proposed DP formulation and policies to real de-identified driver data. Using this DP model our industry collaborator is assessing how much capacity EVs have to potentially power a home or give back to the grid. Through this collaboration, we quantify not just the savings that individual drivers can earn from participating but also show the capacity EV’s have to act as inventory for the grid.

**Pavithra Harsha**

**Title:** Decision-Focused AI for Real-World Operations

**Abstract:**

Machine learning is increasingly used to guide operational decisions in supply chains, pricing, and recommendations. Yet improvements in predictive accuracy do not always yield proportional gains in decision outcomes. Because optimization depends on forecasts in complex, nonlinear, and often discrete ways, even small prediction errors can trigger costly shifts in the final decision space.

This talk introduces decision-focused AI, an approach that trains models directly on decision objectives rather than prediction accuracy. Using examples from pricing, electricity generation, and inventory management and fulfillment, I will highlight advances across single- and two-stage stochastic optimization problems, under both exogenous uncertainty andendogenous uncertainty where decisions shape the data we observe. While these methods reduce costs and often require only point forecasts—an important practical advantage—challenges remain in computation, scalability, and interpretability. Addressing these gaps is key to building robust, decision-aware ML systems. This is joint work: Rares Cristian, Georgia Perakis and Brian Quanz.

**Gonzalo Romero**

**Title:**

Managing Payment Flexibility in Rent-to-Own Contracts for Off-Grid Energy Products

**Abstract:**

The diffusion of technological innovations in developing economies has been facilitated by the use of Rent-To-Own (RTO) business models, which give flexibility to consumers by allowing them to make incremental payments over time. Understanding how to best manage this flexibility is a fundamental problem for firms operating in these markets. Motivated by an application of RTO to the distribution of solar lamps in developing countries, we examine the drivers and impact of payment flexibility on repayment performance and consumer behavior in RTO contracts. We formulate a stochastic dynamic programming model that characterizes an important dimension of payment flexibility, i.e., the ability of consumers to make bundled payments (multiple installments paid at once, in advance). We show that consumers may bundle payments because of uncertainty about budget in future periods, which leads to a non-monotonic impact of income uncertainty on repayment performance. We further show that bundled payments are more likely to occur closer to the end of the ownership cycle. We examine different flexibility levers that the firm can adjust as part of its contract design (repayment frequency/installment amount, grace period), accounting for the impact of bundled payments. Our results show that an intermediate level of flexibility could benefit both the firm and consumers under some conditions. Whereas payment flexibility is a fundamental component of RTO contracts in the developing world, our findings indicate that a moderate level of flexibility can go a long way in helping firms and consumers in these environments. Hence, RTO firms may not need to offer extreme degrees of flexibility in order to achieve desirable outcomes.

**Wei Sun**

**Title**  
Compute-Aware Routing and Test-Time Scaling: Optimizing Cost-Accuracy Trade-offs

**Abstract**  
In this talk, I will describe a series of projects at the intersection of decision-making and efficient LLM inference. The first focuses on causal model routing, where we leverage observational data to minimize regret in selecting the right model or inference strategy for each query. The second centers on latency- and token-aware test-time scaling, where we dynamically allocate compute to balance accuracy with user experience and cost. Together, these projects illustrate a unified perspective on compute-aware decision-making, showing how principled optimization can guide routing and scaling policies that deliver superior accuracy-latency-cost trade-offs under real-world deployment constraints. Beyond efficiency, this line of work also highlights a path toward greater AI sustainability, ensuring that advanced inference strategies not only improve performance but also reduce computational waste and environmental footprint.