High-stakes decisions from low-quality data: AI decision-making for planetary health

Lily Xu (Harvard University)

I develop and deploy machine learning methods for making reliable decisions in high-stakes settings when data are incomplete. Guided by research questions that emerge from my close collaboration with the public sector, my work enables practitioners to take efficient and robust actions necessary for planetary health. Particularly for biodiversity conservation, I work closely with NGOs and government stakeholders to deploy AI for on-the-ground conservation.

Planetary health is an emerging field which recognizes that *human health and the health of our planet are inextricably linked* [20]. Core challenges include **biodiversity conservation**, for which I have developed algorithms to protect wildlife from poaching in protected areas around the world, and **public health**, for which I have partnered with an NGO in India to connect expecting mothers with lifesaving preventative care information. Addressing these planetary threats effectively requires making time-sensitive decisions. Unfortunately, the set of possible actions is often enormous and the available data is incomplete, making it difficult to enumerate or anticipate the impact of our actions.

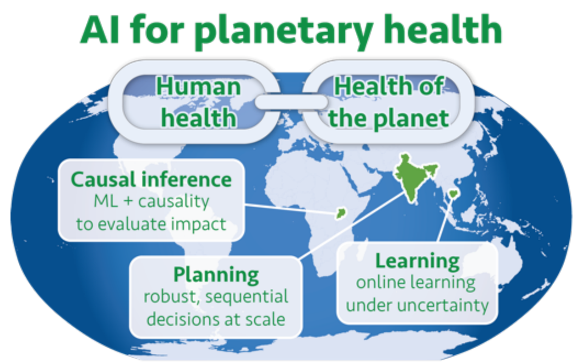

Figure 1: My research develops AI decision-making techniques for planetary health, addressing the intertwined challenges threatening human health and the health of our planet. My work is grounded in concrete applications in Uganda, India, and Cambodia.

My research agenda aims to resolve the large scale and uncertainty underlying these decisions. I develop AI methods to make data-driven decisions that are efficient at the scale necessary to act on our planet's most urgent challenges, while also being robust to uncertainty as these high stakes demand (fig. 1). My research agenda interweaves three main thrusts:

1. **Learning under uncertainty.** Online learning enables us to proactively collect more data to improve our models, trading off *exploring* actions with large uncertainty and *exploiting* actions with high observed reward. To avoid unnecessarily exploring low-reward actions which harms our performance, I draw inspiration from immersion in the domains with which I work to integrate problem structure into algorithm design. The methods I've developed reduce exploration to achieve higher reward more quickly for multi-armed bandits [3, 7] and reinforcement learning [15, 12].

2. **Sequential planning that is robust and efficient at scale.** Resource allocation problems such as maternal health interventions involve multi-step sequential decisions, combinatorial actions, and discrete optimization — introducing exponentially large action spaces and NP-hard optimization problems. I develop algorithms with provably strong guarantees to make these challenging problems more tractable, integrating advances from game theory [4, 5, 11, 6] and mixed-integer programming [16].

3. **Causal inference for impact evaluation.** Establishing causation is difficult when we cannot conduct randomized controlled trials and the available data is messy and full of confounders. I show that machine learning can help overcome these missing data challenges, using this approach to study ranger patrol data from a national park in Uganda with one of the highest levels of poaching in the world. Our results provide the *first causal evidence* for poaching deterrence, showing that ranger patrols reduce poaching by 46% [10].

Key to my work is my commitment to **deployment**. I work in close partnership with stakeholders to identify bottlenecks in existing algorithmic solutions and deploy AI to achieve measurable impact on the ground (fig. 2). My machine learning algorithms are being *scaled to 1,200 protected areas around the world* to predict poaching hotspots, in partnership with conservation NGOs including WWF [1], and integrated with the world's largest mobile health program to reduce maternal mortality in India [11].

My research has been recognized with the *INFORMS Doing Good with Good OR award* and *AAAI best paper runner-up*, and I have received a *Google PhD Fellowship* and been named a *Siebel Scholar*.

Figure 2: For my research, I have conducted field visits to protected areas in three continents to build domain knowledge and strengthen partnerships. I've (*a*) gone on patrol with frontline rangers in Srepok Wildlife Sanctuary, Cambodia; (*b*) organized a workshop in Rio Bravo, Belize for 13 rangers to discuss AI predictions of poaching threats that I developed; and (*c*) presented my causal inference work to government officials in Murchison Falls National Park, Uganda.

1 Learning under uncertainty

Traditional online learning algorithms may spend excessive time exploring suboptimal actions, compromising short-term performance for long-term optimality. Applied to conservation, for example, these short-term losses would result in animals poached and forests destroyed unnecessarily. We therefore aim to limit unnecessary exploration to converge faster. Towards this goal, I've developed provably faster techniques in online learning for both stochastic (multi-armed bandits) and sequential (restless bandits) settings.

Efficient exploration in online learning (**AAAI best paper runner-up, IJCAI**) Planetary health problems are often spatial planning problems, where each action is context-specific, corresponding to a physical region on Earth. The geographic features associated with each action enables us to use *smoothness* assumptions to relate actions. For example in wildlife conservation, we expect that regions with similar geographic features will also have similar poaching patterns. Mathematically, we can encode this similarity in the reward of two actions as Lipschitz-continuity. Additionally, each action — patrolling a set of regions — is combinatorial in the number of regions, but can be separated into constitute regions with *additive, decomposable reward*. We therefore introduce a new multi-armed bandit variant that incorporates decomposability and smoothness, enforcing Lipschitz-continuity assumptions on those reward estimates for demonstrably faster convergence [3]. We show that this approach leads to *improved regret bounds that have no dependence on the number of regions N*, achieving $\tilde{O}(T^{2/3})$ regret with time horizon T compared to the state-of-the-art $\tilde{O}(T^{\frac{N+1}{N+2}})$ — a proven lower regret bound for bandits that only incorporate Lipschitz-continuity [24].

Motivated by real-world needs, I've extended this work to a multi-objective setting to account for the disparate impacts that an action has on different groups (such as different wildlife species) [7] and to continuous action spaces with adaptive discretization [8], where we introduce a meta-algorithm with *provably better sample, storage, and computational complexity* than uniform discretization or kernel regression methods.

Online learning for sequential decisions (**AAAI**) Sequential planning methods often assume that environmental dynamics are perfectly known, which is unrealistic in many practical applications. In public health, for example, we may not know in advance how individual patients will respond to an intervention. I have thus worked on integrating online learning into *restless bandits*, a generalization of multi-armed bandits that introduces a "state" for each arm, where each arm models one patient as an independent Markov decision process. Our approach maintains confidence bounds for the transition probabilities of each arm based on observed transitions, then solves a quadratic program to solve for what we call *optimistic transition probabilities* — those that yield the highest future reward rather than just the highest transition probability, as would be used in a naive application of standard upper confidence bounds [15]. Our work provides the first sublinear frequentist regret guarantee for restless bandits, bounding the worst-case regret from *unknown* transition dynamics. In contrast, prior work requires that the distribution over all possible transitions is known a priori. Specifically, we achieve a sublinear regret of $O(H\sqrt{T \log T})$ with T episodes and episode length H. I've also extended this work to consider features associated with each arm, such as demographic information about individual mothers that inform their behavior, integrating this *context* in a Bayesian hierarchical model to share information across arms and learn faster [12].

2 Sequential planning that is robust and efficient at scale

Addressing planetary challenges requires repeatedly acting in dynamic settings to consider long-term effects. Sequential planning is already hard as policies are exponential in size, and harder still when transition dynamics are unknown and the action space is combinatorial. To overcome these challenges, I've developed methods for robust planning for reinforcement learning (RL) [4], sequential planning with NP-hard combinatorial optimization [16, 13], and strategic game theory settings with multiple agents [6].

Robust planning under uncertainty (UAI, AAAI) RL typically requires that we have access to a simulator with perfect specification of the underlying Markov decision process (e.g., representing mothers' behavioral patterns), which is not realistic in many practical settings. More realistically, we may have an uncertainty *interval* over the true transition probabilities. Standard robust RL addresses uncertainty by aiming to maximize the minimum possible reward (minimax reward), which often leads to overly conservative policies that narrowly target worst-case scenarios. I instead consider a *minimax regret* objective, which considers the counterfactual reward of other policies *across the entire uncertainty space*, not just the worst-case scenario. Minimax regret is intractable to solve for directly, so we take an iterative approach by posing the problem as a zero-sum game between an agent and nature. I prove this method converges to an ϵ-optimal policy in a finite number of steps, despite the exponentially large and continuous action space. This work is the *first framework for calculating minimax regret–optimal policies using RL*, broadening RL-based planning for settings with uncertainty [4]. We have extended this work on robust planning to restless bandits [5, 11], where the large number of arms is a challenge, developing practical techniques through clustering to improve tractability when the number of arms is large, necessary as ARMMAN has thousands of beneficiaries enrolled simultaneously.

RL with combinatorial actions (under review) The existing literature at the intersection of RL and combinatorial optimization has been limited to solving *single-shot* combinatorial optimization problems with RL, by decomposing the decision into iterative choices from a small action set (e.g., iteratively choosing the next node in a traveling salesperson problem). I provide the *first approach to solve for per-timestep combinatorial actions in **sequential** settings* where the reward comes not from a single-shot action but is incurred after enacting a policy over many timesteps [16]. We leverage recent advances in integrating deep learning with mathematical programming to directly embed a trained neural network into a mixed-integer program [23]. Incorporated into an RL training procedure, this technique enables our approach to estimate the long-term reward using a deep learning approximation of the Q-function, while tractably optimizing this function over the combinatorial action space at each step with a mixed-integer program.

3 Causal inference for impact evaluation

Studying intervention impact is necessary for building cause-and-effect models for sequential planning and supporting policy decisions. Unfortunately, messy data often impedes direct observations of treatment effects or outcomes and randomized controlled trials are often infeasible in high-stakes, complex settings. I've developed techniques to integrate machine learning into causal inference to help overcome incomplete data, enabling us to provide the first causal evidence that ranger patrols deter poachers [10].

Partnering with the Uganda Wildlife Authority and WCS Uganda, I studied the impact of ranger patrols on poaching in Murchison Falls National Park in Uganda, a 5,000 km^2 park with among the world's richest biodiversity but the greatest recorded density of poaching. There are over 300,000 rangers worldwide patrolling to protect wildlife [25], but the impact of patrols on deterring poaching has never been conclusively demonstrated. Patrol efficacy is difficult to study due to the complexity of ecosystem and human dynamics. To overcome noisy data and imperfect detection, we use machine learning to learn smooth patterns for baseline poaching and patrol predictions for matching, and we embed domain intuition of poaching detection by enforcing monotonicity in the machine learning predictions and using domain-specific data augmentation. Our work provides quantitative evidence for policymakers to advocate for increased funding for ranger patrols and guidance for conservation managers to optimize planning within a protected area.

4 Deploying AI for planetary health

Alongside technical advances, my research also contributes frameworks and open questions to enable others to design and deploy algorithmic tools for real-world impact. To encourage the computational community to think more critically about the ethical implications of impact-oriented work, I co-led a paper to offer concrete recommendations to address criticism of the field of "AI for social good", where we recommend that researchers adopt Amartya Sen's *capabilities approach* to assess the impact of their research and to use participatory design to engage stakeholders [2]. I aim to model these practices in my own work: I iterate with conservation managers to formulate research questions; I traveled to Cambodia, Belize, and Uganda to join rangers on patrol (fig. 2); and **I conducted months-long field tests of my AI algorithms** for poaching prediction in Cambodia and Uganda [1]. My partnership with conservation NGOs is the subject of a Harvard Business School case [26].

I engage conservationists in my scholarship as well as my applied work, where I am the sole computer scientist among coauthors from ecology and sociology. In one article, we discuss analogues between RL and "adaptive management" in conservation and encourage conservation managers to incorporate concepts such as model-free planning, off-policy evaluation, and curriculum learning [9]. In a book chapter, we explore how the growing use of machine learning for environmental management exacerbates political inequality and widens the global division of labor across data collection, infrastructure, and analysis [14].

My academic service complements this mission to bridge research and practice: since 2021, I have co-organized the Mechanism Design for Social Good (MD4SG) research initiative to enable others to work on interdisciplinary research that improves equity and justice. In 2022, I co-organized an AI for conservation workshop with 25 participants across AI, operations research, ecology, industry, and NGOs. Our resulting white paper identifies open research questions in AI for conservation, to help orient new algorithmic decision-making approaches towards addressing biodiversity challenges [17].

5 Future vision

The United Nations Development Programme calls for diverse sectors to come together to achieve planetary health [21]. With my future lab, I intend to respond to that call, continuing to develop algorithmic techniques with cross-disciplinary partners to better translate AI methods to high-stakes problems, building on the partnerships and domain knowledge I already have across conservation and public health.

I aim to design techniques that are data-driven while also integrating knowledge from human experts. One major shortfall of algorithmic decision-making systems is that we require an objective function to be specified in advance, but we may not be able to concretely specify what we aim to maximize. In biodiversity conservation, for example, ecologists often use population sizes of keystone species such as tigers and elephants to measure success, but these populations are hard to measure and do not offer a complete picture of ecosystem health. My work on robust planning brings us closer to effective planning; moving forward, I plan to investigate reward shaping and inverse reinforcement learning to design better policies for these hard-to-quantify decision-making settings, using human feedback to overcome data deficiencies.

I also aim to make algorithmic decision-making methods more practical for sequential settings, where missing counterfactual information is a key obstacle. Off-policy RL with trajectory stitching can serve as counterfactual evidence for evaluating policies we have not yet implemented. However, our ability to reason about hypothetical policies is constrained by their similarity to policies that have been acted upon. I would like to incorporate domain-specific structural causal models [18, 19, 22] to improve counterfactual reasoning and reduce sample complexity in offline, off-policy RL — necessary for adaptively managing complex ecosystems and maintaining healthy populations. Overall, my research will continue to help decision-makers and policymakers make more informed decisions on the best strategic interventions across critical, low-resource settings, guided by immersion in the most urgent challenges to biodiverse ecosystems and healthy planet.

Asterisk (*) indicates joint first co-authorship.

References to Lily's papers

[1] **Lily Xu**, Shahrzad Gholami, Sara Mc Carthy, Bistra Dilkina, Andrew Plumptre, Milind Tambe, Rohit Singh, Mustapha Nsubuga, Joshua Mabonga, Margaret Driciru, Fred Wanyama, Aggrey Rwetsiba, Tom Okello, and Eric Enyel. "Stay Ahead of Poachers: Illegal Wildlife Poaching Prediction and Patrol Planning Under Uncertainty with Field Test Evaluations". *International Conference on Data Engineering* (*ICDE*). 2020.

[2] Elizabeth Bondi*, **Lily Xu***, Diana Acosta-Navas, and Jackson A. Killian. "Envisioning Communities: A Participatory Approach Towards AI for Social Good". *Artificial Intelligence, Ethics, and Society* (*AIES*). 2021.

[3] **Lily Xu**, Elizabeth Bondi, Fei Fang, Andrew Perrault, Kai Wang, and Milind Tambe. "Dual-Mandate Patrols: Multi-Armed Bandits for Green Security". *AAAI Conference on Artificial Intelligence* (*AAAI*). 2021.

[4] **Lily Xu**, Andrew Perrault, Fei Fang, Haipeng Chen, and Milind Tambe. "Robust Reinforcement Learning Under Minimax Regret for Green Security". *Uncertainty in Artificial Intelligence* (*UAI*). 2021.

[5] Jackson A. Killian, **Lily Xu**, Arpita Biswas, and Milind Tambe. "Restless and Uncertain: Robust Policies for Restless Bandits via Deep Multi-Agent Reinforcement Learning". *Uncertainty in Artificial Intelligence* (*UAI*). 2022.

[6] Kai Wang, **Lily Xu**, Andrew Perrault, Michael K. Reiter, and Milind Tambe. "Coordinating Followers to Reach Better Equilibria: End-to-End Gradient Descent for Stackelberg Games". *AAAI Conference on Artificial Intelligence* (*AAAI*). 2022.

[7] **Lily Xu**, Arpita Biswas, Fei Fang, and Milind Tambe. "Ranked Prioritization of Groups in Combinatorial Bandit Allocation". *International Joint Conference on Artificial Intelligence* (*IJCAI*). 2022.

[8] (alphabetical) Siddhartha Banerjee, Sean R. Sinclair, Milind Tambe, **Lily Xu**, and Christina Lee Yu. "Artificial Replay: A Meta-Algorithm for Harnessing Historical Data in Bandits". *arXiv*. 2023.

[9] Melissa Chapman, **Lily Xu**, Marcus Lapeyrolerie, and Carl Boettiger. "Bridging adaptive management and reinforcement learning for more robust decisions". *Philosophical Transactions of the Royal Society B* 378.1881 (2023), p. 20220195.

[10] Rachel Guo*, **Lily Xu***, Andrew Perrault, Luke W. Miratrix, Andrew J. Plumptre, Joshua Mabonga, Herbert Kitimbo, Fredrick Wanyama, Simon Nampindo, Milind Tambe, and Andrew B. Davies. "Ranger patrols deter poaching: first causal insights for improving protected area management". *under review* (2023).

[11] Jackson A. Killian*, Arpita Biswas*, **Lily Xu***, Shresth Verma*, Vineet Nair, Aparna Taneja, Aparna Hegde, Neha Madhiwalla, Paula Rodriguez Diaz, Sonja Johnson-Yu, and Milind Tambe. "Robust Planning over Restless Groups: Engagement Interventions for a Large-Scale Maternal Telehealth Program". *AAAI Conference on Artificial Intelligence* (*AAAI*). 2023.

[12] Biyonka Liang, **Lily Xu**, Milind Tambe, and Lucas B. Janson. "Online Learning for Contextual and Non-Stationary Restless Multi-Armed Bandits". *in preparation*. 2023.

[13] Paula Rodriguez Diaz, Jackson A. Killian, **Lily Xu**, Aparna Taneja, Arun Sai Suggala, and Milind Tambe. "Flexible Budgets in Restless Bandits: A Proximal Primal-Dual Algorithm for Efficient Budget Allocation". *AAAI Conference on Artificial Intelligence* (*AAAI*). 2023.

[14] Caleb Scoville, Hilary Faxon, Melissa Chapman, Samantha Jo Fried, **Lily Xu**, Carl Boettiger, J. Michael Reed, Marcus Lapeyrolerie, Amy Van Scoyoc, and Razvan Amironesei. "Environment, Society, and Machine Learning". *Oxford Handbook of the Sociology of Machine Learning*. Ed. by Christian Borch and Juan Pablo Pardo-Guerra. Oxford University Press, 2023.

[15] Kai Wang*, **Lily Xu***, Aparna Taneja, and Milind Tambe. "Optimistic Whittle Index Policy: Online Learning for Restless Bandits". *AAAI Conference on Artificial Intelligence* (*AAAI*). 2023.

[16] **Lily Xu**, Bryan Wilder, Elias B. Khalil, and Milind Tambe. "Reinforcement learning with combinatorial actions for coupled restless bandits". *arXiv; under review*. 2023.

[17] **Lily Xu***, Esther Rolf*, Sara Beery, Joseph R. Bennett, Tanya Berger-Wolf, Tanya Birch, Elizabeth Bondi-Kelly, Justin Brashares, Melissa Chapman, Anthony Corso, Andrew Davies, Nikhil Garg, Angela Gaylard, Robert Heilmayr, Hannah Kerner, Konstantin Klemmer, Vipin Kumar, Lester Mackey, Claire Monteleoni, Paul Moorcroft, Jonathan Palmer, Andrew Perrault, David Thau, and Milind Tambe. "Reflections from the Workshop on AI-Assisted Decision Making for Conservation". *arXiv*. 2023.

Other references

[18] Ryan Plummer. "The adaptive co-management process: an initial synthesis of representative models and influential variables". *Ecology and Society* 14.2 (2009).

[19] Paul J Ferraro and Merlin M Hanauer. "Advances in measuring the environmental and social impacts of environmental programs". *Annual Review of Environment and Resources* 39 (2014), pp. 495–517.

[20] Richard Horton, Robert Beaglehole, Ruth Bonita, John Raeburn, Martin McKee, and Stig Wall. "From public to planetary health: a manifesto". *The Lancet* 383.9920 (2014), p. 847.

[21] Sarah Whitmee, Andy Haines, Chris Beyrer, Frederick Boltz, Anthony G Capon, Braulio Ferreira de Souza Dias, Alex Ezeh, Howard Frumkin, Peng Gong, Peter Head, et al. "Safeguarding human health in the Anthropocene epoch: report of The Rockefeller Foundation–Lancet Commission on planetary health". *The Lancet* 386.10007 (2015), pp. 1973–2028.

[22] Yi Fan, Jiquan Chen, Gabriela Shirkey, Ranjeet John, Susie R Wu, Hogeun Park, and Changliang Shao. "Applications of structural equation modeling (SEM) in ecological studies: an updated review". *Ecological Processes* 5 (2016), pp. 1–12.

[23] Matteo Fischetti and Jason Jo. "Deep neural networks and mixed integer linear optimization". *Constraints* 23.3 (2018), pp. 296–309.

[24] Robert Kleinberg, Aleksandrs Slivkins, and Eli Upfal. "Bandits and experts in metric spaces". *Journal of the ACM (JACM)* 66.4 (2019), pp. 1–77.

[25] Michael R Appleton, Alexandre Courtiol, Lucy Emerton, James L Slade, Andrew Tilker, Lauren C Warr, Mónica Álvarez Malvido, James R Barborak, Louise de Bruin, Rosalie Chapple, et al. "Protected area personnel and ranger numbers are insufficient to deliver global expectations". *Nature Sustainability* 5.12 (2022), pp. 1100–1110.

[26] Brian Trelstad and Bonnie Yining Cao. "SMART: AI and Machine Learning for Wildlife Conservation". *Harvard Business School Case 323-036* (2022).