

Accelerating Model Driven Discovery with ML

—
Presented by Sekou L. Remy

Work done in solidarity with:

Irene Wanyana, Makerere University

Joyce Nakatumba-Nabende, Makerere University

Susan Kizito, Makerere University

Balla Diop Ngom, Université Cheikh Anta Diop

Ernest Mwebaze, Sunbird AI

Barbara Han, Cary Institute of Ecosystem Studies

Fadel Thior, IBM

Rose Yvette Essomba, IBM

Stephanie Muller, IBM

Cliff Kerr, Institute for Disease Modeling

...

and many, many friends!

Developed as a workshop proposal for ICLR 2023

3000

Researchers

14

Locations

6

Continents

6 Nobel Laureates
10 Medals of Technology
5 National Medals of Science
6 Turing Awards



Our Lab's starting point in Global Health



In 2020, there were **~247 million cases** of malaria, 95% of these cases were in Africa, **~625,000 malaria deaths**, and malaria also increases the risk of death from other diseases.

Approximately **\$3.5B USD** was estimated for elimination and control efforts.

World malaria report 2022. Geneva: World Health Organization

Malaria **can be treated and cured** with anti-malarial drugs (i.e. interventions exist)
Malaria **can be prevented** with effective malaria control.

Our starting point in Global Health



In 2018, there were **~219 million cases** of malaria, 95% of these cases were in Africa, **~450,000 malaria deaths**, and malaria also increases the risk of death from other diseases.

Approximately **\$4.5B USD** was estimated for elimination and control efforts.

World malaria report 2020. Geneva: World Health Organization

Malaria **can be treated and cured** with anti-malarial drugs (i.e. interventions exist)
Malaria **can be prevented** with effective malaria control.

Starting with a Significant Global Health Challenge



Pre-September Questions 2018



Philip Welkhoff Jennifer Gardy David Wilson Damian Walker

Q1 Are there intervention packages that are not being explored today because of lack of tooling?

Q2 Are there overlooked approaches to generate insight from multiple models, and accommodate the inherent uncertainty therein in a principled manner?

Q3 How should model output from multiple models be presented to a policy influencer to enable the best policies to be selected?



David L. Smith
Professor of Health Metrics Sciences



Post September Targets



Jimmy Opigo Agaba Bosco Rutazaana Damian Philip Welkhoff Jennifer Gardy David Wilson Damian Walker

T1 Develop and deploy a dashboard which can harness existing resources (malaria models and data) in a manner to support the selection of intervention policies for Uganda.

T2 Recommend useful intervention packages which would not have been considered by informed policy makers today.

T3 Identify the context under which additional stratification would be useful for policy decisions in Uganda. Focus is the best overall outcomes from a prevalence, cost, and execution complexity perspective.

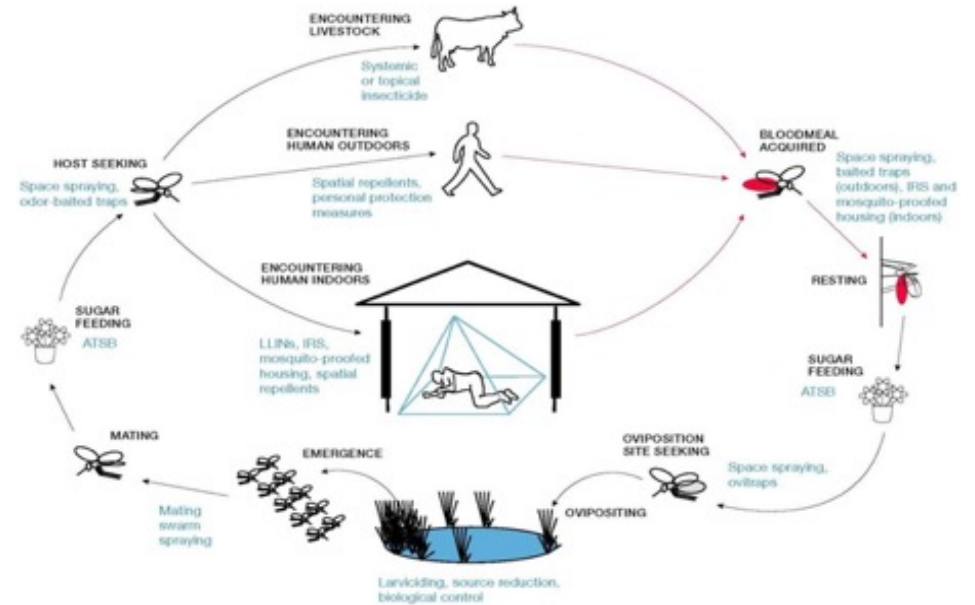
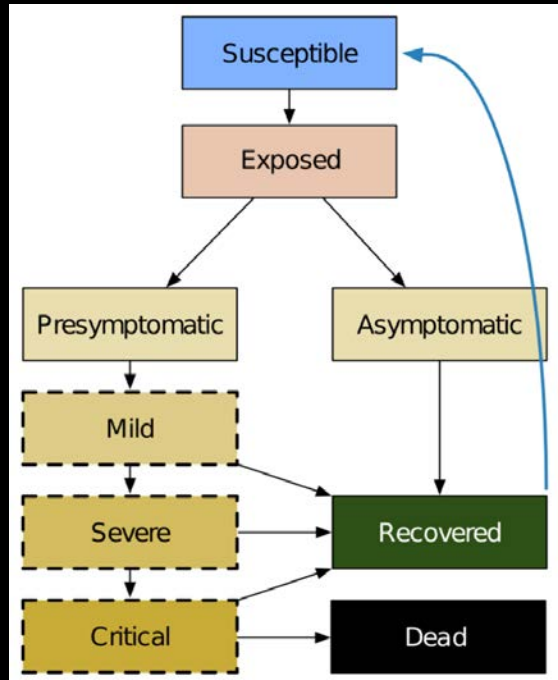


THE UGANDA MALARIA REDUCTION STRATEGIC PLAN 2014-2020

New Goal

By 2025, drive malaria parasite prevalence in all 134 districts of Uganda to 5% or less (from 9%).

(Parameterised) Disease Models

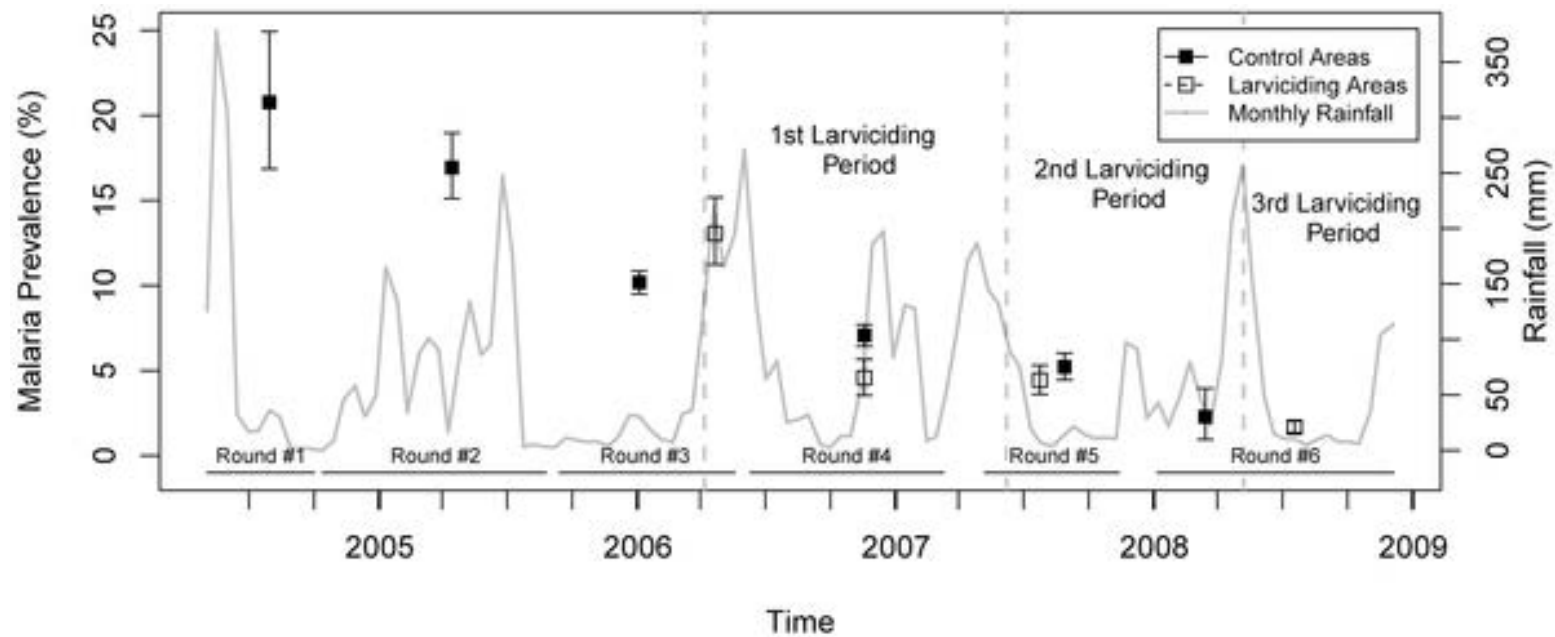


Kerr, C.C., Stuart, R.M., Mistry, D., Abeysuriya, R.G., Cohen, J.A., George, L., Jastrzebski, M., Famulare, M., Wenger, E. and Klein, D.J., 2022. Python vs. the pandemic: a case study in high-stakes software development. In *Proc. of the 21st Python in Science Conf.(SCIPY)*.

Kiware SS, Chitnis N, Tatarsky A, Wu S, Castellanos HMS, et al. (2017) Attacking the mosquito on multiple fronts: Insights from the Vector Control Optimization Model (VCOM) for malaria elimination. *PLOS ONE* 12(12): e0187680. <https://doi.org/10.1371/journal.pone.0187680> <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0187680>

What you do, and when you do it matters

Crude prevalence of malaria infection stratified by survey round and larviciding status.

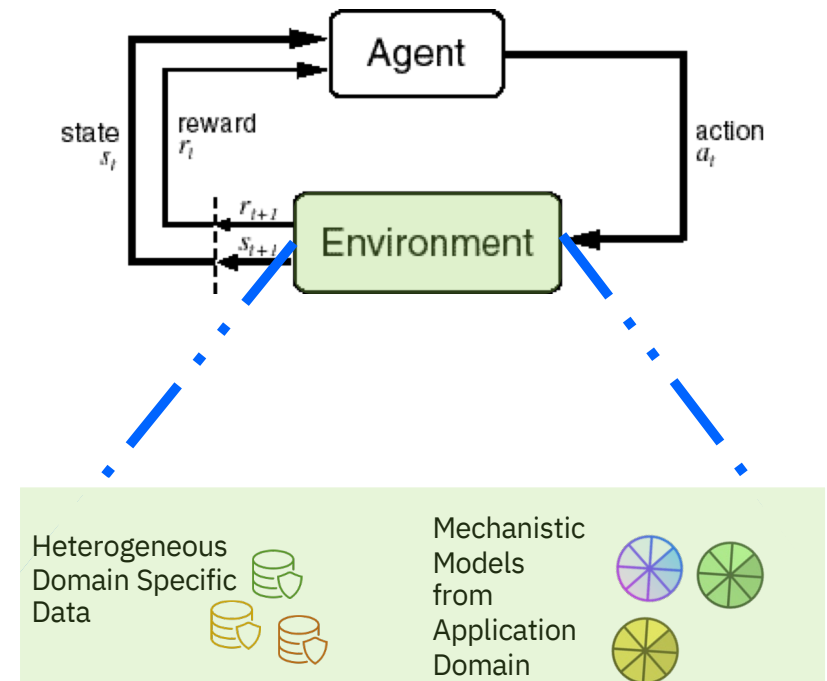


Supporting Decisions with ML

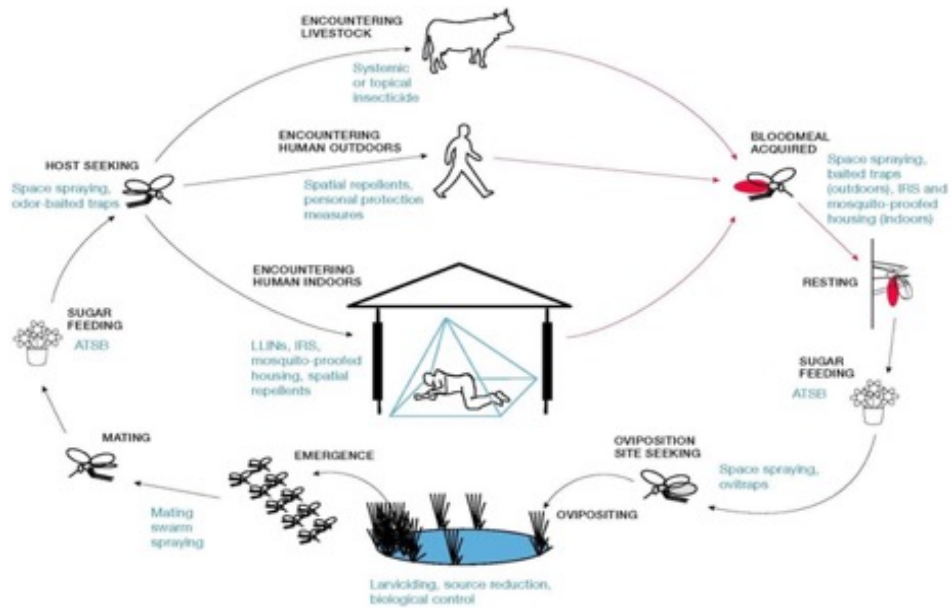
O. Bent, et. al. *Novel exploration techniques (NETs) for malaria policy interventions*. In Thirty-Second AAAI Conference on Artificial Intelligence (IAAI 2018)

S.L. Remy, et. al. *Reshaping the use of digital tools to fight malaria*. Multilateral Initiative on Malaria - Workshop on Using Digital Tools to Strengthen the Malaria Supply Chain (MIM 2018)

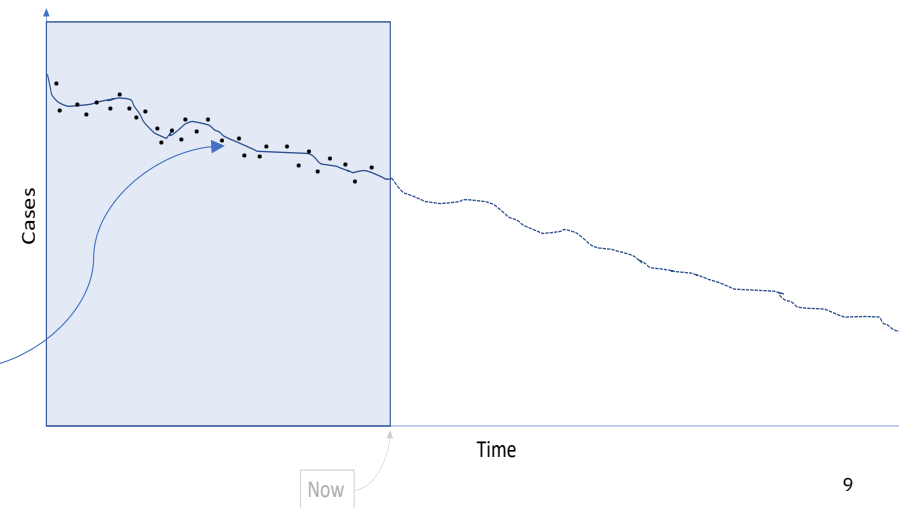
S.L. Remy, *Innovations for Global Health and Development*. ITU/WHO Workshop on Artificial Intelligence for Health (AI4H 2019)



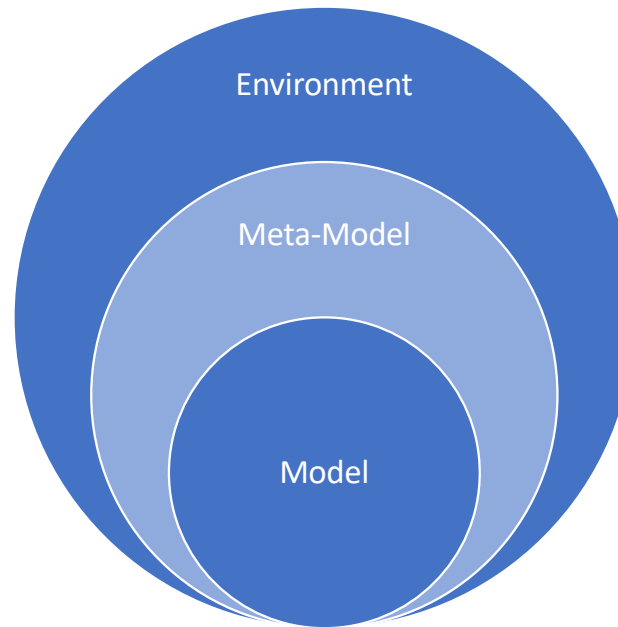
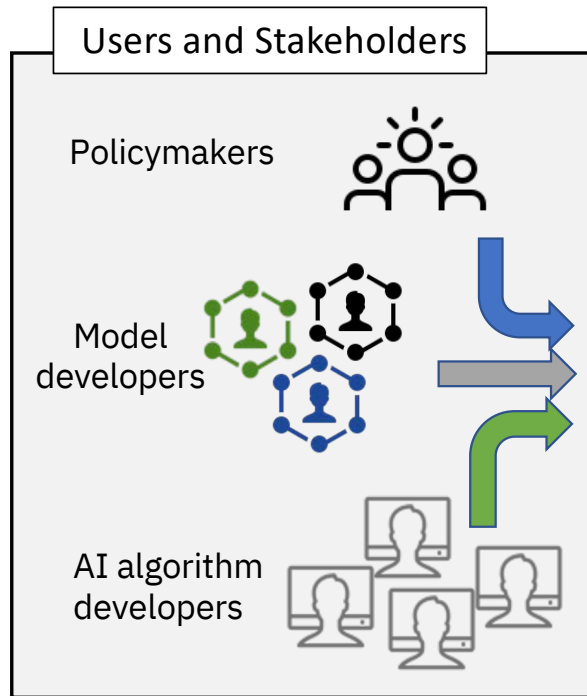
Example Model



Kiware SS, Chitnis N, Tatarsky A, Wu S, Castellanos HMS, et al. (2017) Attacking the mosquito on multiple fronts: Insights from the Vector Control Optimization Model (VCOM) for malaria elimination. PLOS ONE 12(12): e0187680.
<https://doi.org/10.1371/journal.pone.0187680>
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0187680>



Key Design Change



Environments define:

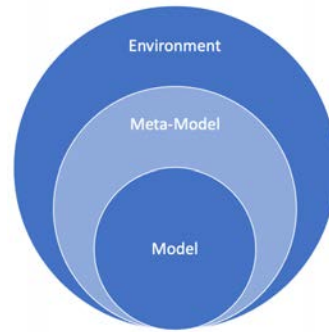
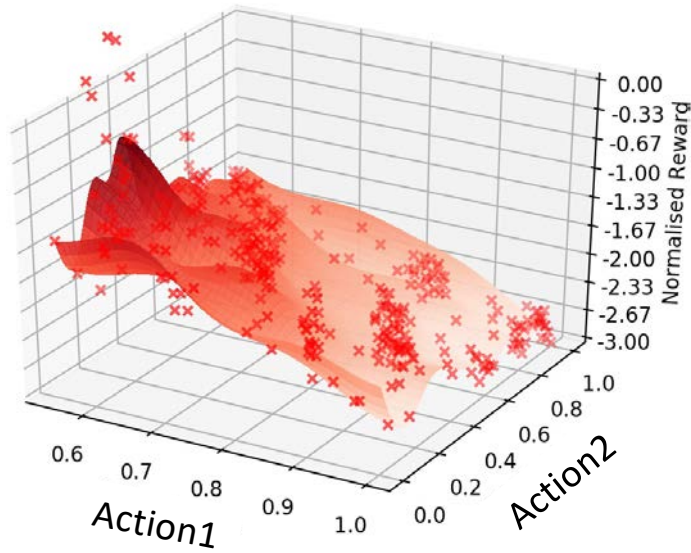
- Available actions
- Available observations
- Action goodness

Meta-Models define:

- Which parameters can be changed
- What data can be observed

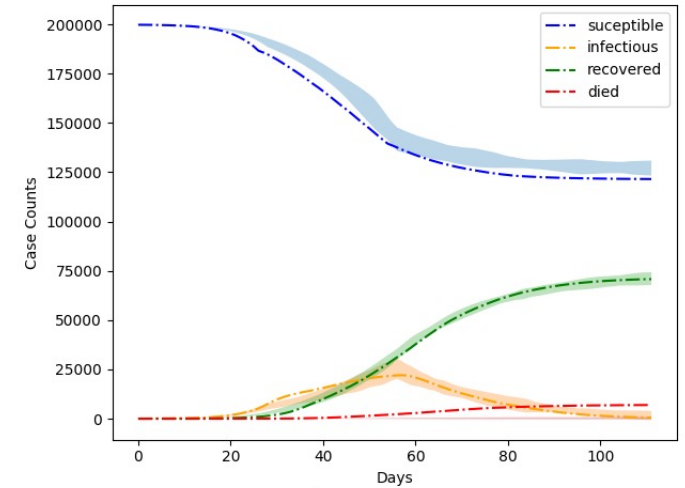
Remy, S.L. and Bent, O., 2020, August. A Global Health Gym Environment for RL Applications. In NeurIPS 2019 Competition and Demonstration Track (pp. 253-261). PMLR

Finding Intervention Programs Which Work Well

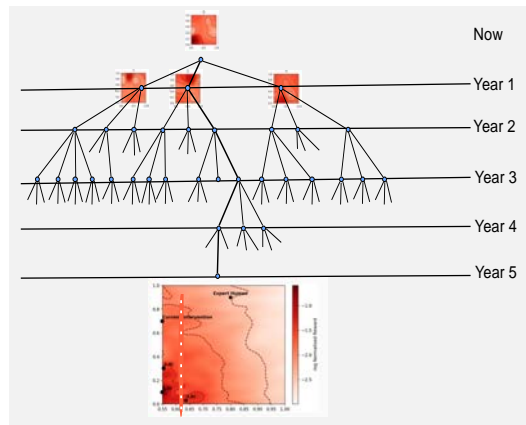
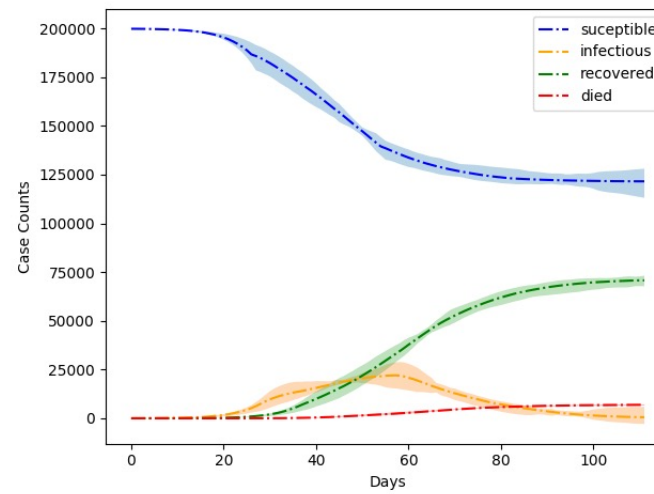


Tuning/Calibrating Models for Specific Use

Model 1



Model 2



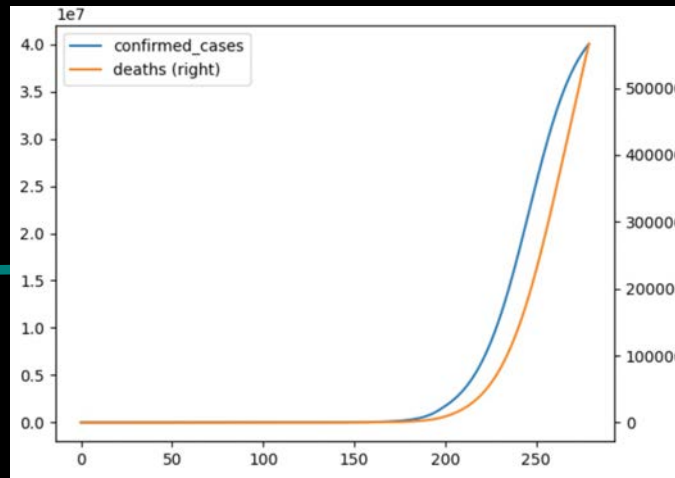
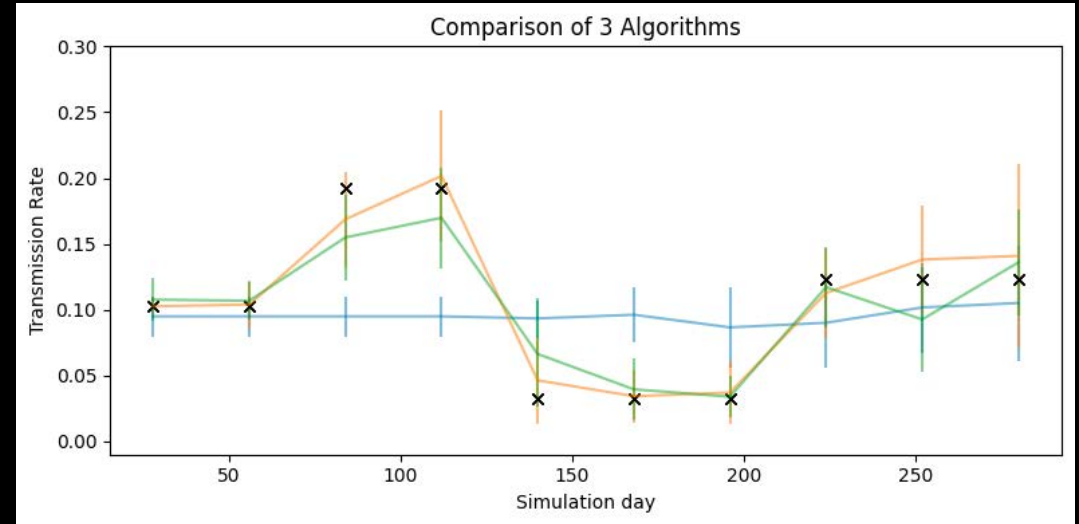
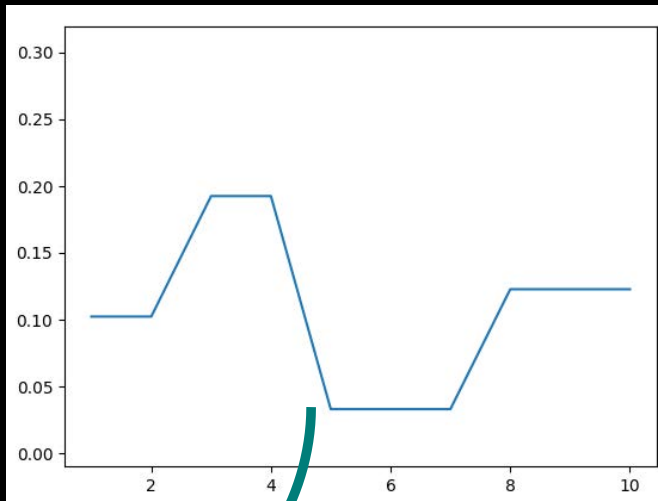
Reward:

$$C_{DA} = \frac{C_{int} + HSC_{int} - HSC_{no\ int}}{DALY_{s_{int}} - DALY_{s_{no\ int}}}$$

Reward:

NRMSE

Building Trust With Sample Data



COVID MODEL

RL with "COVID MDPs"

Pre-Pandemic Engagement with the ML Community

Humanities Track - KDD Cup 2019

Feedback Phase

Begins : April 15th 2019
Ends : June 29th 2019

Check Phase

Begins : June 30th 2019
Ends : July 6th 2019

Verification Phase

Begins : July 7th 2019
Ends : July 19th 2019



247
Teams

295
Competitors

735
Submissions

IBM Research



Z:IND!

IBM Malaria Challenge

Knowledge

Use reinforcement learning to determine the optimal combination of interventions to control the transmission of malaria

47 data scientists enrolled, 21 on the leaderboard

25 August–30 August 2019 12:59



Peri-Pandemic Engagement with the ML Community



iclr.cc

ICLR

In-Person Poster presentation / poster accept

Efficient Deep Reinforcement Learning Requires Regulating Overfitting

Qiyang Li · Aviral Kumar · Ilya Kostrikov · Sergey Levine

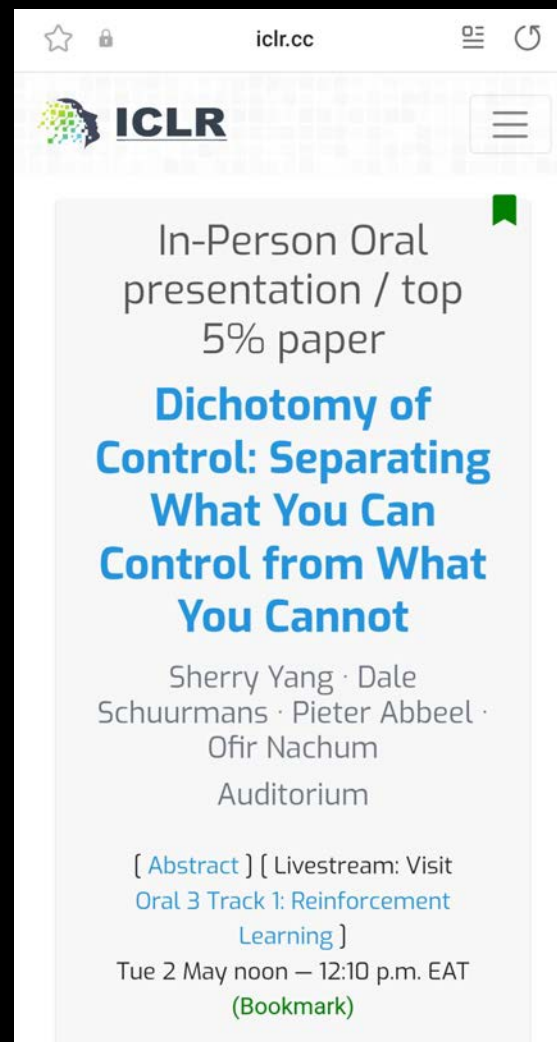
MH1-2-3-4 #96

Keywords: [Reinforcement Learning] [reinforcement learning] [Statistical Overfitting] [Sample Efficient RL]

[[Abstract](#)]

[[Poster](#)] [[OpenReview](#)]

Tue 2 May 12:30 p.m. EAT – 2:30 p.m. EAT ([Bookmark](#))



iclr.cc

ICLR

In-Person Oral presentation / top 5% paper

Dichotomy of Control: Separating What You Can Control from What You Cannot

Sherry Yang · Dale Schuurmans · Pieter Abbeel · Ofir Nachum

Auditorium

[[Abstract](#)] [[Livestream: Visit Oral 3 Track 1: Reinforcement Learning](#)]

Tue 2 May noon – 12:10 p.m. EAT ([Bookmark](#))

“I feel like my research topics are shaped by the current benchmarks which are limited in scope and it’s hard to evaluate ideas that are different...”

Currently most of the benchmarks are focussing on robotics and games, and going outside of that to actual sequential decision-making problems that would have an impact would be really good.”

ICLR2023 Attendee
May 2nd 2023
MH1-2-3-4

Peri-Pandemic Engagement with the ML Community

```
conda create -n iclr23env python=3.10
conda activate iclr23env
pip install git+https://github.com/IBM/ushiriki-policy-engine-library
```

`gym-0.26.2`

Setup

Usage

```
import gym
import ushiriki_policy_engine_library
from gym.wrappers import normalize

e = gym.make('ICLR23Workshop-env1-v0',
             model_name="http://.../ibmracovid19modelv1/",
             driver_data="https://gist.githubusercontent.com/..",
             numdays=28, duration=280,
             apply_api_compatibility=False)
env = normalize.NormalizeObservation(e)
```

```
model = TD3(env, replay_buffer_size = 5e5, hidden_dim = 512)
```

`Your algo!`

Since we're living in the future...

How can ML help a public health professional infer the transmission rate of a disease in a particular location?

Machine learning (ML) can help a public health professional infer the transmission rate of a disease in a particular location by analyzing data related to the disease and the location. Here are some ways ML can be useful:

1. Predictive modeling: ML models...

It seems like you're talking about using ml models to directly infer the transmission rates, do these concerns of they're indirectly used? For example to optimize the parameter of known epidemiological models?

You are correct that the concerns I mentioned earlier may not apply if ML models are indirectly used to optimize the parameters of known epidemiological models, rather than directly inferring the transmission rates. Indirect use of ML in this context can refer to the application of ML algorithms to optimize model parameters and improve the accuracy of existing epidemiological models, such as SIR (Susceptible-Infectious-Recovered) or SEIR (Susceptible-Exposed-Infectious-Recovered) models

Since we're living in the future... (take 2)

How can I infer transmission rate from case data for a disease?

In order to infer the transmission rate from case data for a disease, you can use mathematical models that describe the transmission dynamics of the disease. One commonly used model is the SIR (Susceptible-Infectious-Recovered) model, which divides the population into three groups: those who are susceptible to the disease, those who are infected and can transmit the disease, and those who have recovered and are no longer susceptible.

The transmission rate (also called the reproduction number or R_0) is the average number of secondary infections that result from each infected person during the early stages of an outbreak, when the majority of the population is still susceptible. This value can be estimated by analyzing the case data and fitting the SIR model to the observed data.

...

Is the transmission rate and Reproduction number truly the same thing? I thought one was a probability

The transmission rate and the reproduction number are related concepts, but they are not exactly the same thing.

...

Accelerating Model Driven Discovery with ML

—
Presented by Sekou L. Remy

Work done in solidarity with:

Irene Wanyana, Makerere University

Joyce Nakatumba-Nabende, Makerere University

Susan Kizito, Makerere University

Balla Diop Ngom, Université Cheikh Anta Diop

Ernest Mwebaze, Sunbird AI

Barbara Han, Cary Institute of Ecosystem Studies

Fadel Thior, IBM

Rose Yvette Essomba, IBM

Stephanie Muller, IBM

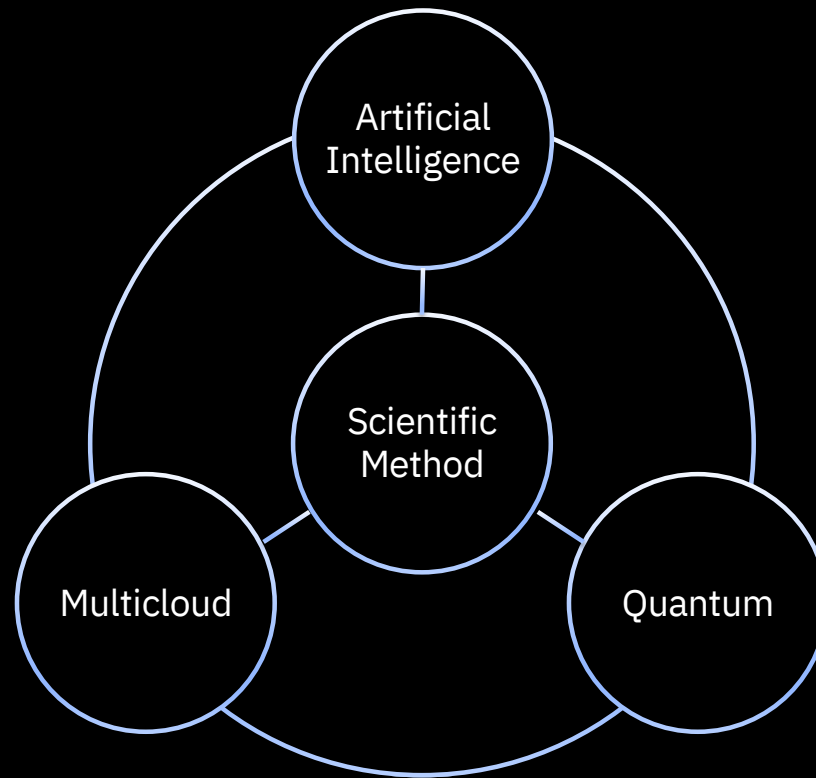
Cliff Kerr, Institute for Disease Modeling

...

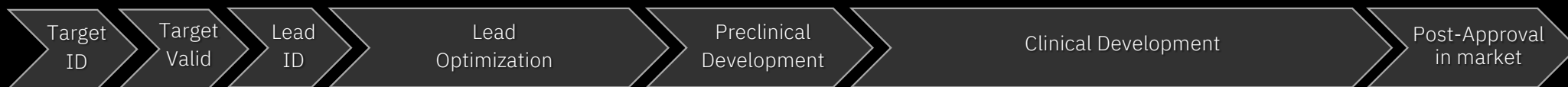
and friends!

Accelerated discovery for
neglected tropical diseases
(NTDs)

IBM is developing partnerships with non-profit and government organizations on new discovery accelerators for NTDs



Two phases of exploration



Accelerated design of new drug molecules



Accelerated generation of intervention program

Tentative

Our Goal for 2025

IBM is developing tech with the goal of overcoming bottlenecks, accelerating discovery

10x

increase the speed of
discovering more effective drugs
for NTDs with partners

Questions

Do you think this can work?

What else do you think we need?

What could go wrong/right?

Who do you think might want to join us in the journey?

- Or whose journey can we join?

Data Sources

Worldwide Non-pharmaceutical Interventions Tracker for COVID19

Updated: Aug 6, 2020 0:00 GMT+3 Disclaimer (Beta) Feedback

182

Territories Covered

5,937

Total No. of Events



Overview

Data

How to

Date or Days (X-axis)

Date

Cases, Deaths or Recoveries (Y-axis)

Confirmed Cases

Linear or Log (Y-scale)

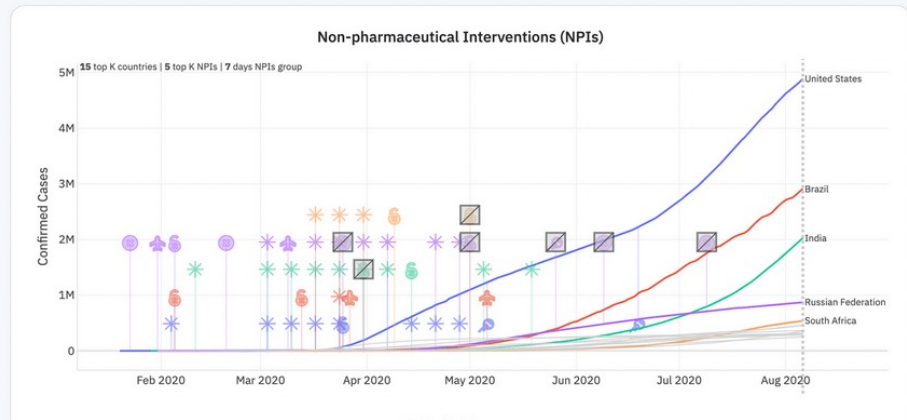
Linear Scale

Normalise Y-axis with Population Size

None



2020-08-06



JOHNS HOPKINS UNIVERSITY OF MEDICINE | CORONAVIRUS RESOURCE CENTER

Tracking Home Data Visualizations Global

COVID-19 Dashboard

by the Center for Systems Science and Engineering

Last Updated at (M/D/YYYY)
11/30/2022, 10:20 AM

Total Cases
642,731,461

28-Day Cases
11,778,591

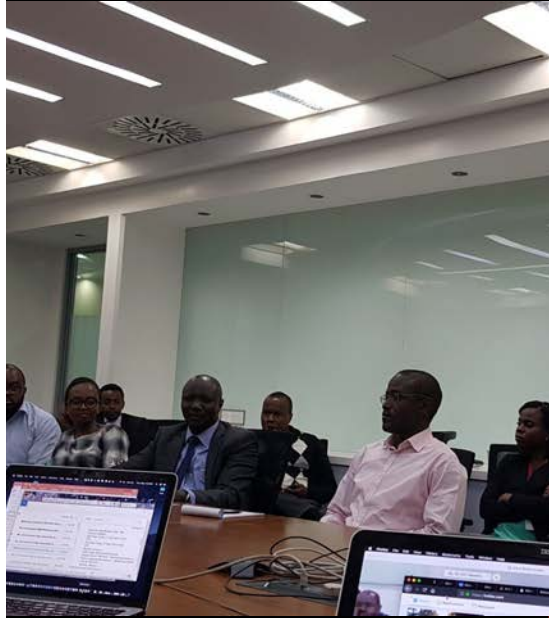
Deaths by Country/Region /Sovereignty

Japan	28-Day: 2,287,000 2,625	Total: 26,478,812 49,442
Korea, South	28-Day: 1,418,377 1,267	Total: 29,098,534 30,506
US	28-Day: 3,119,399 8,812	Total: 38,476,688 1,079,888
France	28-Day: 945,686 1,881	Total: 37,078,048 159,915
Germany	28-Day: 813,821 3,977	Total: 36,483,882 157,791
China	28-Day: 748,091 328	Total: 8,268,128 15,970
Italy	28-Day: 729,631 1,997	Total: 8,268,128 181,098
Taiwan*	28-Day: 348,687 1,421	Total: 8,268,128 14,297
Czechia		

#	Country, Other	Total Cases	28-Day Cases	Change	Recoveries	Deaths	Active Cases	Population	
1	USA	35,268,255	689,717		163,769	3,203	63,776,166	296,146	
2	India	27,098,734	+67,415	30,506	+52	527,933	594	15,804,065	307,892
3	France	24,793,166	+138,396	49,644	+210	197,422	395	81,441,509	648,498
4	Germany	24,260,660		181,098	402,581	3,005	257,057,363	4,265,608	60,262,77

<https://ibm.github.io/wntrac/>
<https://www.worldometers.info/coronavirus/>
<https://coronavirus.jhu.edu/map.html>

Motivation



Artificial Intelligence

Mimic Humans

Machine Learning

Learn with Experience

Reinforcement Learning

Interactively
Learn from Data

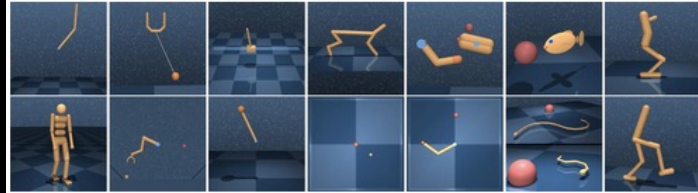
Supervised Learning

Learn from
Labeled Data

Unsupervised Learning

Learn from
Unlabeled Data

... Deep Learning?
... Data Science?



Source: <https://arxiv.org/abs/1801.00690>.

Usage

