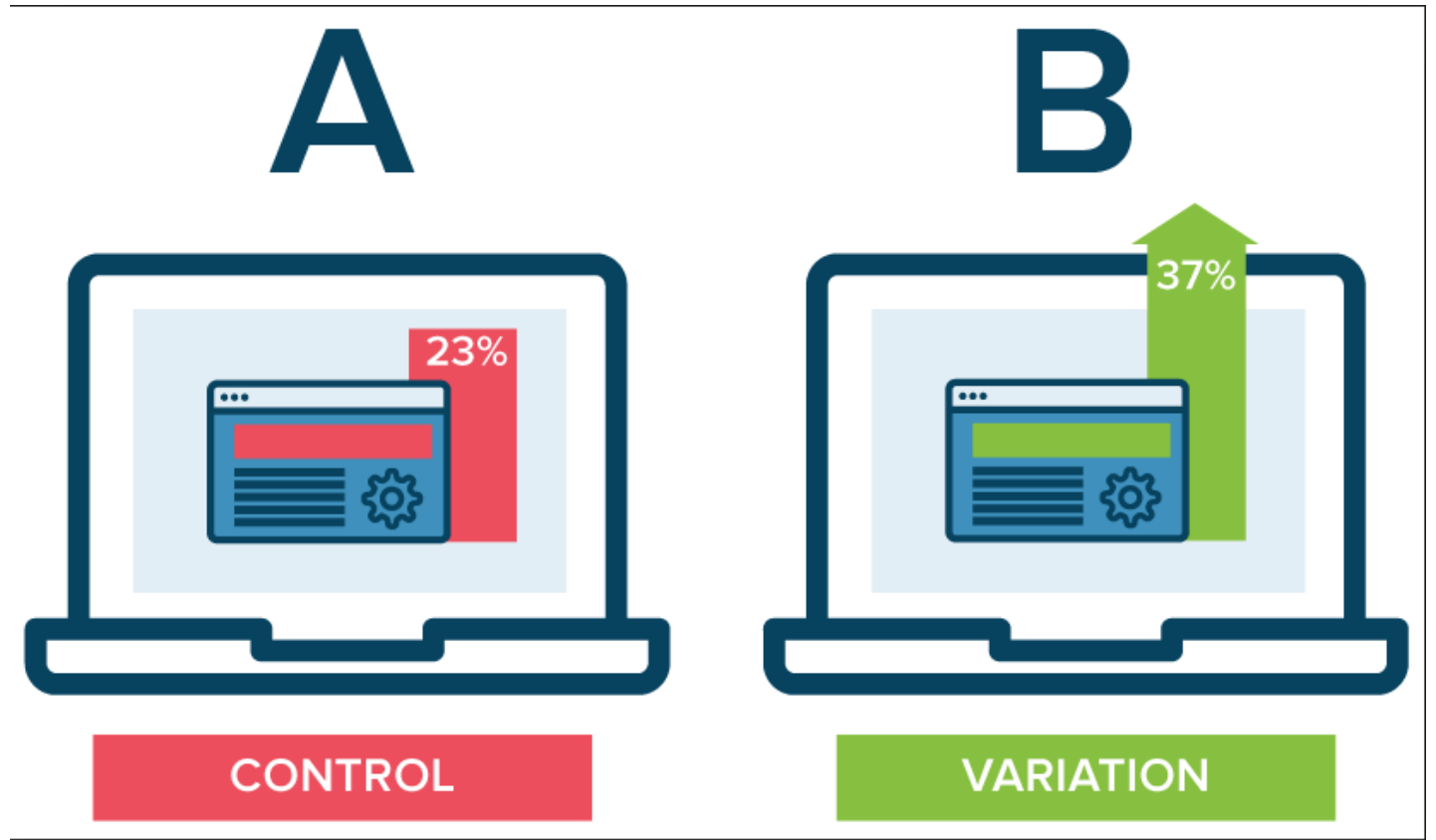


# COUNTER FACTUAL EVALUATION

PAVITHRA SRINATH



# A/B TESTING



# THE BANE OF A/B TESTING

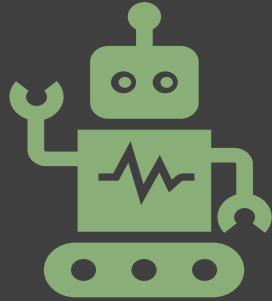
- You have a new idea
- You have multiple ideas
- You pitch it to your manager
- Expensive
  - Money
  - Time
- Multiple iterations
- Results ... Ummmm
- WHAT IF ..... BETTER WAY



# COUNTER FACTUAL EVALUATION



**What If  
Analysis**



**Only using online  
experimentation data**

- With a small twist
- Deploy  
explore/exploit  
models

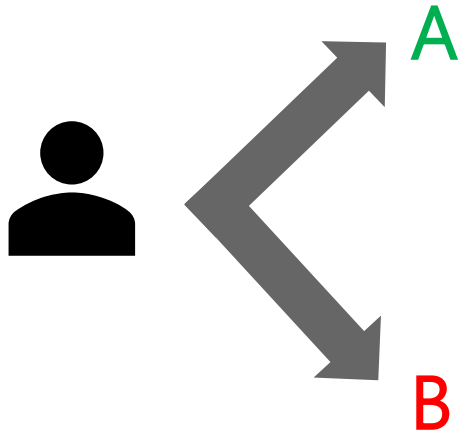


**Visualize  
different  
policies**



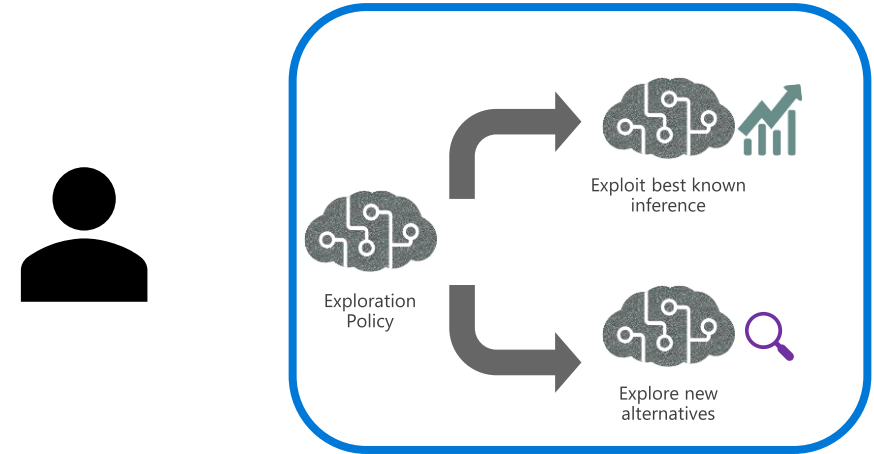
**Find a policy that  
works well  
with confidence  
bounds**

# A/B VS. COUNTERFACTUAL



## A/B Test:

1. Design the Right Experiment
2. Test online once
3. Start over



## Offline Experiment:

1. Use models that exploit and explore
2. Record User Interaction
3. Find the policy and model that fits reality

# USE CASE REVIEW

	Tom	Anna
Monday	Politics	Sports
Sunday	Music	Politics



	Tom	Anna
Monday	Politics	Sports
Sunday	Sports	Sports

## MODEL

Function that we learnt

## POLICY

Model that acts in the world

Dashboard > Cognitive Services > icml-demo - Evaluations

### icml-demo - Evaluations

Cognitive Services

Search (Ctrl+ /)

+ Create evaluation Delete evaluation

**We are actively creating a dedicated instance for you. Please be patient while provisioning completes.**

The effectiveness of the Personalizer Loop can be evaluated with a technique called Counterfactual Evaluation.  
[Learn more about Counterfactual Evaluations.](#)

NAME	START DATE	END DATE	STATUS
Click Create Evaluation to create an evaluation.			

Overview

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

RESOURCE MANAGEMENT

Settings

Model and Policy

Evaluations

Keys

Quick start

Pricing tier

Billing By Subscription

Properties

Locks

Export template

Monitoring

Alerts

Metrics

Diagnostic settings

# CREATE A COUNTERFACTUAL EVALUATION

[Dashboard](#) > [Cognitive Services](#) > [icml-demo - Evaluations](#) > Create an evaluation

## Create an evaluation

\* Name

evaluation1

\* Start date ⓘ

2019-06-06

\* End date ⓘ

2019-06-06

### Additional evaluation options

#### Optimization Discovery

Discover a more optimal learning policy? ⓘ

Yes

No

#### Compare additional learning policies

You can optionally compare the performance of your model with the performance of additional learning policies.

Add Policy

LEARNING POLICY

Ok

# CREATE A COUNTER FACTUAL EVALUATION



# COUNTER FACTUAL EVALUATION

Dashboard > icml\_demo - Evaluations

icml\_demo - Evaluations  
Cognitive Services

Search (Ctrl+/)

+ Create evaluation - Delete evaluation

The effectiveness of the Personalizer Loop can be evaluated with a technique called Counterfactual Evaluation.  
[Learn more about Counterfactual Evaluations.](#)

NAME	START DATE	END DATE	STATUS
evaluation1	6/6/2019	6/6/2019	completed

Overview  
Activity log  
Access control (IAM)  
Tags  
Diagnose and solve problems

RESOURCE MANAGEMENT

Settings  
Model and Policy  
**Evaluations**  
Keys  
Quick start

Dashboard > icml\_demo - Evaluations > Evaluation Summary

## Evaluation Summary

Name  
evaluation1

Status  
completed

Start Date  
6/6/2019

End Date  
6/6/2019

### Evaluation Results

Comparison of different learning policies

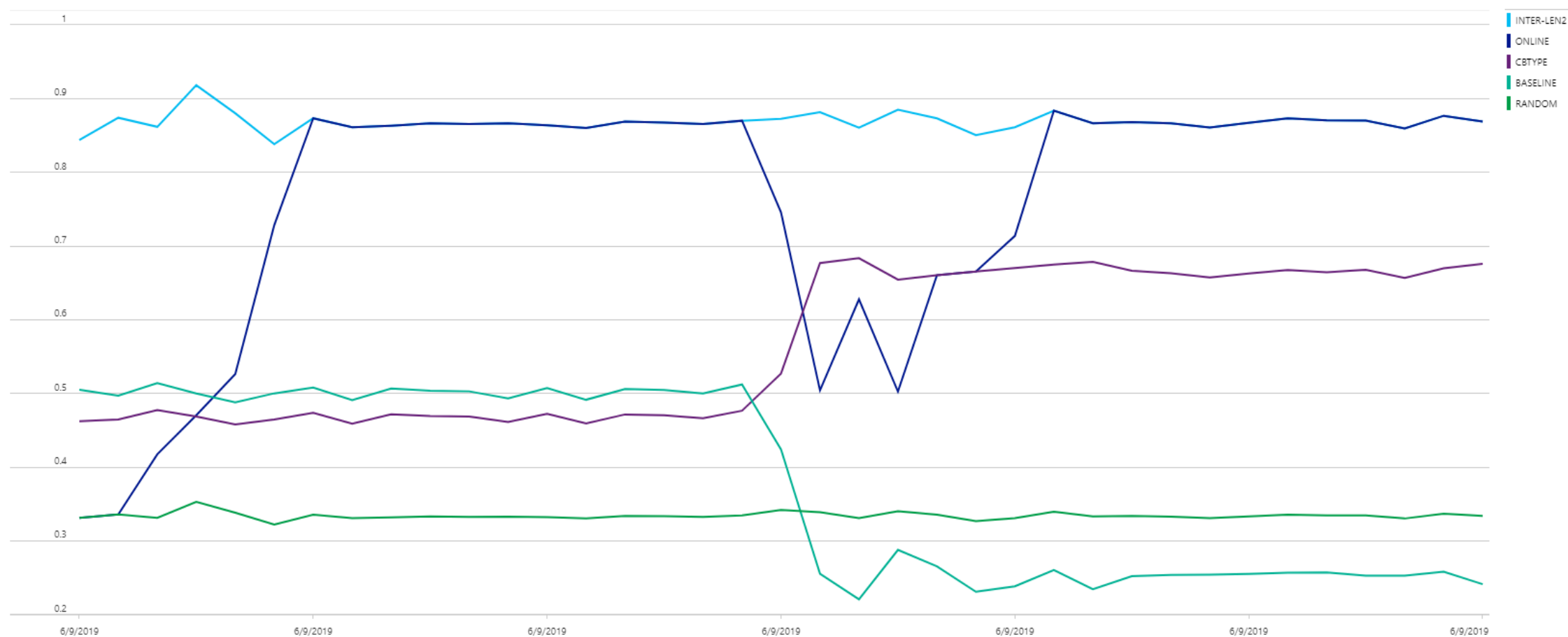
# Evaluation Results

## Learning Policy Comparison

LEARNING POLICY	SOURCE	AVERAGE REWARD	CONFIDENCE INTERVAL (P95)	DOWNLOAD
Inter-len2	Uploaded	0.86102	0.85150 - 0.87054	<a href="#">Download</a>
Online	Current Learning Policy of this Personalizer Loop	0.67561		
Marginals	Uploaded	0.52820	0.52281 - 0.53359	<a href="#">Download</a>
cbType	Uploaded	0.52790	0.52245 - 0.53334	<a href="#">Download</a>
Hyper1	Uploaded	0.52732	0.52185 - 0.53279	<a href="#">Download</a>
Baseline	First action sent to Rank by the app	0.41478	0.40897 - 0.42060	
Random	Randomly generated ranks	0.33160	0.32793 - 0.33527	

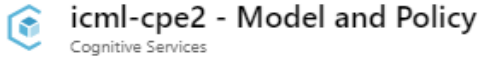
# COUNTER FACTUAL EVALUATION RESULTS

# COUNTER FACTUAL EVALUATION DASHBOARD



# IMPORT POLICY

[Dashboard](#) > icml-cpe2 - Model and Policy



- Overview
- Activity log
- Access control (IAM)
- Tags
- Diagnose and solve problems

## RESOURCE MANAGEMENT

- Settings
- Model and Policy
- Evaluations
- Keys
- Quick start
- Pricing tier
- Billing By Subscription
- Properties
- Locks
- Export template

### Model

The Personalizer model is a file that captures what the service learnt with the current learning policy, since the model was created.

Model created  
6/9/2019, 8:56:37 AM

Model last updated  
6/9/2019, 1:53:14 PM

[Export model](#)

### Learning policy

Learning policies are machine learning settings that determine how the service learns from events.

[Export learning policy](#)

Import learning policy

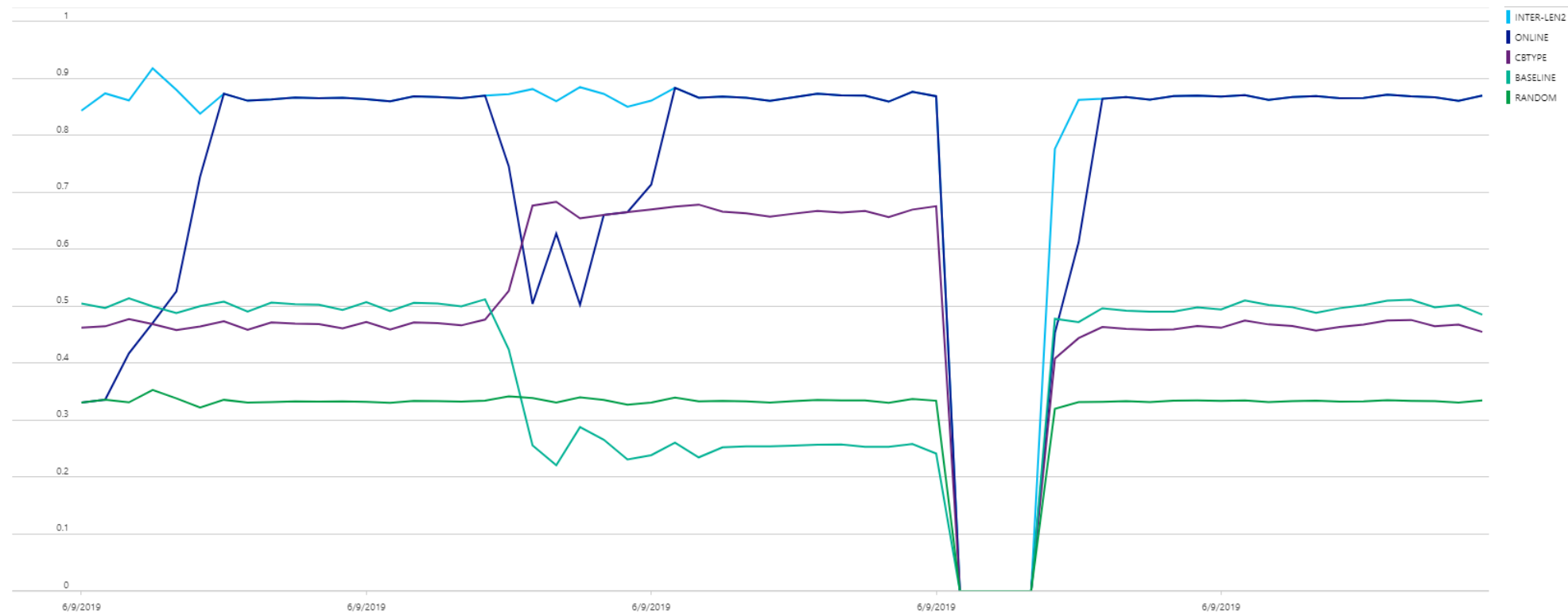
[Upload](#)

### Data

[Clear data](#)

# COUNTER FACTUAL EVALUATION DASHBOARD

Average Reward vs Time (UTC)





QUESTIONS??

THANK YOU!