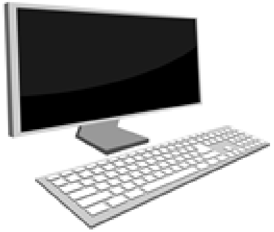
Carnegie Mellon University
Electrical & Computer Engineering

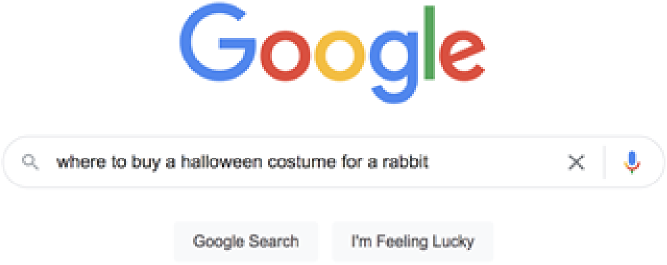
Designing Equitable Scheduling Systems

Sahana Rangarajan
Xuesi Chen, Pratyush Patel, Sara Mahdizadeh Shahri, Jaylen Wang
Akshitha Sriraman

Where do we see web services day to day?



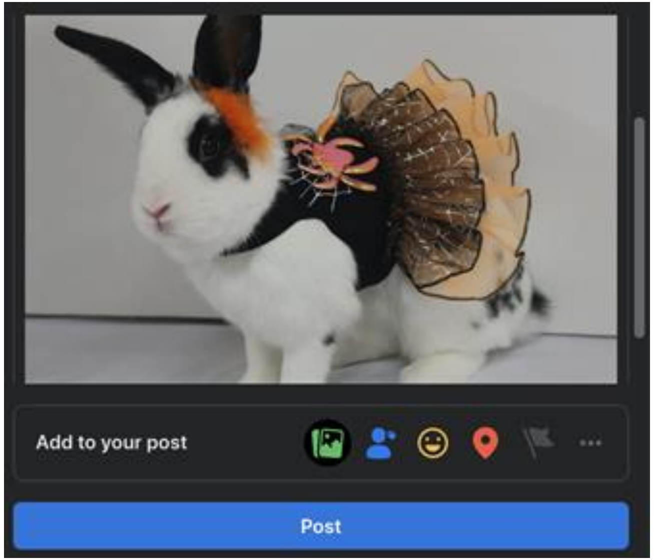
Search



where to buy a halloween costume for a rabbit

Retail

Social Media



Is latency a big deal?

Yes!

100 ms delay → 1% drop in sales

0.5 s delay → 20% drop in traffic

Result: Stringent latency constraint (300ms)

Where do we stand today?

increasingly ML-based

Task prioritization by **schedulers**

^

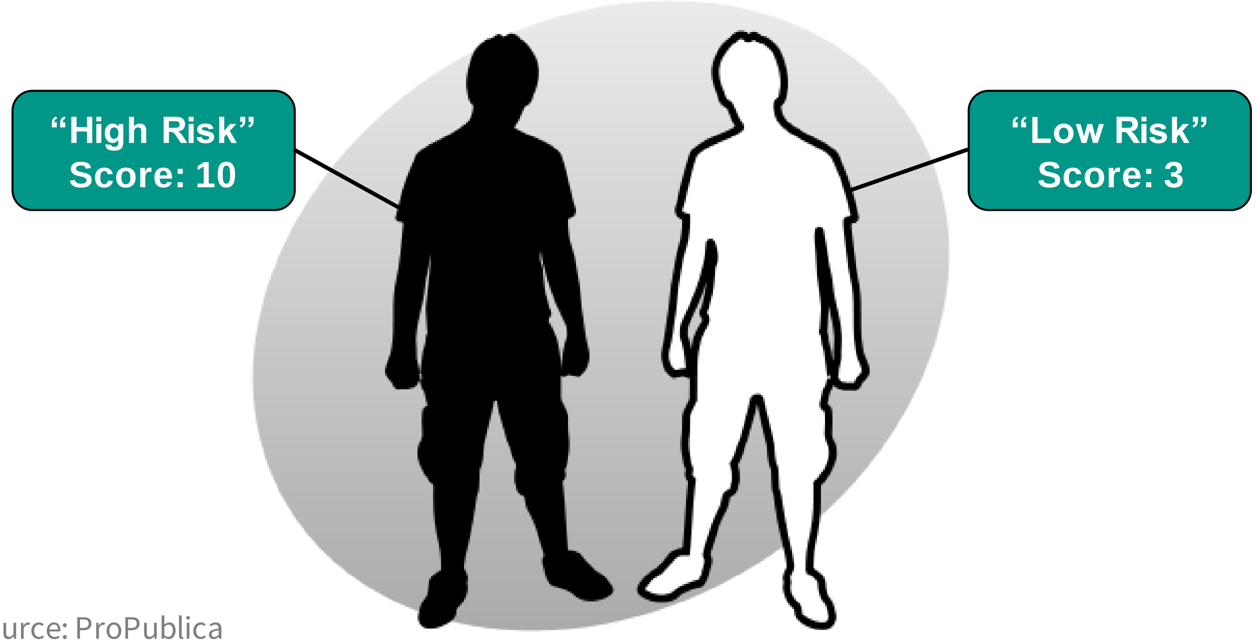
Examples:

- Decima
- Quasar
- DeepRM
- Paragon

What’s a possible pitfall with ML?

Bias

Example: COMPAS Algorithm for Recidivism Risk



“High Risk”
Score: 10

“Low Risk”
Score: 3

Source: ProPublica

How’s this relevant to us?

User 1
Younger, less patient
Stricter latency requirement

Can I optimize performance by scheduling user 1’s task before user 2’s?

User 2
Older, more patient
Relaxed latency requirement

How could this go wrong?

Case study: Varying perceptions of Wikipedia QoS

Created with mapchart.net

Lower user satisfaction:
Germany
#4 global GDP (nominal)

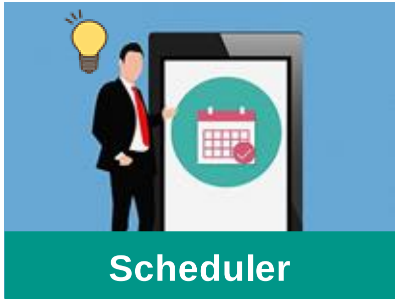
Higher user satisfaction:
Belarus
#90 global GDP (nominal)

Optimizing on user satisfaction is a slippery slope

Source: Analyzing Wikipedia Users' Perceived Quality of Experience: A Large-Scale Study (Salutari et al.)

Proposal: A Bias-Free Scheduling Framework

Do different **demographics** exhibit different **latency tolerances**?



Can a scheduler **capitalize** on demographic differences?

Build a **bias-free scheduling framework** with **under-the-hood real-time auditing**

Carnegie Mellon University
Electrical & Computer Engineering

Designing Equitable Scheduling Systems

Sahana Rangarajan
Xuesi Chen, Pratyush Patel, Sara Mahdizadeh Shahri, Jaylen Wang
Akshitha Sriraman