Practical Lessons from Conversion Ads at Pinterest

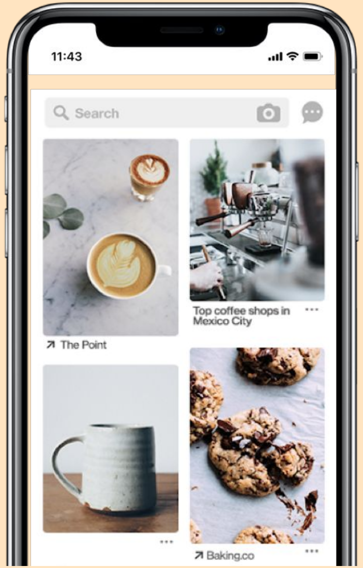
Aayush Mudgal

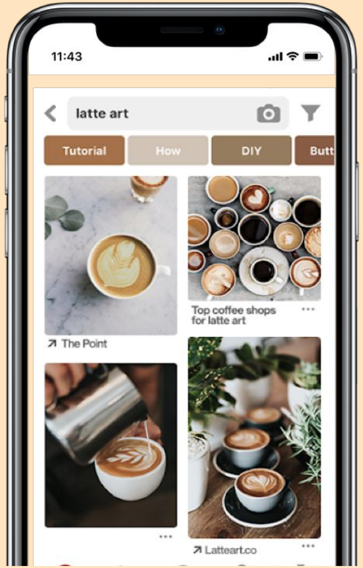
AI Conference
September 27th, 2023

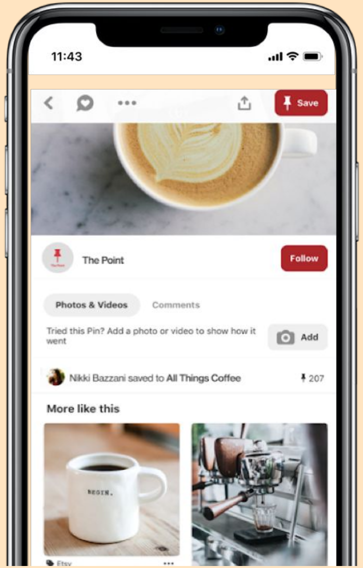
Overview

- Pinterest and Ads @ Pinterest
- How Conversion Optimization Works?
- Unique Challenges to Conversion Optimization
- Brief History of the Conversion Ranking Model
- Privacy Landscape Changes and its impact to Ads Personalization
- Deep Dive into some Industry Standard Solution to Enhance Personalization

Bring everyone the **inspiration** to **create the life** they love



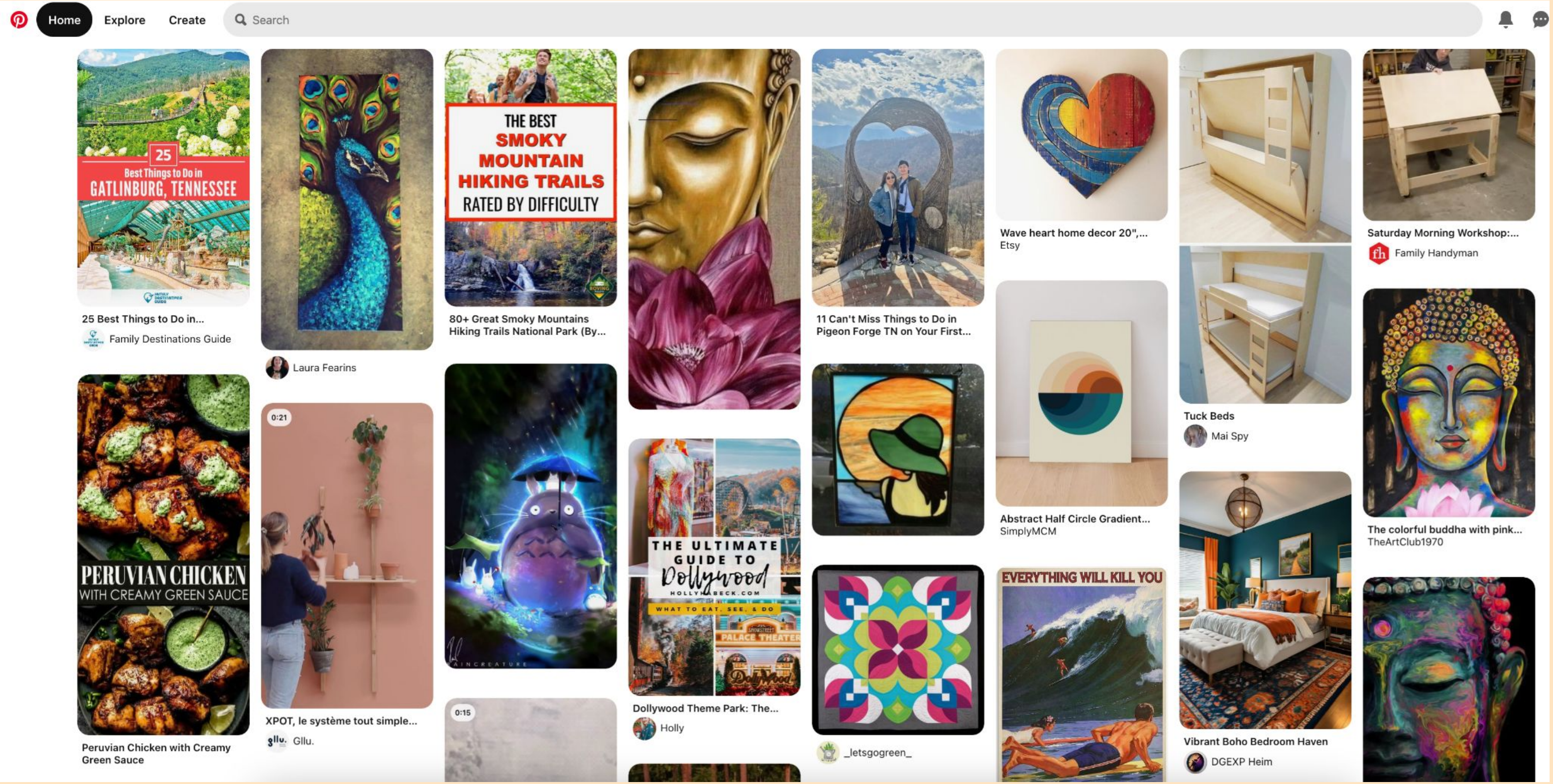




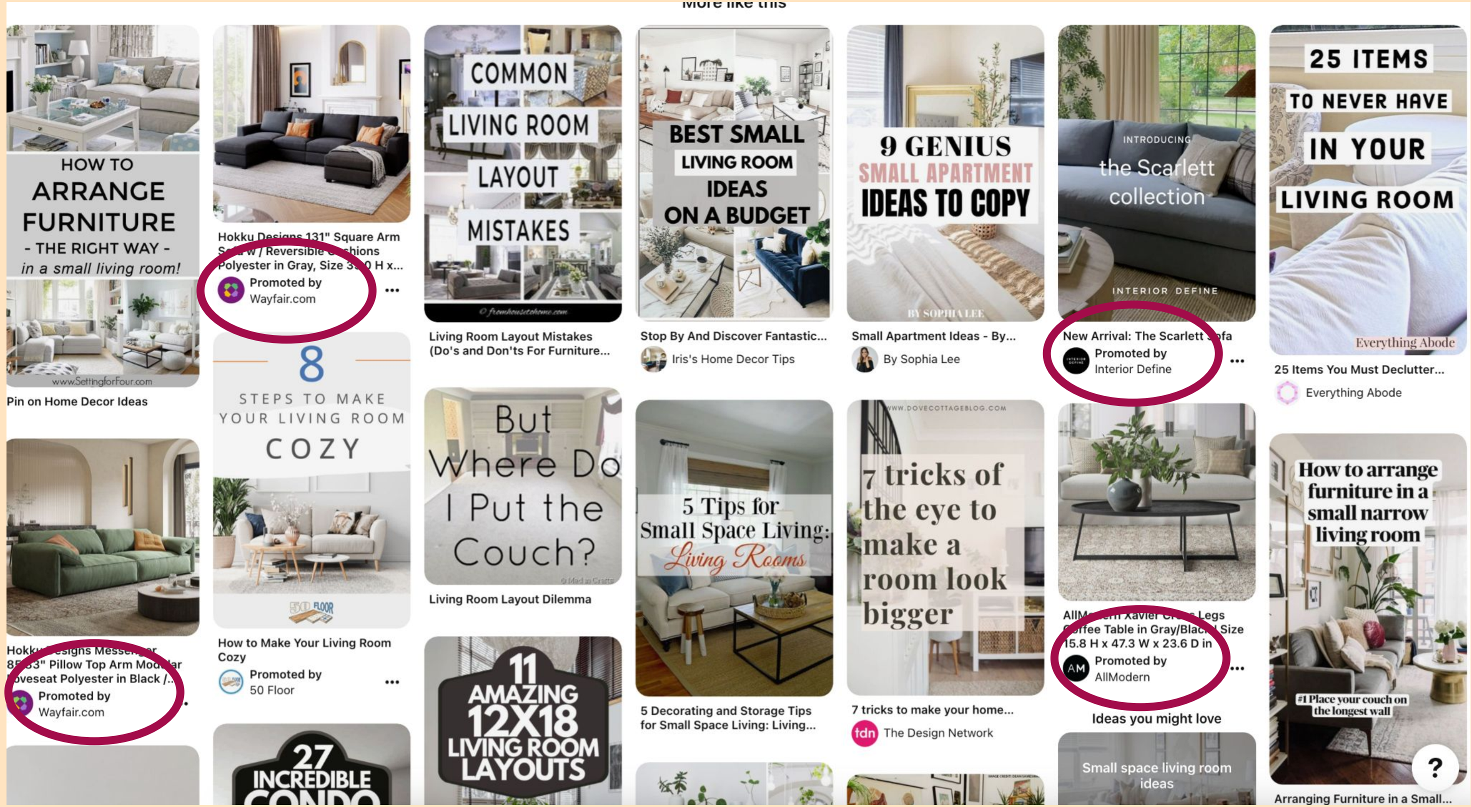
Home Feed & Search

Aayush Mudgal

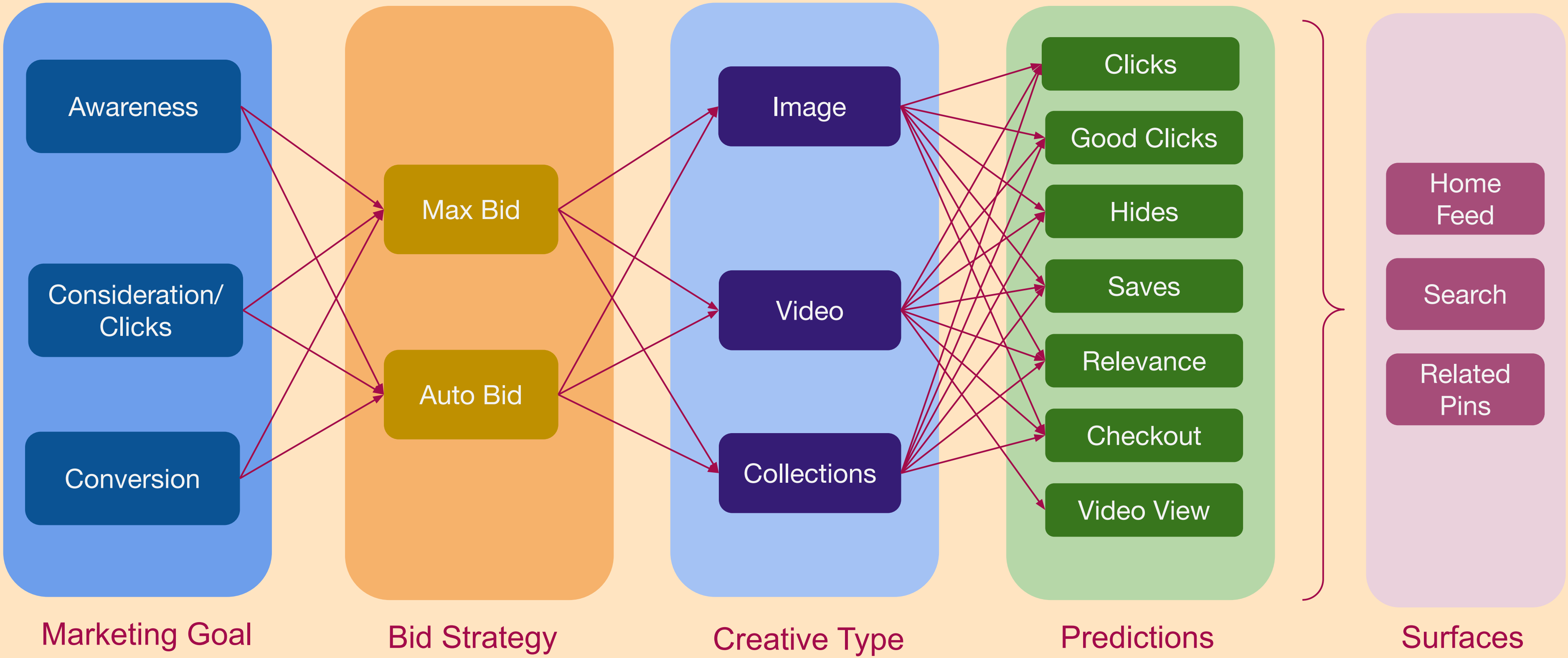
@mudgal01



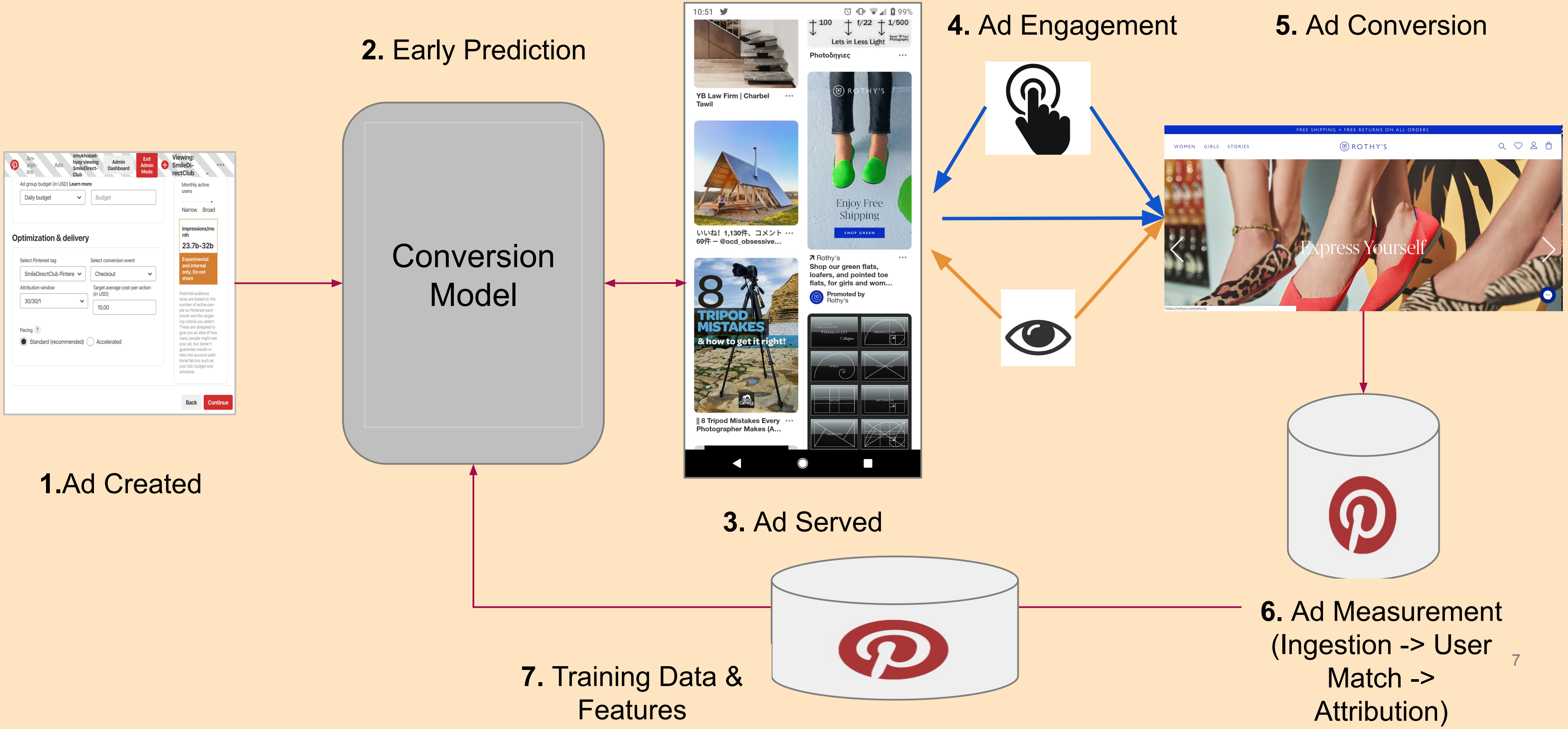
Ads @ Pinterest



Ads Product in a nutshell
Managing Complexity



How Does the Conversion Pipeline Work



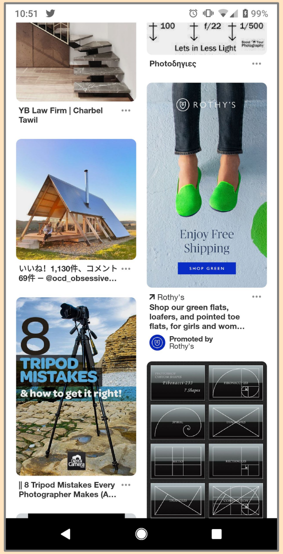
Conversion Optimization Campaign

- Conversion Event Type
  - Event that happens on advertiser website/app
    - Reported via a Pinterest Tag or API
  - Checkout, Signup, Add to cart, Lead (supported for optimization)
  - Page Visit, Search, Video View (reported, used in features)

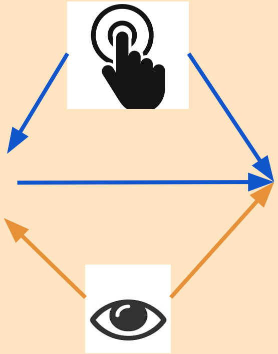
- Attribution Window
  - Time during which the platform can take credit for a conversion
  - E.g. 7/7/1 window - Pinterest gets credit for any conversion on any device that happens up to
    - 7 days after a click
    - 7 days after a save
    - 1 day after a view

Conversion Prediction

Optimized for Conversions but billed on Impressions



Ad Engagement



$$P\left(\frac{Conversion}{Impression}\right) = P\left(\frac{Click}{Impression}\right) \times P\left(\frac{Click\ Attributed\ Conversion}{Click}\right)$$

Challenges Unique to Conversion Modeling

Data quality

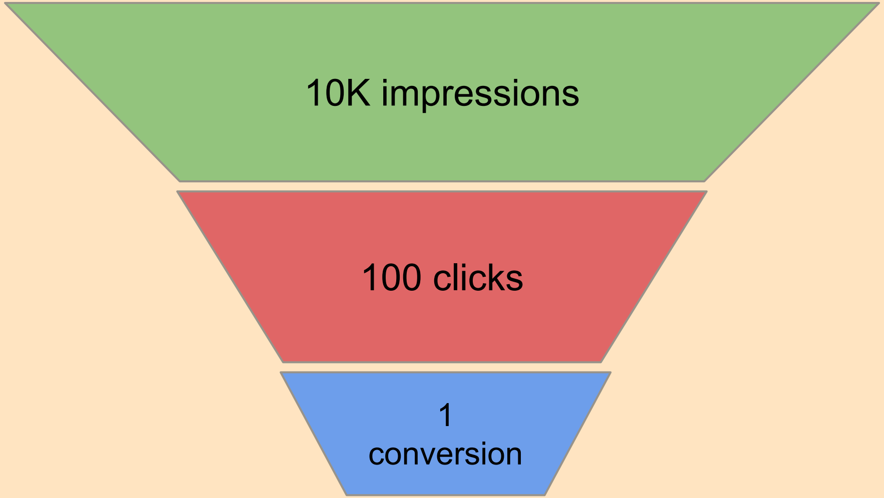
- Controlled by advertisers
- Inaccurate labels and abnormal conversion volume (over-report or under-report)
- User Match is stochastic

Data volume and label sparsity

- Extra constraints on model complexity
- Slower iteration on experiment

Delayed feedback

- Comes from the nature of the attribution window
- Frequency of model update vs. false negative
- Model calibration



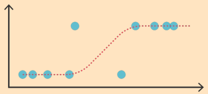
Data Quality based on Downstream Application

- For Example: An Auto-Manufacturer is sending Page-visits as Checkouts
    - Reporting/Ads Manager: Want to Report as Advertiser expects to see events they send
    - Model Training:
        - Can Potentially Filter Events if impacts overall performance
        - Handle through ID Features usage in the model
    - Internal Metrics Reporting
        - Filter such events as they might be outliers

Handling Data volume and Label Sparsity

- Enhance data
  - Multi-task multi-tower: leverage rich dataset from other offsite and onsite tasks

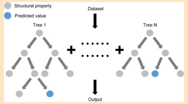
- Improve model efficiency given the limited data volume
  - Efficient architectures: such as feature crossing
  - Efficient features: better user feature embeddings, and interaction features



First ML model

2014

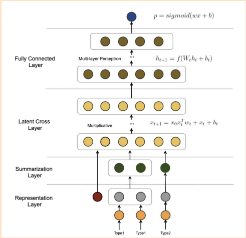
GBDT Model + LR



2017

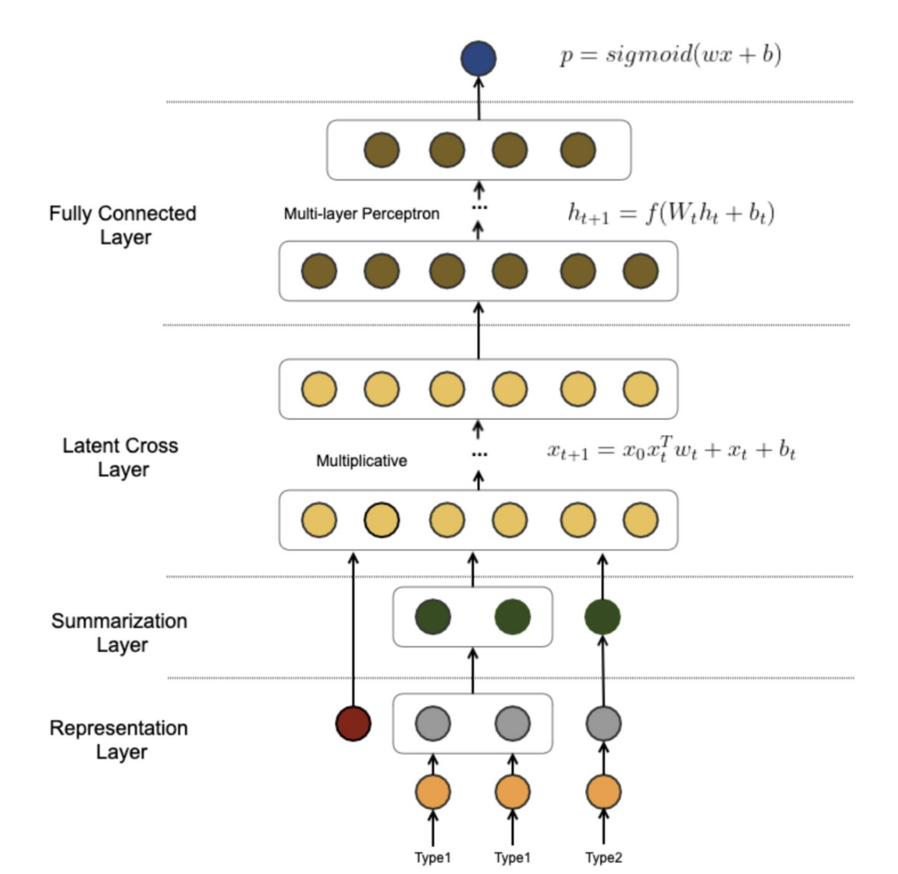
Neural Network Model (Deep and Wide)*

2018



Pinterests' AutoML

2019



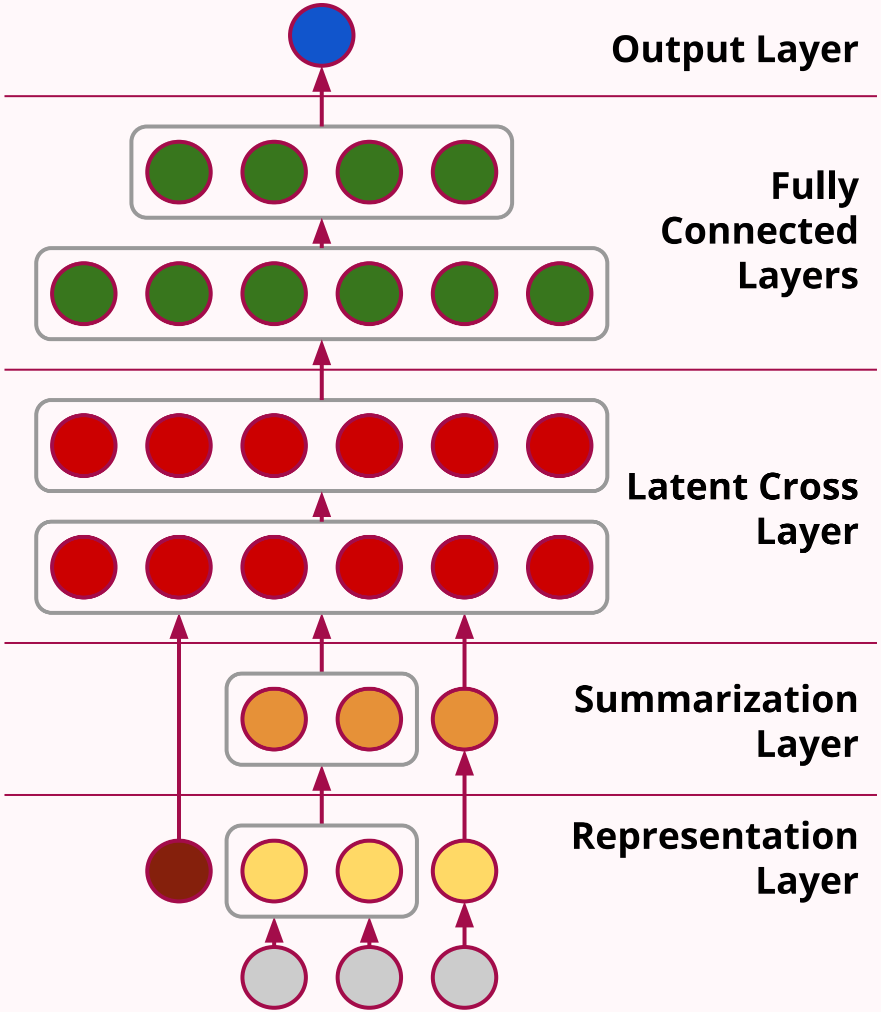
[How we use AutoML, Multi-task learning and Multi-tower models for Pinterest Ads]

Pinterest's AutoML

1.
Raw features as input

2.
Learned feature interactions through summarization and latent cross

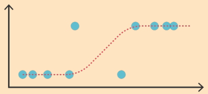
3.
Multi-task objectives for different engagement like repin, click



[How we use AutoML, Multi-task learning and Multi-tower models for Pinterest Ads]

Pinterest’s AutoML

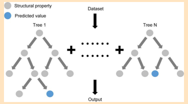
- **Representation layer**
  - Squashing, clipping, hashing projection, normalization, automate feature transformation
- **Summarization layer**
  - Grouping, learn common embedding (category vector for user and pin)
- **Latent cross layer**
  - Multiplicative layer, high degree interactions, force “explicit” feature crossing, (could be DCNv2, low rank DCNV2, Masknet)
- **Fully connected layers**
  - Classic deep neural network



First ML model

2014

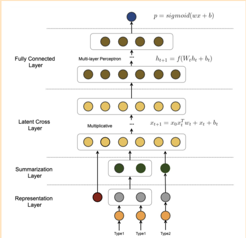
GBDT
Model + LR



2017

Neural Network
Model (Deep
and Wide)*

2018



Pinterests'
AutoML

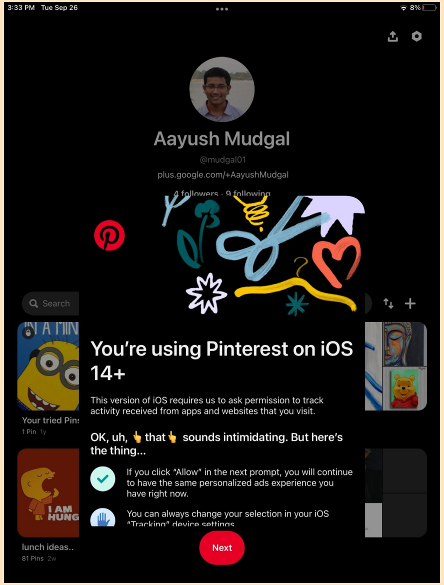
2019

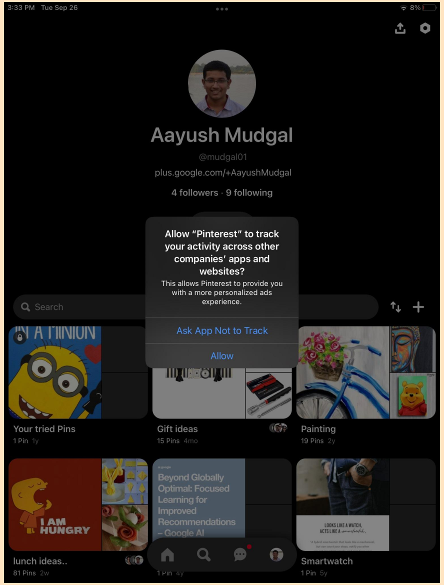
iOS 14 Rollout

2020

Trade-off between
Model Size & Label
Sparsity !!!!

Privacy Changes: More Power To Users to choose how the data is shared





Apple iOS Broad Prompt

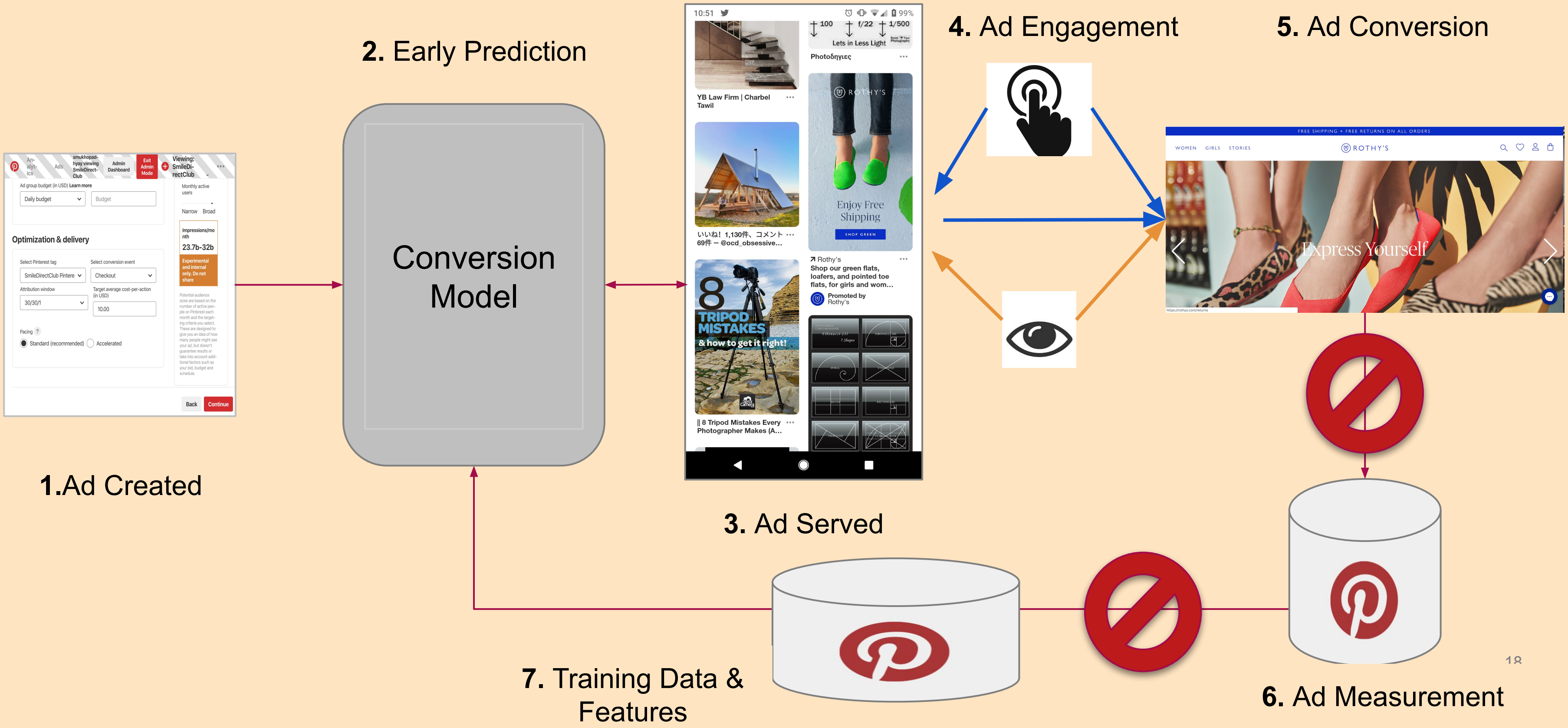
Don't allow Tracking

Unobserved Data

Allow Tracking

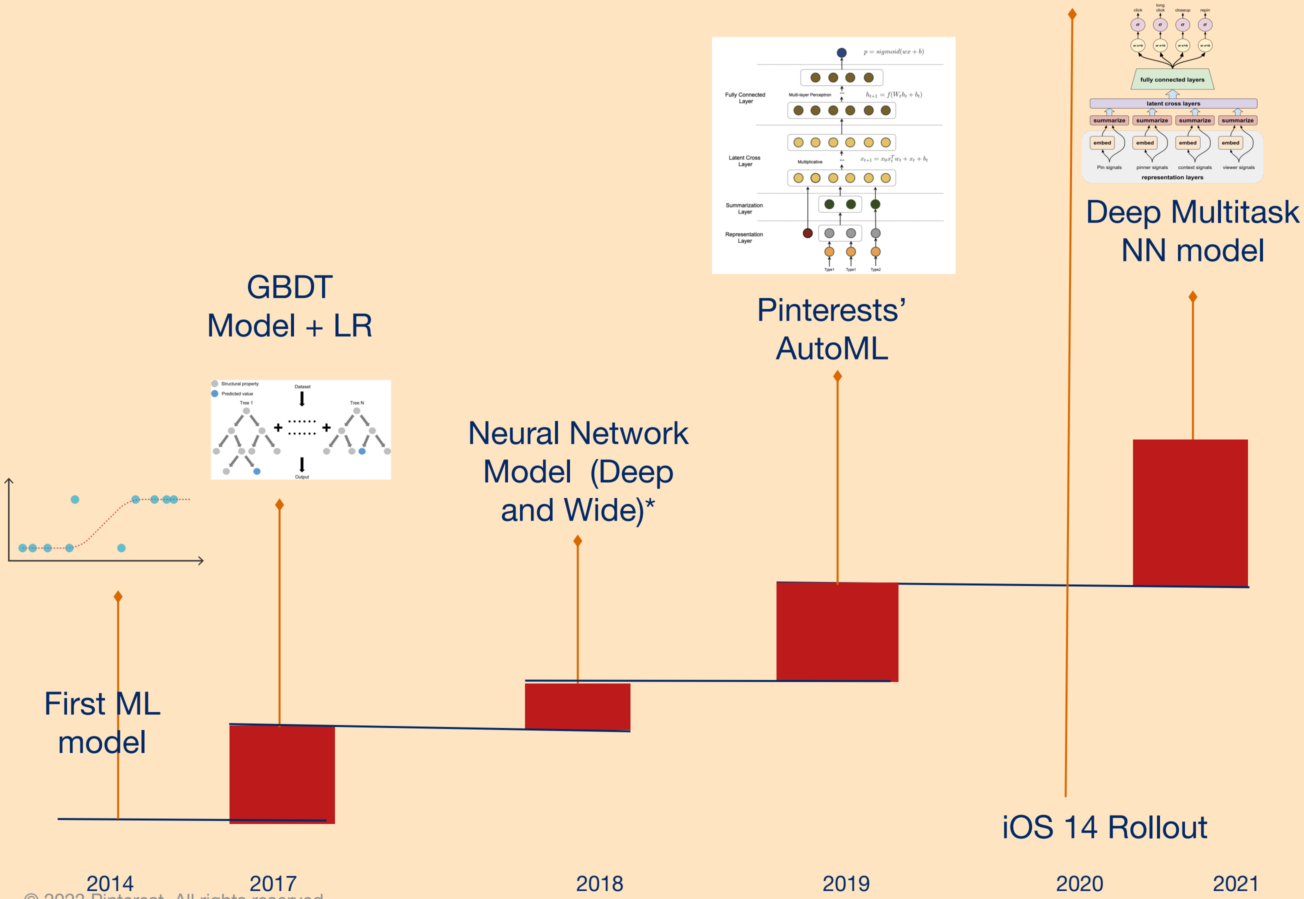
Observed Data

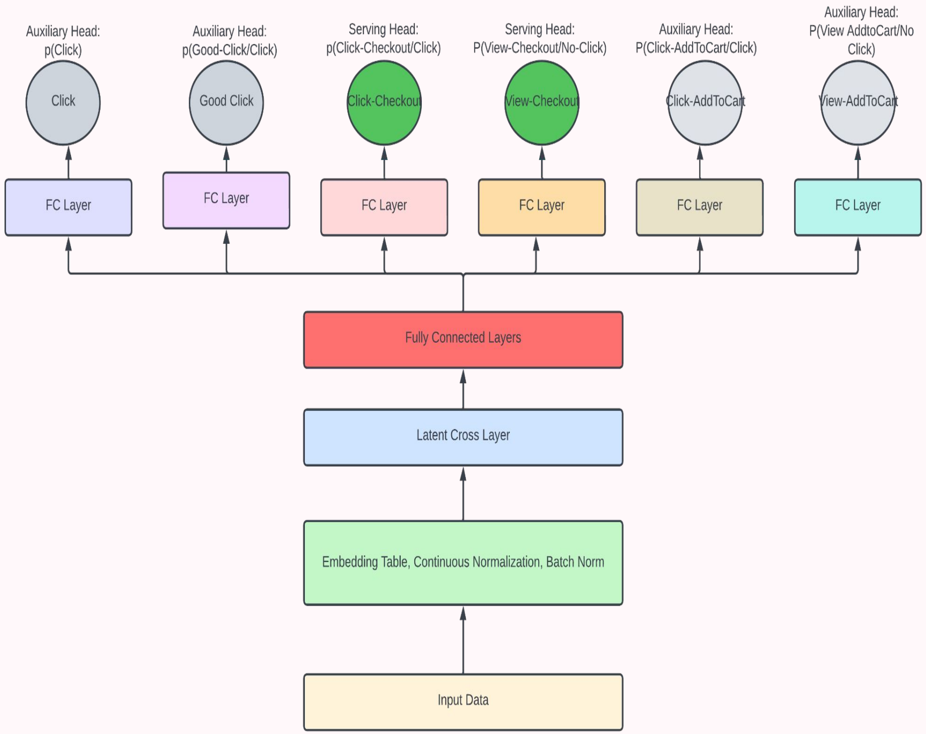
Conversion Optimization: Landscape Changes



Major Impact to Conversion Predictions

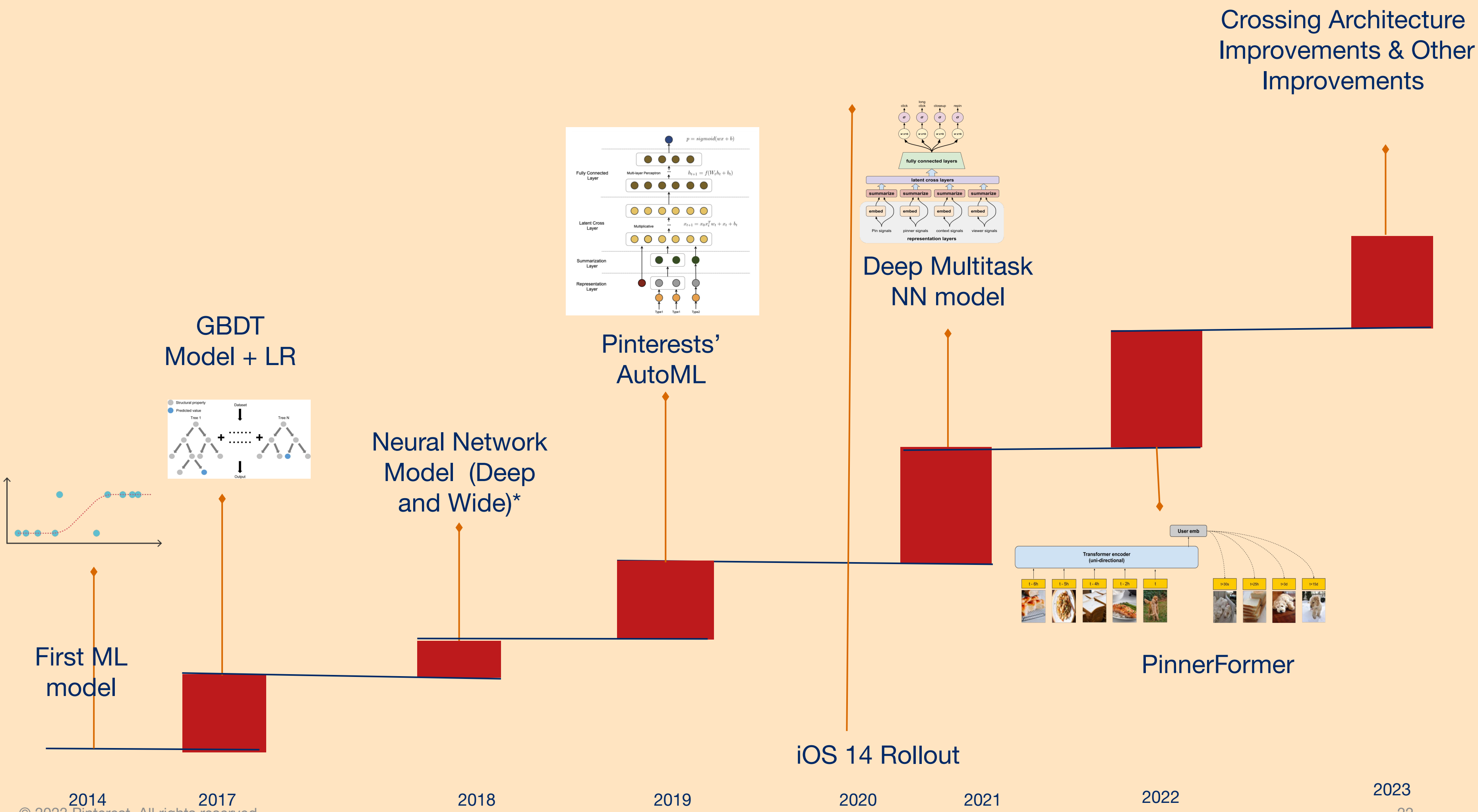
	Users with Visibility 🙋‍♀️	Users without Visibility 🙅‍♀️
Training Data	✓	✗
Offsite Featurization	✓	✗
Ads Serving	✓	✓
User Matching	✓	✗
Experiment Analysis/Reporting Impacted	✗	✗


First ML model
GBDT Model + LR
Neural Network Model (Deep and Wide)*
Pinterests' AutoML
Fully Connected Layer
Latent Cross Layer
Summarization Layer
Representation Layer
Deep Multitask NN model
fully connected layers
latent cross layers
summarize
embed
representation layers
iOS 14 Rollout
2014
2017
2018
2019
2020
2021



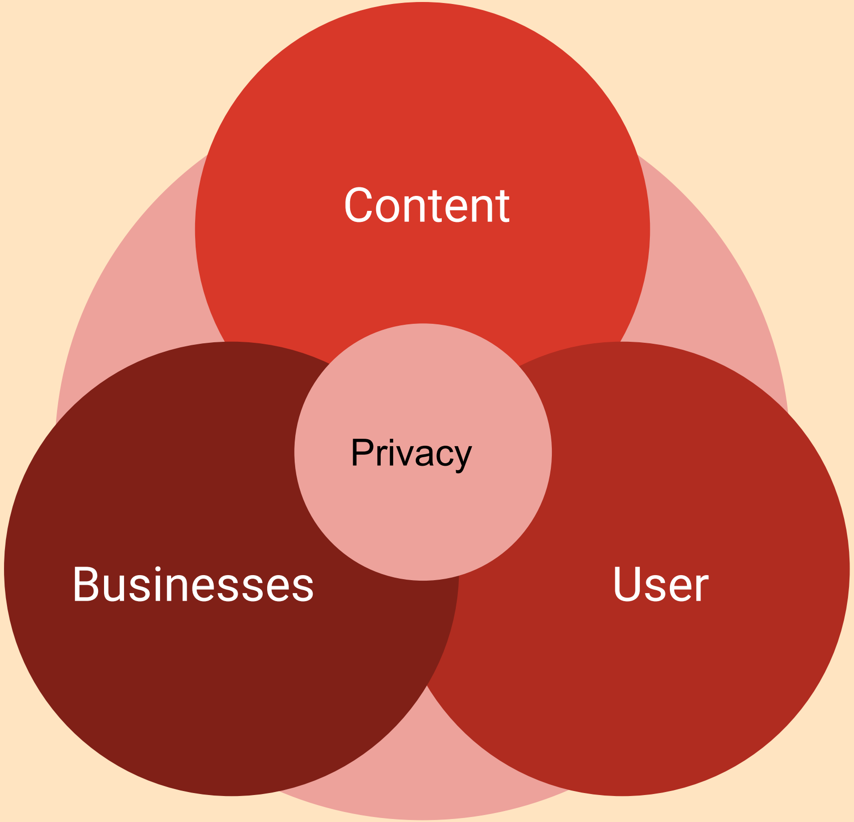
Multi-Task Learning

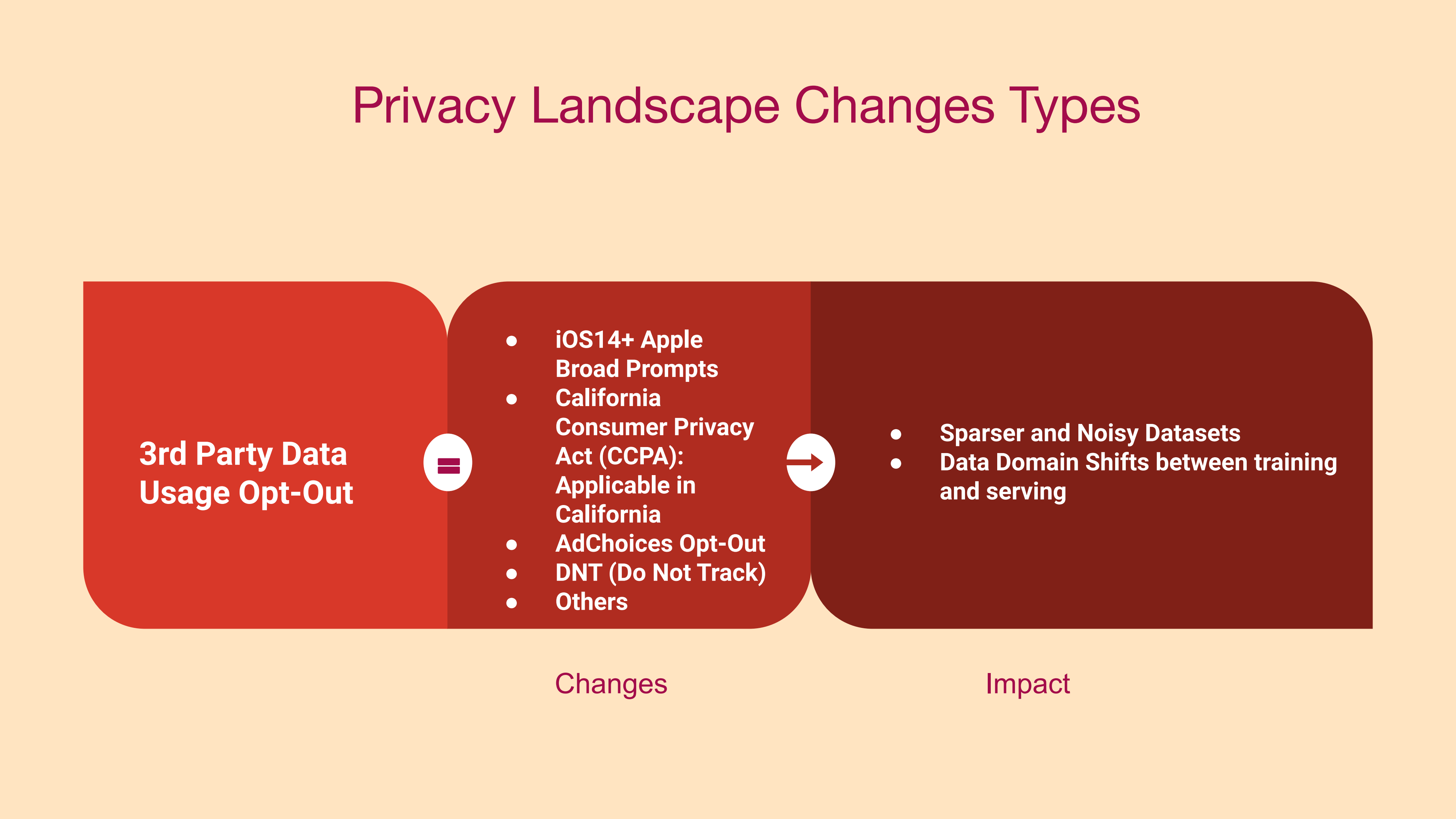
- **Enrich Data**
  - Enrich the training data, warm up embedding from rich onsite signals
- **Auxiliary Heads for Training**
  - Learn more efficiently with other actions as auxiliary head
    - Like Good-Clicks, AddToCart
- **Easier to Maintain**
  - Reduce number of models to maintain from 10+ to 3
- **Reduced Serving Infrastructure Costs**
  - Single Model inference can give multiple predictions



Crossing Architecture Improvements & Other Improvements
First ML model
GBDT Model + LR
Neural Network Model (Deep and Wide)*
Pinterests' AutoML
iOS 14 Rollout
Deep Multitask NN model
PinnerFormer
Transformer encoder (uni-directional)
User emb
fully connected layers
latent cross layers
summarize
embed
representation layers
2014
2017
2018
2019
2020
2021
2022
2023

Personalization doesn’t have to come at the expense of privacy




Privacy Landscape Changes Types
3rd Party Data Usage Opt-Out
iOS14+ Apple Broad Prompts
California Consumer Privacy Act (CCPA): Applicable in California
AdChoices Opt-Out
DNT (Do Not Track)
Others
Sparser and Noisy Datasets
Data Domain Shifts between training and serving
Changes
Impact

Privacy Landscape Changes Types

Featurization Impact

- Digital Services Act (DSA)
- Fair Lending Act
- Control how features are being computed

- Limited Feature to Learn from
- Feature Distribution Shifts impacting model calibration
- Require monitoring for compliance for feature handling

Changes

Impact

Privacy Landscape Changes Types

Identification Headwinds

=

Safari, Chrome and other browsers are actively working to remove identifiers used for ad tracking (like 3rd party cookie, IP address) and propose new privacy safe solutions

→

- Deterioritiating Data Quality: Inability to match to the right user
- A/B experimentation impacted due to noise

Changes

Impact

Privacy Landscape Changes Types

New Browser Based Measurement Solutions

=

Google's Privacy Sandbox: Limit tracking of individuals and provide safer alternatives to existing technology on these platforms

→

- **Noisier Labels**
- **Different Label Distribution**
- **Less Granular**
- **Aggregated Data**

Changes

Impact

Privacy Landscape Changes Types

Data Deletion Requirements

- Flexibility to allow data from the system to be deleted (user and/or advertiser)

- Perform deletion in scalable manner
- Monitoring and Data Flow Requirements
- Model Unlearning

Changes

Investment Directions

Industry Standard Investments

01	**Algorithmic Advancements**	• Architectures and Algorithms resilient w.r.t training with Sparse and Noisy Datasets • Handling Data Domain Shifts between training and serving • Training on encrypted data and aggregated datasets

02	**Data Storage and Privacy Safe Handling**	• Privacy Restrictions are tiered depending on use case
		  ○ Data Flow Monitoring and Alerting of any misuse
		• Access Restrictions enforcement on Use-Case level
		  ○ Columnar vs File Storage Level

03	**Secure Data Sharing**	• Secure Multi-Party Computation:
		  ○ Share data safely between two parties without divulging private data
		• Clean Rooms
		  ○ Neutral party for doing Ads Measurement and Attribution
		• API Conversions: Server to Server

04	**Infrastructural Changes**	• On Device Learning • Federated Learning

Industry Standard Approaches for Better Models

01	**Learn Better from First Party Onsite User Data**	• Post Click Feature Distillation • Onsite Auxiliary Labels • Onsite Feature Engineering
02	**Learning from Unlabeled Data**	• Pseudo Labeling • Semi-Supervised Learning
03	**Learning with Sparse Data**	• Ensemble Technique • Data Augmentation
04	**Handling Bias Between Different Population**	• Reweight/Resample • Domain Adaptation Techniques

Introducing Privacy by Default: K-Anonymity

A dataset provides k-anonymity protection if the information contained for each person contained in the dataset cannot be distinguished from at least k-1 other individuals whose information also appears in the dataset

Name	Gender	Age	Zip	Label
AB	M	23	121*	0
BC	M	25	121*	1
DE	F	40	123*	1
CD	F	41	123*	0

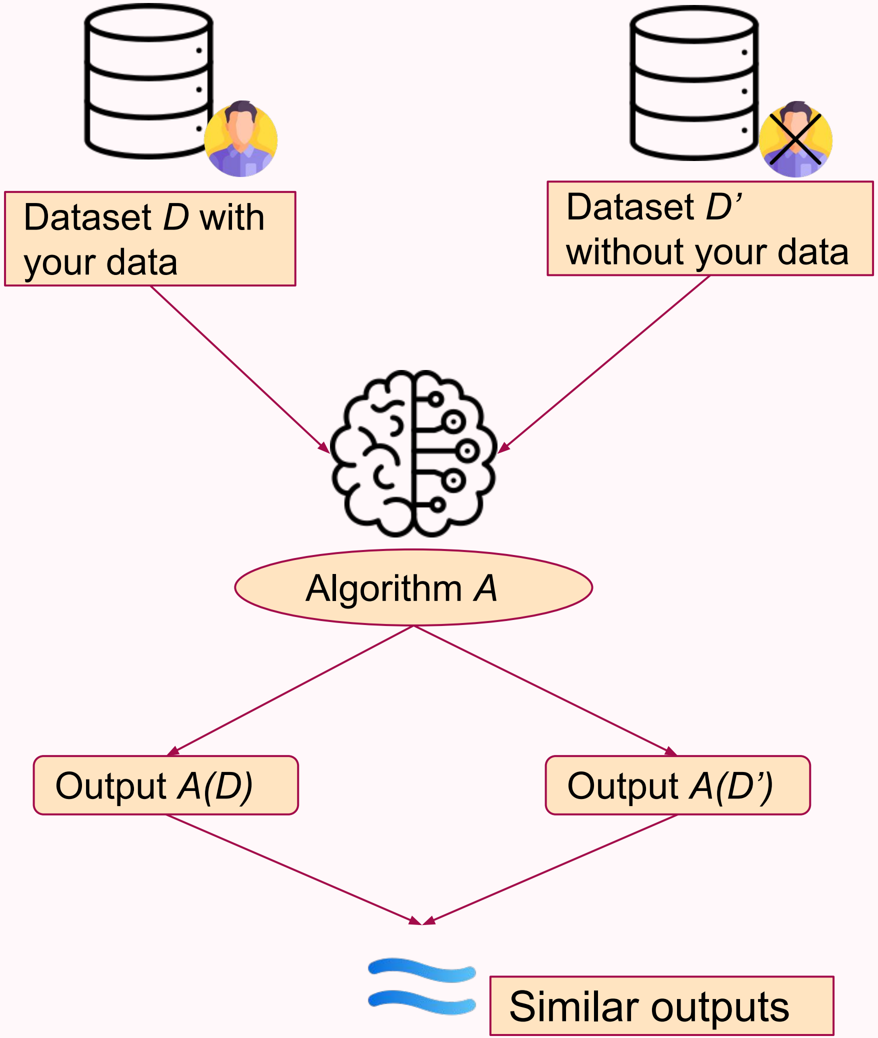
Example: Original Data

Gender	Age	Zip	Label
M	[20-25]	121*	0
M	[20-25]	121*	1
F	[30-50]	123*	1
F	[30-50]	123*	0

K=2 Anonymous Data

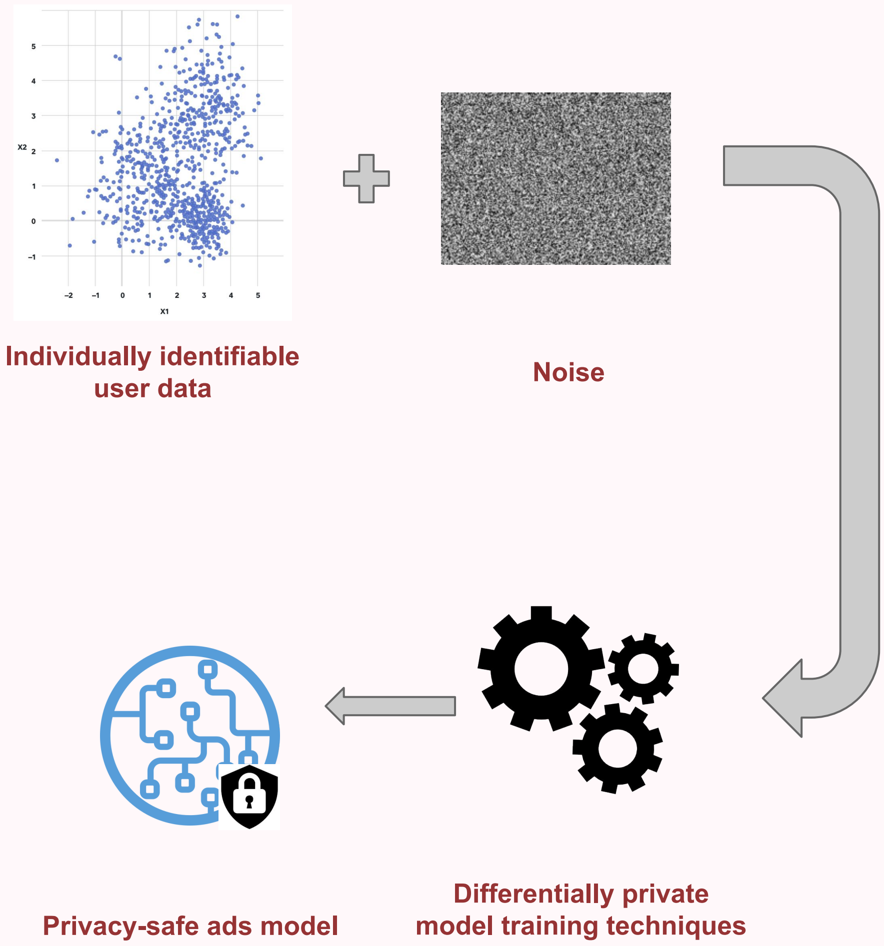
Introducing Privacy by Default: Differential Privacy

- Strong, mathematical definition of privacy in the context of statistical analysis and machine learning.
- DP guarantees that inferences are indistinguishable whether or not a single individual’s private information was or was not in the analysis input.



Introducing Privacy by Default: Differential Privacy

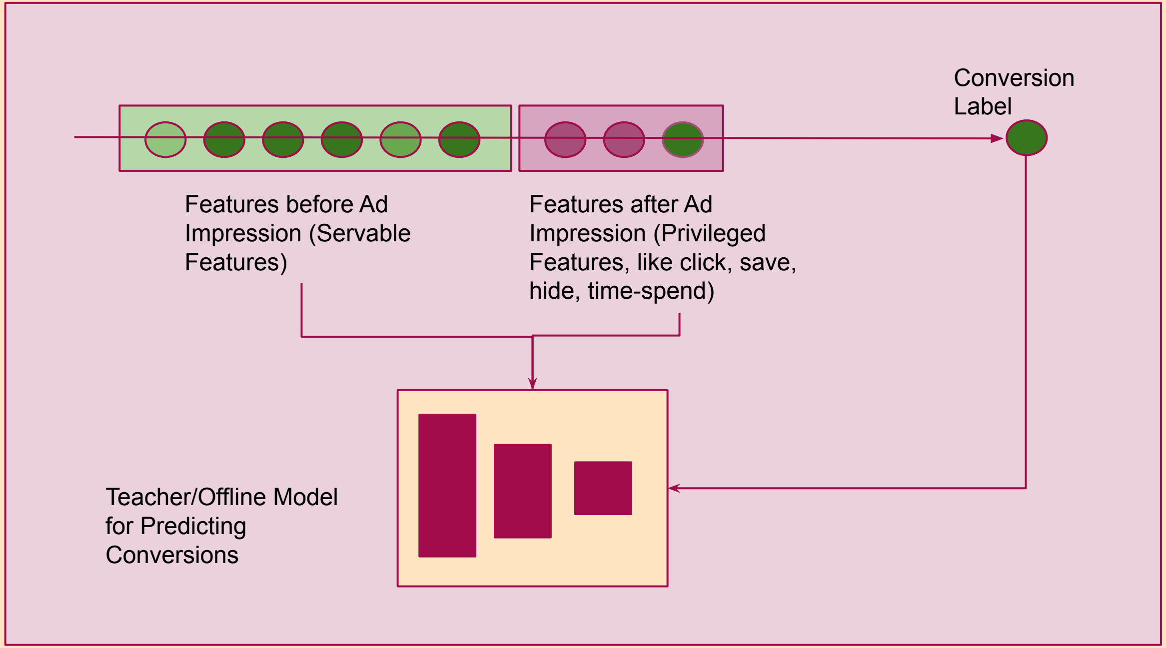
- DP uses randomization (injecting noise) to prevent privacy attacks.
- The noise level of DP is computed based on the privacy budget (epsilon) we have. Larger epsilon means we have more privacy budget.
- DP is robust to cumulative risk from successive data releases



Acknowledgement

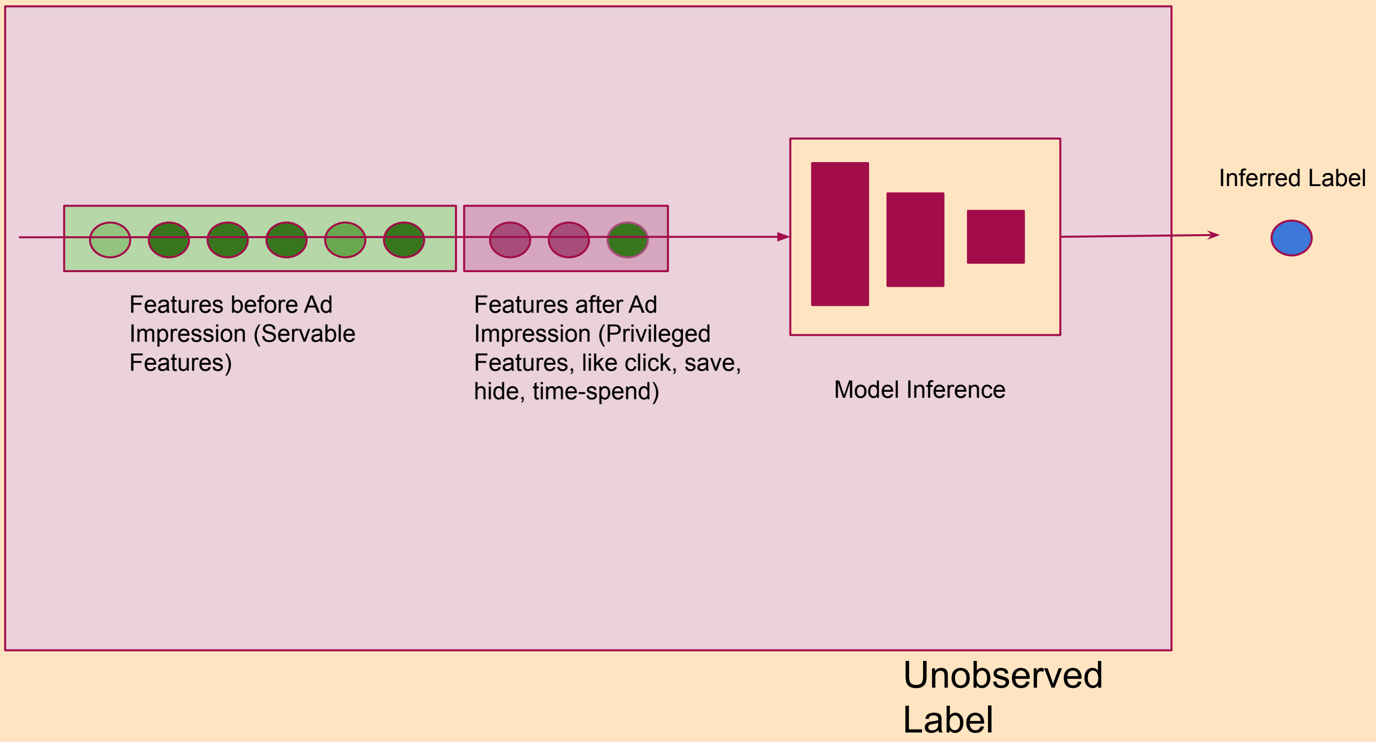
Thanks for XFN collaboration with entire Ads Quality, Ads Infra, Ads Measurement, Advanced Technology Group, Advertiser Solution Group, Content and User Engineering teams

Post Click Featurization as Synthetic Data Generator



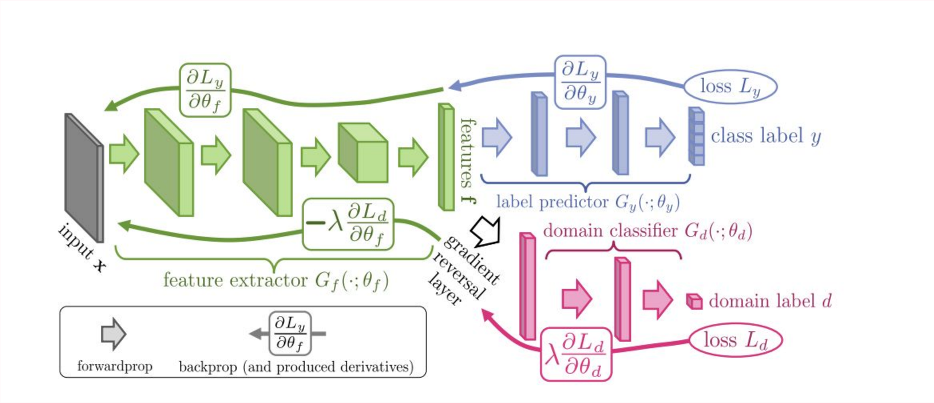
Observed Labels

Post Click Featurization as Synthetic Data Generator



Industry Standard ways to Improve Personalization

- Post-Click Distillation as Synthetic Label Generator
- Domain Adaptation Application to reduce bias between training and serving population
- Ensemble Models to Improve Robustness
- Adding Differential Privacy by Default



Reference: [Domain-Adversarial Training of Neural Networks]