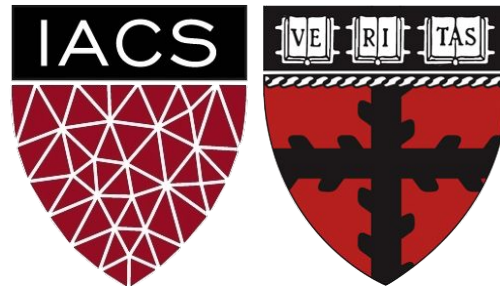


# Privacy Report Card - Project Outline

Advanced Practical Data Science, MLOps

AC295

Chad Stoughton & Christopher Lee



# Outline

---

- Project Scope
- Project Workflow
- Process Flow
- Data
- Models

# Problem Definition

---

Chris is tech-savvy in a tech-oriented world. He likes to receive personalized services. Recommendations based on his preferences, alerts on his spending habits, and discounts for his favorite brands are just a few of the ways he uses tech.

There is just one thing that keeps Chris awake at night - he shares his personal data to receive these personalized services without actually knowing what happens to his data. Usually, he just agrees to the terms and conditions posed by the company providing a service he wants to get.

Would be possible to get a quick and accurate grasp on how the companies he deals with uses his personal data? There is no way he would be able or willing to read through each privacy policy of all the companies he deals with!

# Proposed Solution

---







Chris found an app that examines any privacy policy and generates a report card on the way the corresponding company handles his personal data.

If a company collects his location data, the report card will tell him that his location data may be at risk. If a company shares his data with third party companies, then the report card notifies him with a red-labeled warning, indicating his data may end up under the control of companies he has never heard of.

What if this could become the norm that all companies expect? Wouldn't we see more companies becoming examples of success from being ethical and responsible in their way of business? Wouldn't we feel more secure and confident in deciding to share our data? Wouldn't we be able to use our personal data better for the purposes we want to use them for?

# Proposed Solution

- Chris provides a URL to a privacy policy
- A paragraph in that privacy policy indicates his personal data will be shared with third party partners purely for marketing purposes
- Chris doesn't like being bombarded with advertisements
- We will build an app to identify risks based on personal preference within the text of a privacy policy
- If a user is unknowingly about to share his personal data with a third party he'd rather not share with, then the analysis provided by the app will not only let him know he may not want to deal with that company, but it will tell him exactly where in the privacy policy the questionable statement is and what it says.
- This means even businesses can tell whether another business is responsible and handles their customers' personal data ethically.
- This app can raise the bar on the level of trust citizens of the world are accustomed to before consenting to share their data.

Uses trackers	 FINAL GRADE	<b>TWENTY</b> OUT OF TWENTY STANDARDS MET, INCLUDING: EFFECTIVENESS POLICY PROGRAM EXPENSES TRUTHFUL MATERIALS ANNUAL REPORT WEBSITE DISCLOSURES DONOR PRIVACY AUDIT REPORT BUDGET PLAN FUNDRAISING EXPENSES
Shares your data /w others	 FINAL GRADE	<b>ONE HUNDRED</b> PERCENT IN ACCOUNTABILITY & TRANSPARENCY. 98% OVERALL RANKED 4 OUT OF 4 STARS
Collects your location data	 FINAL GRADE	<b>FIVE</b> PERCENT OF CHARITIES IN THE U.S. RECEIVE "SEAL OF EXCELLENCE." AWARDED SEAL FOR MEETING THE HIGHEST STANDARDS OF PUBLIC ACCOUNTABILITY, PROGRAM EFFECTIVENESS AND COST-EFFECTIVENESS. FEWER THAN FIVE PERCENT MEET OR EXCEED THESE STANDARDS.
Collects your demographics	 FINAL GRADE	<b>ONE HUNDRED AND TWO</b> 5-STAR REVIEWS OUT OF 102 REVIEWS TOTAL. 5 OUT OF 5 STARS
Collects your contacts data	 FINAL GRADE	<b>SECOND</b> PLACE IN SMALL COMPANY CATEGORY 2016 TOP 10 IN SMALL COMPANY CATEGORY 2017 10-49 EMPLOYEES
Uses SSO	 FINAL GRADE	<b>PLATINUM</b> PARTICIPANT

# Project Scope



## Proof Of Concept (POC)

- Scrape privacy policy texts/verify resulting dataset is as expected
- Train baseline models purposed for both fine-tuning and classification
- Check effectiveness of model training:
  - Fine-tuned models: Generate text and check whether fine-tuning is working (our model should generate text resemblant of privacy policies)
- Classification models: Verify the predicted classifications agree with our expectations on a validation dataset

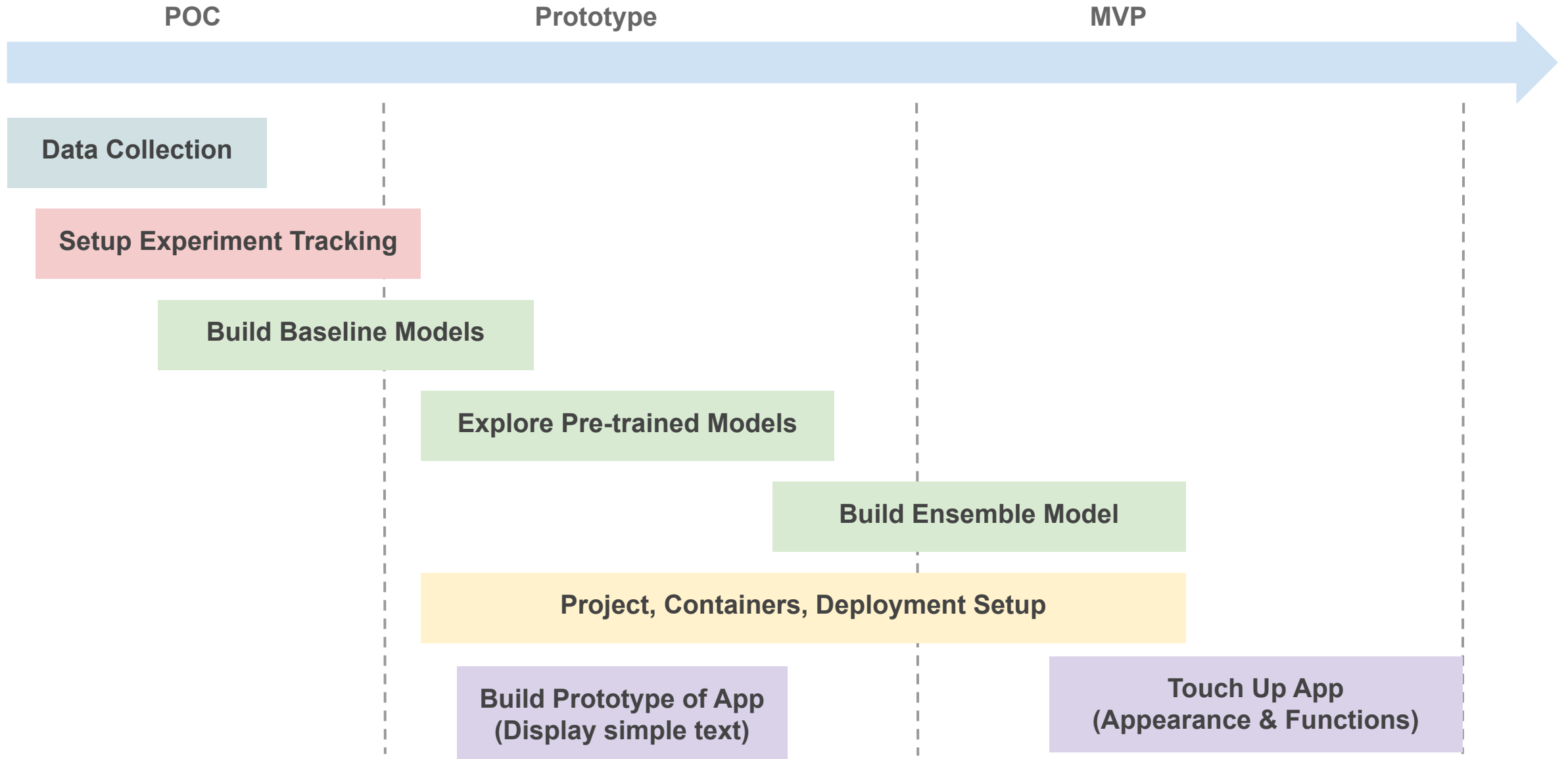
## Prototype

- Deploy one model to Fast API to service model predictions as an API
- Create a demo app using Flutter that:
  - Displays a screen for inputting a url to a privacy policy
  - Prepares a request body containing the URL info to send to our API
  - Calls our API in a POST request with the URL as a request body parameter
  - Waits for a response from our API, and upon receiving it, parses the result
  - Generates a rudimentary report card and displays it on the app screen

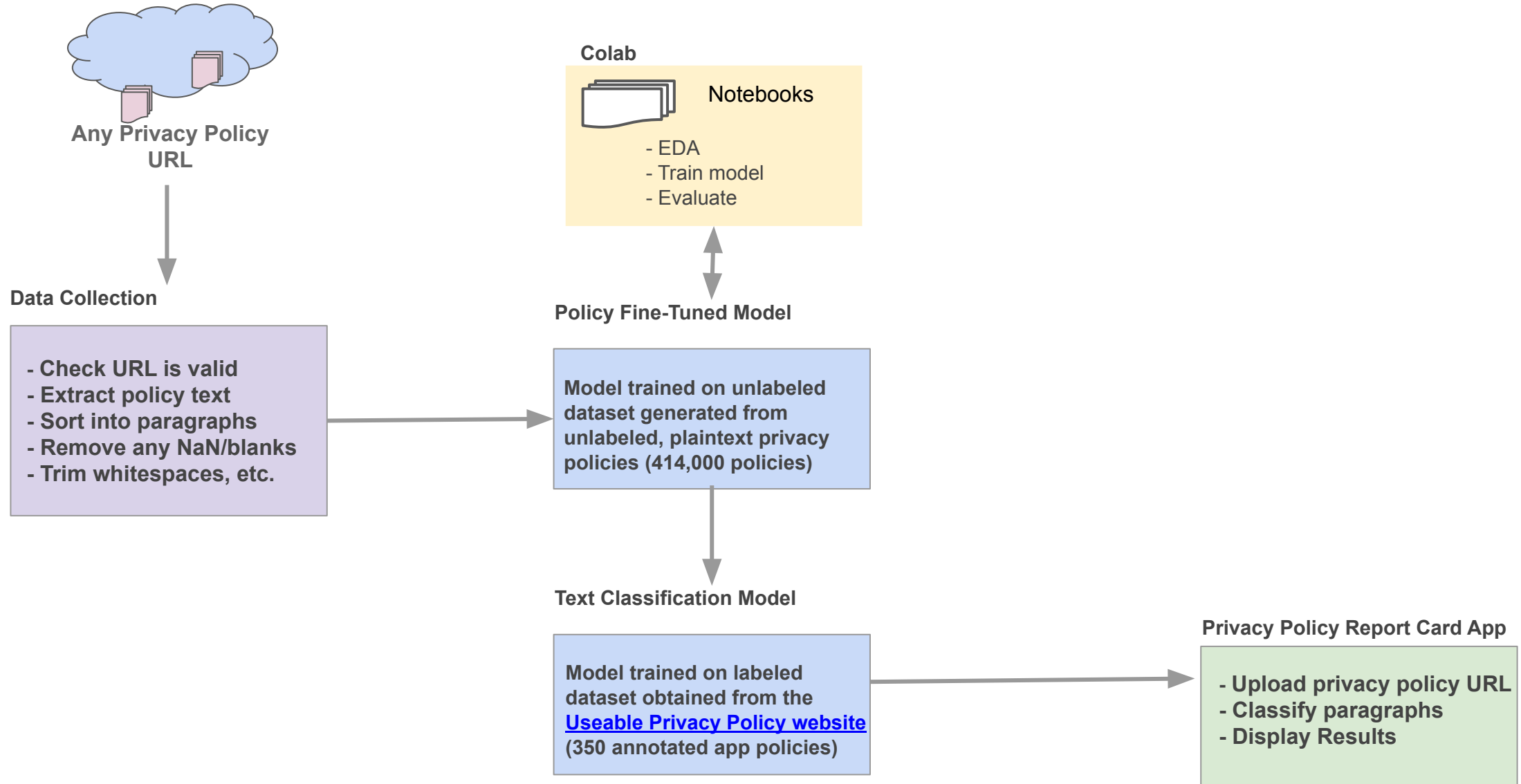
## Minimum Viable Product (MVP)

- Create App that takes an URL to a privacy policy as an input, examines said policy, and generates a report card indicating risks of personal data exposure/mis-handling
- API Server for uploading privacy policy URLs and predicting using best model
- Web app that can take a URL string as input, sends to our API server, and displays the response data by parsing it into a **Privacy Policy Report Card**

# Project Workflow



# Process Flow





# Data

## Example text of paragraphs extracted from our dataset of approximately 414,000 privacy policies

Text: Information that we collect are "NETWORK STATUS INFORMATION", "WIFI STATUS INFOR

Text: Collecting User Information

Text: Samuel J or Eznetsoft uses remarketing with Google AdWords and analytics to disp

Text: 8. DISCLAIMER OF WARRANTIES

Text:

FamilySearch Terms of Use (Updated 2021-09-27) | Privacy Notice (Updated 2021-04-06)

## Example of annotated data from our labeled, annotated dataset of 350 Android App privacy policies

	policy_id	policy_name	segment_id	segment_text	3RD	LOCATION	DEMOGRAPHIC	CONTACT	IDENTIFIER	SSO
4305	243	io.utk.android	2	Passwords The UTK.io staff will NEVER ask for ...	False	False	False	False	False	False
4739	17	com.atomicadd.fotos	11	Location information When you use AtomicAdd se...	False	True	False	False	True	False
7824	98	Xender	7	(d) We may collect and use such data for promo...	False	False	False	False	False	False
6871	336	Viber	39	Here are a few additional important things you...	False	False	False	False	True	False
12536	164	com.eharmony	10	Purchase Information. To process purchases, we...	False	False	False	True	False	False

# Data Details

---

- Total number of unlabeled privacy policies: Approximately 414,000
  - Initial data is provided as list of URLs
  - Each URL is scraped and parsed to generate a list of extracted text by unit of paragraph
  -
- Total number of labeled annotated privacy policies: 350
  - Number of Category Labels: 6
    - 'IDENTIFIER', '3RD', 'LOCATION', 'DEMOGRAPHIC', 'CONTACT', 'SSO'
  - Number of Annotation Labels: 60
    - Two generic, 29 first-party, and 29 third-party related indicators
  - Unit of Annotated Text: Paragraph

# Models

---

## **Fine-Tuning Task:**

- GPT-2
- GPT-3 (fine-tuned via OpenAI's API)
- BERT

## **Classification Task:**

- Simple CNN model With Embeddings
- BERT
- GPT-2
- GPT-3 (predictions obtained via OpenAI's API)

# Models - Training Results

## Fine-Tuning Model Text Generation Results:

### GENERATED TEXT FROM **PRIVACY POLICY** FINE-TUNED MODEL

"Your location data will be used to set the time of your request. \nWhat happens when you send an email to a Facebook Group when the service has been activated, or when your account has been activated. \nYou may leave an account with your password, so that the information that you request is more freely available to you. \nYou may set the time of..."

### GENERATED TEXT FROM **COVID** FINE-TUNED MODEL

"Your location data can save lives," Gerow says. "If there is not a change in the plan, then the only way to avoid it is to have people log in and make a schedule." You can also make a schedule using home data, such as your birthday and your school's total. You can also make the decision to save money or move forward without having..."

## Classification Model Confusion Matrix Results

```
1 confusion_matrix(y_testM, model.predict(x_test)>.5)
```

```
array([[3206,  226],  
       [  47,  398]])
```