DATA PRIVACY
FINAL PRESENTATION

Aryan Jain, Aura Ulloa Ordonez, Ishaan Narain

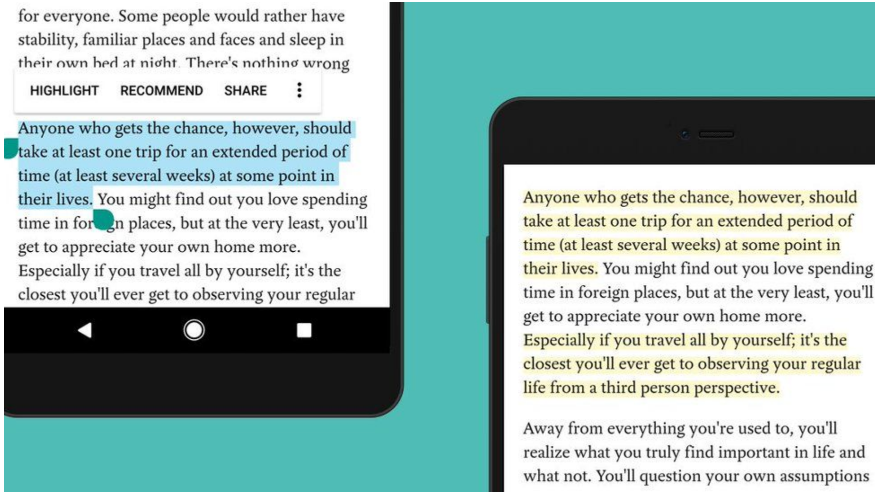
AGENDA
ITEMS

- **MOTIVATION & GOAL**
- **METHODOLOGY & EXAMPLES**
  FRONT-END & BACK-END
- **DEMO & EVALUATION**
- **FINAL THOUGHTS**

Project
MOTIVATION

- Dangers of posting sensitive information grow stronger
- 9% of Americans experience Doxing
- Identity Theft doubled from 2019 to 2020
- 70% of employers use social media to screen candidates

> “**87%** OF **THE US POPULATION IS IDENTIFIABLE** FROM A SET OF INFORMATION CONTAINING **AGE**, **GENDER**, AND **ZIP CODE**.”

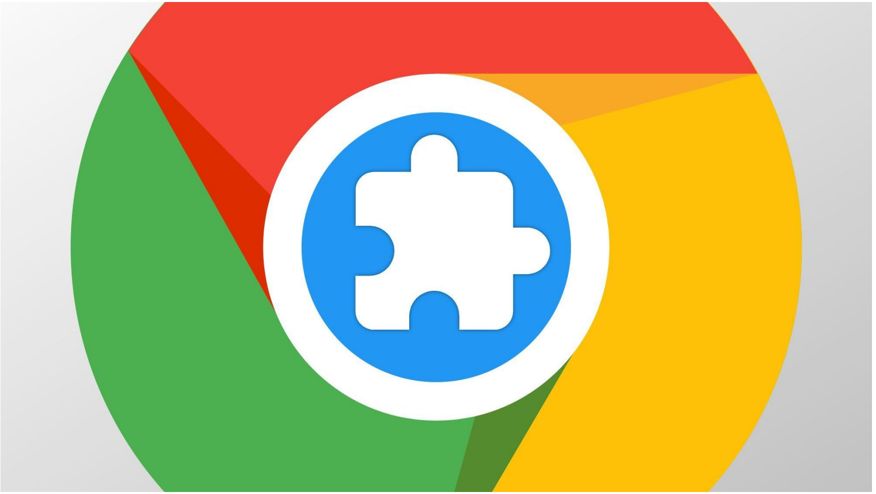


for everyone. Some people would rather have stability, familiar places and faces and sleep in their own bed at night. There's nothing wrong

Anyone who gets the chance, however, should take at least one trip for an extended period of time (at least several weeks) at some point in their lives. You might find out you love spending time in foreign places, but at the very least, you'll get to appreciate your own home more. Especially if you travel all by yourself; it's the closest you'll ever get to observing your regular

Anyone who gets the chance, however, should take at least one trip for an extended period of time (at least several weeks) at some point in their lives. You might find out you love spending time in foreign places, but at the very least, you'll get to appreciate your own home more. Especially if you travel all by yourself; it's the closest you'll ever get to observing your regular life from a third person perspective.

Away from everything you're used to, you'll realize what you truly find important in life and what not. You'll question your own assumptions



Project **GOAL**

This project aims to protect social media users, especially young people, from posting sensitive text-based content and information to avoid safety issues such as those highlighted before.

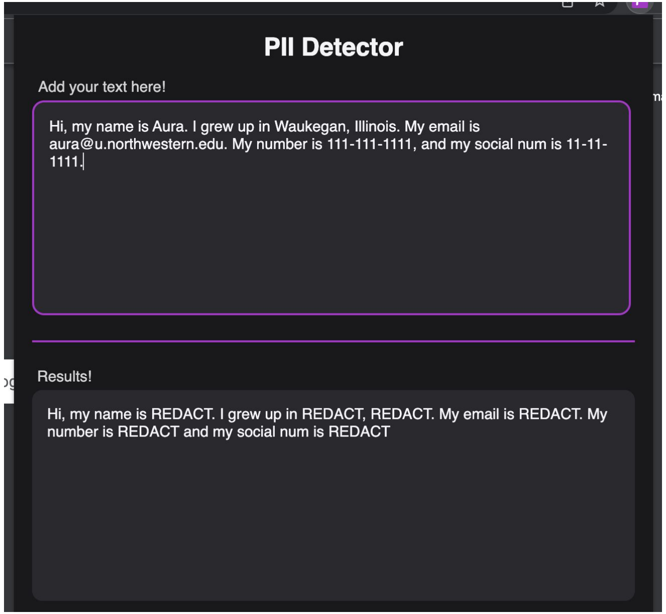
PROJECT **METHODOLOGY**

FRONT END

- Learn the HTML/CSS/Javascript basics needed to write a browser extension
- Python with Flask
- User testing our browser extension to see if users understand its purpose

Project Methodology
FRONT-END

- Users can paste comments and see the modified comments
- Our backend is now a locally run REST API
- Fetch request an analysis of the user's text
- Removed button since feedback is given in real-time



PII Detector

Add your text here!

Hi, my name is Aura. I grew up in Waukegan, Illinois. My email is aura@u.northwestern.edu. My number is 111-111-1111, and my social num is 11-11-1111.

Results!

Hi, my name is REDACT. I grew up in REDACT, REDACT. My email is REDACT. My number is REDACT and my social num is REDACT

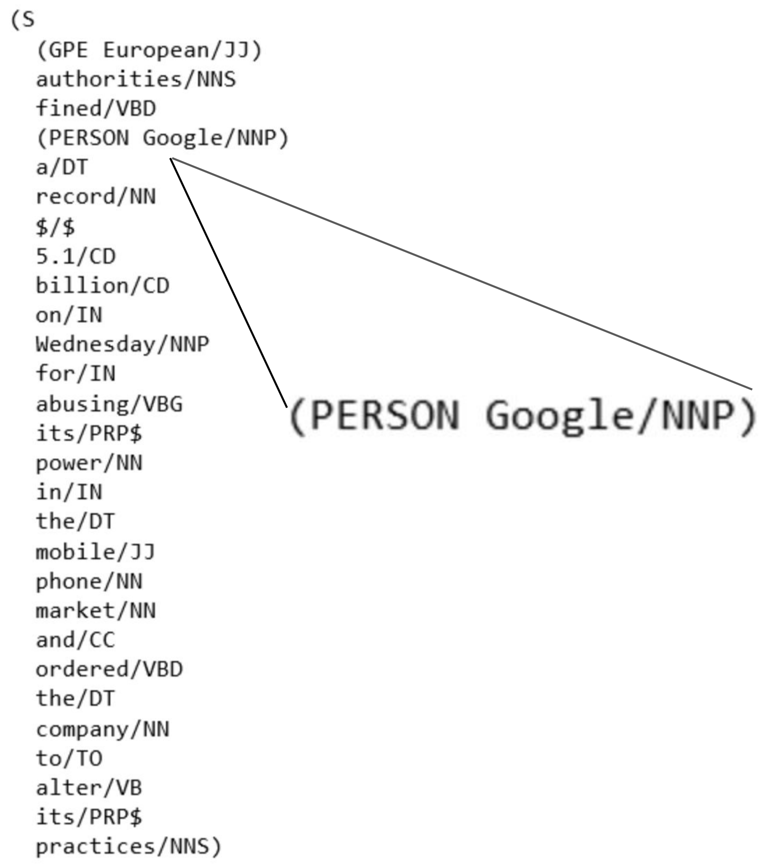
PROJECT **METHODOLOGY**

BACK END

- Develop a language model to characterize and redact private and sensitive data
  - Take in a given text string and remove any key sensitive information according to the model
  - Function should remove PII including personal names, locations, phone numbers, zipcodes

Project Methodology
BACK-END

- Leveraged Named Entity Recognition with the NLTK package
  - Process:
    - Tokenization
    - Chunk Patternization
    - Parse through a classifier to identify named entities
- Utilized regex (Regular Expression) indicators
  - Identify phone numbers, zip-codes, and other sequences of text



```
(S
  (GPE European/JJ)
  authorities/NNS
  fined/VBD
  (PERSON Google/NNP)
  a/DT
  record/NN
  $/$
  5.1/CD
  billion/CD
  on/IN
  Wednesday/NNP
  for/IN
  abusing/VBG
  its/PRP$
  power/NN
  in/IN
  the/DT
  mobile/JJ
  phone/NN
  market/NN
  and/CC
  ordered/VBD
  the/DT
  company/NN
  to/TO
  alter/VB
  its/PRP$
  practices/NNS)
```

```
(PERSON Google/NNP)
```

Example Cases

*“ My name is **Mike**. I grew up in **Evanston, IL**, **60201**, and my house phone number is **3212559002**. Feel free to come over and check out my **3** sailing trophies.*

Example Cases

“ My name is ***REDACT****. I grew up in* ***REDACT****,* ***REDACT****, and my house phone number is* ***REDACT****. Feel free to come over and check out my* ***REDACT*** *sailing trophies.*

Example Cases

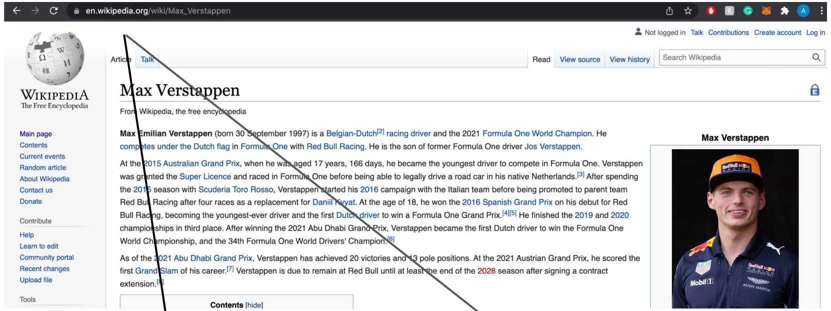
*“ My name is **Mike**. I grew up in **Evanston, IL**, **60201**, and my house phone number is **3212559002**. Feel free to come over and check out my 3 sailing trophies.*

Example Cases

*“ My name is **REDACT**. I grew up in **REDACT**, **REDACT**, and my house phone number is **REDACT**. Feel free to come over and check out my 3 sailing trophies.*

Project Methodology
BACK-END

- Leveraged Python requests library
  - Separates celebrity and famous names from personal identifying names
  - Limited to celebrity full names
  - Checked celebrity status based on presence of wikipedia page



Max Verstappen

From Wikipedia, the free encyclopedia

Max Emilian Verstappen (born 30 September 1997) is a Belgian-Dutch[2] racing driver and the 2021 Formula One World Champion. He competes under the Dutch flag in Formula One with Red Bull Racing. He is the son of former Formula One driver Jos Verstappen.

At the 2015 Australian Grand Prix, when he was aged 17 years, 166 days, he became the youngest driver to compete in Formula One. Verstappen was granted the Super Licence and raced in Formula One before being able to legally drive a road car in his native Netherlands.[3] After spending the 2015 season with Scuderia Toro Rosso, Verstappen started his 2016 campaign with the Italian team before being promoted to parent team Red Bull Racing after four races as a replacement for Daniil Kvyat. At the age of 18, he won the 2016 Spanish Grand Prix on his debut for Red Bull Racing, becoming the youngest-ever driver and the first Dutch driver to win a Formula One Grand Prix.[4][5] He finished the 2019 and 2020 championships in third place. After winning the 2021 Abu Dhabi Grand Prix, Verstappen became the first Dutch driver to win the Formula One World Championship, and the 34th Formula One World Drivers' Champion.[6]

As of the 2021 Abu Dhabi Grand Prix, Verstappen has achieved 20 victories and 13 pole positions. At the 2021 Austrian Grand Prix, he scored the first Grand Slam of his career.[7] Verstappen is due to remain at Red Bull until at least the end of the 2028 season after signing a contract extension.[8]

Contents [hide]

Max Verstappen

https://en.wikipedia.org/wiki/Max_Verstappen

200 | 404

Example Cases

“ My name is ***Mike****. I grew up in* ***Evanston, IL****, and I love* ***Max Verstappen****.”*

Example Cases

*“ My name is **REDACT**. I grew up in **REDACT**, and I love **REDACT**.”*

Example Cases

“ My name is ***Mike****. I grew up in* ***Evanston, IL****, and I love Max Verstappen.”*

Example Cases

*“ My name is **REDACT**. I grew up in **REDACT**, and I love Max Verstappen.”*

PROJECT
DEMO

User Testing

- Test and break on my own
- User testing for intuitiveness
    - Write a greeting
    - A potential tweet they'd write
    - A little about themselves

User Testing

- + Simple to use
- + Real-time/fast
- + Can use anywhere as an extension
- + Good at identifying numbers
- - Only REDACT (make it clearer)
- - Give reason and alternative to word
- - Stronger detection with typos and different writing styles
- - Better with cities and towns

Evaluation of Tool

Cultural names not being identified as easily as more Western names

Context - names, and locations may not always be personal information

Evaluation of Tool

“ My name is ***Ishaan (ORG)*** *and my roomates names are* ***Aryan (RELIGION)*** *and* ***Mike (PERSON)*** *”*

Evaluation of Tool

“ My name is Ishaan and my roomates names are Aryan and Mike ”

PROJECT **ISSUES**

- Named Entity Recognition would highlight and redact famous people’s names
    - Applied Wikipedia API to remove these names but limited to full names
    - Biases in NLP package only highlight non-POC names
- Characterizing gender within language
- Separating numbers denoting age and otherwise

Broader Implications of Tool

- Targeted for those with growing knowledge towards technology and web privacy
- Tool can be used to ensure that PIIs are not put out mistakenly on social media protecting users
- Not limited to certain websites as accessible on browser extension
- Should social media websites integrate this functionality themselves?

FUTURE **TIMELINE**

FINAL Deliverable: **03/16**

- Improve redux expressions for phone numbers
- Write report
- ***NICE TO HAVE***
  - *Add features to allow for altering the text to change private sensitive information with a higher class of knowledge*
  - *Give users reasons why certain words are redacted*
  - *Highlight redacted words in the edit panel for users to write alternatives for (writing assistance-esque)*
  - *Copy/paste button on results panel*

THANK YOU
& QUESTIONS

Evaluation of Tool

“ My name is Mike (PERSON) and my roomates names is Stefan (PERSON). If you want to reach out to me, contact me at 332-255-9002 (PHONE) and ilovestefan@gmail.com (EMAIL), or come to Evanston, IL 60201 (GPE).

“ My name is REDACT and my roomates names is REDACT. If you want to reach out to me, contact me at REDACT and REDACT, or come to REDACT.

Evaluation of Tool

*“ My name is **Ishaan (ORG)** and my roomates names are **Aryan (RELIGION)** and **Mike (PERSON)** ”*

*“ My name is **Ishaan** and my roomates names are **Aryan** and **REDACTED**”*