Dark Patterns Buster Hackathon (DPBH-2023)

Round 3 Grand Finale @ IIT(BHU), Varanasi
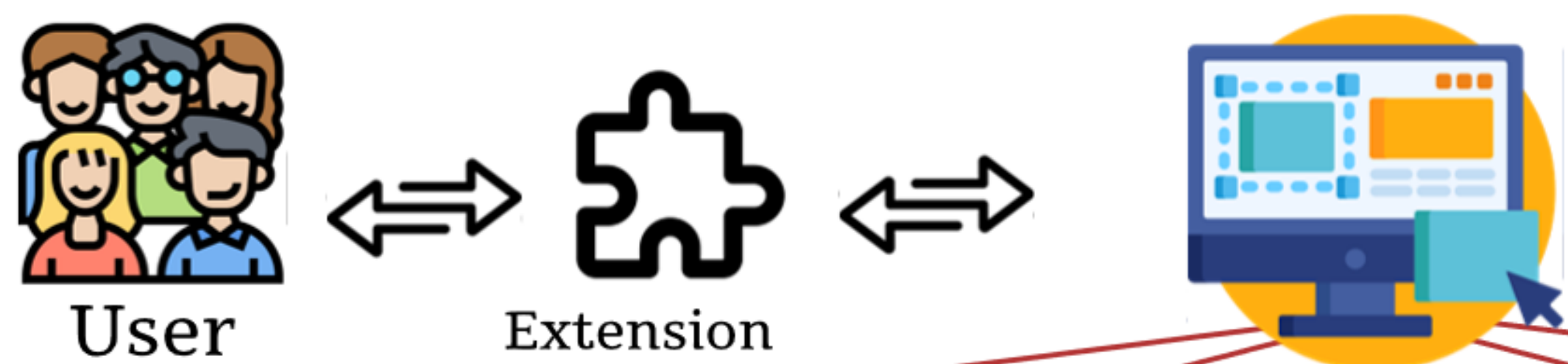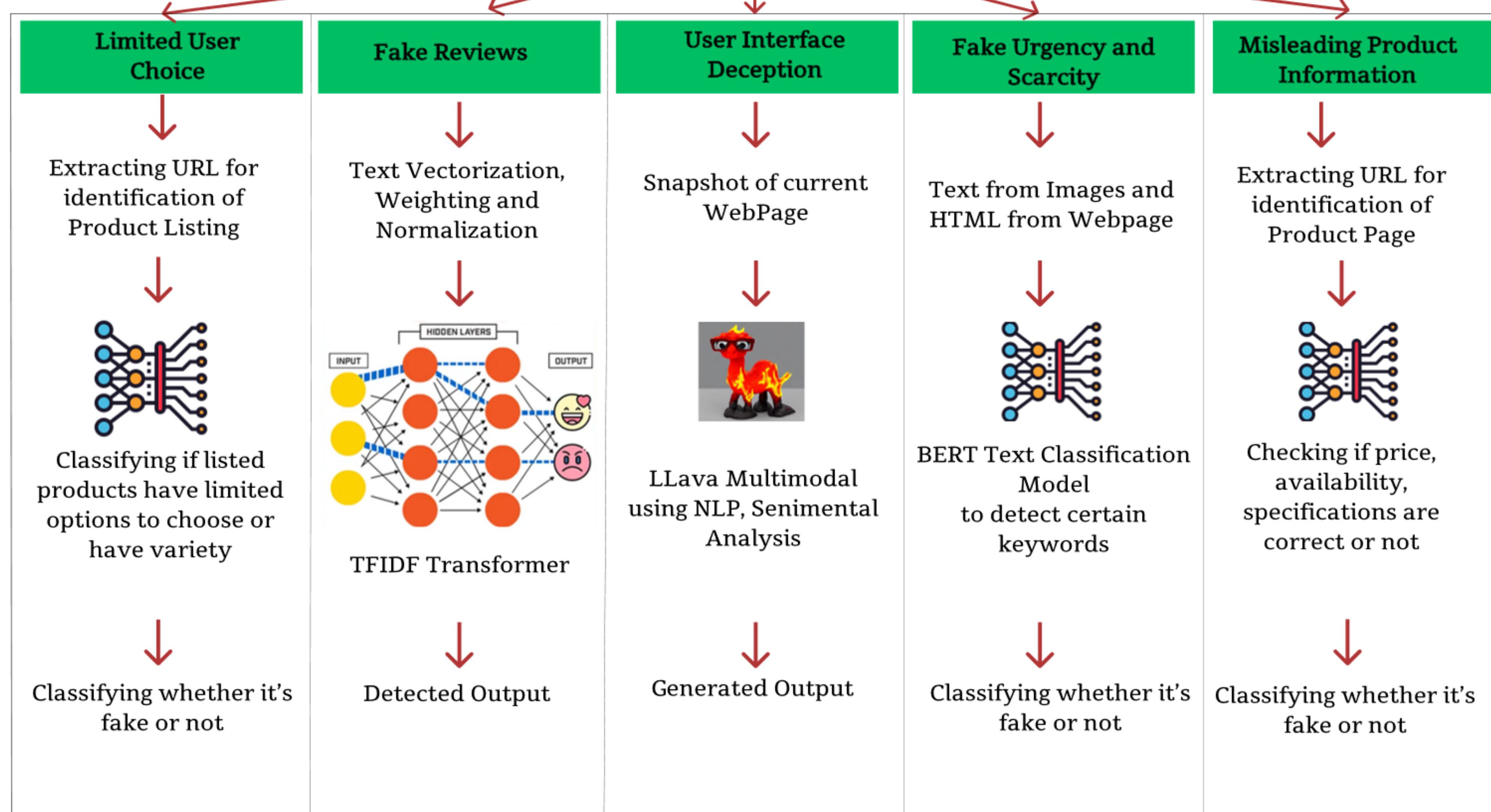
(Feb.17, 2024)

LOGIC GATES

Maulana Azad National Institute of Technology

Achyuta Gupta
Atharva Kumar
Rajneesh Kushwaha
Prabal Pratap Singh
Aniket Jain

clearInfo Web Extension

User ⇔ Extension ⇔

Limited User Choice	Fake Reviews	User Interface Deception	Fake Urgency and Scarcity	Misleading Product Information
Extracting URL for identification of Product Listing	Text Vectorization, Weighting and Normalization	Snapshot of current WebPage	Text from Images and HTML from Webpage	Extracting URL for identification of Product Page
Classifying if listed products have limited options to choose or have variety	TFIDF Transformer	LLava Multimodal using NLP, Senimental Analysis	BERT Text Classification Model to detect certain keywords	Checking if price, availability, specifications are correct or not
Classifying whether it's fake or not	Detected Output	Generated Output	Classifying whether it's fake or not	Classifying whether it's fake or not

Salient Features

- Efficient Webpage Capture
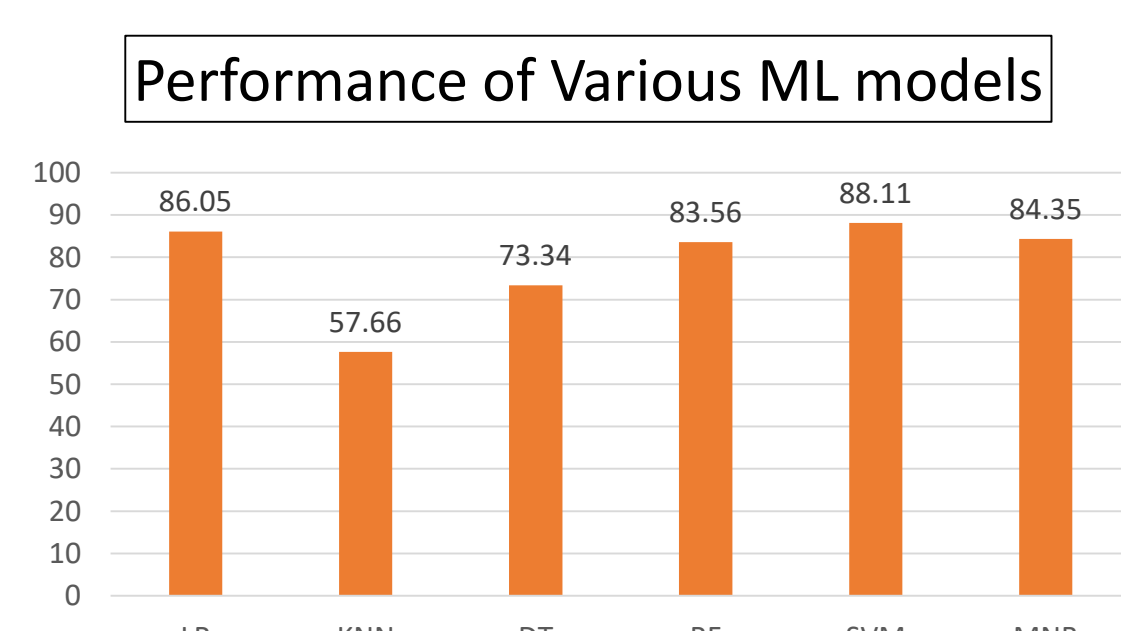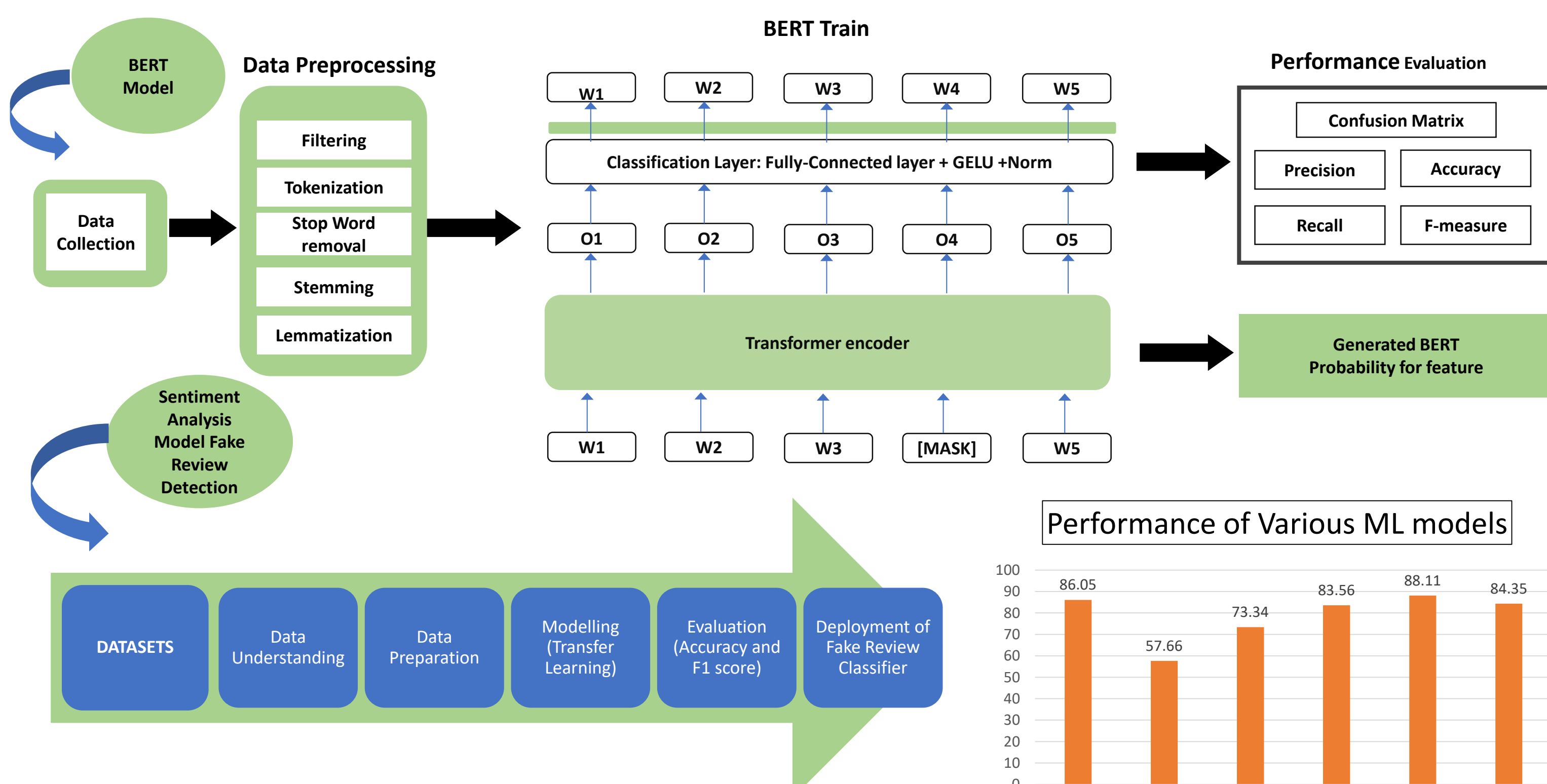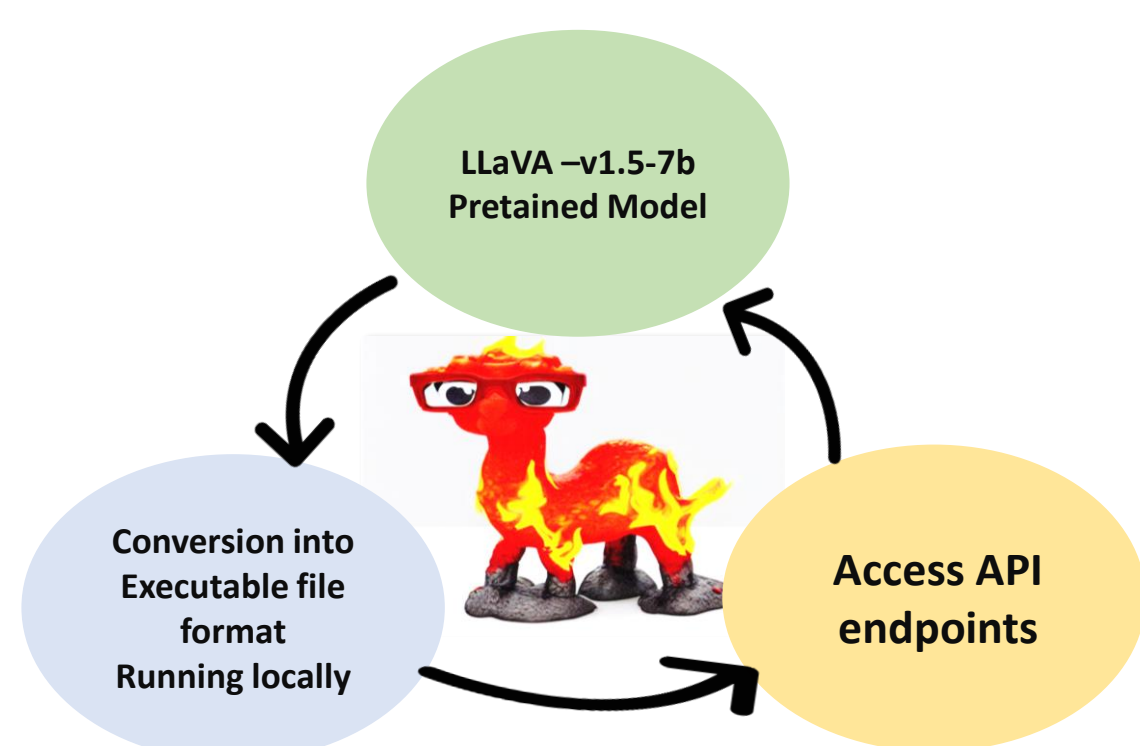- User-Friendly Automation
- User Review Integration

Model Used/ Results

- Start with the LLaVA pretrained model
- Convert the model to a ".lammafile.exe" format
- Execute the model locally on port 8080
- Access API endpoints exposed by the local execution
- Integrate the model into a browser extension
- Allow users to interact with the model via extension

LLaVA –v1.5-7b Pretained Model

Conversion into Executable file format Running locally

Access API endpoints

BERT Model

Data Collection

Data Preprocessing
- Filtering
- Tokenization
- Stop Word removal
- Stemming
- Lemmatization

BERT Train

W1 W2 W3 W4 W5

Classification Layer: Fully-Connected layer + GELU +Norm

O1 O2 O3 O4 O5

Transformer encoder

W1 W2 W3 [MASK] W5

Performance Evaluation

Confusion Matrix	
Precision	Accuracy
Recall	F-measure

Generated BERT Probability for feature

Sentiment Analysis Model Fake Review Detection

DATASETS → Data Understanding → Data Preparation → Modelling (Transfer Learning) → Evaluation (Accuracy and F1 score) → Deployment of Fake Review Classifier

Performance of Various ML models

Model	Value
LR	86.05
KNN	57.66
DT	73.34
RF	83.56
SVM	88.11
MNB	84.35

Some Results

clearInfo

Dark Patterns found in current page:

✓ UI Deceptions ▼
✓ Limited Choices ▼
✓ Misleading Infomations ▼
✓ Urgency & Scarcity Keywords ▼
✓ Fake Reviews ▼

Send Report

clearInfo

Dark Patterns found in current page:

UI Deceptions ▼
Not Found

Do you find this response useful ? 👍 👎

clearInfo

Dark Patterns found in current page:

✓ UI Deceptions ▼

There is a dark pattern on this Webpage. The "Continue" button is green, which is a color that is often used to indicate that an action is safe or positive. However, in this case, the "Continue" button actually leads to a page where the user is asked to sign up for a paid subscription. This is a deceptive practice because it tricks the user into thinking that they are simply continuing to the next page, when in reality they are being asked to sign up for a paid subscription.

Do you find this response useful ? 👍 👎

✓ Limited Choices ▼
✓ Misleading Infomations ▼
✓ Urgency & Scarcity Keywords ▼
✓ Fake Reviews ▼

Send Report

Contact:
Achyuta Gupta
achyutagupta888rg@gmail.com

Dr. N. S Rajput
Convener, DPBH-2023
Email: dpbh2023@iitbhu.ac.in

Abstract

ClearInfo is a Chrome extension designed to empower users by detecting and reporting manipulative design practices known as "dark patterns" on websites.

It identifies five key dark patterns:

- **UI Deception**: Leveraging the LLava language model, ClearInfo uncovers misleading visual elements and interfaces.
- **Scarcity & Urgency**: Utilizing the BERT model, it detects tactics that pressure users into quick decisions based on false limitations.
- **Misleading Product Info**: By extracting URLs and cross-referencing information, ClearInfo exposes inaccurate or exaggerated product claims.
- **Limited User Choice**: Through URL parsing, the extension reveals interfaces that restrict user options and steer them towards specific actions.
- **Fake Reviews**: Employing the TF-IDF transformer, ClearInfo identifies suspicious review patterns suggestive of inauthenticity.

Key Features:

- Efficient Webpage Capture: Captures both visible and hidden page elements for comprehensive analysis.
- User Review Integration: Incorporates user reviews to provide a broader context for the identified dark patterns.
- User-Friendly Automation: Simplifies the process of detecting and reporting dark patterns, empowering users to contribute to a more ethical online environment.

Team Members

Achyuta Gupta
Web Developer, Build the Frontend of Extension and Backend

Atharva Kumar
ML & Figma, poster, designing & implemented basic working of ML Model

Rajneesh Kushwaha
Web Development, Implemented Fake Reviews ML Model

Prabal Pratap Singh
AI and ML, Implemented Fake Reviews ML Model

Aniket Jain
Backend & ML, Backend Implementation & Fine tuning

Impact Assessment & Conclusion

ClearInfo is like a guardian for your e-commerce browsing. It spots sneaky tricks on websites that could trick you. With its clever tools, it uncovers main types of tricks, such as fake reviews and misleading information. ClearInfo makes it simple for you to understand what's happening online, empowering you to make smarter decisions. It's like having a companion who watches your back while you explore the internet, ensuring a safer and more reliable online experience.

Solution Highlights

ClearInfo stands out for its user-friendly features. It efficiently captures webpage details, even hidden ones, ensuring a thorough analysis. By integrating user reviews, it offers a broader perspective on detected dark patterns. Its simplicity makes it easy for users to detect and report deceptive practices, fostering a more transparent online environment. With ClearInfo, users gain the power to make informed choices and contribute to a safer internet for all.