BOLSTER

Blog / Tips & Guides / Research Labs
Research

# Empowering Brand Protection: Large Language Models and Their Role in Webpage Intent Identification for Brand Safety

September 12, 2023 | 6 min read

Adithya Singh

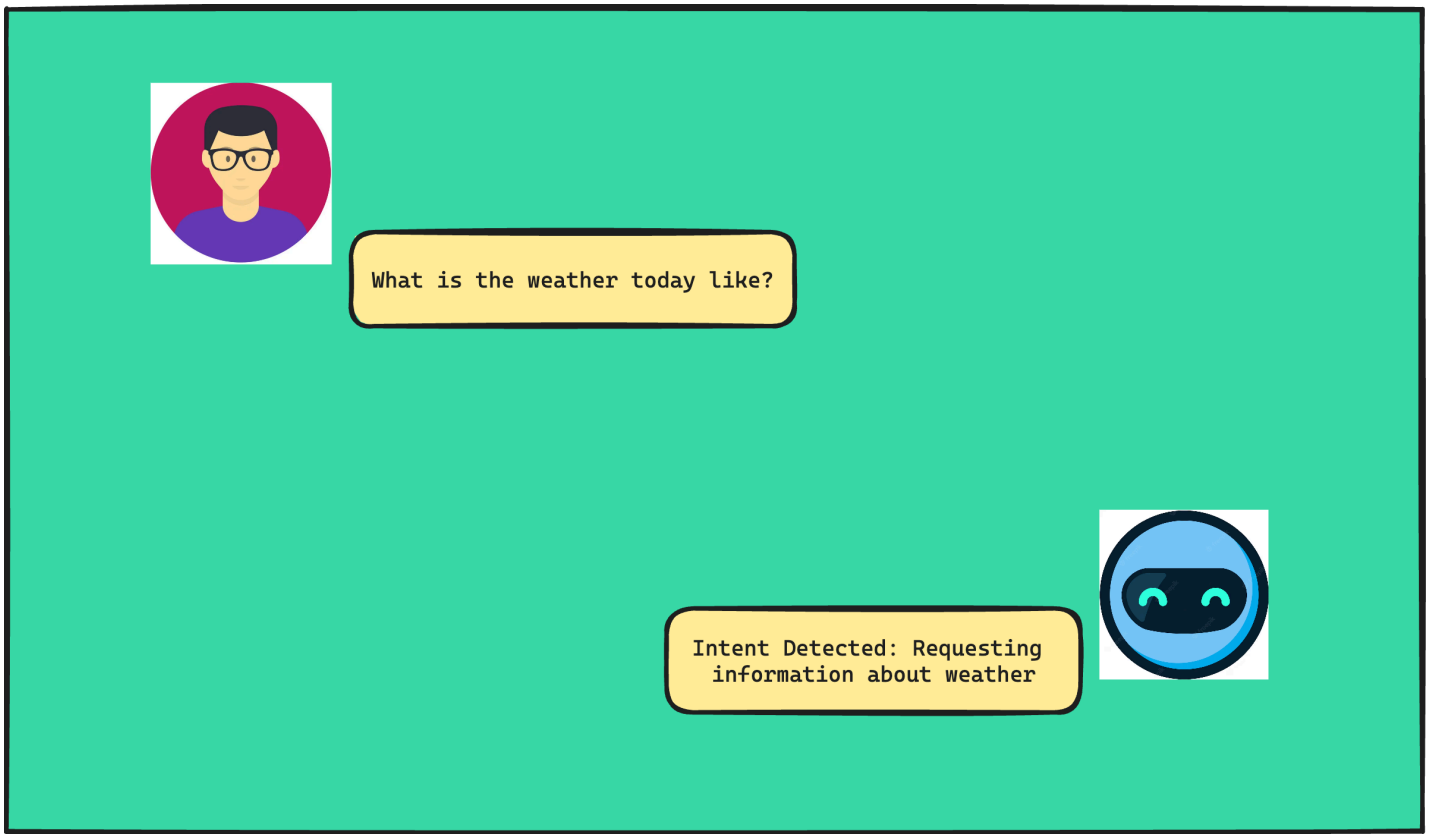
Pattern recognition is a key skill that has enabled humans to climb up the evolutionary ladder. The ability to recognize trends in natural phenomenons perceived by humans in form of visual, sensory and linguistic data has not only led us to develop cognitive dominance, but is also the very basis of some of the most essential technological advancements that affect our lives every day.

**Technology today** is fueled by data, and the systems capable of collecting, organizing and analyzing this exponentially growing volume of data play a critical role in propagating the advancements. One of these advancements is large language models.

## Textual Intent Recognition: An Introduction to Large Language Models

Natural language is one of the most prominent medium of information exchange in both real and virtual world. With online services booming, textual data and large language models are disseminating information every where around us.

Intent recognition system can help process, analyze and classify this vast corpus of text data and and build NLP based machine learning systems at scale.




An example of technology detecting the intent of an online message

**Intent recognition** is a natural language processing paradigm, at intersection with machine learning, which is able to deduce the intent or objective of a textual or verbal linguistic phrase by learning the patterns in labeled examples.

Consider a google search request as an example, “What’s the best Danish bakery in town?”, the intent of the user with this search query is to retrieve information regarding the best Danish bakery nearby. A personalized recommendation system capable of capturing this intent will help Danish bakeries to reach out to such a potential customer and boost their sales.

## How Textual Intent Recognition and Large Language Models Can Help Streamline Brand Protection Workflows?

Large scale commercial brands are constantly targeted by **phishing and scam attacks** by malicious entities (we covered a **large-scale brand impersonation scam** earlier this year).

These brands are targeted for their loyal customer base, where attackers create **websites** or **apps** that look authentic but are placed with fraudulent intentions like stealing personal sensitive information and even the payments made through their gateways. Customers not accustomed to detect such deceptive portals end up becoming the victims.

***Read more about vision language models***

The brands also face losses in terms of user traffic on their platform as well as their loyal customers.

At **Bolster.ai**, we have found intent recognition to be a potent tool in protecting these brands. A webpage which showcases the name of a brand in the site title or page title along with content it displays, and has same intent as the brand's official website but is not owned or hosted by the brand should be **classified as scam**. The intent is often directly correlated to the brand's business context.

As an example, ETHCrypto is an American company which provides a software platform to trade cryptocurrency, the intent of ETHCrypto's webpage platform is contextually related to cryptocurrency.

Any website not owned by ETHCrypto, but uses ETHCrypto's name in its webpage content, and renders text on its webpage with intent to target consumers interested in trading cryptocurrency should be classified as a scam. Such pages should be subjected to a **takedown**.

In this use case, large language models play a subsystem level role and analyzes the content of the webpage to detect the intent. This combined with a similar machine learning system to detect the brand of the website, can be used as **scam website detector to protect brands**.

# NLP's Quantum Leap : Transformers and Large Language Models

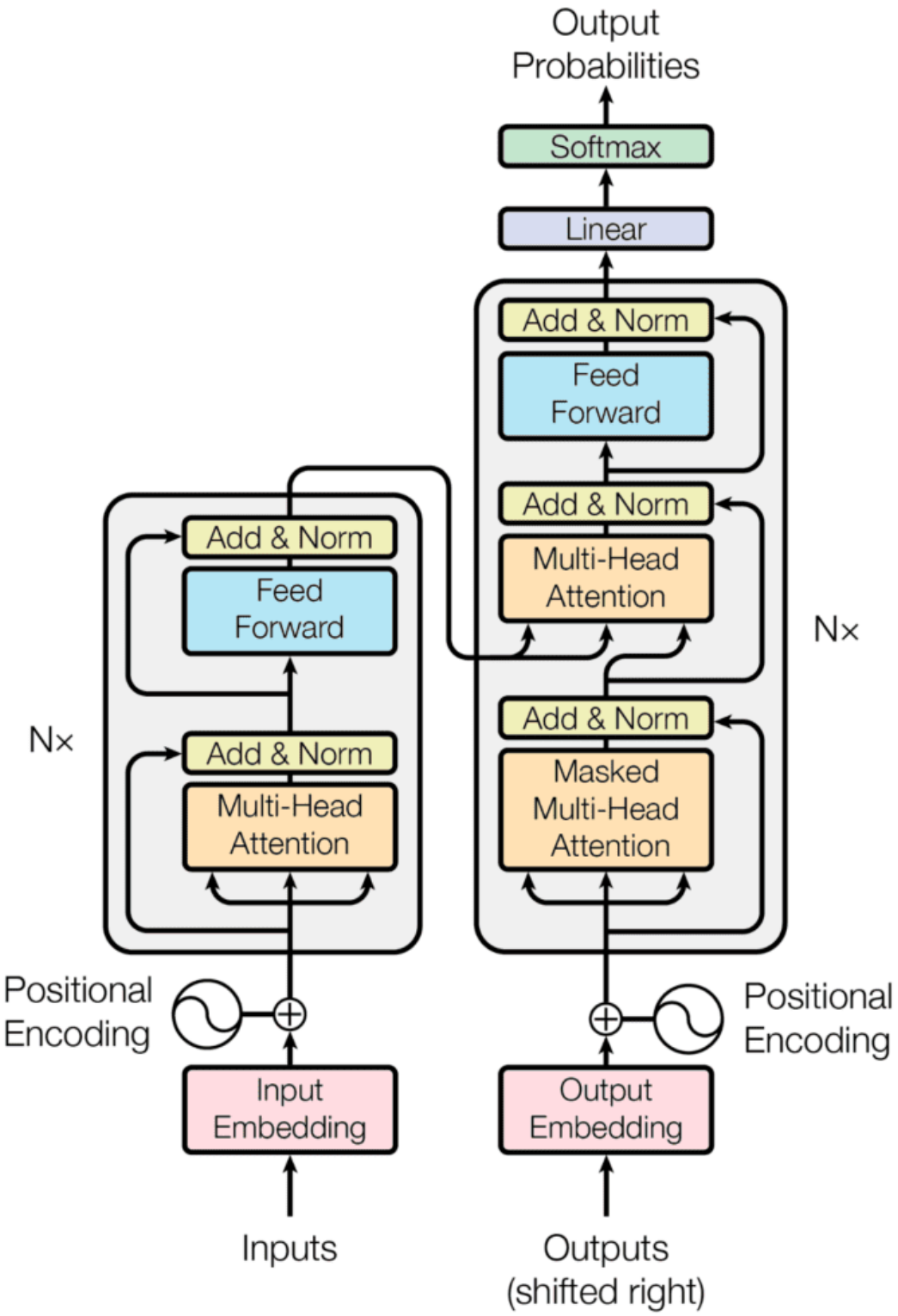



Diagram showing a Multi-Head self-attention Transformer

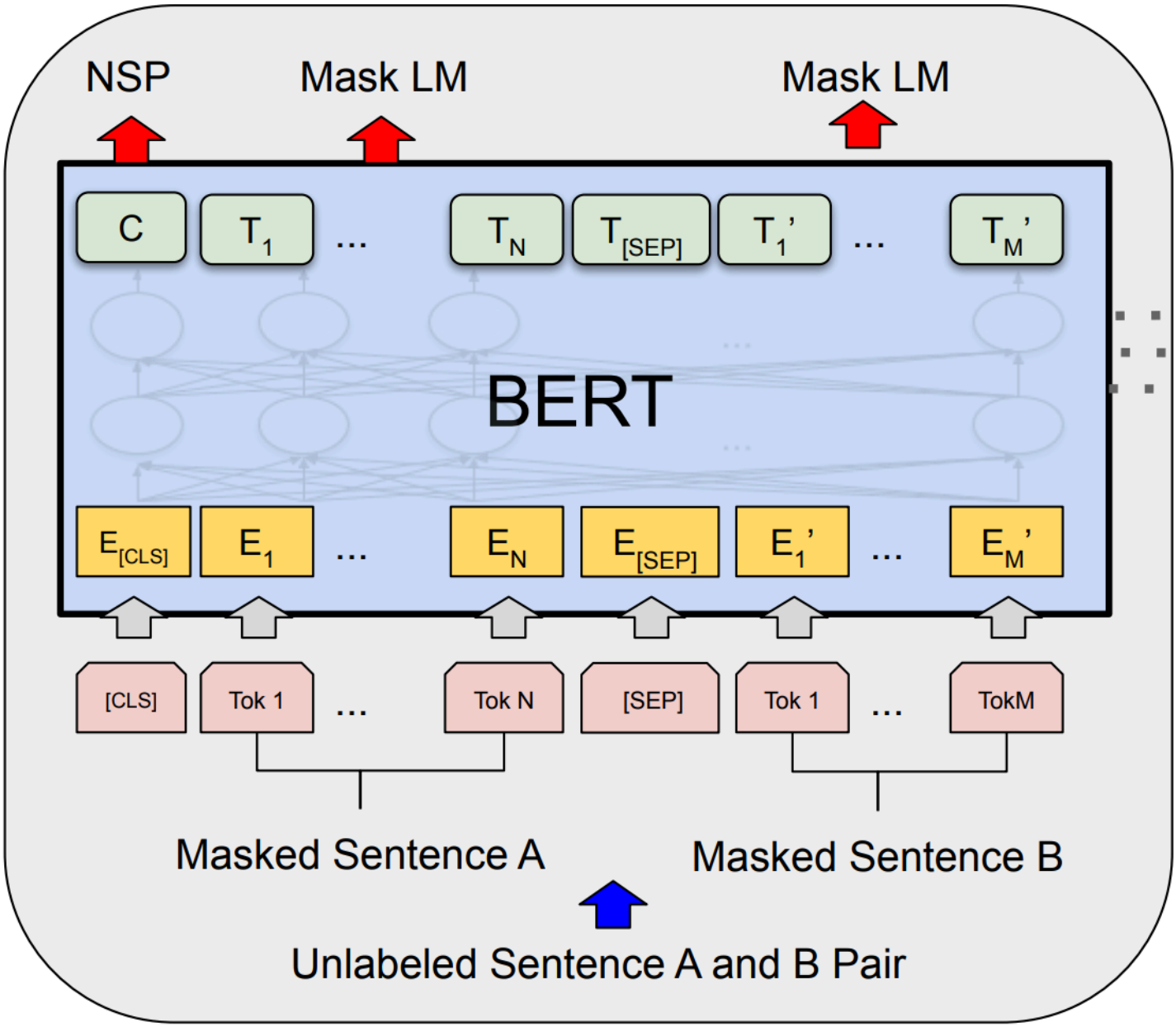
Transformers have emerged as a revolutionary breakthrough in the field of Natural Language Processing (NLP). These sophisticated neural network architectures, introduced in the paper "**Attention Is All You Need" by Vaswani et al.** in 2017, have

fundamentally changed the way we approach language understanding and generation tasks.

At their core, transformers rely on a mechanism known as “attention” to process sequences of data, making them exceptionally effective in handling sequential data like text.

What sets transformers apart is their ability to process input data in parallel, rather than sequentially like traditional recurrent neural networks (RNNs). This parallelization, driven by the attention mechanism, allows transformers to capture long-range dependencies in text, making them highly efficient at tasks like machine translation, text summarization, sentiment analysis, and more.

Transformers can attend to all words in an input sequence simultaneously, giving them a unique advantage in understanding context and relationships between words, even in long and complex documents.

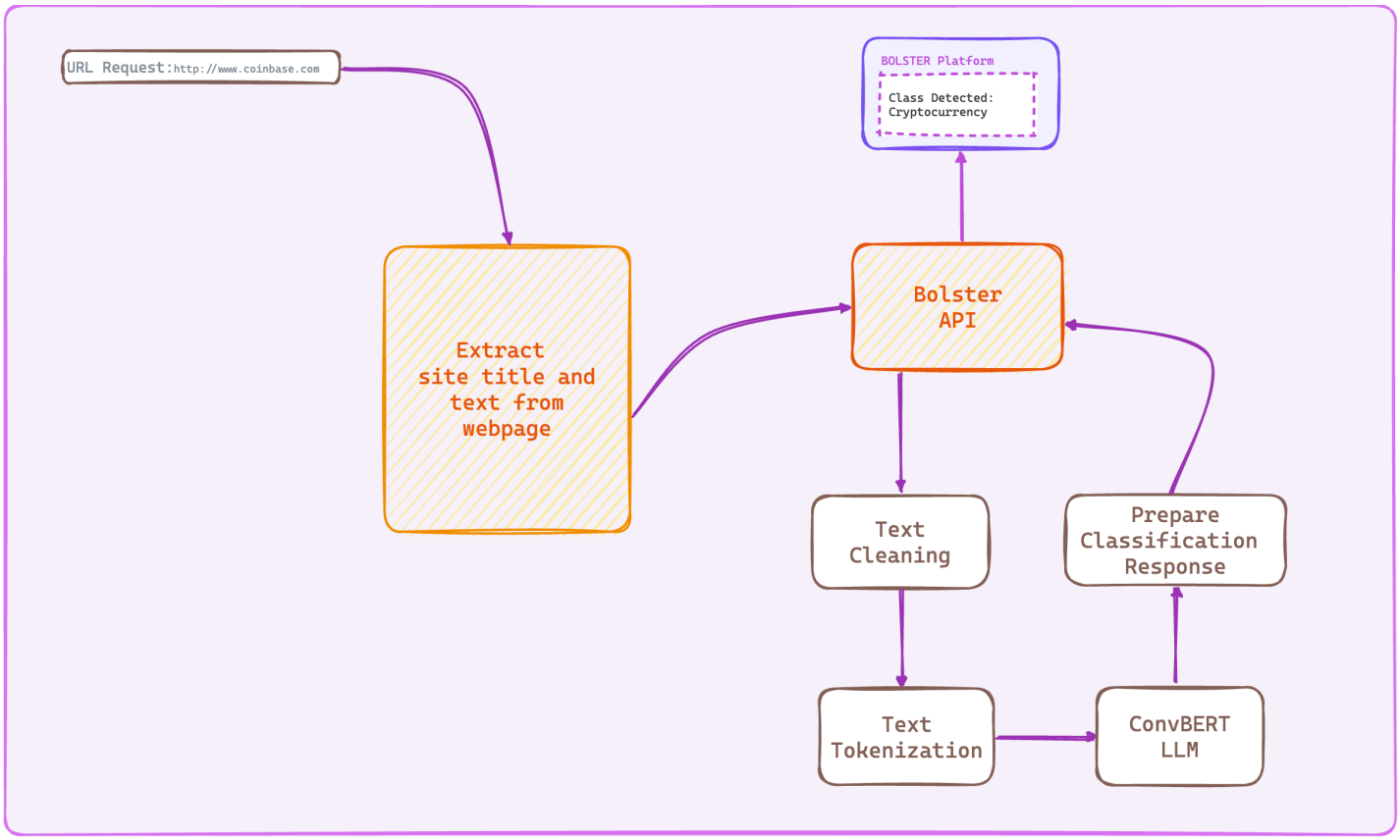



BERT Model (arXiv:1810.04805)

One of the most famous transformer based large language models is BERT (Bidirectional Encoder Representations from Transformers), which pre-trains on massive corpus of text data and has shown remarkable performance improvements for NLP tasks.

Large language models like BERT have become the backbone of many state-of-the-art NLP models and have enabled significant advancements in areas such as question answering, language translation, chatbots, and sentiment analysis. The ability to transfer knowledge learned from pre-training to downstream tasks has simplified the development of NLP applications.

# The Bolster AI Approach to use Intent Recognition to Protect Their Client's Brand Value




The Bolster workflow for intent recognition

At Bolster, we have created a dedicated **machine learning workflow** to extract site title and displayed text from any webpage, and analyze it using transformers-based neural networks to detect the intent of the webpage. Our workflow is representable by a Directed Acyclic Graph which ingests labeled examples, converts them to feature vectors and perform error minimization using back-propagation.

Our latest deployment uses Convolutional BERT (ConvBERT) model, proposed in "ConvBERT: Improving BERT with Span-based Dynamic Convolution" by Jiang et al. in 2020 and adopted from the **HuggingFace model zoo**.

We use embedding based tokenization to convert unstructured text data to numerical tensors suitable for machine learning workflow and error minimization. ConvBERT boasts of following advantages as compared to BERT and its other variants as confirmed by authors in the research literature:

1. Fewer number of parameters
2. Lower training cost
3. Higher performance metrics

Apart from the model architecture, dataset plays a critical role in performance of a machine learning workflow. We have collected our own dataset of labeled examples tailored for the use case of detecting intent of a webpage and classifying it to one of the following classes:

1. Adult websites
2. App Store
3. Cryptocurrency trading
4. Cryptocurrency giveaways
5. Directory Listing
6. Domain Parking
7. Domain Purchase
8. Error Pages
9. Gambling
10. Gift Cards
11. Online Store
12. Warning Pages

The intent recognition is a subsystem in ensemble of NLP and computer vision-based machine learning workflows which consider multiple modalities to identify a malicious webpage attacking the business potential of our clients.

# What Advanced Methods Can be Used for Intent Recognition in Future?

Transformers have further gone on to unlock high performance machine learning models in the domain of Generative AI. Large language models like GPT which empowers OpenAI's ChatGPT, have billions of parameters, trained on gigantic corpus of internet textual data, are capable of engaging in natural and contextually rich conversation and generate human like responses to prompts and requests from users.

These models are general purpose in nature and hence can be called upon to solve custom use-case problems without laying down complex machine learning pipelines. Such models provide the capability to be used in an off the shelf fashion to detect intent of a website as well when presented in the form of a prompt.

These engineered prompts allow us to directly utilize ChatGPT's generalization capabilities to perform classification of text rendered on a website into one of the classes above. Furthermore, OpenAI also provides a machine learning workflow API to fine-tune state of the art GPT3.5 model to the specific use case of detecting the intent.

Bolster AI recognizes the potential of Generative AI in enhancing the security we provide to our clients and is working actively on integrating state of the art models like GPT3.5 and GPT4 to automate our essential processes.

To learn more about how Bolster utilizes AI and machine learning to combat cyber threats in innovative ways, **request a demo** with us today.


## 2024 Presidential Election Report

# Discover the latest phishing and online scams threatening the democratic proces

Read the Report