Whistle AI

Disrupting Corruption

Version 1.4

Whistle AI

Overview **2**

The Problem **3**

New Media **4**

Anonymity **4**

Whistle AI Network **7**

Input Processing 7

Machine Learning 8

Human Analysis 8

Accuracy Oversight 8

Application Security **9**

WISL Coin **10**

Market Size **10**

Using Whistle AI **11**

The Content 11

The Buyer 11

The Seller 11

Transaction Breakdown 12

The Human Element of Authentication **12**

Whistle AI Gig Economy **12**

Three-Part Human Incentive System 13

Meritocracy Governance of Participants 13

A White Hat Network **13**

Conclusion **14**

Overview

Whistle AI seeks to attack corruption on the world stage by creating a decentralized platform wherein whistleblowers can safely and anonymously "blow the whistle" without fear of retaliation. Whistle AI is a decentralized marketplace that provides whistleblowers anonymity and allows them to monetize information that, until now, would never have seen the light of day. For buyers of information, such as journalists, new media, or watchdogs, Whistle AI uses a unique combination of artificial intelligence (AI) and crowdsourcing to verify the accuracy of the information being sold, while keeping the substance of the information private and unknown to anyone but the seller.

The Problem

When crime or corruption take place, there are often witnesses or people who have knowledge of the act. These witnesses fall into two categories: the bystander and the whistleblower.

Being a bystander is easy. Do nothing or say nothing and life will continue on as usual. In many cases, there may even be a financial or other benefit for remaining silent.

Being a whistleblower is hard. There's very rarely a real incentive to expose the corruption of powerful people or organizations, and the prospects of career suicide and even violent retaliation are real.

First, a solution would be to protect the whistleblower's anonymity. That would protect him from retaliation.

How do we protect whistleblowers while financially incentivizing them to do the right thing? How do we build a model that will turn more bystanders into whistleblowers? Until fairly recently, an immediate answer to questions like these was not forthcoming.

First is the problem of anonymity. We must make it so that the whistleblower cannot be identified. Let's take a look at a typical information release where the whistleblower is kept anonymous.

An employee of XYZ corporation discovers evidence that XYZ Corp is illegally dumping hazardous materials. The corporation is saving millions of dollars by doing this, but the whistleblower can't stand what is going on, so she decides to reveal her information to the world. How does she do this while protecting her identity?

In the current paradigm, our brave whistleblower can go to the press. She can take documents that prove the fraud beyond doubt and show them to a reporter. However, the whistleblower is now faced with the fact that she is risking everything: her career, her reputation and possibly her safety. Everything depends on the reporter, a person whom she has never met. Will this person protect her identity at all costs? Would this person turn down a million dollar payment from XYZ Corp in return for revealing the source? Would the reporter go to jail for the whistleblower if necessary?

Unfortunately, most whistleblowers decide that putting their lives in the hands of another human being is too risky. While we do hear about the occasional reporter who is willing to do jail time to protect a source, this is the exception, not the rule.

There are also many situations today where a newspaper may not have any interest in covering the information the whistleblower is trying to expose. This is where Whistle AI can have far more impact than what's currently available. Whistle AI can act as a connector between whistleblowers and a worldwide community of prospective buyers. In our scenario, perhaps the journalist isn't interested in covering the story, but there are environmental groups that would be extremely interested. Perhaps it's a vlogger or an online magazine who wants the story.

While media organizations and watchdog groups are clear examples of interested parties, we believe the real, true market potential for whistleblowers' information remains largely untapped.

New Media

This untapped market is new media. Today, traditional or legacy media are being disrupted before our eyes. Millions of citizen journalists now exist throughout the world, many of whom specialize in thousands of different subject matters. Many garner more views online than their legacy media counterparts. We believe that new media is ultimately a force for good and another step toward a more decentralized world. Whatever the subject matter is, there an alternative outlet searching for information that can break a story. We believe that Whistle AI can empower new media in ways never before seen. It will give them a marketplace to shop for stories and information that has always existed, but has never been readily available until now.

Whistle AI stands to be a conduit between two massive markets that have never connected. As of right now, there are millions of new media journalists and content creators all over the world. The subjects they create content around are almost infinite. Some deal with politics, some with various social issues, others with food, others with animal rights and the list goes on. This represents a massive market of untapped buying potential for information that is relevant to their passion or cause. We see new media as a smoldering bundle of sticks, waiting for fuel to fully ignite the fire. Whistle AI, in this scenario, will be gasoline. By connecting these content creators with millions of people in everyday situations where they witness wrongdoing, we can give them potentially endless sources of content to build their followings. The age-old market of gossip finally has a buyer. It's easy to see the potential for Whistle AI at a higher level of crime, Wall Street, government, big business, but think about the possibilities on a more micro scale. Think about the waiter at the restaurant who repeatedly sees the cook intentionally dropping people's food on the ground before serving it. A relatively small issue in the grand scheme of things. But what if that waiter, who's tired of seeing this happen, can now sell pictures he took of this to a local food blogger for $50? Don't confuse micro for small however, when it comes to impact. Imagine this scenario playing out all over the world, in every industry? The untapped market potential that exists between new media and everyday witnesses to wrongdoing is enormous.

Anonymity

One of the problems with making whistleblowing safe and lucrative is that once the damaging information is revealed to anyone, it loses its value. How do whistleblowers expose information to the marketplace while simultaneously retaining anonymity, and also being able to sell this information to the highest bidder, or outlet of their choice - all while keeping the information secret to any one person until the information has been successfully sold?

At Whistle AI, we believe we have the framework for a solution to this problem. A peer-to-peer network serves three powerful functions that will both protect and compensate whistleblowers. First, it can protect the anonymity of a whistleblower more securely than any existing method through a blockchain-based coin geared toward privacy. Second, through a combination of artificial intelligence and crowdsourcing, we can verify the information without any one source having access to the totality of the information. And third - and most important - this anonymous verification process can be used to provide powerful financial incentives for exposing corruption.

Having sensitive information exposed to multiple potential buyers presents a big problem though. It's not like selling a car or a laptop. You can show your car online and let people bid on it, but you can't do the same thing with a document proving malfeasance. Once the document's information is disclosed - especially if that information is newsworthy - the document loses most of its value. How does a whistleblower shop this document around without destroying its value? The answer lies in *minimum disclosure authentication*.

The idea behind minimum disclosure is the concept of breaking up a secret into fragments, and allowing anonymous community members to verify the secret piecemeal, so that no one person knows the secret, but the secret is still verified and its value authenticated.

Crowdsourcing algorithms work well for authenticating general information, but what if that data is sensitive and requires privacy? In these cases, we need to figure out how to prove authenticity without full disclosure of the information.

In computer science, this problem set is referred to as a *zero-knowledge proof*. The idea behind the zero-knowledge proof is to create a verification mechanism that allows a separate party to verify underlying information without having to view the information in its entirety.

Minimum-disclosure protocols, or minimum-disclosure proofs of knowledge, are a subset of zero-knowledge proofs. This protocol takes information and chops it into smaller discernible yet incomplete pieces of information. The algorithm then discloses each piece, one at a time, to a verifier until confidence in the underlying information is reached.

Using a peer-to-peer network, we take the idea of minimum disclosure and combine it with crowdsourcing to achieve a decentralized, private information authentication algorithm. The algorithm splits information into chunks prior to any encryption of the information and sends those pieces to different nodes within the peer-to-peer network, asking for independent verification of each chunk. No one user will be able to verify the object is absolutely what the asset owner states it is; however, if any chunk is obviously different from what the owner states it should be, then the entire asset's authenticity is invalidated.

Here’s a simple example.

> *Suppose a whistleblower has a photo of John Doe stealing a cookie out of a jar. For the purposes of this hypothetical, let’s assume that the cookie theft is of great importance to a lot of people and that the fact that John Doe is the culprit is a valuable secret.*
>
> *Once the world knows that John stole a cookie, the information is no longer valuable and will be impossible to sell. But if the network can verify the secret without actually revealing it, then the photo can be sold at a good price, thus keeping the financial incentive for whistleblowing intact.*
>
> *The network would have built-in protocols that would be able to parse the expression “This is a photograph of John Doe stealing a cookie from a jar.” It would then break that statement up into two parts: “This is John Doe” and “This is a cookie being stolen.”*
>
> *It would then create two versions of the photo. The first would blur out John Doe’s face. Members of the community would be compensated with cryptocurrency for merely answering the question: “Is this a photo of someone stealing a cookie from a jar?” If a sufficient number of users answer yes, then the first part of the secret is then considered authentic.*
>
> *The second version would blur out everything but John Doe’s face, and ask the members of the community “Is this a photo of John Doe?” If a sufficient percentage answers yes, then the second part of the secret will be considered authentic.*
>
> *If both parts of the secret are successfully authenticated, then all of a sudden we have a verified secret with great monetary value, but without having revealed the secret to any one person. This secret would then be offered to a network of buyers with the simple statement “Whistle has verified a photo of John Doe stealing a cookie from a jar.” With the network having authenticated the photo, the monetary incentive for disclosing criminality will have been preserved.*

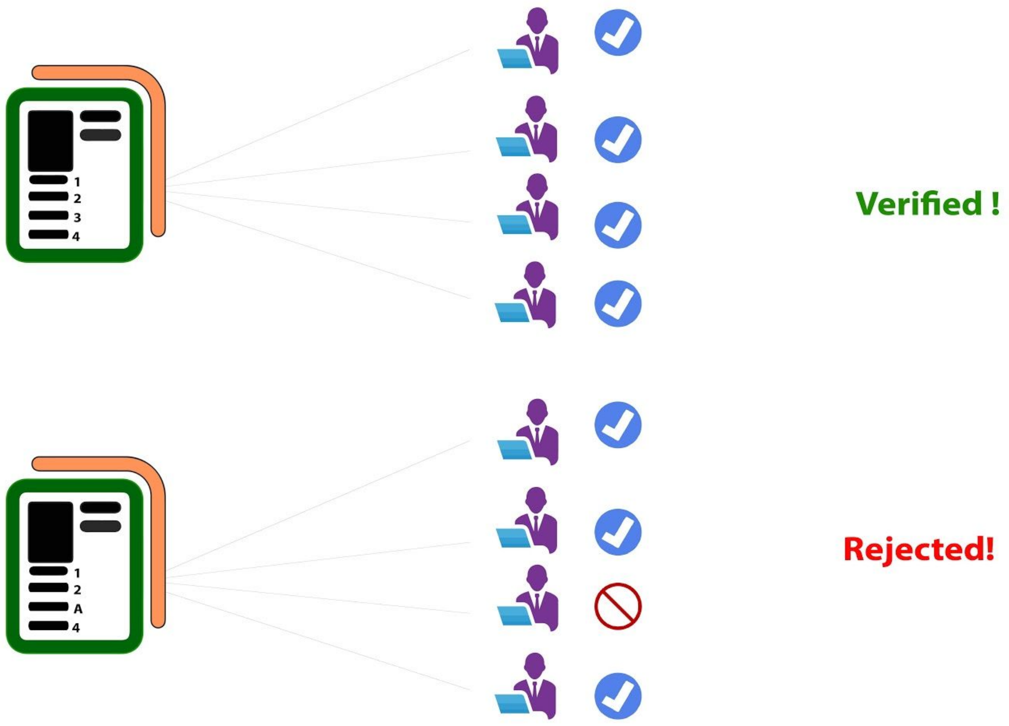

Figure 1: A Minimum Disclosure Protocol Example Breaking Information into Four Pieces

Whistle AI Network

Traditional whistleblowers give value to information by disclosing their personal identity as well as their relationship to the information so that it can be branded as trustworthy. With the anonymous sale of information, we must find another way to replace this idea of trustworthiness to extract the value from the information without attaching it to any individual as the source. The answer is the Whistle AI Network.

Under the hood, it is is a peer-to-peer network that connects machine learning with human crowdsourcing to facilitate the mechanics of the minimum disclosure protocol algorithm.

The network itself is comprised of a group of nodes communicating together to authenticate data that has been submitted to the network. The authentication process is made up of four components: input processing, machine learning, human analysis and accuracy oversight.

Input Processing

A request to authenticate a piece of information must be submitted through a peer-to-peer node. The information must be an accepted file format (png, pdf, mp3, mp4, avi ,docx, etc.) in order for the request to be processed. Based on the digital file type, the network then creates a sequence order of required authentication steps to be taken and keeps track of the progress of the request.

Machine Learning

In the first phase of authentication a digital file will be passed through different machine learners using algorithms to filter for specific things on specific types of data. These filters can range from relevancy filters to determine the piece of information will be valuable on the network, all the way to facial or voice recognition that helps identify an individual's identity.

Human Analysis

As part of the minimum disclosure protocol, the digital data file is altered for privacy and anonymity of the individuals in the pieces of information by machine learning algorithms. It is then passed to our human nodes whose function is to group analyze the same piece of information and extract a yes or no answer about the piece of information.

Not all participating human nodes will vote on every piece of information that comes through the system. Human work assignments are delegated based on an ordering algorithm, called Priority Rank, which takes into account current time spent on the platform as well as an historical accuracy score of work done on the network.

Accuracy Oversight

As the final step before each piece of information can be deemed fully verified, it must pass through the oversight measures we have put in place in order to increase our rate of accuracy. The primary means of oversight is by a group of pre-defined moderators, called Merit Officers. A credentialed Merit Officer has earned her participation eligibility through consistently high-quality authentication work on the network. The main function of a Merit Officer is to monitor for flagged authentication requests as they are processed whether it is by a machine learning algorithm or a human verifier.

Additionally, we include an escrow and arbitration service for added assurance that a user gets exactly what she was expecting when she purchases a piece of information. In these cases, a buyer spends an additional 1% in cost for the option to use an arbiter if the piece of information she bought is fraudulent. After the purchase, the funds of the sale are held in escrow for 48 hours before the seller can claim it. During that time, if the information is deemed "not acceptable" by the buyer, the buyer can open a complaint. Once a complaint is opened, a Merit Officer is randomly assigned to the case through a round-robin algorithm. The Merit Officer then looks at the information in question to determine whether or not the seller acted in good faith. The Officer's arbitration ruling on the complaint is final. Both parties must acknowledge this finality ahead of time.

A more detailed explanation of the technical details can be found in the separate Whistle AI Full Technical white paper.

Application Security

Peer-to-peer networks face different security challenges than the centralized client/server networks such as Amazon, Facebook, and other modern web applications. Client/server application architectures have a central security model which means all data remains on a protected server. This centrality generally creates a scenario where only a single access point needs to be secured. The common web protocol used to secure these application access points is HTTPS.

Peer-to-peer networks have no central server, so security is a much harder problem to solve. Every communication between users needs to be independently secured so that it can be trusted. The current best practice of secured peer-to-peer networks is asymmetric encryption, which is also known more commonly as public key encryption. In this model, two unique keys are created to encrypt and decrypt the messages being sent through the network. If one of these keys remains private to an individual peer within the network, and the other shared with all other users, communications can be privately encrypted by other users. Also, messages can be guaranteed to have been generated by the source. These properties create two useful security protections.

First, a user can send a message to another user encrypted with the public key so that only she will be able to decrypt it with the unknown private key. Second, messages encrypted with the private key are verifiable to anyone with the public key. This reversal is useful if one wants to verify that the message originated from the expected source. This latter capability is called a digital signature, so that all messages can be verifiably pointed back to their original destinations. When requiring user-to-user authentication instead of user-to-server authentication, digital signatures are key to secure communications.

Security of peer communications and anonymity of payments are both crucial to the success of Whistle AI. Vital information maintains its value only while it remains private. To protect the integrity of the information's value for the seller, it must be stored securely on our network and can be passed along only to trusted nodes. Within the network, the only nodes that should receive a full piece of information are the node of the owner of the information herself, as well as the machine learning computational nodes that obfuscate the information prior to human analysis.

To ensure that the information remains only within the intended nodes, we use digital signatures to ensure accuracy of the source and public key encryption to guarantee that the information is only decryptable by the intended receiver.

The coin used by Whistle AI is deployed using blockchain technology that has been production-tested for security as well as proven anonymizer algorithms to protect the identity of the users.

WISL Coin

In order to preserve a seller’s anonymity and allow her to be paid for her contributions, Whistle AI requires an anonymous compensation method. A precedent already exists for private payments in the form of privacy based cryptocurrencies like Monero, Dash, Zcash, Verge, and most recently PIVX. The Whistle AI Platform uses its own privacy coins, named Whistles (WISL). WISLs serve a dual purpose, as they will be the single payment method for buying and selling assets on the network, but they will also be used by human crowdsourcing participants while they help authenticate information on the network.

Market Size

Corruption never seems to be in short supply; in fact, it seems to be one of the worlds most abundant, renewable resources. And while there has always been an interest in exposing corruption, a true market for it has never been able to form. Legacy media, the traditional buyer for such information, has always represented a centralized “bottleneck” of information, which limited the size of the potential buying market for information. In the mid-20th century, Journalist A.J. Liebling stated that "freedom of the press is guaranteed only to those who own one." Meaning that only those voices powerful enough to produce a book or a print a newspaper could be heard. Today, the power of the press is being disrupted and redistributed among the people. This redistribution creates a massive shift in the economy of information. We are witnessing the creation of a global market for information the likes of which the world has never seen. The potential size of this market is unknown yet, but we estimate it to be in the billions of dollars, with exponential growth thereafter.

There are, however, still challenges facing this new media. “Fake News,” as it’s been labeled, has plagued New Media and now Legacy Media is even coming under fire. The problem is that “trust” in the information being provided by media has traditionally been granted through the authority of the oligarchies that control it. As New Media is coming online, there is no authority to vouch for the information being provided, as in the 20th century. As New Media and Legacy Media create conflicting content, both are feeling the pinch of an erosion in trust by the consumer.

We have seen a similar phenomenon with the evolution of e-commerce. A shift in trust from the big box retailer to the individual seller of goods had to take place. In the e-commerce world, it happened through a combination of mechanisms like online reviews and frameworks, like Ebay, which lent legitimacy to the transaction, without actually being the buyer or the seller in the transaction. Buying something before physically inspecting it is a very risky proposition. Allowing potential buyers to sample trusted customer reviews before purchasing created enough of a comfort level to allow consumers to buy things online. Buying something online will always be more convenient than having to go to a brick and mortar store, as long as one can trust the seller. Customer reviews of products supplied that trust, and as the public’s trust in e-commerce has grown, it has become over a trillion dollar per year market.

Today there are two massive markets, information sellers and New Media, that need a conduit through which to transact. A framework that allows buyers and sellers to transact without being directly involved. Just as Ebay first provided the validation and trust needed for online transactions between individuals to take place on a massive scale, Whistle AI provides the framework and verification for information to be transacted between New Media/Legacy Media buyers and the world of everyday people who witness wrongdoing and corruption.

While it's difficult to put an exact number on this emerging market, many similarities hold true between it and the birth of e-commerce. We inject trust for online transactions when it existed previously only face-to-face. We also create buyer convenience and save content creators and other buyers days or weeks of research.

Using Whistle AI

At its core, Whistle AI is a marketplace for information that connects buyers with sellers.

The Content

When content is uploaded to the system, it includes a digital file, description, initial list price for the sale and any other supporting documents required for authentication.

The description is used to make sure the content follows the intentions of the network per the founding principles to help govern the network. It can also be used as the basis of understanding for the users within the authentication algorithm to make a judgment call to flag the piece of information as different from what a buyer would expect it to be. If more than 50% of crowdsourced participants flag a piece of information, then it fails authentication and is not sellable.

The Buyer

Buyers within Whistle AI have the option to disclose themselves and have their identity verified, or to remain anonymous. Verified buyers have a distinct advantage because sellers will be able to seek them out with offers for relevant information. Anonymous buyers must search through the network for interesting information themselves. Making it even harder is sellers may require the buyer of their information to be verified, making the possible pool of information smaller.

The Seller

The only information stored about the seller through the network is a unique numerical ID. This ID must be different from the wallet address to help preserve a seller's anonymity. All other information about the seller should never leave her own peer node, and will permanently delete itself after a transaction has been completed.

The seller is ultimately in control of what happens to the information. For example, she can restrict sales only to buyers who are known and verified so that she can maintain an understanding of what will happen after the sale.

Transaction Breakdown

Before a piece of information is sellable on the network, it must first be authenticated. The cost to get a piece of information verified by the network is a flat up-front fee, plus a portion of the sale price going to the network profit pool.

The flat fee will be designated in WISL and can be adjust for every new cycle based on the market volatility of WISL as well as the demand for authentication. The goal is to allow the fee to increase based on demand, but remain stable despite the volatility of the price of the coin. The flat fee proceeds are split by granting 60% to the human validators and 40% to the other nodes as broken down in the Incentive and Fees section above.

Once a buyer is found and a piece of information is sold, then the proceeds of the sale are split into multiple initial buckets. The first is the seller, who receives 92% of the buy price. The remaining 8% is split evenly four ways among the development fund, the merit officer payment fund, the pool of machine learning resource holders and the pool of human authenticators.

The human authenticators do not receive a direct kickback from the deals they approve. Instead, all approved proceeds go into a general pool that pays out weekly, proportional to each participant's Priority Rank for that week. This mechanism is to help remove the temptation for participants to approve bad information in hope of receiving an allocated share of the profits. If all profits are shared in a way that is directly correlated toward positive accuracy and time spent on the network, then the end result should be a more truthful authentication and oversight process.

The Human Element of Authentication

Humans will interact with an interface that allows "swipe right" for a positive verification and "swipe left" for negative vote. For each human attempt at authentication, there will be a set number of voters assigned to a given task. The voters must collectively approve the asset at a percentage above a set threshold for authentication. Each user's voting result will have either a positive or a negative effect on her accuracy score based on whether she was among the majority. The higher a user's accuracy, the faster she will receive another task from the routing nodes and the higher rank she will receive.

Whistle AI Gig Economy

Human verifiers are an essential aspect of machine learning in the Whistle AI Network. A three-part incentive system is built in to reward the human nodes for their time and participation. This work utility will add another dimension to the gig economy made prevalent by companies like Uber and AirBnB. The goal is to have a network of over 500,000 people collectively interacting with and teaching the AIs on the platform by 2020. A large human network of intelligence will thus be a game changer for the progress of Artificial Intelligence.

Three-Part Human Incentive System

Money alone can't motivate every type of person. Recognizing this, we aim to deliver a three-part reward system based on merit, to both incentivize performance and to add the elements of accomplishment and progress to the human verifier experience. A person participating in the Whistle AI gig economy has three incentives for performing: WISL**,** Ranking and Promotion. When a person joins the Whistle AI Platform, she begins earning WISL for accurate verifications. As she verifies more and more tasks accurately, she begins to earn badges and ranks within the system. The more badges and the higher her rank, the more important her role becomes in the Whistle AI ecosystem. Subsequently, the more she earns per verification. In this model, long-term loyal workers get priority access to premium work, but new participants can hang out on the network and raise their ranking in order to compete with longstanding participants by building up current active participation. A participant's access to quicker and higher paying work is directly correlated to that participant's rank relative to other users on the network at that time.

Meritocracy Governance of Participants

Moderation positions to help govern both the crowdsourcing participants and the quality and relevancy of the information being submitted to the network. These moderation positions are called *Merit Officers* and are paid positions in the Whistle AI Network. Merit Officers are the decision makers for the Whistle AI Network and are responsible for keeping the network "White Hat". They are compensated through verification fees. To become a Merit Officer, a set number of badges and ranks must be achieved. Of those who achieve the proper rank, the top 10% of both total network earnings and work accuracy score are eligible to become Merit Officers.

A White Hat Network

Whistle AI is determined to be a force for good, and will be designed to stay that way. We are aware that sensitive information can be used for good or for evil. Blackmail is a crime for a reason: society doesn't benefit when one citizen threatens to expose another's embarrassing activities - especially activities that harm no one. With that in mind, the Whistle AI network contains ground rules that prevent it from becoming a breeding ground for “Black Hat” uses.

The Whistle AI constitution is a small set of founding principles that set present and future moderating guidelines for Merit Officers to enforce and modify. Like any good constitution, the guidelines can be amended by a two-thirds majority vote of special governing nodes called Masternodes. The voters are instructed to make sure the guidelines always follow and maintain the ethos of the founding principles.

Founding Principles of The Whistle AI Constitution

1. Whistle AI exists to expose harm being inflicted on others
2. Whistle AI does not exist to inflict harm on others

Founding Set of Amendable Guidelines for Merit Officers Enforcement

1. Abuse of Power is always fair game
2. Whistle AI is not a tool for individual blackmail, so disclosures of affairs, as long as it is not an abuse of power, is forbidden.
3. Disclosure of a person's sexual orientation or gender identity are private and not harmful to others so they are strictly forbidden.
4. Pedophilia is fair game, but it is forbidden to sell this type of information to the intended target as a form of blackmail.

When content is uploaded to Whistle AI, it must be accompanied by a description. The network identifies and flags descriptions that contain words and phrases that could potentially represent “Black Hat” material. At this point the content goes to a Merit Officer for review.

As part of the verification process, Whistle AI compares the content to the description, to ensure that there is a match. If the AI cannot confirm, then the file is flagged and sent for review by Merit Officers.

Conclusion

Whistle AI truly has the potential to change the world for the better. At the same time, it creates a massive new market for information that has never had an outlet. From multimillion dollar pieces of information on the wrongdoings of the powerful to the $400 piece of evidence that the construction company is cutting corners. Whistle AI has the potential to impact the world for the better, on a global level. Until recently, there was never much hope for this kind of change, but today with the decentralization of blockchain, the advent of cryptocurrency and its ability to enable secure, private payments as well as the progression of Artificial Intelligence, we are finally reaching a point where technology can provide the anonymity, incentives and verification to allow the truth to come to light. Louis D. Brandeis once said, “sunlight is said to be the best of disinfectants.” We believe that Whistle AI can bring sunlight to darkest corners of the world, by incentivizing whistleblowers and simultaneously protecting them from the criminals they expose.