

# Amplify

October 2021

## Abstract

We propose the Amplify Protocol: a decentralised marketplace for work, where individuals working on tasks ("workers") can be compensated by individuals and companies seeking the completion of tasks ("requesters"). The core Amplify Protocol will be task-agnostic and the community will be able to build on it to deliver many possible types of works. The initial use cases that will be built out by the Amplify team focus on data labelling and content moderation. Data labelling is a critical service for the growing fields of AI and machine learning. A decentralised solution allows the creation of a trustless, efficiently priced and scalable labelling service that will supercharge the growth of the AI industry by making accurately labelled data cheaper and more flexible. Content moderation is a similarly critical service required by all online platforms, for whom a trustless, efficiently priced and scalable moderation service can also be a significant enabler.

## 1 Introduction

Specialised online services are becoming integral to the way that we live our lives, the future of technology, and how we operate as a society. These services often require extensive human labour input, which is undertaken either by firms' dedicated internal resources, or large organisations that have large and diverse resources for labour-intensive tasks (eg. data labelling services provided by Amazon Mechanical Turk, or content moderation provided to Facebook by 5,000 Accenture employees).

There are two possibilities that can lead to issues when it comes to accessing affordable and high-quality services:

1. Smaller service providers operating in niche industries are able to command a higher price due to the lack of competition. This is evident in practices such as a "talk to sales" option on some data-labelling service's website, indicating that there is no incentive to advertise a price, and that the market is not efficient.
2. Larger service providers using cheap labour will often not provide quality services. For example, one user of Amazon Mechanical Turk found that in 27% of cases, workers providing

data labelling services disagreed, meaning 1/4 of all results needed to be checked in house to ensure accuracy.<sup>1</sup>

The two key use cases we propose for the Amplify Protocol (data labelling and content moderation) are among the most important in today's society; the first driving the development of AI models, and the second dictating how some of the most ubiquitous services of the 21st century are used.

In addition to the issues of accessing quality services, the centralisation of these services means that even in the case of freelance service providers, a single entity controls the pay and rewards. This combinations means that both the service providers and receivers lose out. Expanding the scope of decentralisation in the market for this kind of work will allow for smaller players to access higher quality services, and for workers to be rewarded in a more effective and transparent way. To facilitate this expansion, a platform needs to be developed that can host this collaboration in an efficient and trustworthy way.

## 2 Key use cases: data labelling and content moderation

### 2.1 Data labelling

Accurately labelled data sets underlie almost all recent successes in the fields of AI and machine learning. This has become the case because many applications of machine learning techniques, including traditional neural networks, require supervised learning approaches. Attaining labelled data is an expensive and labour intensive endeavour

Key users of data labelling services today include:

- Academia: data labelling is required for many AI research projects
- Companies building AI products: most AI projects rely on labelled data and most companies for whom AI is not a core competency choose to outsource labelling

#### 2.1.1 Market size

The market for data labelling alone may have grown to more than \$1 billion USD in 2020. We expect that at least 50% percent of this market is of a nature that could be served by the Amplify Protocol. We also project that the Amplify Protocol will enable use cases that would not have been served by the current data labelling market, and therefore could materially grow the market.

---

<sup>1</sup>(<https://www.zerofox.com/blog/five-things-to-consider-before-using-mechanical-turk/>)

### 2.1.2 Why has the problem not been solved in Web 2.0?

Although a market has already developed for third party contracted data labelling, there are limitations on the centralised Web 2.0 model that prevent these companies from serving the full range of opportunities.

Key issues with the centralised model for data labelling services include:

- Opaque pricing: Presently, the data labelling market is highly inefficient with most large players operating with no public-facing price transparency. This inhibits accessibility to services and effective competition, keeping prices higher to further reduce accessibility.
- Failure to serve long tail use cases: Most companies attempting to serve the centralised data labelling market
- Low ability to tap global labour markets: The centralised nature of current services means that they normally must find and hire employees to undertake labelling work
- Data ownership: data must pass fully into the hands of a data labelling company that will be responsible for the labelling, with no technical means for its protection

## 2.2 Content moderation

Content moderation has become a central challenge for all online networks that allow user generated content. No online platform with user generated content can choose not to have a content moderation policy as there are types of content including spam and illegal content that must be removed. Companies have adopted a range of strategies in relation to this issue, with most deploying AI tools to flag content and highly labour intensive review processes undertaken by a mix of internal and outsourced reviewers.

Content moderation can be seen as a subset of data labelling, as the fundamental task involves using human intelligence to assess characteristics of types of information. The current market for content moderation is badly served for similar reasons to data labelling.

One notable potential use of a decentralised network for content moderation is moderation of a decentralised social network. There are a variety of projects working toward various blockchain-based decentralised networks that will invite user generated content. These networks will not be immune to issues of spam, illegal and undesirable content and therefore would be well served by a decentralised way to undertake content moderation that is external to the internal incentives many are working toward.

## 3 The Amplify Protocol: A Decentralised Solution

Amplify is proposing a platform that will host a decentralised market for data-labelling activities, built on the Solana blockchain. This will allow small scale or minimally resourced researchers to amplify they themselves and their projects.

The core Amplify Protocol is a decentralised solution for linking "requesters", who seek the completion of scalable tasks requiring human intelligence, and "workers", who complete tasks. The Amplify Protocol will connect these groups and organise their interaction, including allowing the completion of work and the payment of workers.

### 3.1 Amplify Protocol Elements

The core of the Amplify Protocol is an on chain Solana smart contract. The Amplify development team will build a centralised web application on top of the on-chain program but the program itself will have openly documented APIs to allow access to anyone wanting to build on top of it.

The core features of the smart contract will be:

- Submission of tasks
- Distribution of tasks
- Payment for tasks

#### 3.1.1 Submission of tasks

Requesters will submit to the smart contract series of tasks, a price they are willing to pay for the work and the minimum rating for workers whose work they are willing to accept (eg. only workers with ratings above 4 stars out of 5). To assist requesters with deciding how much to offer to pay, it will be possible to see all transactions in real time, allowing requesters to assess to the market price for workers of various ratings. Front end services will allow requesters to upload tasks in bulk.

Task metadata will be stored on-chain on the Solana blockchain as a series of accounts (key/value pairs) but the underlying data (eg. images for labelling) will be stored off-chain. The on-chain Amplify Protocol will be neutral as to how the off-chain data is stored and different front-ends can choose whether to store the data using suitable decentralised networks (eg. FileCoin, Arweave) or centralised services.

#### 3.1.2 Distribution of tasks

Workers will be able to request from the smart contract a full list of currently requested tasks, the prices that requesters have offered to pay for the completion of those tasks, and the minimum rating

that the requester needs to have to accept the task. The smart contract will allow workers to accept tasks, removing them from the list of tasks to be accepted. This will be done at an individual task level: a requester may request that 1000 images are labelled and each will be an individual tasks. Front-end services will allow workers to accept tasks in bulk.

### **3.1.3 Payment for tasks**

Payments on the Amplify Protocol will initially be made in SOL but this could be generalised in the future. 5% of all the fees paid for work done on the Amplify network will be diverted to the Amplify Protocol Treasury, to be paid out to holders of the Amplify Protocol Governance token.

### **3.1.4 Privacy**

The Amplify Protocol will make data publicly available by default, but provide a mechanism for requesters to encrypt their data with the public key of the requester. This will need to be handled on chain because the encryption will need to occur at the point of acceptance of the task - the requester will not know which worker will be undertaking the task until the point of acceptance and therefore would not know which public key to use to encrypt the data.

Front end applications will need to be built to specifically support the encryption feature so they can undertake the encryption in real time at point of acceptance. Computational limits on real time encryption can be solved by encrypting the data itself at time of upload, and simply encrypting a symmetric encryption key with the public key of accepting worker at time of acceptance.

This encryption mechanism can be undertaken alongside a cap on the percentage of tasks within the requester's overall tasks that a single worker is allowed to undertake, preventing one worker from having a material proportion of the requester's decrypted data inputs.

### **3.1.5 Incentives**

There are several key economic incentives that will be built into the platform. The team acknowledges that these incentives will require ongoing refinement to achieve an optimally functioning marketplace. For the Amplify Protocol to succeed, requesters must trust that they will receive work completed to a reliably high standard and at a reasonable price. Creating a network where this is true will be the ongoing responsibility of the community

There are three key elements of the Amplify Protocol's incentive structure:

1. Workers will be paid for all work completed and requesters will have to pay for all work completed: this eliminates the potential for abuse by requesters of any ability to not pay for work

2. The potential for workers to abuse the system by submitting low quality work will be mitigated through a rating system and the ability for requesters to only make work available to workers of a sufficient rating. The rating system has been carefully designed to avoid abuse, with key elements including workers starting off with no rating and their rating being biased downward until they have completed a sufficient number of tasks to earn a higher rating. The rating system will include recency bias and pattern bias, including ensuring that the impact of a single bad rating is minimised to protect workers from a single requester having a negative opinion.
3. To align the incentives of the Amplify Protocol’s developers and its key users, the overall stewardship of the protocol will be managed by a governance token (details below).

### **3.1.6 Governance token**

The Amplify Protocol will be governed by a governance token known as the AMP Token. This token will allow for on-chain governance by the community, allowing the Amplify Protocol to be truly decentralised without giving up the ability to coordinate future progress. The governance token will also drive community involvement - this can be enabled through one of the many mechanisms explored by other projects, including creating a space where only authenticated token holders can discuss community decisions.

The governance token will have value because a small fee for use of the core Amplify Protocol will be paid to all holders of the token. A fee, initially set at 5% will be taken by the Amplify Protocol smart contract. The amount of the fee will be controlled by the holders of the governance token and can be changed over time to ensure that it provides a sufficient incentive for work on the protocol without acting as an encumbrance to driving uptake and allowing the community to achieve their goals.

AMP tokens will be issued to Amplify Protocol developers and other team members as compensation for their work on the protocol. This will allow a fully decentralised model where the governance token allows compensation of developers and a sustainable model for ongoing work.

Holders of the AMP token will have the opportunity to vote on key decisions for the protocol, including all changes to the smart contract. Core functionality, and incentives for workers and requesters will be able to be enhanced over time without defeating the decentralisation of the protocol.

## **4 Amplify Protocol off-chain web apps**

The user interface for the marketplace will be a web-based front end. This will be where the requests for labelling are generated from, and where the labelling tasks can be accessed from.

”Requesters” will create requests via the front-end, which will then send a transaction to the program on the blockchain, creating an account that stores the labelling instance. The information

to be labeled is then uploaded to a storage location (eg Arweave). Once the labeling task has been completed, the results will then be stored in the on-chain account and can be pulled by the requested through the front-end.

”Labellers” will be able to query the on-chain program to see a list of outstanding labeling tasks, and then select those they wish to complete. The data will be downloaded from the storage location, and the results of the labelling uploaded to the on-chain account.

To facilitate accurate and reliable labelling, labellers will continually be benchmarked against other labelers on the tasks they complete. This benchmarking will allow for rating of labellers, which will in turn assist with determining fees, number of labellers required per task, etc. Initial benchmarking will be done using a standard set of labelling tasks. In the future, an AI model will perform the cross-validation.

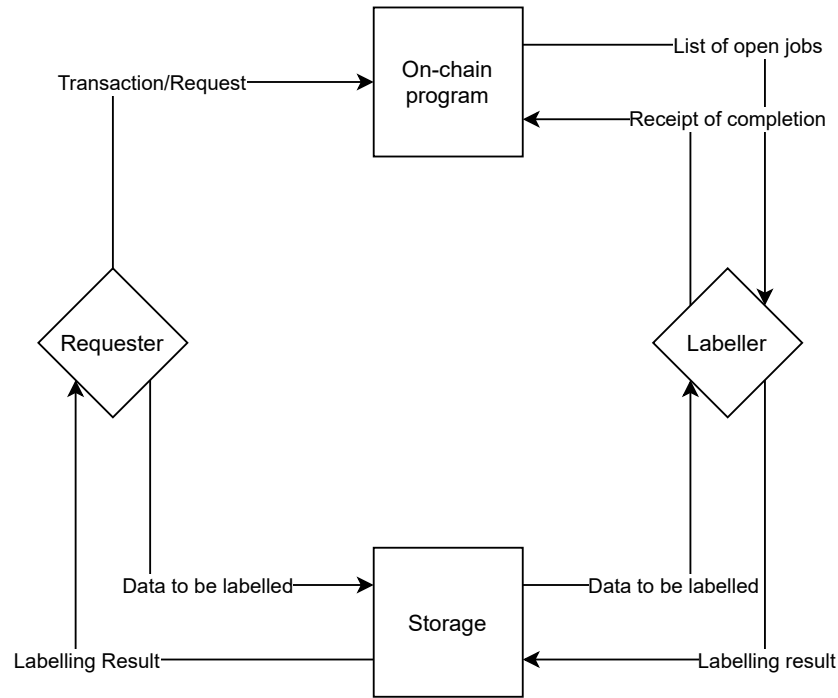


Figure 1: Process flowsheet for Amplify marketplace

#### 4.1 Process flow

1. Researchers/Engineers post labelling tasks programmatically.
2. Labellers can log on ad-hoc to select and complete tasks in real time, their quality is assessed via accordance with other labellers, and % of accepted data by the poster.

3. Community members upload ML models trained on the input/output of posted tasks, which can themselves compete for task completion or labeller validation. Ensembles of these models will provide ‘first guesses’ to accelerate labelling, or entirely replace human labellers.
4. Labellers and model uploaders are rewarded not only for the task they complete, but rewarded a share of total system revenues. This means prolific labellers are rewarded for quality work by long-term rewards, and incentivises contributing powerful models (or easy to use labelling interfaces) to the network. This is crucial to ensuring the proceeds of powerful AI models are spread more widely

## 5 Further use cases for the Amplify Protocol

We have proposed data labelling and content moderation as the initial use cases for the Amplify Protocol but we strongly believe that the protocol has a large range of further use cases, including many that may not become obvious until the protocol is available and the community can experiment with what new opportunities it makes possible.

A few initial ideas for further use cases that could be served by the Amplify Protocol include:

- Expanding the scope of the marketplace to ”content generators” offering the services of their AI systems for a range of tasks
- IP protection (identification of IP breaches, particularly for visual artists)
- Specialist tasks (legal etc)
- Moderation of the platform is already being looked at through a partnership with Grape Protocol