## **1. The problem our project solves**

The current influx of corona-related fake news is unprecedented in terms of how widespread such content is being shared (geographically) and in terms of the real harm it is causing, right now, to real people and communities all over the world.

There are many potential strategies for tackling fake news, but before any other action can be taken, **step 1. is identifying fake news**.

Current approaches for identifying fake news rely mainly on teams of professional fact checkers. The problem is:

* these teams can't cope with the huge influx of fake news - meaning a lot of content is gaining traction before being flagged
* these teams are typically based in only a handful of (highly developed) countries and can only check content in a limited number of languages

**FakeMash addresses these issues.**

****

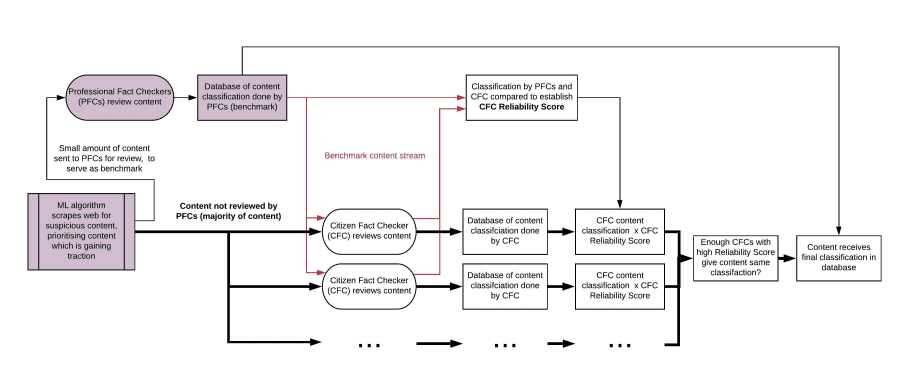
## **2. What is FakeMash?**

FakeMash is a community-driven platform for identifying fake news, where anyone can contribute to tackling misinformation! The platform will:

* allow us to vastly expand efforts / personnel available for identifying fake news, thereby ensuring that such content is flagged quickly, before gaining widespread traction;
* ensure this effort is carried out worldwide and in all languages, rather than being concentrated in the most developed nations.

FakeMash works as follows:

1. Users **sign up** to the platform, and indicate what “categories” of fake news they want to review (coronavirus, climate change, …) and the languages in which they want to review content.
2. Users are taken to the platform’s Fact Check page, where they receive a short interactive tutorial, and can **select their first online content to review**. The content which we offer to the user to review, is determined by existing Machine Learning algorithms for identifying “suspicious content” (ML algorithms are already good at flagging suspicious content, but not as good as humans at accurately determining whether content is in fact fake news).
3. After taking their time to read and fact check the content, they **decide whether the content should be classified as “Legitimate”, “Satire”, “Misleading” or “Fake News”**.
4. Their submission is added to our database, and the user can pick the next content to review. Unbeknownst to the user, some content which he/she reviews has already been fact checked by a professional fact checker. This is “Benchmark Content” which is used to determine the **user’s “Reliability Score”** (based on how well their content classifications correlate with the content classifications chosen by professional fact checkers). As such, the intention is \_ not \_ to let users “democratically decide” what content should or shouldn’t be considered as fake news, since this would lead to community bias, and would expose the platform to trolls wanting to ‘game’ the system. Rather, **the intention is for the platform to serve as an extension / amplification of the work already being carried out by professional fact checkers.**
5. Once different users have reviewed the same article, their classifications are accumulated. Based on each user’s Reliability Score, and on the variance between various users’ responses, **our algorithm determines whether we have enough reliable reviews to give the content its final classification**. If so, it is moved to the final content database, and will no longer be shown to users to review.



That is FakeMash's core functionality. Additional platform features include:

* **Gamification**: users can climb the leaderboards, obtain badges for different achievements, make teams, unlock new features as they progress, ...
* Allowing users to **submit “new” fake news content** which they have found.
* A **Platform Metrics page** where users can see how they rank compared to other users, and see what the platform has accomplished so far.
* A **Community Resources page** which users can use to improve their fact-checking skills.

## **3. Our project's impact to the crisis**

Over the past couple of months, the Internet has been flooded by fake news and conspiracy theories regarding the coronavirus. We have all seen claims that the virus was intentionally developed by Bill Gates for personal profit, or that medical professionals from Europe and North America are using the situation to test harmful vaccines on African populations, or that the virus does not even exist, but is a story fabricated by the supposed “deep state” as a premise to exert ever more control over its helpless citizens.

These sorts of false claims undermine trust in legitimate governments trying their best to tackle the crisis and they lead to a further erosion of confidence in real science and experts. Ultimately, corona-related misinformation is putting thousands of lives at risk.

At the same time, if the past few weeks have shown us anything, it’s that in times of crisis communities come together and individuals step up to the challenge in order to help in any way they can, whether it be through the [sewing of homemade face masks](https://www.forbes.com/sites/tjmccue/2020/03/20/calling-all-people-who-sew-and-make-you-can-help-solve-2020-n95-type-mask-shortage/#48b8339d4e41), volunteering their time to [health services](https://www.bbc.com/news/uk-52029877) or to [community support groups](https://www.bbc.com/news/uk-england-51978388), putting their computer resources at work to [calculate potential remedies to the coronavirus](https://www.forbes.com/sites/jasonevangelho/2020/03/16/the-graphics-card-in-your-pc-can-help-fight-coronavirus/#6b98418c1704), or indeed, [monitoring online groups in order to remove harmful content](https://www.bbc.com/news/blogs-trending-52149568).

Our platform empowers regular citizens to take action and contribute to tackling this "infodemic", in order to ensure that fake news is identified quickly, and action can be taken to curb its spread - before it causes harm.

#weareinthistogether

The current influx of corona-related fake news is what inspired this project, and it is an issue which must be tackled with urgency. However, other types of fake news - such as fake news related to elections or to climate change - will continue to be a problem long after this pandemic has passed, and our platform can be used to identify fake news related to any topic. Furthermore, the platform gives users the option to choose what “category” of fake news they are interested in reviewing, thereby ensuring that content related to the “most pressing issue” at any given time, will be the content which is reviewed with most urgency.

As our fake news database grows, it could also be used to train Machine Learning algorithms to become better at accurately identifying fake news content (rather than just flagging “suspicious content” which must then be reviewed by humans), thereby progressively reducing our reliance on human fact checkers.

## **4. What’s next?**

Our project provides a relatively simple solution to a complex problem. We participated in the Hackathon with the intention of putting this solution in the spotlight, and demonstrating community interest. Now that we have won the Hackathon, there are several potential paths for taking this project forward and making the platform viable long-term. In the current Matchathon phase we look forward to exploring these options with potential partners (see “What partnerships are we looking for?” below).

**Option 1: Implementation of this solution by a large tech company (or consortium of companies)**

Companies such as Google, Facebook and Twitter have received a lot of criticism in recent months regarding the amount of corona-related fake news which continues to spread through their platforms (see for instance [these articles](https://www.ft.com/content/4d9a408c-79a4-4407-83e5-0ce9e49a0a1a;%20https://edition.cnn.com/2020/05/08/tech/covid-viral-misinformation/index.html;%20https://www.buzzfeednews.com/article/ryanhatesthis/the-most-popular-youtube-videos-about-the-coronavirus-are)). At the same time, in a [joint statement](https://mashable.com/article/google-facebook-twitter-coronavirus-fake-news/?europe=true), they have expressed their willingness to address this issue together. This is momentum they can build upon.

There are several reasons why implementing a solution such as the one we propose would be in these companies’ interest:

* It would demonstrate to users, to shareholders and to the public that these companies are in fact serious about tackling fake news.
* These companies are understandably concerned about being perceived as the "policemen" of the Internet - large tech companies deciding what users should or shouldn't see online. In this regard a community-driven platform for classifying fake news (whose users receive a Reliability Score based on their correlation with professional fact checkers) would be a perfect "middle ground". The companies could say: look, our policies for mitigating the spread of fake news are informed by the classifications provided by a diverse and decentralised, worldwide community of fact checkers, not by some "black box" algorithm which we developed.
* As we all know, the relationship between “big tech” and the EU isn’t always ideal. Partnering up with a young Hackathon team, with the backing of the EU, would be an easy PR win for these companies.

Furthermore, in terms of ensuring the long-term viability and usefulness of this sort of platform, there are several advantages to involving large tech companies in its implementation:

* These companies already have all the “components” needed to make the platform functional, such as a) algorithms which scrape the web for content, and which can identify "suspicious content" which is gaining traction and which needs to be reviewed by humans and b) partnerships with professional fact checking organisations, whose input the platform would need in order to provide "benchmark data" for determining public users' Reliability Scores.
* These companies have the financial and human resources needed to keep the platform running and to scale it as required.
* Having the solution implemented by a large tech company would ensure that the platform has a lot of visibility and a large user-base from the get-go, thereby amplifying its impact.
* These companies are potential “end users” of the platform, in the sense that they could use the resulting “fake news database” to inform their strategies for tackling the spread of fake news.

**Option 2: Integrating this solution with an existing fact-checking website**

Various (non-profit) organisations, such as the International Fact-Checking Network, Full Fact, NewsGuard and Snopes, are already hard at work fact-checking news. Our solution would amplify / accelerate / streamline the excellent work they are already doing. Additionally, integrating the platform with an existing organisation would again help to ensure the platform has more visibility and a larger user-base from the start. Furthermore, these fact-checking organisations already have one of the components needed to make the platform functional, namely benchmark data which has been reviewed by professional fact checkers (to determine public users’ reliability score). They also have significant experience and insight working with fake news, and will have a better idea of for instance which content classification system (reliable, misleading, parody, fake news, etc.) should be used.

**Option 3: Establishing a non-profit charged with developing and maintaining the platform**

The non-profit could fund the operation and upkeep of the platform through donations or (government) grants. Additionally, we could consider limited monetisation of the platform. Certain brands or companies also suffer the consequences of fake news, so we could charge corporate entities for access to our fake news database so they can implement more targeted strategies for tackling fake news related to their brand or company + we could generate ad revenue by allowing non-intrusive ads on our platform (such as ads linking to legitimate news sources such as the WHO, or ads related to other initiatives for tackling fake news).

**Option 4: Continue developing this platform as an open-source project**

Rather than establishing a dedicated non-profit entity, another approach would be to open-source the project. We have developed it further in the past 2 weeks, and will continue to update our GitHub repository as we develop, but we can also make "requests" on GitHub for people to look at specific aspects of the project. Members of existing fact-checking communities have already expressed interest in participating. A community-driven development process for a community-driven platform for solving a public problem - there is a certain elegance to this approach.

## **5. What partnerships are we looking for?**

In consultation with potential partners, we look forward to making our future plans more concrete in the coming weeks. Partnerships we are looking for include:

* The platform would need an **"input feed" of content** (scraped from the web) **which has been flagged as "suspicious"** and which is gaining traction (which then needs to be fact-checked by humans). We can try to develop this from scratch, but we would first like to reach out to **companies such as Google, Twitter and Facebook**, since they already have such algorithms in place.
* The platform would need a **database of content which has already been reviewed by professional fact checkers** (to use as benchmark data to establish the Reliability Score of our public users). **Organisations such as the International Fact-Checking Network, Full Fact, NewsGuard and Snopes** could help us with this.
* In terms of "**end-users**" of our content database (who can then use the content classifications to inform their strategies for dealing with fake news) we are thinking of **social media companies, news websites, content aggregators (such as news feeds), as well as other legitimate parties such as NATO, the WHO, government entities and researchers from academia** who might have an interest in using / studying the resulting “fake news database”.
* **We would welcome input from academia** regarding issues such as a) which content classification system would be best to use (at the moment we picked "reliable", "parody", "misleading" and "fake news" - but these can be changed, and the final classification system should be grounded in evidence), b) best-practices regarding algorithms for determining user reliability score / determining when certain online content has gotten enough reviews from reliable users to receive its final classification in the content database, c) any insights from fake news / conspiracy theory researchers or people with experience in "user engagement" (how to ensure that people keep using the platform and form an active community).
* **In terms of funding**, we are interested in getting in touch with **any company / organisation / government entity willing to contribute to tackling this "infodemic", without necessarily expecting a Return on Investment** (since at the moment we consider monetisation of the platform only as a "last resort" option). The EU has already informed us that most of the Hackathon partners they have talked to are not looking for ROI, and just want to help out, which is great. Many parties stand to gain from less fake news (whether corona-related or otherwise), but obtaining funding to provide such a “public good” isn’t always easy.
* In terms of concrete support we would like to request from official Hackathon Partners, this includes: **Adobe Stock credits** (for front-end design), **AWS credits and AWS support** (for back-end implementation) and **Google Ad credits** (for marketing).

## **Evaluation Criteria (Hackathon phase)**

**Impact Potential**

Compared to current approaches for identifying fake news, our solution would ensure that fake news is identified significantly faster (before gaining widespread traction) and that this identification is carried out worldwide, in all languages (rather than being concentrated in the most developed nations).

Any number of users could sign up to the platform, and given that millions of people are currently stuck at home (due to lockdowns), itching at any opportunity to help out, we believe interest would be high (as demonstrated by enormous public engagement with other crowd-sourced solutions to tackling the corona pandemic, such as [Folding@Home](https://www.anandtech.com/show/15661/folding-at-home-reaches-exascale-1000000000000000000-operations-per-second-for-covid-19)).

The resulting database of fake news content would enable companies such as Google, Facebook or Twitter, as well as news websites and content aggregators, to then take appropriate and timely action to limit the spread of such harmful content. Furthermore, as our fake news database grows, it could also be used to train Machine Learning algorithms to become better at accurately identifying fake news content (rather than just flagging “suspicious content” which must then be reviewed by humans), thereby progressively reducing our reliance on human fact checkers.

Despite the apparent simplicity of the concept, a community-driven approach to identifying fake news has received surprisingly little attention. Projects that have followed this approach have run into issues with a) community bias, which raises questions concerning the validity of the classification results and b) lack of user engagement. Our platform overcomes these issues by a) correlating user input with professional fact checker input in order to determine user reliability scores and b) having a very simple review submission procedure (no need to write a report) + platform gamification.

**Technical Complexity & Novelty**

Rather than opting for an unnecessarily complex solution, we are offering a relatively simple solution to a complex problem. The main conceptual innovation is the simple idea of correlating user input with professional fact checker input, in order to determine user reliability scores. This effectively makes the platform an extension / amplification of the excellent work already being carried out by professional fact checkers (rather than it being a forum for “democratically deciding” whether any given content should be considered legitimate or fake news).

**Prototype Completion**

We have a functioning front-end UI prototype, as well as solid back-end algorithms for a) determining user reliability scores (from their input’s correlation with expert fact checker input) and b) determining when we have enough user input (from users with a sufficiently high reliability score) to give content its final classification as either legitimate news, satire, misleading or fake news.

What is missing is the connection between front-end and back-end. We have built a preliminary implementation of the UI in Drupal, and have a good idea of how to integrate front-end and back-end on this platform, however we are open to exploring other options (notably in consultation with our potential future teammates from “DetectiveCollective”). Ideally, rather than relying on a platform such as Drupal, we would work with a front end developer to port the UI design to HTML / CSS and then integrate it with the back-end.

Update: We have now engaged several front-end designers and developers. The design will be implemented in HTML / CSS, and integrated with the necessary databases on AWS.

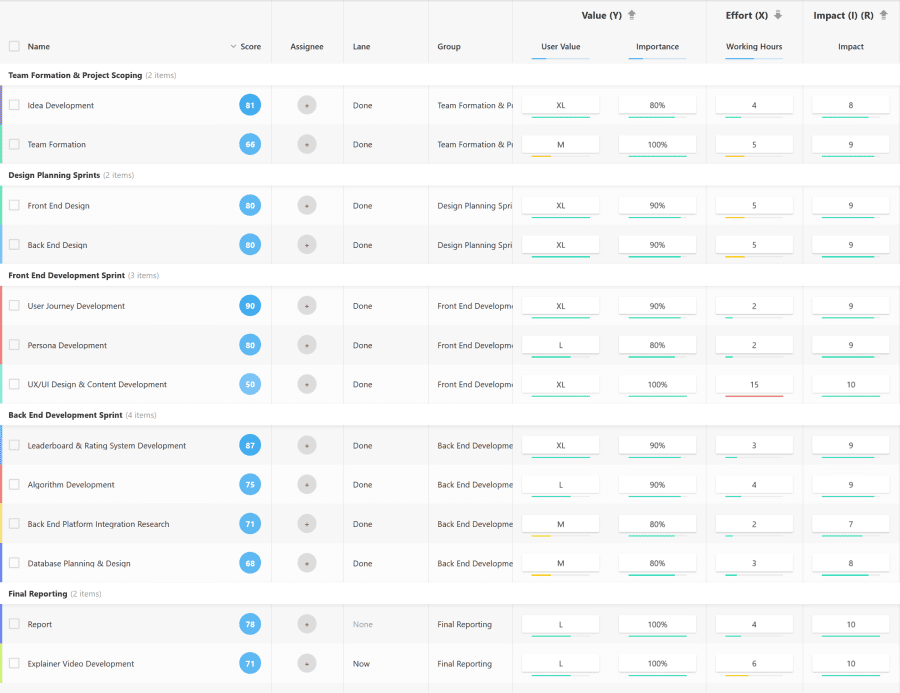
**Business Plan**

(see: “What’s next?” above)

## **What we did during the weekend**

The core concept was brainstormed beforehand and the team was assembled through the Slack channels on Friday. Everything else was done over the weekend: fleshing out the concept, front end and back end design planning, user persona creation, user journey creation, algorithm development, database design, UI design and content development, deciding on a platform name, designing a logo, concept testing, creating the final submission post, creating the submission video.

Tools used: front-end user interface designed in Adobe XD / algorithms for determining user Reliability Score + final article classification coded in Python, SQLlite / libraries: Pandas, Faker / preliminary front and back-end integration done in Drupal



## **Main Critiques (and Responses) to a Community-Driven Approach for Identifying Fake News**

**A. Political or ideological bias will pervade a community-driven response if community members collectively lean one way or the other on a political or ideological spectrum.**

=> Response: Correlating user-input with professional fact checkers would ensure that a community-driven approach is no more, or less, biased than the current approach.

**B. Fake news is produced so abundantly, and spreads so rapidly to different platforms that it is impossible to curb.**

=> Response: Producing believable fake news does in fact require quite a lot of creative thinking and time. Fake news is being spread by a relatively small number of individuals, and the means at their disposal for spreading such content are relatively basic. Fake news content is spread to other platforms simply by copy pasting content, or by reproducing it with minimal alterations. The "bad side" thus has a relatively labour-intensive task at hand, and has relatively unsophisticated means for spreading fake news. Meanwhile, the number of "good people" capable of recognising fake news vastly outnumber the people creating fake news, and identifying (often glaringly obvious) fake news is a lot less labour intensive than producing it. As such, if the Internet Giants throw their capital and resources (such as machine learning algorithms) behind a community-driven approach to identifying fake news, the means that the "good side" has available for combating fake news are vastly greater and more sophisticated than the resources the "bad side" has available. By all measures, this should be an unfair fight, with the advantage being for the "good side".

**C. Who are we to decide what users should or shouldn't view? What about free speech?**

=> Response 1: We currently already accept that professional fact checkers can to some extent decide what content should be promoted, and what content should be limited. It is not unethical for companies such as Facebook, Google or Twitter to take the stance that they wish to promote reliable news, and limit the spread of fake or harmful corona-related fake news.

=> Response 2: A community-driven approach does not have to be black or white: there are many gradations between obviously malicious and harmful fake news, and completely reliable news. Other designations such as "conspiracy theory", "parody", "misleading news", ... can be used, with different response strategies being appropriate for each category.

## **Built With**

* adobexd
* [animoto](https://devpost.com/software/built-with/animoto)
* drupal
* faker
* [pandas](https://devpost.com/software/built-with/pandas)
* [python](https://devpost.com/software/built-with/python)
* [sqlite](https://devpost.com/software/built-with/sqlite)