## **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.**

****

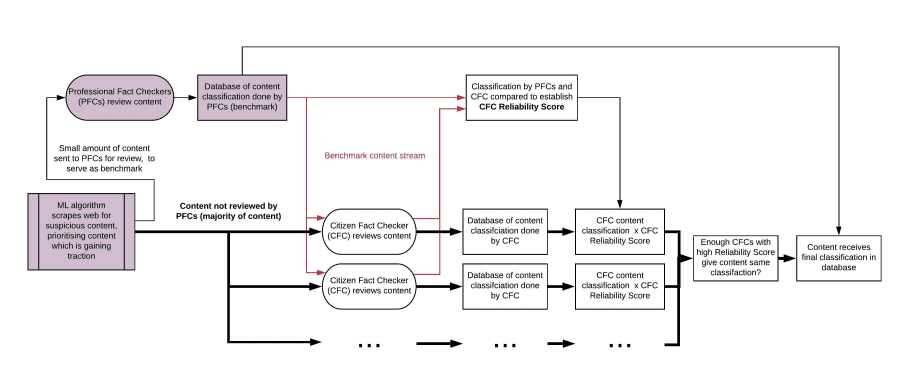
## **2. What is FakeMash?**

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.



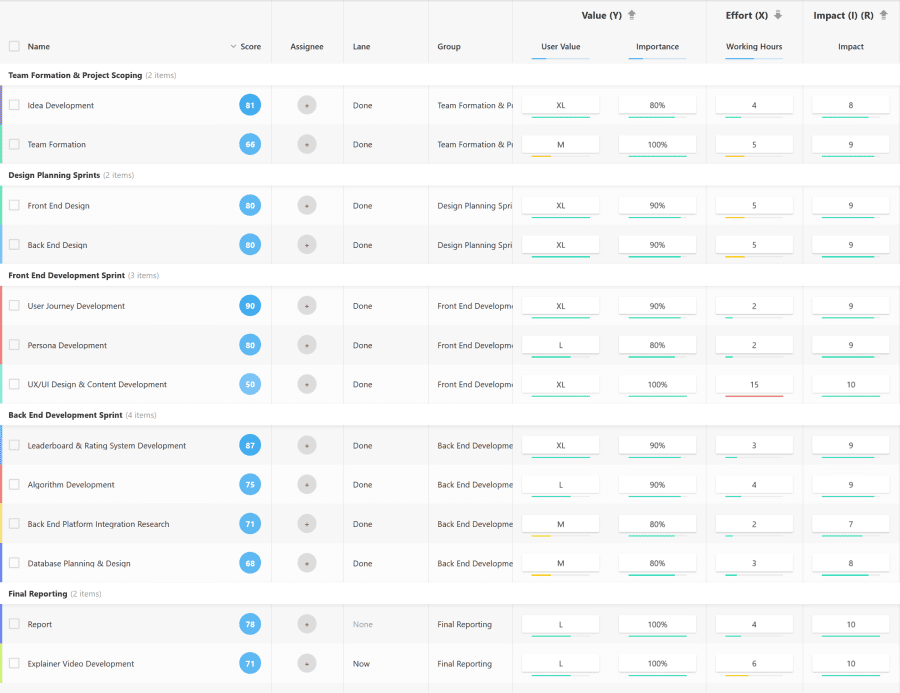
That is FakeMash's core functionality. Additional platform features include:

* **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.

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

## **3. 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.



## **4. 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 corona virus. 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, they lead to a further erosion of confidence in real science and experts, and ultimately, they may result in thousands of additional corona deaths, which could have been prevented.

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

## **5. What's next for our project?**

We have been in contact with the team at “[DetectiveCollective](https://devpost.com/software/detective-collective)”, who are developing a similar community-driven solution to identifying fake news. Both teams want to collaborate on developing this concept further, beyond the Hackathon.

Once the platform is ready, it will be hooked up to existing Machine Learning algorithms which flag “suspicious content” (content which can then be presented to our users for reviewing), as well as to databases of content which has already been reviewed by professional fact checkers (to provide benchmark data in order to establish public users’ reliability score).

Regarding databases of content checked by professional fact checkers, as a starting point we have been in touch with members of the “[Journalistes Solidaires](https://devpost.com/software/journalistes-solidaires)” team, and they have expressed interest in providing such data.

## **6. The value of our solution after the crisis**

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.

## **Evaluation Criteria**

**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, 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.

**Business Plan**

Once developed, there are three possible paths for keeping the platform viable long-term:

* The most straightforward path is for a company (or partnership of companies), such as Google, Facebook or Twitter, to take charge of the platform. They have the capital and developers needed to keep the platform running and to scale it as required. Furthermore, they already have the necessary components which must be hooked up to the platform, namely a) Machine Learning algorithms for flagging suspicious content and b) teams of professional fact checkers whose input can be correlated with that of public users. These companies are currently receiving a lot of criticism for the amount of (corona-related) fake news which continues to spread through their platforms. Launching an effective solution for identifying fake news would therefore be in their interest, as it would allow them a) to more efficiently limit the spread of fake news on their platforms, and b) demonstrate to their users, their investors and to the public that they are making an effort to step up their response to fake news.
* The second path is to establish a non-profit charged with maintaining the platform. The non-profit could fund the operation and upkeep of the platform through donations or (government) grants.
* The third path is to fund the platform by charging a monthly fee to third parties wishing to use our extensive, up-to-date, worldwide fake news database (parties such as social media websites, news websites or governments). However, we consider this the least desirable option, as a) it might alienate public fact checkers, since the platform would effectively be generating income thanks to their volunteer work and b) the fake news database would only be accessible to paying customers, thereby limiting its impact.

## **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)