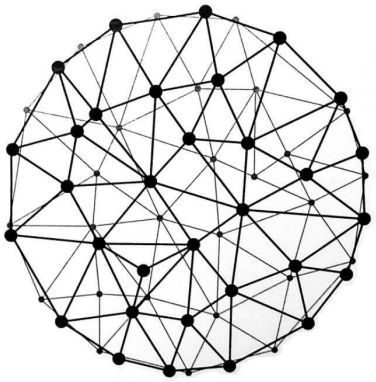
OUTLINE OF PRODUCT

Sav Sidorov, Johann Cooper

1. PROBLEM

The internet has brought with it some fantastic breakthroughs when it comes to the way we communicate with each other. It's has made information sharing easier than ever before. Since then, some individuals and organizations have emerged and become more popular than others, and as a result have gained access to a larger audience. They've become influential beacons of information. All interconnected like nodes in a giant network.



Despite all the good the advent of mass communication has brought us, it's becoming harder to parse truth from falsehood, bias and credibility. There is no accountability or incentives to optimize for truth and good reputation, if it comes at the cost of clicks and user engagement. If we can't make sense of what's going on in the world, progress on the technological front becomes difficult. The information haze also poses an existential risk: if we have tools that can deal great damage to humanity (nukes, AI, etc...) but are unable to have an effective public conversation, things get risky pretty fast.

2. THE 10X IMPROVEMENT

Proposal: a 10x improvement over our current methods of sensemaking. A system that tracks credibility of individuals (Nodes) and publications through a crowdsourced process. The hope

is that a system like this provides the right incentive structure for Nodes and publications to improve their coverage and commentary on world news. This in turn improves the quality of our collective sensemaking. This looks something like:

10x Improvement → Incentive structure towards greater Node credibility → Healthier collective sensemaking.

This leads to the mission statement:

> *"To provide a positive incentive structure for propagating information, with the goal of facilitating healthier collective sensemaking"*

Our current methods include a) Organizations like Politifact, The Knife Media b) Organizations like Rootclaim c) Algorithmic solutions à la Google and Facebook d) Committee as suggested by Andrew Yang. All of these have flaws that we would like to fix.

To have a 10x improvement, the product has to have (at least) the following features:

- **Scalability** — Politifact and TKM operate(d) by having a team of editors. Solely relying on trained editors parse information for bias is not scalable.

- **Agenda-free** — Curating solely by a team of editors lets in the biases of the group of editors (eg. Politifact, TKM, Yang's proposal). You run into the same issue if you have an algorithm analyze fake news and credibility for you. Not only is it hard to determine exactly how it works, but what is there to ensure that the ideological biases of the algorithm managers don't seep in?

- **Intuitive and easy to use** — Rootclaim, similar to the Wikipedia backend has a complicated UI. Simplicity attracts users.

- **Accessibility should not sacrifice depth** — one of the strengths of a website like Rootclaim is the depth of information that it allows. Info must be presented in an intuitive way, low res to high res, and not just a low res at-a-glance view.

- **Emergent editor board** — at some point you need a team of people to curate content on the website (look to Wikipedia). The team needs to emerge from the rules of the

system (i.e some kind of score) instead of being appointed.

- **Gamification** — could be a very promising way to maintain engagement and potentially monetize the service. The Duolingo model comes to mind: streaks, profile costumes and other cosmetics.

- **Accountability** — a means by which the individuals (Nodes) who rate other Nodes are held accountable. Rating cannot by anonymous. Raters themselves can be rated.

3. DETAILED SOLUTION

The proposed app has several core features:

1. A Rating System
2. An Editorial Board
3. Browser Extension
4. Fact Gathering and Evaluating System (down the line)

They all work together in the following way:

Everything is crowdsourced. If we manage to get a diverse cross section of the public using the website, this would help eliminate bias as seen in other approaches, and is at the same time incredibly scalable.

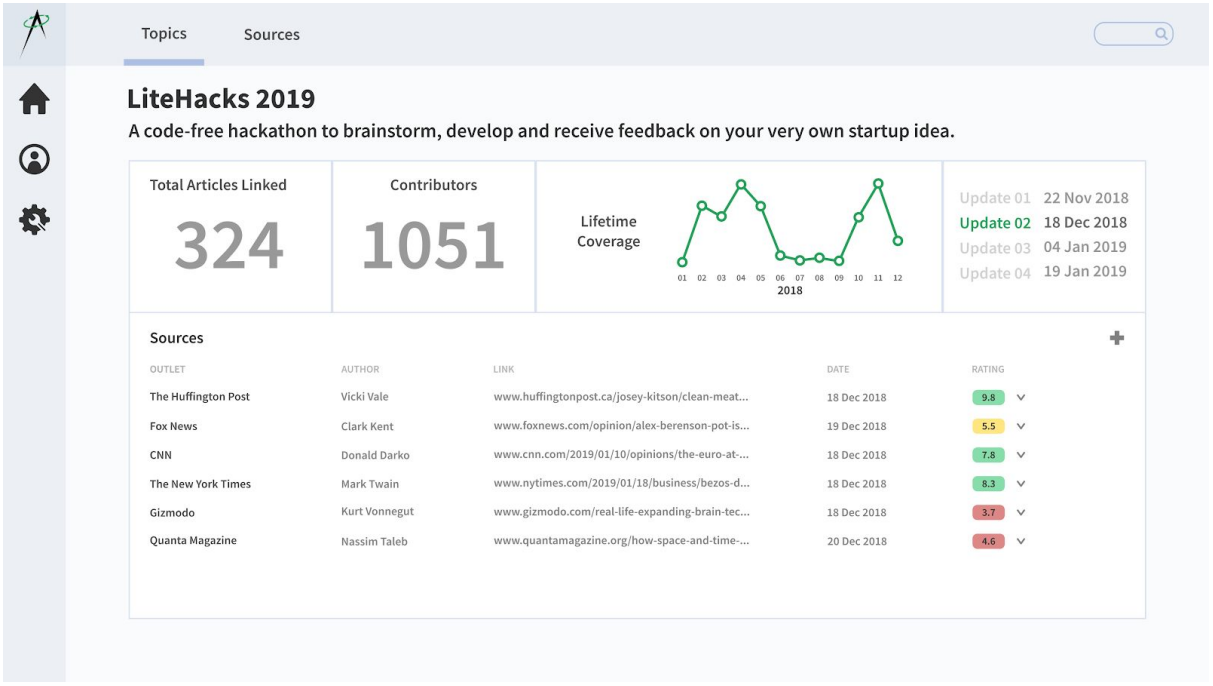
Users can create new Nodes, articles, topics and publications. Emphasis to be placed on Nodes and articles. Articles can be rated, and the aggregate article rating affects the credibility of the Node and publication. Users are themselves Nodes and can also be rated.

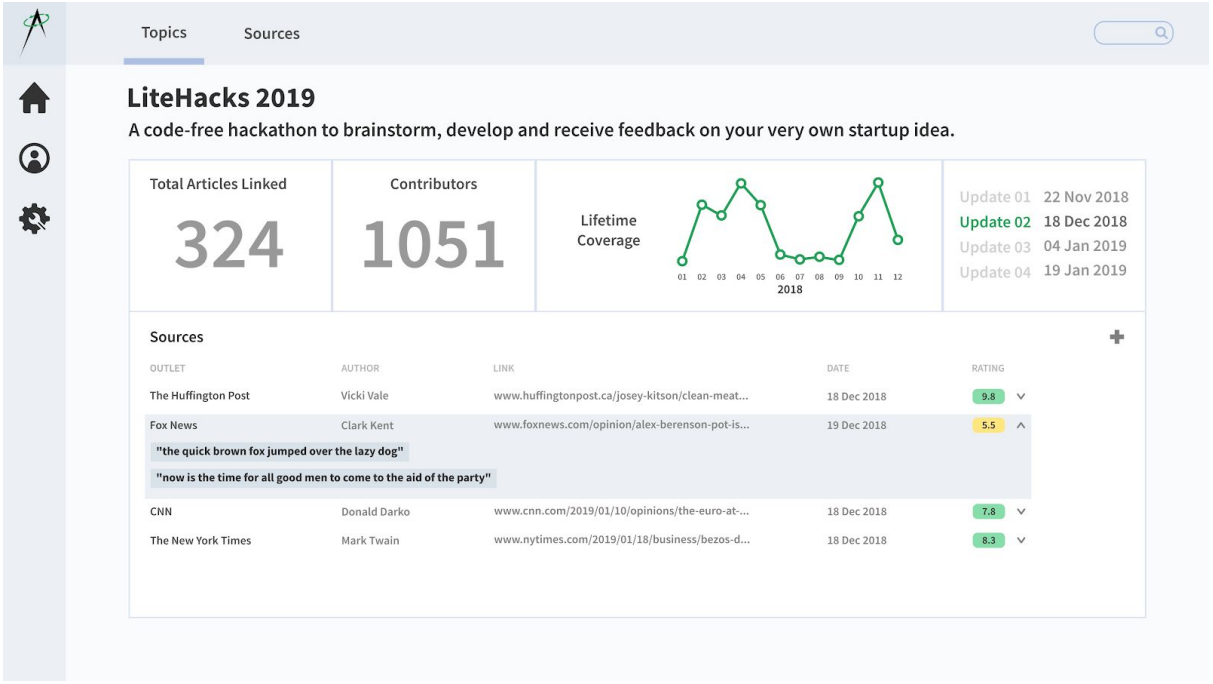
An editorial board curates content and prevents vandalism, similar to the way Wikipedia operates.

The browser extension is used to highlight Fact Coverage and Bias snippets, and shows the current rating of the web page you're on.

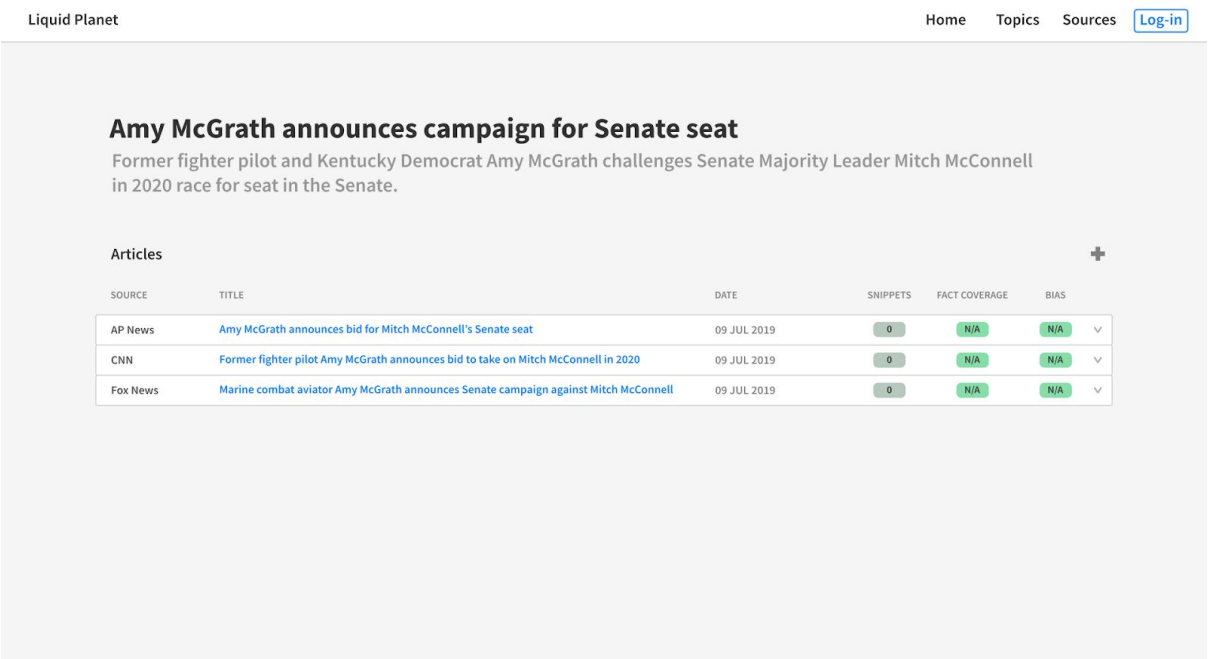
Eventually, it might be worth having an advanced system of fact evaluation to promote high-quality theorizing and analysis of information.

Examples of topic and source page UI designs below:

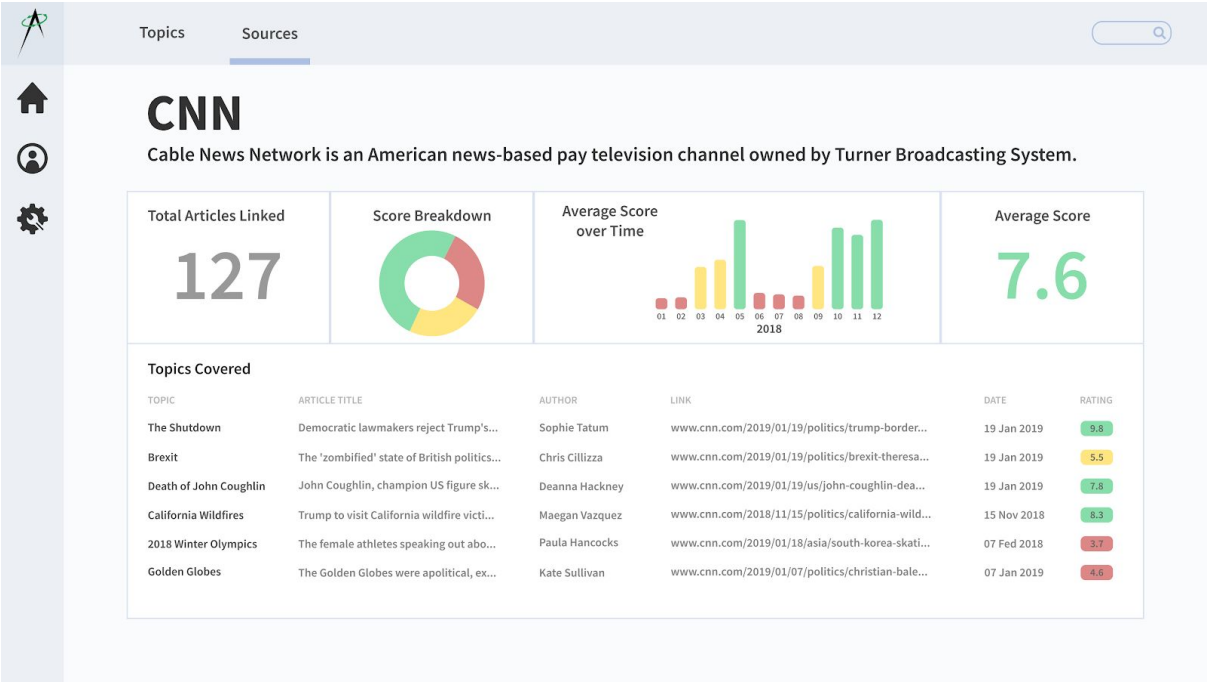




Topic page with a list of articles (showing dropdown functionality, displaying snippets).



Topic page with a list of articles (simplified alternative).



Publication source page.

3.1 THE RATING SYSTEM

THE THREE PILLARS

There are three main components to my model: **Rating, Fact Coverage, Bias**. There need to be different options for how users interact with the system—more surface level and more

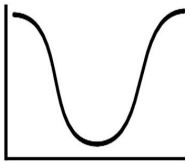
in-depth. There is some value to be gained from users who want to simply rate articles, but do not want to go through the whole process of finding snippets, evaluating on bias, etc.

3.1.1. RATING

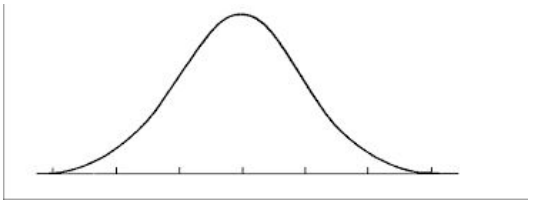
A simple 1-5 or 1-10 or 1-100 rating system for users who do not have the time to utilize more in-depth features. It's important to account for power laws—for example, something like 0.3% of Wikipedia users are active editors. Therefore we need a feature that is quick, simple, and encourages more casual users to contribute.

A couple things to note with this:

- We need to somehow distinguish between "divisive" and "consensus" ratings. In other words, two items can have the same average but completely different distributions. A "divisive" item has fat tails where ratings stack up on either side, a "consensus" item looks more like a conventional bell curve. Maybe taking the median into account can solve this?



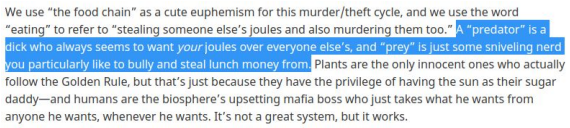
"Divisive", Average: 50%



"Consensus", Average: 50%

3.1.2 FACT COVERAGE

Imagine that, when you hold down a key—say for example **[2]**—you enter Fact Coverage mode. Similar to Portal's aesthetics, when you highlight with the left mouse button (LMB), you designate the highlighted text as a fact (blue highlights). When you highlight with the right mouse button (RMB), you designate the following as a falsehood (orange highlights).



Fact

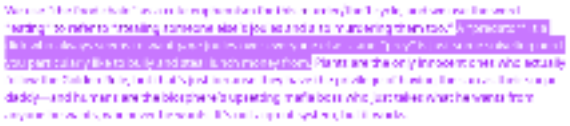
Falsehood

Fact coverage can be based on a **scale of -1 to 1**, where -1 is completely false and 1 is completely factual. Perhaps the ratio of highlights to the total word count can be converted into a metric. But, there are probably smarter ways of quantifying this information.

There might be technical limitations to this (can you ever use RMB to highlight?) but we can come up with workarounds.

3.1.3 BIAS

Bias highlighting can work in a similar fashion. If you hold down **[3]**, you enter Bias mode. Highlights with LMB are designated as positive emotive conjugates (purple highlights), highlights with RMB are designated as negative emotive conjugates (red highlights).



Positive Emotive Conjugation

Negative Emotive Conjugation

Bias can be based on a **scale of 0 to 1**, and it’s lower the fewer highlights there are on a page. Perhaps the bias rating on our website can have a background with a purple-red gradient to symbolize the overall slant, positive or negative? This color metric would be independent from the 0 to 1 rating scale. Examples:

0.75	0.3	0.58

3.1.4 OVERALL SCORE

All three components will then be somehow be put together and contribute to the overall score of the source. This score will be time sensitive, where recent ratings for Nodes will be weighed heavier than older ones. Article ratings will not be time sensitive.

AN EXAMPLE ANALYSIS BY THE RATING SYSTEM CAN BE SEEN IN THE FURTHER READING SECTION.

3.2 EDITORIAL BOARD

There needs to be some kind of board—not unlike websites like Wikipedia or Genius—that curates content and prevents vandalism. Content curation will involve verifying topics, articles, snippet highlights, etc to make sure that these features are used as intended.

It's important that there be a system in place for selecting user eligible for editor board membership. The ingroup should not be allowed to gatekeep and prevent new users from joining. What's nice about the crowdsourcing approach is that the editor board can scale with the size of the user base. As more users perform actions and become more respected, they get promoted to editor status.

Potentially user accounts could operate on a points system, where the more points you have the more rights you have on the website. This would prevent vandals from creating new accounts, and would help verify people as who they say they are. It's hard for a bot to accumulate a high point score.

3.3 BROWSER EXTENSION

The browser extension will facilitate Fact Coverage and Bias snippet highlighting (for example how **Highly** does it), and will also give you information about the page you're currently on.

In the right hand corner, it will show: the current status of the page you're on (aggregate Rating, Fact Coverage and Bias), and the credibility rating of the author (Node).

3.4 FACT EVALUATION

A more down-the-line feature is an advanced system of crowdsourced fact evaluation. Intended to promote high-quality theorizing and analysis of information, similar to **Rootclaim** but with a better UI (might even be worth reaching out to the Rootclaim guys for help). Once the website has an established user base, a feature like this can be made available for the users to try.

Could be a very good way of piecing together a news story. There are barriers in online filter bubbles, as well as in different countries (people speak different languages, etc) that prevent people from seeing all facets of what's going on. With a large enough global user base, a system like this could prove very beneficial.

However, while very useful in the long run, I don't think it's a good idea to have a feature like this available out of the box. You need a big, diverse user base for this to work. Besides, this is somewhat tangent to our mission statement of determining Node credibility.

4. THE BUSINESS SIDE

4.1 MARKETING

- Word of mouth
- Social media and viral campaigns
- Posters
- Reach out to influential people concerned with this issue

4.2 BUSINESS MODEL

- Donations
- Fundraising
- Selling gamified cosmetics à la Duolingo (?)

4.3 MARKET RESEARCH

Is there demand for this? Yes, appears so (see pictures).

Elon Musk @elonmusk · 23 May 2018

Create a media credibility rating site (that also flags propaganda botnets)

88% Yes, this would be good

12% No, media are awesome

681,097 votes • Final results

4.5K 13K 39K

WikiLeaks @wikileaks · 23 May 2018

Replying to @elonmusk

WikiLeaks, which has a 100% record of authenticity, C-Span and other primary source sites should be at the top of such rankings. We have long wanted to do this also, but such a system is likely to be gamed by the most powerful player in inter-state conflicts.

97 481 2.9K

WikiLeaks @wikileaks · 23 May 2018

Here is a very well documented example, where the Israeli state uses soldiers and paid student troll farms against the much weaker Palestine: israelpalestinenews.org/israel-partisa…

Russia vs Ukraine is another well documented example. See also the UK's JTRIG.

How Israel and its partisans work to censor the Internet

Numerous well funded, organized projects by and for Israel work to flood social media with pro-Israel propaganda, while blocking facts Israel dislikes

israelpalestinenews.org

24 190 489

5. UNRESOLVED ISSUES

Listed in no particular order. Important to note that we're always going to have unresolved issues. The goal is to keep moving, implement features and solve these problems along the way.

- **The User Base Problem:** How do we get a healthy cross section of the population to use this product? If we have a slanted, or biased, user base then the product is no good (just look to Gab). This is one of the hardest problems we face.

- **Negative Rating Asymmetry:** From the Rating System post: "It's ***maybe*** worth considering negative rating asymmetry, where negative ratings hurt a source more than positive ratings help. This goes back to the analogy that "a drop of poison in water" is different than "a drop of water in poison", or something like a convicted murderer is the same as a regular citizen *except* for one day. Not sure though, my thoughts aren't that concrete on this issue."

    - Especially considering the fact that—for example—most news on a website are passable and mundane, yet there are rare high-profile fuck-ups. These rare events is what we want to be focussed on.

    - Shifting the focus from publications to individual Nodes might solve this issue without the need for asymmetry.

- How do we **handle different mediums** of information, such as: podcasts, videos and pictures of text?

- Would **short form content** (such as Tweets) and **long form content** (Books, Reports) need to be handled differently? Does scale matter?

- **How do you explain *why* you tagged a piece of information** as biased, or factual, etc..? I'm of the opinion that we should keep commenting and text-based replies to a minimum (it's more intuitive that way, and minimizes potential for vandalism). Could we have users select from a variety of tags already in the system, such as "bad science",

"one-sided", etc..? Would having a text box where people explain their thoughts be the way to go (this is harder ro police)?

- How do you combat the **negative side effects that come with metrics**? I don't remember the exact quote or who said it, but it's something like "When you're using metrics to measure your progress to a goal, the moment that optimizing the metric *becomes the goal* it stops being a good metric". How do we set up an incentive structure that leads people to optimize for good sensemaking and good journalism instead of the metrics?
    - An extreme example of this happened a few years back with the game Fallout: New Vegas. The developers were promised a cash bonus if they reached a certain score on Metacritic. They ended up being one point shy, and did not receive the cash bonus. *Not* something you want to facilitate or incentivise.

- How should **quotes by other people** in a given article be analyzed? If the quote is biased, it should not be attributed to the author of the article. However, selecting quotes that favor one side of an argument *should* be attributed to the author, as it is a property of the article itself.

- How to deal with **updates to articles**? How would highlighted content keep up with updates to stories? Let's say a highlighted string gets deleted from the original article. What then?

- There are **more dimensions to bias than just emotive conjugates**. How do we represent those dimensions, if at all? Things like omitting certain facts, relying on a particular set of sources for information and quotes, and so on.

- How much **emphasis** should be placed on credibility of individuals, as opposed to credibility of outlets and publishers?

- Need a **more robust solution to bots** than just a user points system.

6. FURTHER READING

- Rating System: Example Analysis
- Emotive Conjugation