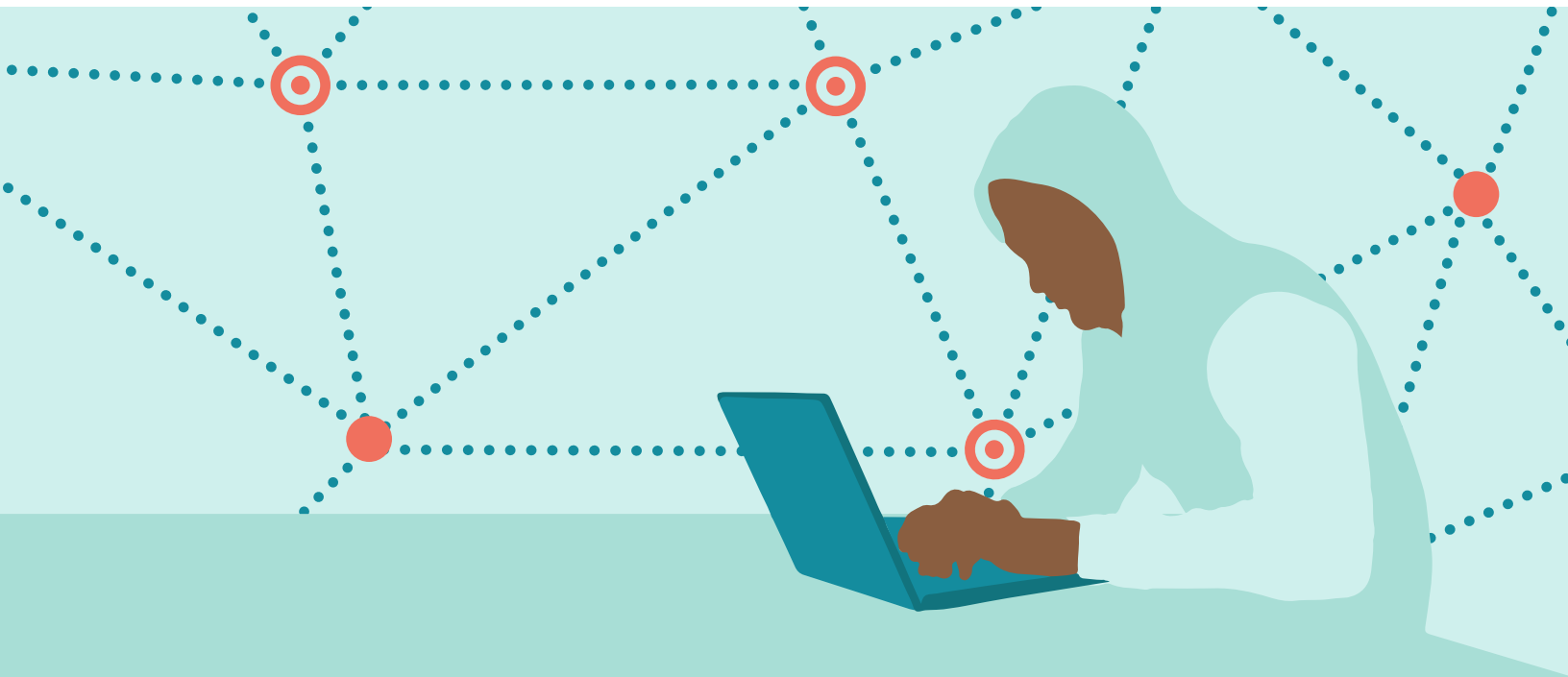


POLICY BRIEF

'Contextualization Engines' can fight misinformation without censorship

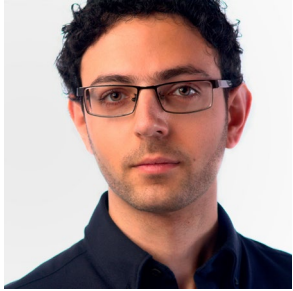
Search engines are nice. But we can do far better with modern AI.



—
By: Aviv Ovadya

TTP

cybersecure
policy
exchange
Powered by RBC



Aviv Ovadya (@metaviv)

is an internationally recognized expert on the societal implications of emerging technology, with a focus on internet platforms and artificial intelligence. His underlying goal, as aptly described in *The New York Times*, is to “incentivize...systemic changes” — to help create an information ecosystem that facilitates understanding and trust. Aviv writes, builds, advises, and conducts research as the founder of the Thoughtful Technology Project, as a non-resident fellow at the Alliance for Securing Democracy, and through his consulting work.

How to Cite this Brief:

Aviv Ovadya. (2021). 'Contextualization Engines' can fight misinformation without censorship. Cybersecure Policy Exchange.

<https://www.cybersecurepolicy.ca/policy-brief-contextualization-engines>



This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-nc-sa/4.0/). You are free to share, copy and redistribute this material provided you: give appropriate credit; do not use the material for commercial purposes; do not apply legal terms or technological measures that legally restrict others from doing anything the license permits; and if you remix, transform, or build upon the material, you must distribute your contributions under the same licence, indicate if changes were made, and not suggest the licensor endorses you or your use.

This work was supported by Cohere and with an honorarium from the Cybersecure Policy Exchange. You can find an HTML version of this brief at <https://www.cybersecurepolicy.ca/policy-brief-contextualization-engines> and at <http://thoughtfultech.org/context-engines>.

Cybersecure Policy Exchange: <https://www.cybersecurepolicy.ca/>

[@cyberpolicyx](https://twitter.com/cyberpolicyx)

[@cyberpolicyx](https://www.facebook.com/cyberpolicyx)

[Cybersecure Policy Exchange](https://www.linkedin.com/company/cybersecure-policy-exchange)

The Future of Content Platforms. *Hopefully.*¹

Inputs • Criteria • Outputs

Dominant Organizations³

Dominant Nations³

Search engines ~2000s
select relevant content using keywords.²



Recommendation engines ~2010s
select engaging content by correlating content and user behavior.



Contextualization engines ~2020s
select and integrate meaningful content through a semantic understanding of context and users.



[1] Contextualization Engines are not a prediction. They are a necessity. They may not be sufficiently incentivized without public support. But without them, democracy may not survive technological disruptions. We can incorporate contextualization features into existing systems, and that is beginning, but it is unlikely to be sufficient unless the purpose of those systems also changes.

[2] While the primary input is keywords, search can also involve, for example: implicit reputation signals, user behavior, user environment, etc.

[3] Examples of well-known organizations deploying such systems and the nations that have the most power over them.

Executive Summary

Search engines transformed the first decade of the millennium.

Recommendation engines revolutionized the second decade. *Neither in their current form are sufficient for addressing misinformation.* They focus on **discovery** and primarily rely on *relevance*. But they are not particularly helpful at many other important information tasks, particularly **contextualization**.

We need better tools to help people quickly contextualize media that they come across online. This is especially important for supporting busy everyday people needing to rapidly make sense of the misinformation-laden text, images, and videos shared in group chats and online platforms.

Contextualization engines can help do the media literacy grunt work for you, *SIFTing* through the internet to identify what you might want to know to better understand what some 'media object' means. Just as people now use the term 'Googling' to describe using any search engine to find information about some keywords; we might use the unique term *contextify* to describe using a contextualization engine to make sense of a media object.

In an ideal world, one could ‘contextify’ any sort of media object with a tap or two from any app, whether it be a long chain message, an image meme, an article link, a video, a PDF, an audio file, etc. The goal of the contextualization engine would be to help the user answer key questions like: *“What does this mean? How meaningful is this? How does this relate to the things I know about and care about?”* Implicit in those questions is whether or not the claims are accurate — but contextualization at its ideal goes further, helping a person get at what actually matters to them and their relation to society.

Systems built specifically for contextualization might not only support media literacy; they could also provide the data needed for fact-checkers to determine what to focus on, and could even help support the emotional literacy relevant to avoid harmful reactions to misinformation (from lashing out at loved ones, to terrorist radicalization).

Moreover, contextualization systems do not involve any sort of external censorship or message monitoring — they embody **the “more speech” approach** to countering misinformation, by *providing a place for users to optionally access context and authoritative speech when they want it*. This means that such functionality could even be required by law, just as nutrition facts are for foods. It might be required for all messaging and social media apps over a certain size, built into operating systems, or mandated through Apple App and Google Play stores. At minimum, the public sector could help **fund** the rapid development of contextualization engines, given the current market failure and the potential of such systems to protect against current and emerging threats to democracy and the financial system.

Recent advances in artificial intelligence have made powerful contextualization engines possible — but the same advances also enable terrifying new methods for spreading misinformation. We must race against time to **build contextualization capacity** given the stronger geopolitical and market incentives to use these new AI technologies for self-serving propaganda and profit.

What a contextualization engine might look like in practice

Imagine that you were forwarded a terrifying message in a group chat. Or saw a post shared on Facebook which made you furious at some news organization. But something seems a tiny bit fishy...

Option A: Without a contextualization engine

While you would like to know if the claims are really true — and you may “want” to look it up ... you just don’t have time for that sort of thing. It’s easier to just go with the flow. It’s also a giant pain to copy and paste things or type out many search terms trying to figure out if someone else is just confused — especially on a phone. *So you don’t check.*

It remains in your memory, as something *perhaps* true — but you may forget the ‘perhaps’ with time. If you see enough similar messages, maybe you start to instinctively believe them — and then you may start sharing those messages also.

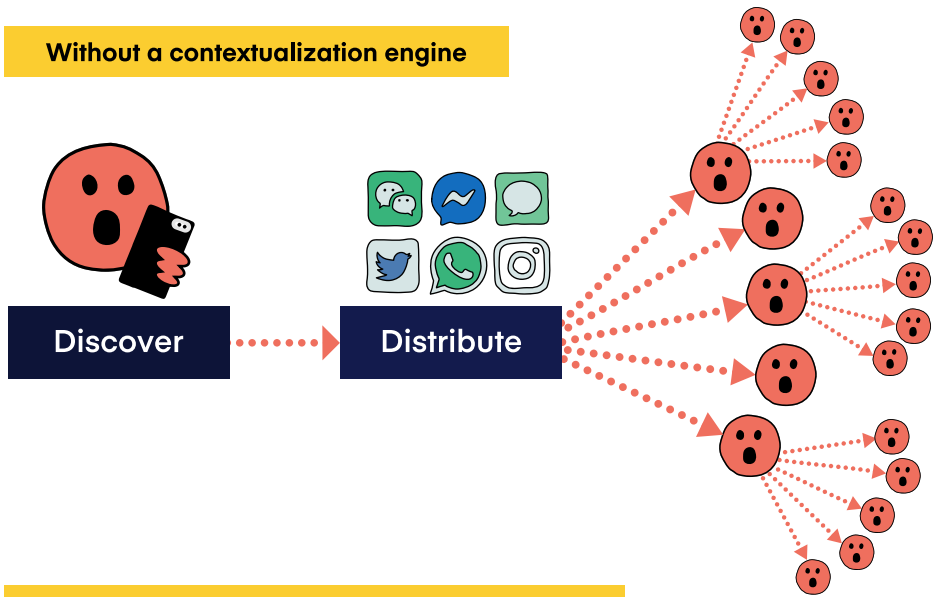
Option B: With a (very basic) contextualization engine

You see something that looks fishy — and tap a button to ‘contextify’ it...

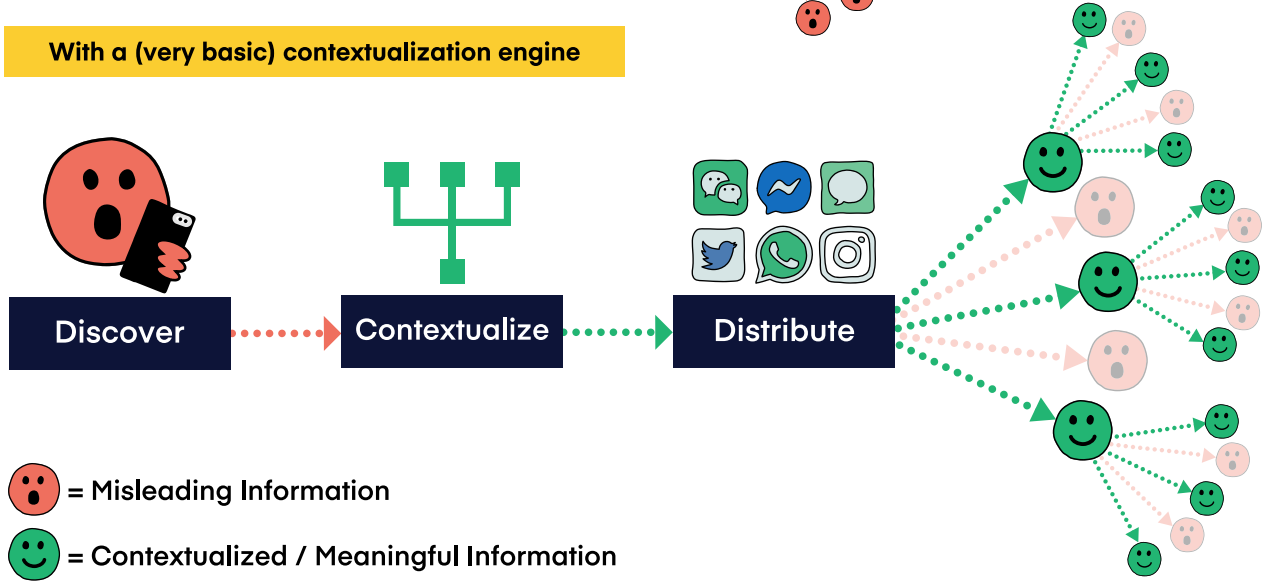
1. The contextualization engine compares the content being shared with that from *authoritative* sources and provides articles or other media results that are sufficiently related. This might be in a search result style interface, through a chatbot, or a hybrid.



2. If it finds no close enough matches, it *warns the user* and potentially identifies the [most likely relevant keywords](#) that the user can run a more traditional search with if they would like (with another tap).
3. It adds the media object to a triage queue for relevant organizations to potentially evaluate (e.g., fact-checkers).

Without a contextualization engine



With a (very basic) contextualization engine



 = Misleading Information
 = Contextualized / Meaningful Information

The 'contextify button' is a drop-in for [WhatsApp's magnifying glass feature](#) — but the results are very different. WhatsApp creates a keyword search for a traditional search engine, [which can backfire badly due to data voids](#).

This magnifying glass feature on WhatsApp was a valuable step forward. But it doesn't currently work in practice in [many cases](#). It makes it easier to look up messages on Google, but keyword search doesn't work for long messages, images, videos, audio, with data voids, etc. We need more tools **designed for contextualization**.

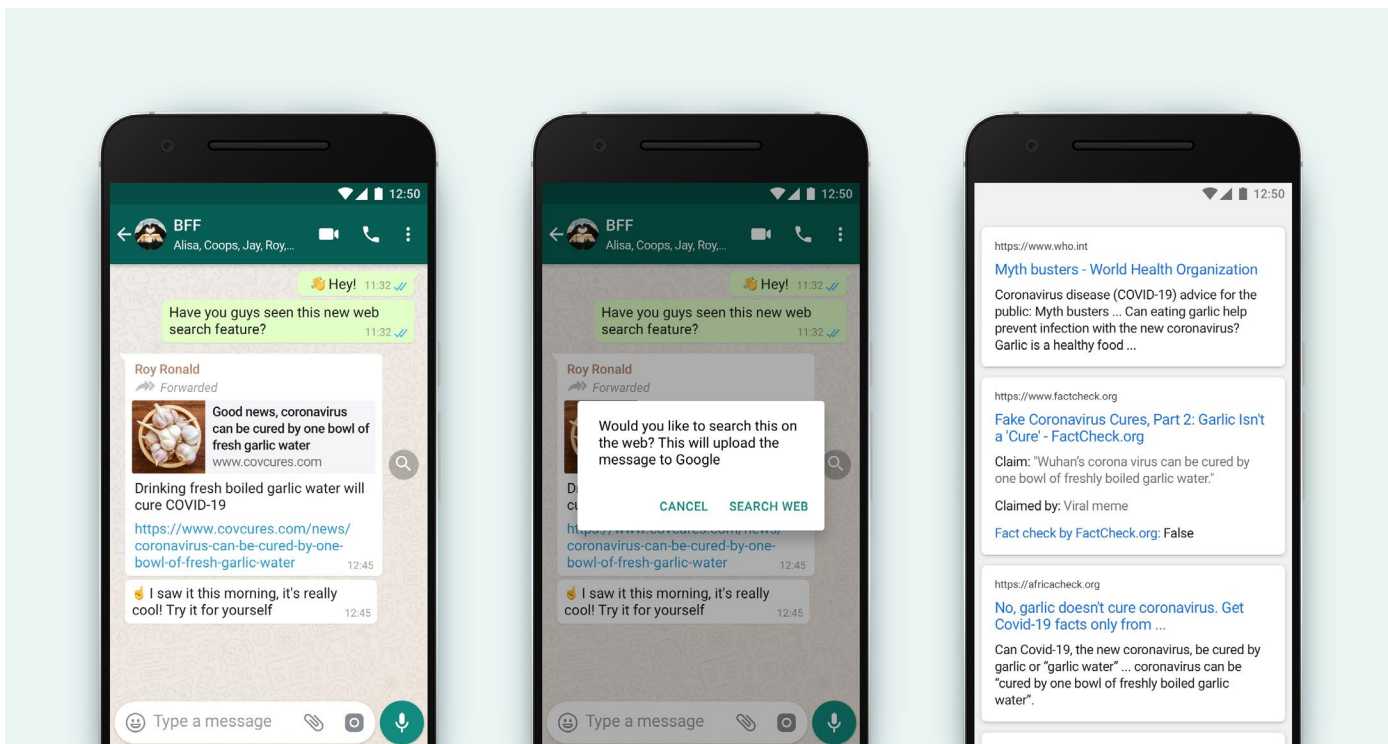


Image Source: [WhatsApp](#)

Why Even the Basic Contextualization Engine Helps

Key Insights: Unlike a Google keyword search, ‘contextifying’ does several crucial things:

- *Analyzes* complete ‘media objects’ — to see how likely they are to be related to one another; e.g., the entire chain message, entire fact-check articles.
- *Focuses* on authoritative sources — likely initially using whitelist certification through recognized 3rd parties, such as the [International Fact-Checking Network](#) (IFCN), [First Draft](#), [News Guard](#), standards organizations, etc.
- *Warns* about [data voids](#) — lets the user know if the system can’t find good information on the topic.
- *Supports* the people doing deeper investigations — provides the human fact-checkers and other organizations with information about what is important to explore — and potentially revenue from web traffic in ways that are directly aligned with the users’ goals.

These all support the “**F**ind better coverage” component of the SIFT media literacy method — an approach developed by [Michael Caulfield](#), inspired by research at Stanford, and taught by many educational and civic organizations, from the University of Washington in the US, to [Civix](#) across Canada.

Contextualization systems can be even more helpful

This is just the beginning of the potential for contextualization engines and interfaces. A contextualization system might also support the remainder of the SIFT method:

- **Stop (SIFT):** The contextualization engine flow can provide educational support for executing other aspects of media literacy. For example, it can help remind users to pause and notice their emotional reactions to the content. It might even provide tips on how to bring up the potential misinformation in a delicate way in a group chat or comment thread.
- **Investigate the source (SIFT):** If the contextualization system already has information on why a source might be considered authoritative, it can provide that information to the user — *showing why they might trust it* (e.g., this source is certified by IFCN).
- **Find better coverage (SIFT):** Building on the ‘analyze’ component described earlier, a more fully featured contextualization engine would not only auto-generate audio and video transcripts from media, but also automatically interpret any imagery and captions in order to better understand the content and find contextually relevant sources.
- **Trace claims, quotes, and media to the original context (SIFT):** Finally, the contextualization engine can *do the tracing for the user*. It can essentially scour the web for the original context of any content.

None of this requires any new technology — this isn’t science fiction — though it is only recently that this sort of analysis has become effective and practically feasible. Some aspects of these suggestions have started being integrated in small ways into existing platforms, for example by excerpting contextual snippets from Wikipedia, but contextualization still does not appear to be their focus.

The potential — and risks — of artificial intelligence advances

While recent advances in artificial intelligence make a ‘contextify button’ possible, imminent advances will also make contextualization systems critically important to address threats to democracy and financial systems. Deepfake videos, incredibly effective [AI-optimized phishing attacks](#), and automated troll armies may become pervasive — and *indistinguishable* from the real thing by an ordinary person.

Thankfully, the same technology that is creating these threats — powerful new language understanding and generation systems — can also help support contextualization to counter them. These AI advances will enable software to directly answer those key questions for users: “*What does this mean? How does it relate to the things I know about and care about?*” These powerful language systems can be used to help **translate** jargon, such as from scientific papers and legal documents, into writing and images that everyday people can understand and apply. AI advances are even enabling the creation of systems that could automatically **integrate** content from multiple authoritative sources to generate helpful mini-essays and engaging animated videos (this could be technically possible within the year—with very significant investment).

Such systems will need to navigate a challenging terrain of bias, information quality, misuse, and privacy, especially as they extend beyond the domains of authoritative sources. We must fund research and responsibility infrastructure to ensure that this revolutionary potential is applied wisely — while maintaining a bias for action given the clear negative impacts of moving too slowly.

How can we make this happen?

A “contextify button” to push media to a contextualization engine could be built into *everything* — just a normal and expected part of the interfaces for viewing and sharing content.

But such systems do not quite exist yet — they face a chicken-and-egg problem where it is challenging to get traction unless existing platforms buy into them, but platforms will not adopt them until the contextualization systems have traction. Funders and investors know this, and so it is difficult to raise the funds to hire the necessary talent. This has left us **many years behind where we need to be given *current* and *emerging* threats.**

Recent developments, such as [Meedan](#)’s work with WhatsApp to develop chatbots for fact-checking and contextualization are a valuable step in the right direction — but platform integration and funding for such work pales in comparison with platform integration and funding for systems that (often unintentionally) facilitate deception. **To accelerate the development of contextualization systems, policymakers may need to create *usage mandates and provide rapidly-deployable public sector funding*** — ideally including dedicated funding for responsible deployment.

The first two decades of the millennium were dominated by search and recommendation engines, bringing Google, Facebook, and Amazon to prominence. We now have a chance to innovate — building the contextualization engines that could define this third decade of the millennium — and perhaps help address the harms of the blundering tech giants.

SUMMARY

Opportunity, Threats and Recommendations

Opportunity:

- Contextualization engines can help people make sense of the information that they come across online by making ‘media literacy work’ less time-consuming.
- A “contextualize button” to push media to a contextualization engine could be built into everything — from social media and messaging apps to operating systems; a normal and expected part of software interfaces.
- New AI technology allows computers to *understand* media in ways that were not possible before, allowing contextualization engines to more effectively sort through information.

Emerging Threats

- The same AI advances are also making it much easier to generate misleading media, threatening democracy and financial systems.
- Deepfake videos, AI-optimized phishing attacks, and automated troll armies could become pervasive — and have been shown to be almost indistinguishable from the real thing by ordinary people.

Government Recommendations

- Rapidly deploy public sector funding to support the development of contextualization engines.
- Include dedicated funding for responsible deployment and related research.
- Potentially require platforms, app stores guidelines, or operating systems to provide an interface (“button”) to allow users to quickly send media to a contextualization engine for evaluation.