***Below is a blog post I wrote with Sara Gerke to be published on the*** [***Bill of Health***](http://blog.petrieflom.law.harvard.edu/)***, the Petrie-Flom Center for Health Law Policy’s blog***

Walking her bike across an Arizona road, a woman stares into the headlights of an [autonomous vehicle](https://www.bloomberg.com/news/articles/2018-03-21/for-self-driving-cars-seeing-everything-isn-t-always-enough) as it mistakenly speeds towards her. In a nearby health center, a [computer program analyzes](https://www.healthcareitnews.com/news/university-iowa-healthcare-rolls-out-first-autonomous-ai-diagnostic-system-cleared-fda) images of a diabetic man’s retina to detect damaged blood vessels and suggests that he be referred to a specialist for further evaluation – his clinician did not need to interpret the images. Meanwhile, an unmanned [drone](http://www.flyzipline.com/) zips through Rwandan forests, delivering life-saving vaccines to an undersupplied hospital in a rural village.

From public safety to diagnostics to the global medical supply chain, artificial intelligence (AI) systems are increasingly making decisions about our health. Legislative action will be required to address these innovations and ensure they improve wellbeing safely and fairly.

In order to draft new national laws or international guidelines, we will first need a definition of what constitutes artificial intelligence. While the examples above underscore the need of such a definition, they also illustrate the difficulty of this task: What is uniquely common between self-driving cars, diagnostic tools, and drones?

**What is AI?**

In 1955, four giants of 20th century computer science⎯[John McCarthy](https://en.wikipedia.org/wiki/John_McCarthy_(computer_scientist)), [Marvin Minsky](https://en.wikipedia.org/wiki/Marvin_Minsky), [Nathaniel Rochester](https://en.wikipedia.org/wiki/Nathaniel_Rochester_(computer_scientist)), and [Claude Shannon](https://en.wikipedia.org/wiki/Claude_Shannon)⎯proposed that they could simulate intelligence in a machine. This proposition led to the [*Dartmouth Summer Research Project on Artificial Intelligence*](http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf), a summer-long brainstorming session on the topic, thus coining the term “artificial intelligence”.

Since then, the term AI has entered the common [lexicon](https://books.google.com/ngrams/graph?content=artificial+intelligence&year_start=1800&year_end=2018&corpus=15&smoothing=3&share=&direct_url=t1%3B%2Cartificial%20intelligence%3B%2Cc0). But if you ask someone to define AI, you will get no universal answer. One of the key challenges in legislating AI will be to define it.

[Several bills related to AI](https://www.congress.gov/search?q=%7B%22congress%22%3A%22115%22%2C%22source%22%3A%22legislation%22%2C%22search%22%3A%22artificial%20intelligence%22%7D&searchResultViewType=expanded) have been introduced into Congress over the past 15 months. Most discuss AI without defining it. However, the [FUTURE of Artificial Intelligence Act of 2017](https://www.congress.gov/bill/115th-congress/house-bill/4625/text), the [AI JOBS Act of 2018](https://www.congress.gov/bill/115th-congress/house-bill/4829/text), and the [National Security Commission Artificial Intelligence Act of 2018](https://www.congress.gov/bill/115th-congress/house-bill/5356/text), contain explicit definitions. Strikingly, the three offer similar explanations.

**A Modern Approach**

The three bills base their definitions largely on Stuart Russell’s and Peter Norvig’s textbook [*Artificial Intelligence: A Modern Approach*](https://digitaletext.com/shop/artificial-intelligence-a-modern-approach-3rd-edition-ebook-pdf/?gclid=CjwKCAjw39reBRBJEiwAO1m0OdIaUi0uBdhkxmCfRyxrhJDl7ACDCX_i4Pnwn_UDYhpuObwRigWJ5RoCmhQQAvD_BwE). Russell and Norvig classify AI into four categories:

1. Thinking Humanly,
2. Thinking Rationally,
3. Acting Humanly, or

Acting Rationally.

The first and third categories follow a *human-centered approach*, whereas the second and fourth categories follow a *rationalist approach*. In contrast to humans who (unfortunately for us) make many mistakes, systems are rational if they do the “right thing”, given what they know. Both approaches deal on the one hand with *thought processes and reasoning* (the first two categories) and on the other hand with *behavior* (the last two categories).

However, these four categories are insufficient.

**Thinking humanly or rationally**

Russell and Norvig explain that a machine would think like a human if a correct theory of the human mind were implemented in code. Alternatively, a machine would think rationally if it used logical reasoning to determine its behaviour.

We argue that neither case can truly be described as thinking. The major challenge in defining AI in terms of how machines think is that such a claim is unfalsifiable. In the famous essay, “[Computing Machinery and Intelligence](https://www.csee.umbc.edu/courses/471/papers/turing.pdf)”, [Alan Turing](https://en.wikipedia.org/wiki/Alan_Turing) concludes that since we can never actually get inside someone else’s head, the assumption that other humans think is simply a “polite convention”. If we cannot prove that another human is thinking, how could we do so for a machine?

We will not go into the years of controversy following Turing’s paper. The point is that, regardless of whether machines can actually think or not, the burden to prove such activity is too high since doing so appears to be impossible. More concretely, self-driving cars, diagnostic tools, and drones cannot be said to be thinking, but they should still be classified as AI.

The bills cited above provide [neural networks](https://www.youtube.com/watch?v=aircAruvnKk) as an example of systems that think like humans, but this is incorrect. While a neural network may have been inspired by brain cells, they are much closer to fitting a line to data than anything resembling cognitive activity. Even if one were to perfectly replicate the brain in code, the machine would be imitating the *behaviour* of the brain rather than *thinking* like a human. One would still have no method of proving that the neural network thinks.

**Acting humanly**

While defining AI in terms of acting instead of thinking is more verifiable, it suffers from similar problems. Following Russell and Norvig, if we make the strong requirement that a computer’s actions must be indistinguishable from human actions, we arrive at Turing’s definition of computer intelligence (known as the [Turing test](https://en.wikipedia.org/wiki/Turing_test)). No system has passed this test to date despite several AI developers who [claim the opposite](https://www.cnet.com/news/alphabet-chairman-says-google-duplex-passes-turing-test-in-one-specific-way-io-2018/).

If we lower our standard to say that a machine must act like a human *only in one respect*, then we arrive at the surprising conclusion that most software constitutes AI. For example, consider the calculator. A calculator acts like a human in that it can perform simple arithmetic tasks. In fact, it is much better than us at doing so. Does this mean that a calculator is intelligent? Maybe – but this would mean that future regulation applying to AI would also include calculators, which might not be the legislator’s intention.

**Acting Rationally**

Russell and Norvig define an agent to be “just something that acts” and a rational agent to be one that “acts so as to achieve the best outcome or, when there is uncertainty, the best expected outcome.”

Once again, this criterion is unverifiable. To assess whether an agent is acting to achieve the actual or expected optimal outcome, one would have to know this outcome (or in the case of expectation, the distribution of outcomes) beforehand. Finding the best result in any situation is impossible since there is always the possibility that another course of action would have led to a better result. We can never be sure that autonomous vehicles, diagnostic tools, and drones are acting rationally because we lack a standard of true rationality to compare them to.

**What Should Be Regulated?**

The definitions of AI provided by the U.S. government are first attempts to nail down an elusive concept. Defining AI in terms of thinking humanly, thinking rationally, or acting rationally is too narrow, while identifying AI with respect to acting humanly is too broad.

Though autonomous vehicles, diagnostic tools, and drones may all have some semblance of intelligent behaviour, this post demonstrates that defining precisely what constitutes this intelligence is challenging. Under the precision and clarity demanded by the law, our usual understanding of AI falls apart. Legislators need to consider whether we should anchor regulation on a definition of AI or if a different framework, such as algorithmic decision making or machine learning, would be a better approach.

*This blog post was inspired by the presentation “AI in Drug Discovery and Clinical Trials” by Sara Gerke held on October 24, 2018, at the Conference* [*“Drug Pricing Policies in the United States and Globally: From Development to Delivery.”*](http://petrieflom.law.harvard.edu/events/details/drug-pricing-policies-in-the-united-states-and-globally)

***Below is a memo I wrote for the class Practical Solutions to Technology’s Public Dilemmas at the Harvard Kennedy School***

To: Senator Mark Warner

From: Joshua Feldman

Date: Oct 8, 2018

Subject: Social Media’s Polarization of America

Foreign interference is not our greatest concern when considering social media’s negative effects on democracy — social media hurts American democracy most by **polarizing the electorate**. Social media facilitates the spread of misinformation and promotes divisive content, driving voters further apart.[[1]](#footnote-1),[[2]](#footnote-2) Polarization precludes meaningful discourse and compromise, which are essential to a functioning democracy. These divisions make the U.S. vulnerable to foreign propaganda campaigns, such as those conducted by Russian actors in the 2016 presidential election. To prevent the chasms between Americans from increasing, I make four proposals:

1. **Revise §230 of the Communications Decency Act (CDA)** to make technology companies responsible for what trends on their platforms;
2. **Incentivize moving away from the advertising model** to encourage social media companies not to solely optimize for engagement;
3. **Require increased transparency in curatorial algorithms** to help non-profits, academics, and journalists hold social media firms accountable;
4. **Encourage social media companies to develop voluntary codes of ethics** to provide an official alternative to solely economic decision-making.

Updating §230 to Include Curators

*The Problem:* A driving force of polarization is that political groups increasingly operate with contradicting sets of facts. The incubation and spread of **misinformation** via social media make these divergences possible.[[3]](#footnote-3) Spreading misinformation on social media is easy because §230 of the CDA frees social media companies from legal responsibility for third-party content.

*The Solution:* **Amend §230** of the CDA to **treat social media as a publisher** when a post has **reached a threshold number of users**.

*Arguments:*

* Amendment would force social media to take responsibility for the most **popular** content, which would in turn affect the **greatest number of users**.
* Making social media companies liable for *all* content would encourage the removal of “risky” posts, thus **limiting freedom of speech**. In contrast, the current proposal would encourage critical consideration of what gets **promoted**on these platforms, thus **protecting freedom of speech** while encouraging **responsible curation**.
* Restricting social media liability to the most popular content, as opposed to all content, would make content control **cheaper and easier**.
* The government **cannot control speech**. Controlling curation is the **extent** to which you can attack this problem with direct regulation.[[4]](#footnote-4)

Tax Cuts to Encourage Alternates to the Advertising Model

*The Problem:* Social media’s **advertising-centric business model** leads to the most **extreme content being promoted**.[[5]](#footnote-5)To sell ads, social media companies need users to spend time on their platforms. To entice this user base, social media promotes content that generates the most engagement. Just like with other forms of addiction, short term rewards are often at odds with long term benefits. Since polarizing ideas create larger immediate responses, the need to sell ads leads social media algorithms to promote radical content.

*Solution:* Offer **tax incentives** encouraging **non-advertising business models** for social media.

*Arguments:*

* **Technologists are master optimizers**. If you provide incentives to reduce their reliance on engagement, they will create solutions that we cannot imagine today.
* The detrimental effects of social media on democracy is an externality of the advertising model. Tax incentives would help the **market correct for these external costs**.
* A by-product of having consumers pay for social media is that it would hold these companies **accountable to users rather than advertisers**.

Open Social Media Data

*Problem*: We **do not understand** how social media algorithms respond to misinformation, disinformation, and radical content because these companies **do not willingly share their data**. This lack of transparency makes it difficult for journalists, not-for-profit organizations, and academics to study and identify harmful practices on these sites.

*Solution*: Create legislation that requires large social media companies to **provide certified individuals access to their data**. Such data could include third party content/advertising statistics, anonymized user behaviour data, and information on what causes posts to trend.

*Arguments:*

* Instead of creating costly government watchdogs, this approach **leverages the existing expertise** at journalistic, academic, and public interest organizations.
* Many large social media companies **already have developed APIs** to release data to researchers, although they currently ensure the motive of those individuals are aligned with corporate interests. Moreover, these results are often kept secret.[[6]](#footnote-6),[[7]](#footnote-7)
* Policies such as **robust anonymization**, **updated research ethics**, and **careful approval of access** would be needed to protect user privacy while still allowing for transparency.

Code of Ethics

*Problem*: Employees at tech companies are forced to **prioritize business objectives over ethics**.

*Solution*: Convene tech company leaders and employees in order to **develop a code of ethics**.

*Arguments*:

* An explicit code of ethics would **encourage self-regulation** by technology companies
* It would **empower employees** to speak up against unethical business decisions.
* Compared to other regulation, this option would be **inexpensive** for the government.
* Though this policy alone is **not sufficient** to reign in the polarizing effect of social media, involving the technology companies in the governance process is a **necessary** step.

Conclusion

Implementing these four policies will **catalyze a variety of actors** to prevent social media from **further dividing the American electorate**.

***Below is a link to Insights, an app I helped develop at BlueDot, where I worked most on the natural language processing supporting the infectious disease event identification.***

[*https://insights.bluedot.global/Home/Index*](https://insights.bluedot.global/Home/Index)

*Username:* [*test@test.com*](mailto:test@test.com)*, Password: password*

1. Bakshy E, Messing S, Adamic LA. “Exposure to ideologically diverse news and opinion on Facebook”. *Science*. (2015) [↑](#footnote-ref-1)
2. Sunstein C. *#Republic: Divided Democracy in the Age of Social Media.* (2017) [↑](#footnote-ref-2)
3. Lazer D et al. “The science of fake news”. *Science.* (2018) [↑](#footnote-ref-3)
4. See*, e.g.*, *Reno v. ACLU*, 521 U.S. 844 (1997) (holding government control over “indecent” social media postings as unconstitutional). [↑](#footnote-ref-4)
5. Tufekci Z, “YouTube, The Great Radicalizer”. The New York Times. (2018) [↑](#footnote-ref-5)
6. https://developer.twitter.com/en/docs.html [↑](#footnote-ref-6)
7. https://developers.facebook.com/docs/public\_feed/ [↑](#footnote-ref-7)