

Submission of Evidence to the Enquete-Kommission „Künstliche Intelligenz – Gesellschaftliche Verantwortung und wirtschaftliche, soziale und ökologische Potenziale“

Ben Scott, PhD
Director, Policy & Advocacy
Luminate
10 February 2020

Deutscher Bundestag
Enquete-Kommission
Künstliche Intelligenz
Projektgruppendrucksache
19(27)PG 6-19
TOP 2, 10.02.2020
07.02.2020

Introduction

The social impact of Artificial Intelligence (AI) applied in the digital information marketplace is among the most troubling issues in contemporary politics. There is credible evidence that machine learning systems create profoundly negative effects on the quality and integrity of public debate – spreading conspiracy, hate speech, and organized disinformation campaigns across the Internet. This kind of deception, distortion and corruption in the digital forums for news have the potential to undermine deliberation and decision-making on every issue in democratic society. The technologies of digital media that were once celebrated for empowering new political movements and giving voice to marginalized communities are now blamed for turning the Internet into a vast machine of social manipulation. The stakes are high.

Have we embraced technology companies as enablers of modern democracy only to see them poison it in exchange for monopoly profits? And if this is even partially true, what should governments do about this? These are the questions posed and answered in this submission.

The Critique

The critique of AI in digital media requires some explanation. It begins with the structure of contemporary information markets that are increasingly concentrated in the hands of a small number of companies. Google and Facebook in particular have achieved astonishing dominance over the advertising markets that support information production and the distribution channels that deliver it to consumers. They have accomplished this through the practice of what Harvard Professor Shoshana Zuboff calls “surveillance capitalism.”¹ Through the near-total surveillance of digital behavior online, they have created two valuable assets: 1) detailed data profiles of the preferences and behaviors of billions of people around the globe compiled over many years; and 2) machine-learning algorithms that feed on these profiles as training data to shape information flows in their products.

The objective of the algorithms for information curation across these digital media products is to capture the maximum amount of user attention to sell to advertisers. That is a major change from traditional media systems. The public service editorial judgements that (at least partially) guide most professional newsrooms are not a part of this equation. There are no humans and no suggestion that any part of the goal is to provide people the information they might need as

¹ Shoshana Zuboff, *Surveillance Capitalism*, Profile Books, 2019.

citizens. The entire mission of these automated system is to optimize the time spent on the platform -- the number of clicks, shares, and pageviews that drive advertising sales. In a kind of pincer movement on the user, the social media platforms curate noncommercial (i.e. organic) information flows to optimize for engagement even as they sell targeted advertising designed to maximize influence over preferences. The unimaginable size of the training data, the world-leading investment in AI experts, and the scale of the testing interface delivers efficient results and ever-increasing quarterly returns for the Silicon Valley giants.

These products are free of charge to the individual, but the price we pay as a society is high. The content that is curated in these systems is not simply indulging the worst instincts of every individual without any attempt at serving the collective good. It is actively shaping their behavior and changing their views to make them more predictable and more likely to stay on the platform and keep clicking. Not surprisingly, what keeps people clicking is sensationalism, outrage, and titillation. Over time, this steady diet of extreme views gradually becomes normalized. And once we are habituated to see and engage with this kind of content on a daily basis, we become reliable inputs into a machine that sells access to our attention. Zuboff summarizes the business model poetically: "It is no longer enough to automate information flows about us; the goal now is to automate us."²

Zuboff is not alone in this conclusion. Renowned expert on AI, Stuart Russell, a professor of computer science at University of California at Berkeley, argues that the application of machine learning algorithms in social media is a prime example of an AI system whose objectives are not aligned with the public good. In a recent interview, Russell elaborated on the point:

"The algorithms are not trying to show you the stuff you like. They're trying to turn you into predictable clickers. They seem to have figured out that they can do that by gradually modifying your preferences and they can do that by feeding you material. That's basically, if you think of a spectrum of preferences, it's to one side or the other because they want to drive you to an extreme. At the extremes of the political spectrum or the ecological spectrum or whatever image you want to look at. You're apparently a more predictable clicker and so they can monetize you more effectively. So this is just a consequence of reinforcement learning algorithms that optimize click-through. And in retrospect, we now understand that optimizing click-through was a mistake."³

The industry rejects these kinds of claims, arguing that information curation algorithms merely deliver the content that is most appealing and relevant to the desires of the user. If people click on extreme content, it is because that is what they want and who they are. However, the internal memos and marketing materials from Facebook tell a different story. In a recently

² Quotation from Zuboff in John Naughton, "The goal is to automate us," *The Guardian*, 20 January 2019, <https://www.theguardian.com/technology/2019/jan/20/shoshana-zuboff-age-of-surveillance-capitalism-google-facebook>

³ See transcript of the interview with Lucas Perry, AI Alignment Podcast, 8 October 2019, <https://futureoflife.org/2019/10/08/ai-alignment-podcast-human-compatible-artificial-intelligence-and-the-problem-of-control-with-stuart-russell/?cn-reloaded=1>

leaked document, a corporate executive argued that his company's advertising tools were responsible for Donald Trump's election victory in 2016.⁴ This is not the first claim from Facebook that its combination of targeted ads and AI-curated audiences can manipulate people. In 2017, another leaked Facebook document from its Australian business operation bragged to advertisers that it could predict and respond with the right commercial pitches to the moods of teenagers gleaned from behavioral data.⁵ As far back as 2014, data scientists conducted "emotional contagion" experiments on unsuspecting Facebook users that altered their attitudes and steered them toward posting positive or negative content.⁶

These allegations of hidden manipulation through data profiling and AI-targeted streams of information are disturbing. Add to these the evidence that digital media's algorithms spread falsehood far faster than truth⁷, foster harassment and misogyny⁸, and endanger children⁹ by increasing the frequency of exposure to predatory and inappropriate content.

International Public Policy Response

This analysis is not presented with a claim that it is undeniably accurate. Although there is persuasive evidence for all of this critique, it is not unquestionable. The point here is simply that these arguments rise well above the standard of plausible. And the threshold that demands action from government is much lower. If these arguments meet the test that they COULD be true, then it is imperative that government act on behalf of its citizens first to determine without doubt whether they ARE true, and if so, then to apply appropriate remedies. Put simply – it is not acceptable to acknowledge that AI is capable of inflicting this kind of negative social impact without simultaneously constructing a system of legal oversight to protect the public from harm.

Consider the way in which other complex, concentrated industries are overseen by government regulation. It is not permitted for pharmaceutical companies to begin selling new drugs on the medical marketplace without first being subject to rigorous independent review to ensure that the chemical compound is what they say it is and does what they say it does. The automotive

⁴ "Facebook ad campaign helped Donald Trump win election, claims executive," BBC, 8 January 2020, <https://www.bbc.co.uk/news/technology-51034641>

⁵ Sam Levin, "Facebook told advertisers it can identify teens feeling 'insecure' and 'worthless'," *The Guardian*, 1 May 2017, <https://www.theguardian.com/technology/2017/may/01/facebook-advertising-data-insecure-teens>

⁶ Robinson Meyer, "Everything We Know About Facebook's Secret Mood Manipulation Experiment," *The Atlantic*, 28 June 2014, <https://www.theatlantic.com/technology/archive/2014/06/everything-we-know-about-facebooks-secret-mood-manipulation-experiment/373648/>

⁷ Peter Dizikes, "Study: On Twitter, false news travels faster than true stories," *MIT News*, 8 March 2018, <https://news.mit.edu/2018/study-twitter-false-news-travels-faster-true-stories-0308>

⁸ Amnesty International, "Toxic Twitter", March 2018, <https://www.amnesty.org/en/latest/research/2018/03/online-violence-against-women-chapter-1/>

⁹ See, K. G. Orphanides, "Children's YouTube is still churning out blood, suicide and cannibalism," *Wired UK*, 23 March 2018, <https://www.wired.co.uk/article/youtube-for-kids-videos-problems-algorithm-recommend>; and Max Fischer and Amanda Taub, "On YouTube's Digital Playground, an Open Gate for Pedophiles", *New York Times*, 3 June 2019, <https://www.nytimes.com/2019/06/03/world/americas/youtube-pedophiles.html>

industry is subject to extensive rules and independent checks of safety and environmental standards. The financial services sector is required to submit to regular audit and oversight to hold them accountable for the responsible management of money. These systems of regulation do not always work, but there is little debate that they are necessary to protect the public interest. Why are the digital platform giants – with all of the social impact we can see and that which we cannot – omitted from this list?

Many governments have already started to address this deficiency in various ways. Here are several examples worthy of attention.

France – Algorithmic Accountability: A white paper produced by the French government in May 2019 offers an analysis of the problems posed by AI-driven social media for democracy.¹⁰ The central proposal for reform calls for a strong regulation that requires transparency and auditing of the algorithmic decision-making systems that curate digital media content by an independent body. The report concludes:

“Algorithms are tools that may be misused or misappropriated. The importance they have gained on social networking platforms and the abuses they may cause (promotion of hate speech, ineffective moderation, interference by a sovereign state in the public debate, etc.) have made state intervention vital. This involves transparency, i.e. the means to make the underlying logic intelligible, the main processing principles applied by the algorithms” (p 25).

Great Britain – Duty of Care: The UK government has been engaged in several parallel efforts to address the unaccountable power of AI-driven media businesses. The most prominent is a public consultation on the topic of online harms¹¹ that presages the introduction of legislation by the government this year. The proposal contains a framework for regulatory oversight known as a “Duty of Care” that will govern large digital media platforms. If new law follows this formula, it will apply a legal obligation on the industry to be transparent about how automated decision-making works as well as a requirement to reduce or eliminate harms that are identified on their services by an independent regulator. Meanwhile the UK’s data protection agency – the Information Commissioner’s Office – has recently published a ground-breaking interpretation¹² of the GDPR that instructs digital media companies to offer a version of their products to users under the age of 18 that are stripped of data profiling and other features that feed information curation by AI systems. Most recently, the quasi-governmental think-tank –

¹⁰ “Creating a French framework to make social media platforms more accountable: Acting in France with a European vision,” May 2019,

https://minefi.hosting.augure.com/Augure_Minefi/r/ContenuEnLigne/Download?id=AE5B7ED5-2385-4749-9CE8-E4E1B36873E4&filename=Mission%20Re%CC%81gulation%20des%20re%CC%81seaux%20sociaux%20-ENG.pdf

¹¹ UK Government, Online Harms White Paper, April 2019,

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/793360/Online_Harms_White_Paper.pdf

¹² Information Commissioner’s Office, Age Appropriate Design Code, 22 January 2020,

<https://ico.org.uk/media/for-organisations/guide-to-data-protection/key-data-protection-themes/age-appropriate-design-a-code-of-practice-for-online-services-0-0.pdf>

the Center for Data Ethics and Innovation – published a comprehensive report¹³ analyzing the targeting and recommendation algorithms of digital media giants. This analysis takes a strong position recommending that the UK government require transparency, auditing, and harm reduction overseen by an independent regulator.

Canada – Digital Charter: In the spring of 2019, the Canadian government offered a comprehensive plan¹⁴ to reform policies across a broad array of issues in the digital economy. The delivery on the promises in this “Digital Charter” now forms a part of the agenda for the newly re-elected government. Among the specific commitments in the Charter is a set of new regulations that would require opening up AI-decision-making systems in digital media to allow for greater user control and privacy protection. These specific requirements were written into the mandate letter sent to the Minister of Innovation, Science and Industry.¹⁵

There are numerous other international examples. The EU is developing a new set of rules and standards for the technology industry that is expected in late 2020 as the Digital Services Act.¹⁶ In 2019, the Australian Competition and Consumer Commission published a comprehensive report on the harms arising from concentrated power in the digital media industry and proposed a wide variety of remedies.¹⁷ And in the United States, there are several pending inquiries and bills in Congress pertaining to market concentration, exploitation of online political advertising, and data privacy. In addition, the Attorneys General of more than 40 states have begun a joint antitrust investigation into Google and Facebook.¹⁸ Meanwhile the state of California has implemented an unprecedented new privacy law.¹⁹

Concluding Recommendation

The clear evidence of public harms created by the power of AI-driven digital media platforms has opened a wide-ranging public policy debate in many parts of the world. Common themes include harmful content moderation, data privacy, election security, taxation, competition and antitrust, online political advertising, digital literacy and more. This is, of course, also true in

¹³ “Review of Online Targeting, Final Report and Recommendations”, February 2020, Centre for Data Ethics and Innovation, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/863030/CDEI_7836-Review-of-Online-Targeting-04022020-final.pdf

¹⁴ Canada’s Digital Charter, 2019, https://www.ic.gc.ca/eic/site/062.nsf/eng/h_00108.html

¹⁵ Mandate Letter to the Minister of Innovation, Science and Industry, Federal Government of Canada, 13 December 2019, <https://pm.gc.ca/en/mandate-letters/2019/12/13/minister-innovation-science-and-industry-mandate-letter>

¹⁶ New oversight of the technology industry through a Digital Services Act were proposed in the agenda of EU Commission President, Ursula von der Leyen, “Political Guidelines for the Next European Commission,” https://ec.europa.eu/commission/sites/beta-political/files/political-guidelines-next-commission_en.pdf, p 13.

¹⁷ ACCC, Digital Platforms Inquiry, Final Report, 26 July 2019, <https://www.accc.gov.au/publications/digital-platforms-inquiry-final-report>

¹⁸ Steve Lohr, “Google Antitrust Investigation Outlined by State Attorneys General,” *New York Times*, 9 September 2019, <https://www.nytimes.com/2019/09/09/technology/google-antitrust-investigation.html>

¹⁹ California Consumer Privacy Act, Effective 1 January 2020, <https://oag.ca.gov/privacy/ccpa>.

Germany. The Bundeskartellamt in particular has played an active role in calling attention to the relationship between market power and data privacy. And the NetzDG has been a world leading intervention in content moderation law. Meanwhile, a series of court challenges calling for stricter implementation of the GDPR are winding their way through the judicial system.

Ultimately, the rebalancing of interests between democracy and the technology industry will be a whole-of-government project that will stretch over many years. It will continue beyond the current configuration of technology products and services popular in the market, and it will be driven by growing public awareness of the problems and opportunities that technology provides. It must be flexible and adaptive to the inevitable changes to come.

The concluding recommendation of this submission is not a detailed and comprehensive policy agenda. That is beyond the scope of this document. But there is a logical starting point – a first order policy change that will enable the German government to address many of the other harms that have drawn the attention of policymakers. **This is a policy mandating access to data and requiring independent audits of AI-systems in digital media in order to assess social impact, measure harms, and create appropriate remedies.** Without access to the data and AI systems that guide information flows in these markets, there is no obvious way to make good policy that will be adaptive and durable as the industry evolves. None of the issues at the center of this debate can be adequately addressed without this auditing function: democratic election integrity, child online safety, anti-competitive practices, consumer fraud and abuse, harassment and hate speech, and much more.

At present, there is a massive asymmetry of information. The harms are easily observed as specific incidents, and they do in fact appear to form a pattern. But the companies that hold the data that could verify these patterns and measure their scope hold all the data, and they do not make it available for independent review under any circumstances. This lid is kept tightly shut. Without access, regulators are forced to rely on the companies to police themselves through ineffective codes of conduct. This impasse cannot hold much longer.

This is not a radical proposal. It is the minimum necessary to begin building effective policy. The threshold for governments demanding access to the data to measure social impact and design remedies is NOT absolute proof of systemic public harms. The threshold is the *possibility* of systemic public harms. That line was crossed long ago.