Algorithm [draft] [#digitalkeywords]

TARLETON GILLESPIE · JUNE 25, 2014

The following is a draft of an essay, eventually for publication as part of the Digital Keywords project (Ben Peters, ed). This and other drafts will be circulated on Culture Digitally, and we invite anyone to provide comment, criticism, or suggestion in the comment space below. We ask that you please do honor that it is being offered in draft form — both in your comments, which we hope will be constructive in tone, and in any use of the document: you may share the link to this essay as widely as you like, but please do not quote from this draft without the author's permission. (TLG)

Algorithm — Tarleton Gillespie, Cornell University

WITH SUPPORT FROM:

#CULTD

algorithm as a technical solution to a technical problem

For their makers, “algorithm” refers specifically to the logical series of steps for organizing and acting on a body of data to quickly achieve a desired outcome. MacCormick (2012), in an attempt to explain algorithms to a general audience, calls them “tricks,” (5) by which he means “tricks of the trade” more than tricks in the magical sense — or perhaps like magic, but as a magician understands it. An algorithm is a recipe composed in programmable steps; most of the “values” that concern us lie elsewhere in the technical systems and the work that produces them.

FOUNDERS

Tarleton Gillespie
Hector Postigo
Josh Braun
Sam Srauy

AUTHORS

Adam Fish
Adrienne Shaw
Alice Marwick
Alison Harvey
Andrés Monroy-Hernández
Anne Helmond
Annette markham
Alison Novack
Aphra Kerr
Arturo Arriagada
Aynne Kokas
Ben Peters
Elizabeth Wissinger
Brooke Erin Duffy
Burcu S. Bakioglu
C.W. Anderson
Carla Ilten
Carrie Rentschler
Casey O'Donnell
Christina Dunbar-Hester
Christopher Boulton
Chuck Tryon
Daniel Kreiss
Daren Brabham
Dawn Nafus
Daniel Guagnin
Fenwick McKelvey
Finn Brunton
Fred Turner
Gabriella Coleman
Gina Neff
Greg Lastowka
Graham Meikle
Guobin Yang
Hope Forsyth
Hugh Gusterson
Ilana Gershon
Jan Fernback
James Bennett
Jean Burgess

For its designers, the "algorithm" comes after the generation of a "model," i.e. the formalization of the problem and the goal in computational terms. So, the task of *giving a user the most relevant search results for their queries* might be operationalized into a model for *efficiently calculating the combined values of pre-weighted objects in the index database, in order to improve the percentage likelihood that the user clicks on one of the first five results*.[1] This is where the complex social activity and the values held about it are translated into a functional interaction of variables, indicators, and outcomes. Measurable relationships are posited as existing between some of these elements; a strategic target is selected, as a proxy for some broader social goal; a threshold is determined as an indication of success, at least for this iteration.

The "algorithm" that might follow, then, is merely the steps for aggregating those assigned values efficiently, or delivering the results rapidly, or identifying the strongest relationships according to some operationalized notion of "strong." All is in the service of the model's understanding of the data and what it represents, and in service of the model's goal and how it has been formalized. There may be many algorithms that would reach the same result inside a given model, just like bubble sorts and shell sorts both put lists of words into alphabetical order. Engineers choose between them based on values such as how quickly they return the result, the load they impose on the system's available memory, perhaps their computational elegance. The embedded values that make a sociological difference are probably more about the problem being solved, the way it has been modeled, the goal chosen, and the way that goal has been operationalized (Reider).

Of course, simple alphabetical sorting may be a misleading an example to use here. The algorithms we're concerned about today are rarely designed to reach a single and certifiable answer, like a correctly alphabetized list. More common are algorithms that must choose one of many possible results, none of which are certifiably "correct." Algorithm designers must instead achieve some threshold of operator or user satisfaction — understood in the model, perhaps, in terms of percent clicks on the top results, or percentage of correctly identified human faces from digital images.

This brings us to the second value-laden element around the algorithm. To efficiently design algorithms that achieve a target

John Carter McKnight
Jeff Drouin
Jennifer Lena
Jessa Lingel
Jessica Beyer
John Banks
John Durham Peters
Jonathan Sterne
Joshua McVeigh-Schultz
Jeff Pooley
Julia Sonnevend
Karine Nahon
Katherine D. Harris
Katy Pearce
Kevin Driscoll
Kate Miltner
Laura Forlano
Lee Humphreys
Lee Vinsel
Limor Shifman
Lisa Nakamura
Lonny J Avi Brooks
Lisa Silvestri
Lucas Graves
Mari Castaneda
Mark Chen
Mark Deuze
Mary Gray
Matthew Crain
Matt Ratto
Mél Hogan
Mel Stanfill
Mélanie Millette
Neal Thomas
Nicholas John
Nick Couldry
Niki Strange
Rasmus Nielsen
Ray Fouche
Rena Bivens
Robert Gehl
Rodrigo Zamith
Rosemary Avance
Samuel Woolley
Sara Kingsley
Sarah Banet-Weiser
Saugata Bhaduri
Scott Brennen
Seth Lewis
Shira Chess
Stephanie Schulte
Steven Schrag
Tamara Shepherd
Ted Striphas
Tim Highfield
Thomas Streeter

goal (rather than reaching a known answer), algorithms are "trained" on a corpus of known data. This data has been in some way certified, either by the designers or by past user practices: this photo is of a human face, this photo is not; this search result has been selected by many users in response to this query, this one has not. The algorithm is then run on this data so that it may "learn" to pair queries and results found satisfactory in the past, or to distinguish images with faces from images without.

The values, assumptions, and workarounds that go into the selection and preparation of this training data may also be of much more importance to our sociological concerns than the algorithm learning from it. For example, the training data must be a reasonable approximation of the data that algorithm will operate on in the wild. The most common problem in algorithm design is that the new data turns out not to match the training data in some consequential way. Sometimes new phenomena emerge that the training data simply did not include and could not have anticipated; just as often, something important was overlooked as irrelevant, or was scrubbed from the training data in preparation for the development of the algorithm.

Furthermore, improving an algorithm is rarely about redesigning it. Rather, designers will "tune" an array of parameters and thresholds, each of which represents a tiny assessment or distinction. In search, this might mean the weight given to a word based on where it appears in a webpage, or assigned when two words appear in proximity, or given to words that are categorically equivalent to the query term. These values have been assigned and are already part of the training data, or are thresholds that can be dialed up or down in the algorithm's calculation of which webpage has a score high enough to warrant ranking it among the results returned to the user.

Finally, these exhaustively trained and finely tuned algorithms are instantiated inside of what we might call an application, which actually performs the functions we're concerned with. For algorithm designers, the algorithm is the conceptual sequence of steps, which should be expressible in any computer language, or in human or logical language. They are instantiated in code, running on servers somewhere, attended to by other helper applications (Geiger 2014), triggered when a query comes in or an image is scanned. I find it easiest the think about the difference between the "book" in your hand and the

Tim Jordan
Tom Boellstorff
Tony Liao
Torin Monahan
Valerie Belair-Gagnon
Whitney Phillips
Zachary McDowell
Zizi Papacharissi

"story" within it. These applications embody values as well, outside of their reliance on a particular algorithm.

To inquire into the implications of "algorithms," if we meant what software engineers mean when they use the term, could only be something so picky as investigating the political implications of using a bubble sort or a shell sort — setting aside bigger questions like why "alphabetical" in the first place, or why train on this particular dataset. Perhaps there are lively insights to be had about the implications of different algorithms in this technical sense,{2] but by and large we in fact mean something else when we talk about algorithms as having "social implications."

algorithm as talisman

The information industries have found value in the term "algorithm" in their public-facing discursive efforts as well. To call their service or process an algorithm is to lend a set of associations to that service: mathematical, logical, impartial, consistent. Algorithms seem to have a "disposition towards objectivity" (Hillis et al 2013: 37); this objectivity is regularly performed as a feature of algorithmic systems. (Gillespie 2014) Conclusions that can be described as having been generated by an algorithm come with a powerful legitimacy, much the way statistical data bolsters scientific claims, with the human hands yet another step removed. It is a very different kind of legitimacy than one that rests on the subjective expertise of an editor or a consultant, though it is important not to assume that it trumps such claims in all cases. A market prediction that is "algorithmic" is different from a prediction that comes from an expert broker highly respected for their expertise and acumen; a claim about an emergent social norm in a community generated by an algorithm is different from one generated ethnographically. Each makes its own play for legitimacy, and implies its own framework for what legitimacy is (quantification or interpretation, mechanical distance or human closeness). But in the context of nearly a century of celebration of the statistical production of knowledge and longstanding trust in automated calculation over human judgment, the algorithmic does enjoy a particular cultural authority.

More than that, the term offers the corporate owner a powerful talisman to ward off criticism, when companies must justify themselves and their services to their audience, explain away errors and unwanted outcomes, and justify and defend the increasingly significant roles they play in public life. (Gillespie 2014) Information services can point to "the algorithm" as having been responsible for particular results or conclusions, as a way to distance those results from the providers. (Morozov, 2013: 142) The term generates an entity that is somehow separate, the assembly line inside the factory, that can be praised as efficient or blamed for mistakes.

The term "algorithm" is also quite often used as a stand-in for its designer or corporate owner. When a critic says "Facebook's algorithm" they often mean Facebook and the choices it makes, some of which are made in code. This may be another way of making the earlier point, that the singular term stands for a complex sociotechnical assemblage: Facebook's algorithm really means "Facebook," and Facebook really means the

people, things, priorities, infrastructures, aims, and discourses that animate them. But it may also be a political economic conflation: this is Facebook acting through its algorithm, intervening in an algorithmic way, building a business precisely on its ability to construct complex models of social/expressive activity, train on an immense corpus of data, tune countless parameters, and reach formalized goals extremely efficiently.

Maybe saying “Facebook’s algorithm” and really meaning the choices and interventions made by Facebook the company into our social practices is a way to assign accountability (Diakopoulos 2013, Ziewitz 2011). It makes the algorithm theirs in a powerful way, and works to reduce the distance some providers put between “them” (their aims, their business model, their footprint, their responsibility) and “the algorithm” (as somehow autonomous from all that). On the other hand, conflating the algorithmic mechanism and the corporate owner may obscure the ways these two entities are not always aligned. It is crucial that we discern between things done by the algorithmic system and things done in other ways, such as the deletion of obscene images from a content platform, which is sometimes handled algorithmically and sometimes performed manually. (Gillespie 2012b) It is crucial to note slippage between a provider’s financial or political aims and the way the algorithmic system actually functions. And conflating algorithmic mechanism and corporate owner misses how some algorithmic approaches are common to multiple stakeholders, circulate across them, and embody a tactic that exceeds any one implementation.

Endnotes

1. This parallels Kowalski's well-known definition of an algorithm as "logic + control": "An algorithm can be regarded as consisting of a logic component, which specifies the knowledge to be used in solving problems, and a control component, which determines the problem-solving strategies by means of which that knowledge is used. The logic component determines the meaning of the algorithm whereas the control component only affects its efficiency." (Kowalksi, 424) I prefer to use "model" because I want to reserve "logic" for the underlying premise of the entire algorithmic system and its deployment.

2.See Kockelman 2013 for a dense but superb example.

3.See Brian Christian, "The A/B Test: Inside the Technology That's Changing the Rules of Business." *Wired*, April 25. http://www.wired.com/2012/04/ff_abtesting/

References

Barocas, Solon, Sophie Hood, and Malte Ziewitz. 2013. "Governing Algorithms: A Provocation Piece." *Available at SSRN 2245322*. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2245322

Beer, David. 2009. "Power through the Algorithm? Participatory Web Cultures and the Technological Unconscious." *New Media & Society* 11 (6): 985-1002.

Bucher, T. 2012. "Want to Be on the Top? Algorithmic Power and the Threat of Invisibility on Facebook." *New Media & Society* 14 (7): 1164-80.

Cheney-Lippold, J. 2011. "A New Algorithmic Identity: Soft Biopolitics and the Modulation of Control." *Theory, Culture & Society* 28 (6): 164-81.

Diakopoulos, Nicholas. 2013. "Algorithmic Accountability Reporting: On the Investigation of Black Boxes." A Tow/Knight

Brief. Tow Center for Digital Journalism, Columbia Journalism School. http://towcenter.org/algorithmic-accountability-2/

Downey, Gregory J. 2014. "Making Media Work: Time, Space, Identity, and Labor in the Analysis of Information and Communication Infrastructures." In *Media Technologies: Essays on Communication, Materiality, and Society*, edited by Tarleton Gillespie, Pablo J. Boczkowski, and Kirsten A Foot, 141-66. Cambridge, MA: The MIT Press.

Geiger, R. Stuart. 2014. "Bots, Bespoke, Code and the Materiality of Software Platforms." *Information, Communication & Society* 17 (3): 342-56.

Gillespie, Tarleton. 2012a. "Can an Algorithm Be Wrong?" *Limn* 1 (2). http://escholarship.org/uc/item/0jk9k4hj

Gillespie, Tarleton. 2012b. "The Dirty Job of Keeping Facebook Clean." *Culture Digitally* (Feb 22). http://culturedigitally.org/2012/02/the-dirty-job-of-keeping-facebook-clean/

Gillespie, Tarleton. 2014. "The Relevance of Algorithms." In *Media Technologies: Essays on Communication, Materiality, and Society*, edited by Tarleton Gillespie, Pablo J. Boczkowski, and Kirsten A Foot, 167-93. Cambridge, MA: The MIT Press.

Gitelman, Lisa. 2006. *Always Already New: Media, History and the Data of Culture.* Cambridge, MA: MIT Press.

Hillis, Ken, Michael Petit, and Kylie Jarrett. 2013. *Google and the Culture of Search*. Abingdon: Routledge.

Kockelman, Paul. 2013. "The Anthropology of an Equation. Sieves, Spam Filters, Agentive Algorithms, and Ontologies of Transformation." *HAU: Journal of Ethnographic Theory* 3 (3): 33-61.

Kowalski, Robert. 1979. "Algorithm = Logic + Control." *Communications of the ACM* 22 (7): 424-36.

Kushner, S. 2013. "The Freelance Translation Machine: Algorithmic Culture and the Invisible Industry." *New Media &*

Society 15 (8): 1241-58.

MacCormick, John. 2012. *9 Algorithms That Changed the Future*. Princeton: Princeton University Press.

Mager, Astrid. 2012. "Algorithmic Ideology: How Capitalist Society Shapes Search Engines." *Information, Communication & Society* 15 (5): 769-87.

Morozov, Evgeny. 2014. *To Save Everything, Click Here: The Folly of Technological Solutionism*. New York: PublicAffairs.

O'Reilly, Tim. 2013. "Open Data and Algorithmic Regulation." In *Beyond Transparency: Open Data and the Future of Civic Innovation*, edited by Lauren Goldstein and Lauren Dyson. San Francisco, Calif.: Code for America Press. http://beyondtransparency.org/chapters/part-5/open-data-and-algorithmic-regulation/

Rieder, Bernhard. 2012. "What Is in PageRank? A Historical and Conceptual Investigation of a Recursive Status Index." *Computational Culture* 2. http://computationalculture.net/article/what_is_in_pagerank

Seaver, Nick. 2013. "Knowing Algorithms." *Media in Transition 8*, Cambridge, MA. http://nickseaver.net/papers/seaverMiT8.pdf

Striphas, Ted (2010) "How to Have Culture in an Algorithmic Age" The Late Age of Print June 14. http://www.thelateageofprint.org/2010/06/14/how-to-have-culture-in-an-algorithmic-age/

Striphas, Ted (2012) "What is an Algorithm?" *Culture Digitally* Feb 1. http://culturedigitally.org/2012/02/what-is-an-algorithm/

Uricchio, William. 2011. "The Algorithmic Turn: Photosynth, Augmented Reality and the Changing Implications of the Image." *Visual Studies* 26 (1): 25-35.

Williams, Raymond (1976/1983) *Keywords: A Vocabulary of Culture and Society*. 2nd ed. Oxford: Oxford University Press.

Ziewitz, Malte. 2011. “How to think about an algorithm? Notes from a not quite random walk,” Discussion paper for Symposium on “Knowledge Machines between Freedom and Control”, 29 September 29.
http://ziewitz.org/papers/ziewitz_algorithm.pdf

POSTED IN: UNCATEGORIZED

TAGGED WITH: ALGORITHMS, DIGITAL KEYWORDS, DRAFT

AUTHOR

Tarleton Gillespie

Principal Researcher at Microsoft Research, New England; also an adjunct associate professor in the Department of Communication and Department of Information Science, at Cornell University. On Twitter at @TarletonG