Academic Author How Names Order To*

Florian Tramèr† Nicholas Carlini Daphne Ippolito
Chiyuan Zhang Matthew Jagielski Milad Nasr
Christopher A. Choquette-Choo Katherine Lee

April 1, 2023

Abstract

Groups of the same authors often publish papers together. Many times, there is a clear first author or last author, and then a blob of people who all contributed equally, or whose contributions are too varied and nuanced to offer up an obvious ordering. In this work, we present several methods for choosing an author order given these constraints.

1 Introduction

Author ordering is important [SHIT20]. And yet, there is no consensus on the optimal way to solve this critical problem.

Theoreticians—children at heart—often opt for an ordering that matches the tune of the lovable alphabet song. Empiricists, in turn, prefer a partial order stretching from people who did all the work to those who did none of it (they funded the project though!); ties are routinely indicated using such arcane symbols as †, ‡ or ¶. Finally, true believers (and communists) leave this challenging decision entirely up to the iron-clad impartiality of (pseudo-random) fate (cf. "Author ordering determined by coin flip over a Google Hangout" [KB14]). As a group of researchers that regularly publish together, neither of these approaches is fully satisfactory. [CIJ+23, LIN+22, JTT+22, ZIL+21, JNCC+23, CTW+21, BLM+22, CJCC+23, NHS+23, CHN+23, CJZ+22, CAO+22, TSJ+22][1]

For example, alphabetical author ordering forever relegates Chiyuan Zhang to last place, bringing great dishonor to his family; conversely, Nicholas Carlini is always first regardless of his (lack of) contribution.[2] While we have experimented with *reverse*-alphabetical ordering to alleviate this concern, both configurations preserve Katherine Lee's position in the "no-mans-land" that empiricists associate with authors who neither did any real work nor provided any wisdom or funding.[3]

*Title ordered alphabetically.

†Authors ordered.

[1]Switzerland classifies "unjustified" self-citation as grounds for sanctions [Mat21].

[2]Author Nicholas **C**arlini sees nothing wrong with this.

[3]We have yet to experiment with "middle-out" ordering.

Ordering by contributions is also challenging: often, for many of our papers, none of the authors contributed any meaningful work whatsoever.

Finally, fully randomized ordering is perceived as unfair in the common case where a small number of authors actually did some brilliant work, and all the others just added their names in the hope that this extra paper will finally cure their imposter syndrome.

In this paper, we thus initiate the study of more general author ordering schemes that combine (the worst) aspects of existing schemes.

We formally define the "Author Ordering Problem", which asks to find an impartial ordering of authors subject to specific positional constraints. We invite the reader to aid us in evaluating the methodologies put forth here and continue cracking on this critical and persistent problem in academia.

Problem Statement 1 (Author Ordering Problem) *A group of n author consistently publish papers together. Author names must be placed in some order. On any given paper, there are k authors who should have fixed positions in the order (e.g., first or last author). In some cases $k = 0$. The remaining $n - k$ authors should be ordered so as to make the world a better place in a maximal way.*[4]

As we will see, possible solutions to this problem touch on topics of interest to many different areas of computer science. For example, researchers in trustworthy AI may be interested in studying the subtle discriminatory biases that arise when we ask a Large Language Model (LLM) to solve the Author Ordering Problem. Or Game-theorists may find new applications for mechanism design when we cast the Author Ordering Problem as an *auction*, where authors place monetary bids on different slots in the ordering or as marriage problem. Finally, permuting authors based on pseudo-random values extracted from personally identifiable information (e.g., social security numbers, credit cards, etc.) may present challenges of interest to privacy practitioners.

2 Optimal Author Ordering

Ordering 1 (Optimal author ordering) *The optimal ordering of author names in an academic publication is to place them in order.*

Theorem 1 *The above ordering is optimal.*

Proof. By inspection. □

Corollary 1 *The world is now a better place.*

3 Strategies for Practical Author Ordering

In this section, we present ordering strategies that our own research group has employed or considered employing in our efforts to achieve an optimal ordering.

[4] $n = 0$ may be possible.

Order by contribution. It is very common in Machine Learning and other fields for collaborators to attempt to encode contribution amount into the authorship list. Our own research group has considered several strategies for ordering by contribution. For example, we could order by the number of plots one generates (Katherine), or by amount of compute used (Nicholas), or the number of lines of code written (who knows), or the number of words written on the page (this heavily disfavors German speakers who tend to write asinglenounthatisthelengthofanentiresentence). Contribution-based ordering becomes much more complex when collaborators decide to order by overall contribution, rather than a single measurement. How does one weigh the difficulty of organizing a team, with the technical challenges of running experiments, or the conceptual challenges of situating research amongst related work, all normalized by the number of cups of coffee that each author drank?

Further difficulties arise when collaborators do not agree on how to assess contribution levels, for example, when several collaborators insist on being second-to-last author since they feel like they did the least work. For these reasons, our own research group has chosen to minimize use of contribution-based ordering.

Deterministic orderings. Ordering alphabetically by last name[5] is also a common approach in our field. Unfortunately last-name alphabetic ordering systematically biases in favor of researchers with names that begin early in the alphabet.[6] [EY06] found that in the field of Economics, which commonly uses alphabetical ordering, professors with earlier surname initials were significantly more likely to receive tenure. This trend was not true in psychology, not only because psychology experiments are never reproducible, but also because it does not tend to use alphabetic ordering. Some other name-derived orderings which could reduce the favoritism toward early-initialed authors are reverse alphabetic name ordering, ordering by name length, or ordering alphabetically by the pth letter in each name (where p is chosen once by an impartial party with no knowledge of why they are being asked to pick a number from 1 to 10). Alternative non-name-based deterministic methods include ordering by author's age, height, years from completing their PhD, or distanced lived from the office.

Simultaneous authors. A simple solution to all of the problems with determinstic orderings would be to overlay all author names on top of one another (Figure 1, left).[7] Unfortunately, this has several limitations. For example, it is somewhat harder to read than the standard author styling (but still somewhat easier to read than the average equation in a theory paper). This method also favors authors with longer names, as their name still comes "first"—as in, closest left, and overflowing on the right.

Another option (Figure 1, right) is an "Ouroboros" of author names, where every author can claim to be first, last, or anything in between.

[5] Alphabetical by first name orderings are also known to exist [BAA+15].

[6] Alphabetical ordering also implies that $\mathcal{P} > \mathcal{NP}$, which doesn't make any sense.

[7] Some would argue that you can still *glean* author order from this by investigating which author is put on top of the pile, or by inspecting the tikz code to determine which order was used in the source. Through casual experimentation, we have found that some PDF viewers will preserve the drawing order, so that when selecting and copying the blob of text, one can get the original order back, while other (more secure) viewers do not leak such information.

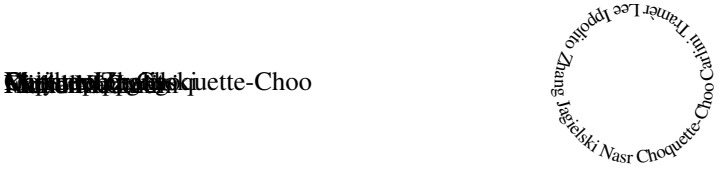

Figure 1: A demonstration of the method of simultaneous authors.

An issue with these schemes is that while they are impartial "on paper" (literally), an actual ordering still has to be chosen for arXiv and conference proceedings, and this is the ordering that will be used when the work is cited.

One way to overcome this limitation would be to use a single author pseudonym, following the example of [Bou39]. We personally recommend a portmanteau of the first couple letters of each of the first and last names of all of the paper's main contributors, which would lead to the set of researchers behind this paper listing "Flonidachimamichrika Zhipjatracholeca" as the author on all their papers together. Unfortunately, this method would likely run counter to academic integrity policies which require authors to exist.

"Band name." Yet another option would be to pick a "band name" for the group of researchers, such as *The Beatles* or *OpenAI*. This gets tricky when the original "drummer"—an affectionate nickname for the author who writes the background section—is inevitably replaced by a better one; or when the lead author joins the "27 club" (authors who reached an h-index of 27 and got promoted into a management role).

Self-correcting orderings. A self-correcting ordering aims to even out credit assignment across the whole team. For example, for each subsequent paper, collaborators could choose an ordering which is least predictable given the author orderings of their previous papers. Alternatively, the person who presently has the fewest number of first-author papers could be listed first on the next paper, or the team could order by who has the fewest citations on their first author papers, or by perceived impact.

The reverse of self-correcting ordering is to reward the senior contributors. After all, the project could not have happened without its most senior member! They probably deserve the bulk of the credit, and they should be in the position[8] that gives that to them. This method would place the authors with the most papers before authors with the least papers.

Author order by auction. Not all papers are equally important to all authors. Given that the majority of us live in the U. S. of A. where capitalism is the law of the land,[9] it may be desirable to force authors to bid on their position in the paper. Authors could place monetary bids using their own money, or use appropriate "gift funds." An alternative is to give each author a virtual bidding budget which could be uniform, or

[8] It is debated whether that position is first or last author.

[9] Switzerland is OK with capitalism too.

derived from each author's current citation count. This might disadvantage those who are less well off, but if that's not capitalism, we don't know what is.

Designing auctions for multiple, possibly fungible goods is tricky[10], and so we leave this as an exercise to the reader or the game theorist.

Artificial Intelligence. We could leverage higher beings to resolve the author ordering challenge for us. Instead of trying to come up with a strategy, we could ask CoPilot to write a function that outputs author order in a "fair" way. When we tried this, CoPilot sometimes would suggest an order that dropped some of the authors all together. Maybe this was CoPilot's way of inferring that these authors really contributed nothing (based on their name alone). We also asked ChatGPT and Bard to suggest good orderings, but both were unwilling to weigh in on the matter.

Stable marriage.[11] In the academic world, author ordering can be a tricky game to play. But fear not, for game theory comes to the rescue! Imagine you and your co-authors are players in a game where the goal is to form the perfect partnership based on your individual preferences. It's like a matchmaking game, but with academic credits on the line! The game is considered stable when no one wants to ditch their current position for someone else. And just like in any good game, there are rules and strategies to be mastered. By using this approach, researchers can gain insights into the dynamics of author ordering and optimize their chances of achieving a successful career. So grab your game face and let's play the academic authorship game with some game theory!

Seeded permutations. In prior work, we developed an algorithm to order authors by searching for the first permutation that satisfies some set of constraints [ITN+22]. For example, in that paper, we enumerated permutations and choose the first where Daphne appeared first (because she did all the work) and where Nicholas appeared last (because he did none of the work).[12] Thus, the problem of ordering authors can be easily reduced to the problem of ordering permutations. Choosing a good ordering of permutations is (as you might guess) left as an exercise for the reader.

Minimum entry permutation. We note that choosing a random permutation has significant limitations. Most importantly, it leaves open several free variables that could have been used to optimize some property. For example, given that arXiv papers are limited to 10MB in total size, something we could optimize is the total file size and choose the ordering that minimizes (or maximizes!) the overall file size. Or instead, we could choose the permutation that maximizes (or minimizes?) the perplexity of some large language model. Choosing whether to maximize or minimize—and what metric to maximize or minimize—is again left as an exercise to the reader.

[10] See `https://www.theguardian.com/business/2000/apr/23/theobserver.observerbusiness2`

[11] Thanks to ChatGPT for writing this paragraph.

[12] The other middle authors also did none of the work (since Daphne did all of it), but they did more of none of the work than Nicholas.

4 Discussion

Authorship serves many roles, and in addition to these roles changing over time, they also vary across disciplines and across lab groups within the same discipline. Authorship order can indicate degree of intellectual contribution, seniority level, who provided funding, or none of these. The field-specific intricacies and ramifications of who does and does not make it onto the author list is a subject of serious academic inquiry [Cla05, SWJ12, MBJ11].

In Mathematics [Soc04], Biology [THR+07], and other fields [HE], the question of authorship order has been fraught enough that explicit recommendations have been made on how paper-writers should determine and explain their orderings and how readers should interpret them.

All authors on this paper identify as part of the Machine Learning community, which is rapidly transitioning toward projects conducted by large teams in which members take on varied and specialized roles. The Machine Learning community has seen a growth in researchers following the recommendations of [BAA+15] and having the role and contributions of each author be specifically recognized. Visualization such as authorship grids have been proposed to make it abundantly clear who contributed what [PLT18, THR+07].

Increasingly, the field is seeing efforts to be very inclusive on the author list. For example, the BIG-bench project offered to list everyone who contributed a task to their new benchmark as an author, which resulted in almost 450 authors [SRR+22]. Other groups, mostly in Industry, have opted to include everyone who works on the larger research team as an author because, in some sense, the paper would not have happened in the way it did without each and every one of them (for example, [PRL+22] from Anthropic and [WS+23] from HuggingFace). Others have opted out of authorship order all together and list only the team name as the author (for example: "OpenAI" [Ope23]).

With the growth in generous author lists and the transition toward explicit contribution sections, the actual order authors end up written on the page is becoming less and less meaningful. We see this as an excellent opportunity to discard meaning entirely, and instead strive to maximize humour and creativity in our orderings. That being said, we understand that it is a privilege to not have to care about authorship order, and unfortunately, the indicators of first or last authorship may still be crucial for securing funding, jobs, or a PhD.

5 Contributions

Florian Tramèr: Responsible for all "—" and "è" in this paper.

Nicholas Carlini: Responsible for all figures in this paper.

Daphne Ippolito: Resposnible for all typos (fixed and missed).

Chiyuan Zhang: Responsible for all "the"s in the paper.

Matthew Jagielski: Responsible for a great attitude.

Milad Nasr: Responsible for evaluations.

Christopher A. Choquette-Choo: Responsible for technical alignment.

Katherine Lee: Responsible.

References

[BAA+15] Amy Brand, Liz Allen, Micah Altman, Marjorie Hlava, and Jo Scott. Beyond authorship: attribution, contribution, collaboration, and credit. *Learned Publishing*, 28(2):151–155, 2015.

[BLM+22] Hannah Brown, Katherine Lee, Fatemehsadat Mireshghallah, Reza Shokri, and Florian Tramèr. What does it mean for a language model to preserve privacy? In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, pages 97–108, Seoul, Republic of Korea, 2022. ACM.

[Bou39] Nicolas Bourbaki. *Éléments de mathématique*. Éditions Hermann, France, 1939.

[CAO+22] Harsh Chaudhari, John Abascal, Alina Oprea, Matthew Jagielski, Florian Tramèr, and Jonathan Ullman. Snap: Efficient extraction of private properties with poisoning, 2022.

[CHN+23] Nicholas Carlini, Jamie Hayes, Milad Nasr, Matthew Jagielski, Vikash Sehwag, Florian Tramèr, Borja Balle, Daphne Ippolito, and Eric Wallace. Extracting training data from diffusion models, 2023.

[CIJ+23] Nicholas Carlini, Daphne Ippolito, Matthew Jagielski, Katherine Lee, Florian Tramer, and Chiyuan Zhang. Quantifying memorization across neural language models, 2023.

[CJCC+23] Nicholas Carlini, Matthew Jagielski, Christopher A. Choquette-Choo, Daniel Paleka, Will Pearce, Hyrum Anderson, Andreas Terzis, Kurt Thomas, and Florian Tramèr. Poisoning web-scale training datasets is practical, 2023.

[CJZ+22] Nicholas Carlini, Matthew Jagielski, Chiyuan Zhang, Nicolas Papernot, Andreas Terzis, and Florian Tramer. The privacy onion effect: Memorization is relative. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 13263–13276. Curran Associates, Inc., 2022.

[Cla05] Larry D. Claxton. Scientific authorship: Part 2. history, recurring issues, practices, and guidelines. *Mutation Research/Reviews in Mutation Research*, 589(1):31–45, 2005.

[CTW+21] Nicholas Carlini, Florian Tramer, Eric Wallace, Matthew Jagielski, Ariel Herbert-Voss, Katherine Lee, Adam Roberts, Tom Brown, Dawn Song, Ulfar Erlingsson, Alina Oprea, and Colin Raffel. Extracting training data from large language models, 2021.

[EY06] Liran Einav and Leeat Yariv. What's in a surname? the effects of surname initials on academic success. *Journal of Economic Perspectives*, 20(1):175–188, 2006.

[HE] Gert Helgesson and Stefan Eriksson. Authorship order. *Learned Publishing*, 32(2):106–112.

[ITN+22] Daphne Ippolito, Florian Tramèr, Milad Nasr, Chiyuan Zhang, Matthew Jagielski, Katherine Lee, Christopher A Choquette-Choo, and Nicholas Carlini. Preventing verbatim memorization in language models gives a false sense of privacy. *arXiv preprint arXiv:2210.17546*, 2022.

[JNCC+23] Matthew Jagielski, Milad Nasr, Christopher Choquette-Choo, Katherine Lee, and Nicholas Carlini. Students parrot their teachers: Membership inference on model distillation, 2023.

[JTT+22] Matthew Jagielski, Om Thakkar, Florian Tramèr, Daphne Ippolito, Katherine Lee, Nicholas Carlini, Eric Wallace, Shuang Song, Abhradeep Thakurta, Nicolas Papernot, and Chiyuan Zhang. Measuring forgetting of memorized training examples, 2022.

[KB14] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[LIN+22] Katherine Lee, Daphne Ippolito, Andrew Nystrom, Chiyuan Zhang, Douglas Eck, Chris Callison-Burch, and Nicholas Carlini. Deduplicating training data makes language models better, 2022.

[Mat21] David Matthews. Switzerland classifies unjustified self-citation as violation of scientific misconduct. Inside Higher Ed, June 2021.

[MBJ11] Ana Marušić, Lana Bošnjak, and Ana Jerončić. A systematic review of research on the meaning, ethics and practices of authorship across scholarly disciplines. *PLOS One*, 2011.

[NHS+23] Milad Nasr, Jamie Hayes, Thomas Steinke, Borja Balle, Florian Tramèr, Matthew Jagielski, Nicholas Carlini, and Andreas Terzis. Tight auditing of differentially private machine learning, 2023.

[Ope23] OpenAI. Gpt-4 technical report, 2023.

[PLT18] Julia C. Phillippi, Frances E. Likis, and Ellen L. Tilden. Authorship grids: Practical tools to facilitate collaboration and ethical publication. *Res Nurs Health*, 41(2):195–208, 2018.

[PRL+22] Ethan Perez, Sam Ringer, Kamilė Lukošiūtė, Karina Nguyen, Edwin Chen, Scott Heiner, Craig Pettit, Catherine Olsson, Sandipan Kundu, Saurav Kadavath, et al. Discovering language model behaviors with model-written evaluations. *arXiv preprint arXiv:2212.09251*, 2022.

[SHIT20] Kuniyasu Saitoh, Takahiro Hatano, Atsushi Ikeda, and Brian P. Tighe. Stress relaxation above and below the jamming transition. *Physical Review Letters*, 124(11), mar 2020.

[Soc04] American Mathematical Society. Statement on the culture of professionalism. `https://www.ams.org/profession/leaders/CultureStatement04.pdf`, 2004.

[SRR+22] Aarohi Srivastava, Abhinav Rastogi, Abhishek Rao, Abu Awal Md Shoeb, Abubakar Abid, Adam Fisch, Adam R. Brown, Adam Santoro, Aditya Gupta, Adrià Garriga-Alonso, Agnieszka Kluska, Aitor Lewkowycz, Akshat Agarwal, Alethea Power, Alex Ray, Alex Warstadt, Alexander W. Kocurek, Ali Safaya, Ali Tazarv, Alice Xiang, Alicia Parrish, Allen Nie, Aman Hussain, Amanda Askell, Amanda Dsouza, Ambrose Slone, Ameet Rahane, Anantharaman S. Iyer, Anders Johan Andreassen, Andrea Madotto, Andrea Santilli, Andreas Stuhlmüller, Andrew M. Dai, Andrew La, Andrew Lampinen, Andy Zou, Angela Jiang, Angelica Chen, Anh Vuong, Animesh Gupta, Anna Gottardi, Antonio Norelli, Anu Venkatesh, Arash Gholamidavoodi, Arfa Tabassum, Arul Menezes, Arun Kirubarajan, Asher Mullokandov, Ashish Sabharwal, Austin Herrick, Avia Efrat, Aykut Erdem, Ayla Karakaş, B. Ryan Roberts, Bao Sheng Loe, Barret Zoph, Bartłomiej Bojanowski, Batuhan Özyurt, Behnam Hedayatnia, Behnam Neyshabur, Benjamin Inden, Benno Stein, Berk Ekmekci, Bill Yuchen Lin, Blake Howald, Bryan Orinion, Cameron Diao, Cameron Dour, Catherine Stinson, Cedrick Argueta, Cesar Ferri, Chandan Singh, Charles Rathkopf, Chenlin Meng, Chitta Baral, Chiyu Wu, Chris Callison-Burch, Christopher Waites, Christian Voigt, Christopher D Manning, Christopher Potts, Cindy Ramirez, Clara E. Rivera, Clemencia Siro, Colin Raffel, Courtney Ashcraft, Cristina Garbacea, Damien Sileo, Dan Garrette, Dan Hendrycks, Dan Kilman, Dan Roth, C. Daniel Freeman, Daniel Khashabi, Daniel Levy, Daniel Moseguí González, Danielle Perszyk, Danny Hernandez, Danqi Chen, Daphne Ippolito, Dar Gilboa, David Dohan, David Drakard, David Jurgens, Debajyoti Datta, Deep Ganguli, Denis Emelin, Denis Kleyko, Deniz Yuret, Derek Chen, Derek Tam, Dieuwke Hupkes, Diganta Misra, Dilyar Buzan, Dimitri Coelho Mollo, Diyi Yang, Dong-Ho Lee, Dylan Schrader, Ekaterina Shutova, Ekin Dogus Cubuk, Elad Segal, Eleanor Hagerman, Elizabeth Barnes, Elizabeth Donoway, Ellie Pavlick, Emanuele

Rodolà, Emma Lam, Eric Chu, Eric Tang, Erkut Erdem, Ernie Chang, Ethan A Chi, Ethan Dyer, Ethan Jerzak, Ethan Kim, Eunice Engefu Manyasi, Evgenii Zheltonozhskii, Fanyue Xia, Fatemeh Siar, Fernando Martínez-Plumed, Francesca Happé, Francois Chollet, Frieda Rong, Gaurav Mishra, Genta Indra Winata, Gerard de Melo, Germán Kruszewski, Giambattista Parascandolo, Giorgio Mariani, Gloria Xinyue Wang, Gonzalo Jaimovitch-Lopez, Gregor Betz, Guy Gur-Ari, Hana Galijasevic, Hannah Kim, Hannah Rashkin, Hannaneh Hajishirzi, Harsh Mehta, Hayden Bogar, Henry Francis Anthony Shevlin, Hinrich Schuetze, Hiromu Yakura, Hongming Zhang, Hugh Mee Wong, Ian Ng, Isaac Noble, Jaap Jumelet, Jack Geissinger, Jackson Kernion, Jacob Hilton, Jaehoon Lee, Jaime Fernández Fisac, James B Simon, James Koppel, James Zheng, James Zou, Jan Kocon, Jana Thompson, Janelle Wingfield, Jared Kaplan, Jarema Radom, Jascha Sohl-Dickstein, Jason Phang, Jason Wei, Jason Yosinski, Jekaterina Novikova, Jelle Bosscher, Jennifer Marsh, Jeremy Kim, Jeroen Taal, Jesse Engel, Jesujoba Alabi, Jiacheng Xu, Jiaming Song, Jillian Tang, Joan Waweru, John Burden, John Miller, John U. Balis, Jonathan Batchelder, Jonathan Berant, Jörg Frohberg, Jos Rozen, Jose Hernandez-Orallo, Joseph Boudeman, Joseph Guerr, Joseph Jones, Joshua B. Tenenbaum, Joshua S. Rule, Joyce Chua, Kamil Kanclerz, Karen Livescu, Karl Krauth, Karthik Gopalakrishnan, Katerina Ignatyeva, Katja Markert, Kaustubh Dhole, Kevin Gimpel, Kevin Omondi, Kory Wallace Mathewson, Kristen Chiafullo, Ksenia Shkaruta, Kumar Shridhar, Kyle McDonell, Kyle Richardson, Laria Reynolds, Leo Gao, Li Zhang, Liam Dugan, Lianhui Qin, Lidia Contreras-Ochando, Louis-Philippe Morency, Luca Moschella, Lucas Lam, Lucy Noble, Ludwig Schmidt, Luheng He, Luis Oliveros-Colón, Luke Metz, Lütfi Kerem Senel, Maarten Bosma, Maarten Sap, Maartje Ter Hoeve, Maheen Farooqi, Manaal Faruqui, Mantas Mazeika, Marco Baturan, Marco Marelli, Marco Maru, Maria Jose Ramirez-Quintana, Marie Tolkiehn, Mario Giulianelli, Martha Lewis, Martin Potthast, Matthew L Leavitt, Matthias Hagen, Mátyás Schubert, Medina Orduna Baitemirova, Melody Arnaud, Melvin McElrath, Michael Andrew Yee, Michael Cohen, Michael Gu, Michael Ivanitskiy, Michael Starritt, Michael Strube, Michal Swedrowski, Michele Bevilacqua, Michihiro Yasunaga, Mihir Kale, Mike Cain, Mimee Xu, Mirac Suzgun, Mitch Walker, Mo Tiwari, Mohit Bansal, Moin Aminnaseri, Mor Geva, Mozhdeh Gheini, Mukund Varma T, Nanyun Peng, Nathan Andrew Chi, Nayeon Lee, Neta Gur-Ari Krakover, Nicholas Cameron, Nicholas Roberts, Nick Doiron, Nicole Martinez, Nikita Nangia, Niklas Deckers, Niklas Muennighoff, Nitish Shirish Keskar, Niveditha S. Iyer, Noah Constant, Noah Fiedel, Nuan Wen, Oliver Zhang, Omar Agha, Omar Elbaghdadi, Omer Levy, Owain Evans, Pablo Antonio Moreno Casares, Parth Doshi, Pascale Fung, Paul Pu Liang, Paul Vicol, Pegah Alipoormolabashi, Peiyuan Liao, Percy Liang, Peter W Chang, Peter Eckersley, Phu Mon Htut, Pinyu Hwang, Piotr Miłkowski, Piyush Patil, Pouya Pezeshkpour, Priti Oli, Qiaozhu Mei,

Qing Lyu, Qinlang Chen, Rabin Banjade, Rachel Etta Rudolph, Raefer Gabriel, Rahel Habacker, Ramon Risco, Raphaël Millière, Rhythm Garg, Richard Barnes, Rif A. Saurous, Riku Arakawa, Robbe Raymaekers, Robert Frank, Rohan Sikand, Roman Novak, Roman Sitelew, Ronan Le Bras, Rosanne Liu, Rowan Jacobs, Rui Zhang, Russ Salakhutdinov, Ryan Andrew Chi, Seungjae Ryan Lee, Ryan Stovall, Ryan Teehan, Rylan Yang, Sahib Singh, Saif M. Mohammad, Sajant Anand, Sam Dillavou, Sam Shleifer, Sam Wiseman, Samuel Gruetter, Samuel R. Bowman, Samuel Stern Schoenholz, Sanghyun Han, Sanjeev Kwatra, Sarah A. Rous, Sarik Ghazarian, Sayan Ghosh, Sean Casey, Sebastian Bischoff, Sebastian Gehrmann, Sebastian Schuster, Sepideh Sadeghi, Shadi Hamdan, Sharon Zhou, Shashank Srivastava, Sherry Shi, Shikhar Singh, Shima Asaadi, Shixiang Shane Gu, Shubh Pachchigar, Shubham Toshniwal, Shyam Upadhyay, Shyamolima Shammie Debnath, Siamak Shakeri, Simon Thormeyer, Simone Melzi, Siva Reddy, Sneha Priscilla Makini, Soo-Hwan Lee, Spencer Torene, Sriharsha Hatwar, Stanislas Dehaene, Stefan Divic, Stefano Ermon, Stella Biderman, Stephanie Lin, Stephen Prasad, Steven Piantadosi, Stuart Shieber, Summer Misherghi, Svetlana Kiritchenko, Swaroop Mishra, Tal Linzen, Tal Schuster, Tao Li, Tao Yu, Tariq Ali, Tatsunori Hashimoto, Te-Lin Wu, Théo Desbordes, Theodore Rothschild, Thomas Phan, Tianle Wang, Tiberius Nkinyili, Timo Schick, Timofei Kornev, Titus Tunduny, Tobias Gerstenberg, Trenton Chang, Trishala Neeraj, Tushar Khot, Tyler Shultz, Uri Shaham, Vedant Misra, Vera Demberg, Victoria Nyamai, Vikas Raunak, Vinay Venkatesh Ramasesh, vinay uday prabhu, Vishakh Padmakumar, Vivek Srikumar, William Fedus, William Saunders, William Zhang, Wout Vossen, Xiang Ren, Xiaoyu Tong, Xinran Zhao, Xinyi Wu, Xudong Shen, Yadollah Yaghoobzadeh, Yair Lakretz, Yangqiu Song, Yasaman Bahri, Yejin Choi, Yichi Yang, Yiding Hao, Yifu Chen, Yonatan Belinkov, Yu Hou, Yufang Hou, Yuntao Bai, Zachary Seid, Zhuoye Zhao, Zijian Wang, Zijie J. Wang, Zirui Wang, and Ziyi Wu. Beyond the imitation game: Quantifying and extrapolating the capabilities of language models. *arXiv preprint arXiv:2206.04615*, 2022.

[SWJ12] Elise Smith and Bryn Williams-Jones. Authorship and responsibility in health sciences research: a review of procedures for fairly allocating authorship in multi-author studies. *Science and Engineering Ethics*, 18:199–212, 2012.

[THR+07] Teja Tscharntke, Michael E Hochberg, Tatyana A Rand, Vincent H Resh, and Jochen Krauss. Author sequence and credit for contributions in multiauthored publications. *PLoS biology*, 5(1):e18, 2007.

[TSJ+22] Florian Tramèr, Reza Shokri, Ayrton San Joaquin, Hoang Le, Matthew Jagielski, Sanghyun Hong, and Nicholas Carlini. Truth serum: Poisoning machine learning models to reveal their secrets, 2022.

[WS+23] BigScience Workshop, :, Teven Le Scao, Angela Fan, Christopher Akiki, Ellie Pavlick, Suzana Ilić, Daniel Hesslow, Roman Castagné, Alexandra Sasha Luccioni, François Yvon, Matthias Gallé, Jonathan Tow, Alexander M. Rush, Stella Biderman, Albert Webson, Pawan Sasanka Ammanamanchi, Thomas Wang, Benoît Sagot, Niklas Muennighoff, Albert Villanova del Moral, Olatunji Ruwase, Rachel Bawden, Stas Bekman, Angelina McMillan-Major, Iz Beltagy, Huu Nguyen, Lucile Saulnier, Samson Tan, Pedro Ortiz Suarez, Victor Sanh, Hugo Laurençon, Yacine Jernite, Julien Launay, Margaret Mitchell, Colin Raffel, Aaron Gokaslan, Adi Simhi, Aitor Soroa, Alham Fikri Aji, Amit Alfassy, Anna Rogers, Ariel Kreisberg Nitzav, Canwen Xu, Chenghao Mou, Chris Emezue, Christopher Klamm, Colin Leong, Daniel van Strien, David Ifeoluwa Adelani, Dragomir Radev, Eduardo González Ponferrada, Efrat Levkovizh, Ethan Kim, Eyal Bar Natan, Francesco De Toni, Gérard Dupont, Germán Kruszewski, Giada Pistilli, Hady Elsahar, Hamza Benyamina, Hieu Tran, Ian Yu, Idris Abdulmumin, Isaac Johnson, Itziar Gonzalez-Dios, Javier de la Rosa, Jenny Chim, Jesse Dodge, Jian Zhu, Jonathan Chang, Jörg Frohberg, Joseph Tobing, Joydeep Bhattacharjee, Khalid Almubarak, Kimbo Chen, Kyle Lo, Leandro Von Werra, Leon Weber, Long Phan, Loubna Ben allal, Ludovic Tanguy, Manan Dey, Manuel Romero Muñoz, Maraim Masoud, María Grandury, Mario Šaško, Max Huang, Maximin Coavoux, Mayank Singh, Mike Tian-Jian Jiang, Minh Chien Vu, Mohammad A. Jauhar, Mustafa Ghaleb, Nishant Subramani, Nora Kassner, Nurulaqilla Khamis, Olivier Nguyen, Omar Espejel, Ona de Gibert, Paulo Villegas, Peter Henderson, Pierre Colombo, Priscilla Amuok, Quentin Lhoest, Rheza Harliman, Rishi Bommasani, Roberto Luis López, Rui Ribeiro, Salomey Osei, Sampo Pyysalo, Sebastian Nagel, Shamik Bose, Shamsuddeen Hassan Muhammad, Shanya Sharma, Shayne Longpre, Somaieh Nikpoor, Stanislav Silberberg, Suhas Pai, Sydney Zink, Tiago Timponi Torrent, Timo Schick, Tristan Thrush, Valentin Danchev, Vassilina Nikoulina, Veronika Laippala, Violette Lepercq, Vrinda Prabhu, Zaid Alyafeai, Zeerak Talat, Arun Raja, Benjamin Heinzerling, Chenglei Si, Davut Emre Taşar, Elizabeth Salesky, Sabrina J. Mielke, Wilson Y. Lee, Abheesht Sharma, Andrea Santilli, Antoine Chaffin, Arnaud Stiegler, Debajyoti Datta, Eliza Szczechla, Gunjan Chhablani, Han Wang, Harshit Pandey, Hendrik Strobelt, Jason Alan Fries, Jos Rozen, Leo Gao, Lintang Sutawika, M Saiful Bari, Maged S. Al-shaibani, Matteo Manica, Nihal Nayak, Ryan Teehan, Samuel Albanie, Sheng Shen, Srulik Ben-David, Stephen H. Bach, Taewoon Kim, Tali Bers, Thibault Fevry, Trishala Neeraj, Urmish Thakker, Vikas Raunak, Xiangru Tang, Zheng-Xin Yong, Zhiqing Sun, Shaked Brody, Yallow Uri, Hadar Tojarieh, Adam Roberts, Hyung Won Chung, Jaesung Tae, Jason Phang, Ofir Press, Conglong Li, Deepak Narayanan, Hatim Bourfoune, Jared Casper, Jeff Rasley, Max Ryabinin, Mayank Mishra, Minjia Zhang, Mohammad Shoeybi, Myriam Peyrounette, Nicolas Patry, Nouamane Tazi, Omar Sanseviero, Patrick von Platen, Pierre Cornette, Pierre François

Lavallée, Rémi Lacroix, Samyam Rajbhandari, Sanchit Gandhi, Shaden Smith, Stéphane Requena, Suraj Patil, Tim Dettmers, Ahmed Baruwa, Amanpreet Singh, Anastasia Cheveleva, Anne-Laure Ligozat, Arjun Subramonian, Aurélie Névéol, Charles Lovering, Dan Garrette, Deepak Tunuguntla, Ehud Reiter, Ekaterina Taktasheva, Ekaterina Voloshina, Eli Bogdanov, Genta Indra Winata, Hailey Schoelkopf, Jan-Christoph Kalo, Jekaterina Novikova, Jessica Zosa Forde, Jordan Clive, Jungo Kasai, Ken Kawamura, Liam Hazan, Marine Carpuat, Miruna Clinciu, Najoung Kim, Newton Cheng, Oleg Serikov, Omer Antverg, Oskar van der Wal, Rui Zhang, Ruochen Zhang, Sebastian Gehrmann, Shachar Mirkin, Shani Pais, Tatiana Shavrina, Thomas Scialom, Tian Yun, Tomasz Limisiewicz, Verena Rieser, Vitaly Protasov, Vladislav Mikhailov, Yada Pruksachatkun, Yonatan Belinkov, Zachary Bamberger, Zdeněk Kasner, Alice Rueda, Amanda Pestana, Amir Feizpour, Ammar Khan, Amy Faranak, Ana Santos, Anthony Hevia, Antigona Unldreaj, Arash Aghagol, Arezoo Abdollahi, Aycha Tammour, Azadeh HajiHosseini, Bahareh Behroozi, Benjamin Ajibade, Bharat Saxena, Carlos Muñoz Ferrandis, Danish Contractor, David Lansky, Davis David, Douwe Kiela, Duong A. Nguyen, Edward Tan, Emi Baylor, Ezinwanne Ozoani, Fatima Mirza, Frankline Ononiwu, Habib Rezanejad, Hessie Jones, Indrani Bhattacharya, Irene Solaiman, Irina Sedenko, Isar Nejadgholi, Jesse Passmore, Josh Seltzer, Julio Bonis Sanz, Livia Dutra, Mairon Samagaio, Maraim Elbadri, Margot Mieskes, Marissa Gerchick, Martha Akinlolu, Michael McKenna, Mike Qiu, Muhammed Ghauri, Mykola Burynok, Nafis Abrar, Nazneen Rajani, Nour Elkott, Nour Fahmy, Olanrewaju Samuel, Ran An, Rasmus Kromann, Ryan Hao, Samira Alizadeh, Sarmad Shubber, Silas Wang, Sourav Roy, Sylvain Viguier, Thanh Le, Tobi Oyebade, Trieu Le, Yoyo Yang, Zach Nguyen, Abhinav Ramesh Kashyap, Alfredo Palasciano, Alison Callahan, Anima Shukla, Antonio Miranda-Escalada, Ayush Singh, Benjamin Beilharz, Bo Wang, Caio Brito, Chenxi Zhou, Chirag Jain, Chuxin Xu, Clémentine Fourrier, Daniel León Periñán, Daniel Molano, Dian Yu, Enrique Manjavacas, Fabio Barth, Florian Fuhrimann, Gabriel Altay, Giyaseddin Bayrak, Gully Burns, Helena U. Vrabec, Imane Bello, Ishani Dash, Jihyun Kang, John Giorgi, Jonas Golde, Jose David Posada, Karthik Rangasai Sivaraman, Lokesh Bulchandani, Lu Liu, Luisa Shinzato, Madeleine Hahn de Bykhovetz, Maiko Takeuchi, Marc Pàmies, Maria A Castillo, Marianna Nezhurina, Mario Sänger, Matthias Samwald, Michael Cullan, Michael Weinberg, Michiel De Wolf, Mina Mihaljcic, Minna Liu, Moritz Freidank, Myungsun Kang, Natasha Seelam, Nathan Dahlberg, Nicholas Michio Broad, Nikolaus Muellner, Pascale Fung, Patrick Haller, Ramya Chandrasekhar, Renata Eisenberg, Robert Martin, Rodrigo Canalli, Rosaline Su, Ruisi Su, Samuel Cahyawijaya, Samuele Garda, Shlok S Deshmukh, Shubhanshu Mishra, Sid Kiblawi, Simon Ott, Sinee Sang-aroonsiri, Srishti Kumar, Stefan Schweter, Sushil Bharati, Tanmay Laud, Théo Gigant, Tomoya Kainuma, Wojciech Kusa, Yanis Labrak, Yash Shailesh Bajaj, Yash Venkatraman, Yifan Xu, Yingxin Xu,

Yu Xu, Zhe Tan, Zhongli Xie, Zifan Ye, Mathilde Bras, Younes Belkada, and Thomas Wolf. Bloom: A 176b-parameter open-access multilingual language model, 2023.

[ZIL+21] Chiyuan Zhang, Daphne Ippolito, Katherine Lee, Matthew Jagielski, Florian Tramèr, and Nicholas Carlini. Counterfactual memorization in neural language models, 2021.