Collaboration & Credit Principles

How can we be good stewards of collaborative trust?

Posted on May 30, 2019

TL;DR: *This essay makes a lot of suggestions, but the most useful/non-obvious/actionable are likely:* ***(1)*** *Be generous.* ***(2)*** *Use author contribution statements.* ***(3)*** *Put "author order not finalized" if it hasn't been.*

A lot of the best research in machine learning comes from collaborations. In fact, many of the most significant papers in the last few years (TensorFlow, AlphaGo, etc) come from collaborations of 20+ people. These collaborations are made possible by goodwill and trust between researchers.

This goodwill and trust is a precious shared resource, and it can be a fragile thing. When people work together, it's easy to have conflict, especially around attribution and credit. If dealt with poorly, attribution issues can fester. I'm aware of several cases of collaborations dieing, or people leaving teams and organizations, where and the underlying issue was hurt feelings and lost trust around collaboration. This strikes me as rather sad.

We often talk about credit issues in kind of binary terms. But if the thing we care about is this shared trust, I think it's not enough to just avoid doing anything wrong. We must also avoid any *feeling or appearance* of unfairness. In fact, we'd ideally actively cultivate the opposite, to behave in ways that add back to the pool of goodwill.

We should also be mindful that credit issues can easily be perceived as, and likely often are, linked to privilege or power gradients. This might be gender or race, but it can also be things like being a remote collaborator (ie. geographically removed from others), being an engineer or designer instead of a researcher, or being at a lower level professionally. A perception that junior collaborators or those from under-represented minorities are taken advantage not only harms the research community, but the larger cause of making sure all humans are treated fairly.

This is a hard problem, and one that lots of people have thought about before. I don't have any perfect answers, but I'd like to share some personal working principles. I'd love your comments and feedback!

Core Principles

- **Always check in with any person who could plausibly be an author** or feel like they should be, *even if you disagree*. Never have authorship or authorship order be decided behind closed doors or without giving people an opportunity to advocate for themselves.
- **Err on the side of sharing credit.** Credit isn't zero sum. It is often in everyone's benefit to be generous with credit, because it creates an incentive for others to help in the future. It also makes sense to be risk-averse to the possibility of not crediting someone who deserves it, because the harm of not crediting someone who deserves it is often greater than the harm of crediting someone who doesn't deserve it.
- **Acknowledge anyone you can remember talking to** about your research significantly. It costs nothing and builds good will. You can still use stronger language to highlight people who helped you more.
- **Avoid diffusion of responsibility**. For example, have someone clearly responsible for checking in with everyone on authorship.
- **Don't reveal someone else's unpublished work or merge it into your own without their consent.**
- **Remember that you are likely overestimating your own contributions.**
- **Act in ways that will make people want to work with you.** Enthusiastic collaborators are one of the most precious thing you can have as a researcher.
- **There's no substitute for emotional labor.** Humans have feelings. No magic bullet will remove the need for us to invest energy understanding them and talking them through together.

Pre-commit yourself to treating others well.

Most attribution issues are mistakes rather than malice. Adopting good collaboration practices is our best defense against making hurtful errors!

More fundamentally, any place you can pre-commit yourself to treating others well is an opportunity to invest in trust and goodwill. If you make ensuring that others are treated well a priority, those around you will notice, and it will foster emotional safety within your working group. It may also make others want to follow your example.

- **Mention collaborators when sending emails about your work or speaking publicly.** This is especially important if you are communicating with leadership or PR.
- **In drafts, put "author order not finalized" under the author list**, use a collective authorship as a placeholder, don't list authors, etc. This avoids an appearance that you've unilaterally determined the authorship order and creates a natural reminder to have a conversation about authorship down the road.
- **In drafts, keep a running list of people to acknowledge.** This reduces the risk of you forgetting to acknowledge someone. It also signals to them that you're taking this stuff seriously.
- **List collaborators in working notes.** If you refer back to them in the future, this will allow you to know who you were talking to and getting ideas from at the time.

Author Contributions / Order

Authorship order can be a tricky topic. It often helps to make credit distribution more nuanced, flexible, and continuous.

- **Consider an "Author Contributions Statement"** in multi-author papers saying what each author did. This is standard practice in many academic fields and required by many academic journals (eg. Nature, PLoS). Often, authors make contributions of different kinds and different magnitudes. A contributions statement can capture this nuance much more clearly than authorship order, allowing for fairer credit distribution. In my experience, this flexible and continuous nature can also makes discussing credit less contentious. Finally, contributions statements may also reduce "honorary authorship" problems, where senior people are on papers they didn't contribute to, by surfacing their level of involvement.
- **Write a contributions statement everyone agrees with before discussing author order.** Writing a contributions statement is often easier and more cooperative than discussing authorship order, because it's more flexible and more objective. It's also kind of positive sum in a strange way, because people are often attached to getting credit for particular things they did that feel important to them, and because listing all the work makes it feel like there's more credit to go around. It helps build empathy for the contributions others made, reduces the stakes of authorship order (because it matters less with a contributions statement), and gives a common starting point for a conversation.
- **Understand people's underlying motivations and needs.** If people disagree on authorship, it may help to understand what each person actually cares about. Sometimes this isn't actually a particular author position, but making sure that they get credit within an organization, or with someone they really respect. Sometimes there's a pragmatic reason why someone really needs a particular position.
- **If there are multiple "primary" authors consider joint first authorship, randomization, or both.**
- **If you can't find an order all authors feel good about, consider a mediated conversation or escalation.**

Mediation / Escalation

Ideally, we'd avoid all attribution conflicts with communication, thoughtfulness, and good collaboration practices. But we're not perfect, and we live in the real world, so conflicts do happen. Bringing someone else in to mediate the conflict is often the best path forward.

- **In cases of conflict, involve a neutral third party.** This is ideally the Lowest Common Manager, or independent respected researchers who are willing to give advice or facilitate conversations on attribution situations.
- **It's essential for the third party to be neutral.** A perception that the conflict was adjudicated by a biased party can make things worse.
- **Managers and senior researchers need to be compassionate, non-judgemental, and take attribution issues seriously.** People are often ashamed to bring up credit issues, because they worry they're being unreasonable or that they'll be perceived as such. This is true even when they're feeling hurt and alienated.

Situations to Look Out For

Be especially careful if any of the following risk factors are present:

- Borrowing or building on unpublished work.
- Collaborators involved in early parts of a project or closely related projects, but not the published result.
- Loss/destruction of code history (eg. submitting code written by someone else).
- Contributions from people who will not be authors.
- Aggregating multiple projects into a single larger one without aggregating authorship.
- Collaborators being in a condition where they can't advocate for their interests (eg. sick, on holiday, dealing with an emergency).
- Remote collaborations, or collaborations with people who have left your group.
- Revealing any part of the project publicly for the first time, such as by open-sourcing related code, or giving a talk about it.

Citing others

Citation serves an important dual role: helping readers find related work, and allocating credit within the research community. This second role makes it especially important to handle well: it can genuinely affect people's careers. In some cases, it may even effect their lives in more profound ways such as immigration status.

- **Cite all work you significantly build on.** This includes infrastructure, such as Theano, Pytorch, TensorFlow, R, or Genome Analysis Toolkit. It can also involve data such as ImageNet.
- **Citation is still necessary when there isn't a "paper version" of something.** If they helped or influenced your work, cite blog posts, artifacts such as code, private correspondence, or unpublished work. (This is what BibTeX `@misc` is for.)
- **Citation is especially important if work isn't highly cited or the author is junior.**
- **Getting citation right is especially important in review papers.**
- **Be open to being corrected if you missed something.**

You can even go a step beyond "citation" in the traditional sense by actively looking for opportunities to publicly praise and draw attention to the work of others. Of course, you don't want to be ungenuine. But if you're like me, you're often impressed by things that others do and never mention it publicly! This is a skill I'd like to practice more.

Taking Care of Yourself

While it's important and admirable to try and assume good faith, it's also important to not have that assumption trap you in unhealthy situations. You don't have an obligation to participate in collaborations that make you feel bad. This is true even if your collaborators are perfectly nice, well-intentioned people, or if you feel uncertain about whether they are. Someone doesn't have to be a malicious plagiarist for the working relationship to not be a good fit.

Example: X- was collaborating with researchers in another city. Over the course of several projects, X- felt like their work was used without them being credited. They felt unable to resolve the situation through conversations. Although X- knew that the cause might be distance and poor communication, they felt hurt and began to experience anxiety sharing unpublished work with these collaborators. In the end, X- decided it would be healthier for them to stop working together.

Acknowledgments

This document also would not be possible without many frank conversations with researchers who prefer to remain anonymous about difficult authorship and credit situations they've experienced.

My thinking on this topic also greatly benefited from conversations with Michael Page and Catherine Olsson, as well as the advice of Michael Nielsen when I was navigating a challenging situation of my own. Finally, I'm grateful for the comments and feedback of Shan Carter, Ian Johnson, Dario Amodei, Holden Karnofsky, Anna Goldie, Smitha Milli, Nick Beckstead, Arvind Satyanarayan, Yomna Nasser, Matt Hoffman, Emma Pierson, Martin Abadi, Greg Corrado, Amy McDonald, Jeff Dean, Samy Bengio, Sagyev Oore, Konstantinos Bousmalis, Peter Liu, Andrew Dai, Jasper Snoek, Delip Rao, and Vincent Vanhoucke. None of these acknowledgments should be seen as that person endorsing the views in this essay.

The wording "pool of goodwill" is inspired by Marilynne Robinson's "reservoir of goodness."

0 Comments