Arvind Narayanan on research careers, professional practice, and AI's impact

Selected posts, talks and essays, 2012–2026.

View online

I became a professor in 2012. Since then I've regularly shared advice on research careers, professional practice beyond academia, and AI's impact on work. I've compiled it here — about a hundred entries, edited lightly and organized thematically.

There's a lot of career advice online. Two things are distinct about this collection:

- Grounded optimism. I'm candid about the structural challenges, twisted incentives, and other limitations of the institutions we inhabit. I share ideas on how we can adapt and thrive in these flawed systems, and how and when we should challenge them.
- Connecting research and practice. Much of my research is on AI agent evaluation and AI's effects on the economy and labor, which lets me share insights into how we can effectively adapt and integrate it into our work.

For updates, follow me on LinkedIn or X.

CONTENTS

Part 1: Research careers

1. Should you do a PhD?
2. Growth during your PhD
3. Picking problems: taste and risk
4. Failure, quitting, and getting scooped
5. Research strategy: projects, portfolios, and distribution
6. Peer review, publishing, and incentives
7. Writing and speaking
8. Becoming and being a professor
9. Breadth, switching fields, interdisciplinarity

Part 2: Work, learning, and AI

10. Learning and professional growth
11. Getting things done
12. AI and the future of jobs
13. Working with AI — tips and strategies

Part 3: My story

Part 1: Research careers

1. SHOULD YOU DO A PHD?

Things to know before applying to a PhD

1. A PhD is not a means to an end. Don't start or suffer through a PhD at the expense of your mental health and life goals.

2. A PhD is a terrible financial decision.

3. A supportive environment for your PhD is much better than a highly ranked school.

More on those three points below.

4. Academic success depends way more on luck than professors like to admit, but there are ways you can manage the risk.

5. Consider getting a couple of years of life and work experience after undergrad before applying to PhD programs. It teaches you project management skills, shows you the opportunity cost of a PhD, and gives you a sense of what matters in the real world.

6. Don't let a few rejections deter you. There are many reasons having nothing to do with you why your application may be rejected, especially in a time of budget cuts. Here's my story.

x.com · 2020-10-28

A PhD is not a backup plan

Students sometimes tell me they're applying to PhD programs as a backup in case they don't get a job as a software engineer or whatever. I get that the job market is rough but this is like saying you'll start a company if you don't find a job. Like starting a company, a PhD is the far *more* risky path and requires incredible grit over a long period. It isn't just that you might realize after a couple of years that a PhD isn't for you. The bigger risk is that you complete a PhD but end up overqualified on a niche topic after having spent 5-7 years on it. The research world, like acting, sports, music, or book publishing, is set up as a tournament, with lots of extremely competent people competing for a small number of highly desirable slots.

We can talk about whether that's fair and whether it's inherent or can be changed, but as a professor I feel the least I can do is loudly and repeatedly warn people about what they're getting into.

x.com · 2026-09-18

Why PhD experiences are so variable and what you can do about it

People who do PhDs seem to have either strongly positive or strongly negative experiences -- for some, it's the best time of their lives, while others regret the decision to do a PhD. Few career choices involve such a colossal time commitment, so it's worth thinking carefully about whether a PhD is right for you, and what you can do to maximize your chances of having a good experience. Here are four suggestions. Like all career advice, your mileage may vary.

1. A PhD should be viewed as an end in itself, not a means to an end.

Some people find that they are not enjoying their PhD research, but decide to stick with it, seeing it as a necessary route to research success and fulfillment. This is a trap. If you're not enjoying your PhD research, you're unlikely to enjoy a research career as a professor. Besides, professors spend the majority of our time on administrative and other unrewarding activities. (And if you don't plan to be a professor, then you have even less of a reason to stick with an unfulfilling PhD.)

If you feel confident that you'd be happier at some other job than in your PhD, jumping ship is probably the right decision. If possible, structure your program at the outset so that you can leave with a Master's degree in about two years if the PhD isn't working out. And consider deferring your PhD for a year or two after college, so that you'll have a point of comparison for job satisfaction.

2. A PhD is a terrible financial decision.

Doing a PhD incurs an enormous financial opportunity cost. If maximizing your earning potential is anywhere near the top of your life goals, you probably want to stay away from a PhD. While earning prospects vary substantially by discipline, a PhD is unlikely to improve your career earnings, regardless of area.

3. The environment matters.

PhD programs can be welcoming and nurturing, or toxic and dysfunctional, or anywhere in between. The institution, department, your adviser, and your peers all make a big difference to your experience. But these differences are not reflected in academic rankings. When you're deciding between programs, you might want to weigh factors like support structures for mental health, the incidence of harassment, location, and extra-curricular activities more strongly than rankings. It is extremely common for graduate researchers to face mental health challenges. During my own PhD, I benefited greatly from professional mental health support.

4. Manage risk.

Like viral videos, acting careers, and startups, the distribution of success in research is wildly skewed. Most research papers gather dust while a few get all the credit -- and the process that sorts papers involves a degree of luck and circumstance that researchers often don't like to admit. This contributes to the high variance in PhD outcomes and experiences. Even for the eventual "winners", the uncertainty is a source of stress.

Perhaps counterintuitively, the role of luck means that you should embrace risky projects, because if a project is low-risk the upside will probably be relatively insignificant as well. How, then, to manage risk? One way is to diversify -- maintain a portfolio of independent research agendas. Also, if the success of research projects is not purely meritocratic, it follows that selling your work makes a big difference. Many academics find this distasteful, but it's simply a necessity. Still, at the end of the day, be mentally prepared for the possibility that your objectively best work languishes while a paper that you cranked out as a hack job ends up being your most highly cited.

An exercise I find useful is to ask myself which of my recent projects I would greenlight with the benefit of hindsight, knowing how they turned out. If the fraction is too low it means I should have used better judgment. But if it's too high it means I'm not taking enough risks.

Conclusion.

Many people embark on a PhD for the wrong reasons, such as their professors talking them into it. But a PhD only makes sense if you strongly value the intrinsic reward of intellectual pursuit and the chance to make an impact through research, with financial considerations being of secondary importance. This is an intensely personal decision. Even if you decide it's

right for you, you might want to leave yourself room to re-evaluate your choice. You should pick your program carefully and have a strategy in place for managing the inherent riskiness of research projects and the somewhat lonely nature of the journey.

A note on terminology. I prefer not to use the terms grad school and PhD student. The "school" frame is utterly at odds with what PhD programs are about. Its use misleads prospective PhD applicants and does doctoral researchers a disservice. Besides, Master's and PhD programs have little in common, so the umbrella term "grad school" is doubly unhelpful.

Why PhD experiences are so variable and what you can do about it · 2018-09-04 x.com · 2018-09-04

The benefit of work experience before a PhD

When evaluating PhD applications, I see it as a big plus when someone has a couple of years of life and work experience after undergrad. It teaches you project management skills, shows you the opportunity cost of a PhD, and gives you a sense of what matters in the real world.

Students who jump from undergrad to PhD tend to struggle with the transition from a rigidly structured to a totally unstructured environment. I played a lot of video games during the early years of my PhD and took a while to get back on track.

x.com · 2020-08-02 x.com · 2019-04-12

What we look for in PhD applicants

Prospective PhD applicants often ask me what we look for when evaluating candidates. What matters more: grades or letters of reference? What to emphasize in the Statement of Purpose? And so on. Here are some answers.

When we look at an application, we are trying to answer one question above all else: “is this person likely to succeed at research?” While there is some difference between departments and institutions, I believe that most of them have the same basic goal for PhD admissions.

In evaluating research potential, grades don’t matter much. Grades below a certain point could be a sign of inadequate preparation for a PhD, but perfect grades aren’t a predictor of research success, nor are GRE scores.

The best way to show evidence of research potential is to have done research. A strong paper or two will carry the day. But there can be a big lag between starting research and publishing it, so letters that describe your research abilities & contributions are extremely

valuable.

Admittedly, this puts students from non-research-focused universities at some disadvantage. But there are ways around this. One is a research internship (check out the NSF REU program in particular). Another is to do a Master's before applying for a PhD.

But it's also important to acknowledge the messy realities of PhD admissions.

x.com · 2018-12-03

2. GROWTH DURING YOUR PHD

That continual feeling of cluelessness when doing research

Hello, Ph.D students! I bring news from professor land. You know how when you're doing research you constantly feel like you have no clue what you're supposed to be doing? Turns out that doesn't go away when you become a professor. We just act all confident when we advise you.

That's just the nature of research — 90% of the time it feels like banging your head against a wall. It makes the other 10% so much more gratifying.

When we write papers we tell the story as if we knew what we were doing all along. We leave out the blind alleys and walls we bumped into. There are good reasons for that, but it does mean we should find other ways of sharing and learning from the real stories behind research.

x.com · 2018-01-05

The term "grad school" is misleading

Calling PhD programs "grad school" beguiles people into thinking somebody will tell them exactly what to do. But figuring out what to do is a big part of a PhD; the university should provide a supportive environment for this inherently somewhat solitary quest.

Calling PhD students graduate researchers is not only more accurate, but also underscores how little they're paid relative to their ability and contribution. In many fields, PhD students (and postdocs) do all the research, and professors are basically managers.

The grad "school" metaphor does further damage by implying that the main goal is to impress your school with your dissertation. But in fact the world is judging you, and the goal is to establish yourself as an expert on your topic. These two goals are only somewhat aligned.

x.com · 2018-01-11

Don't pick a thesis direction too early

The most common mistake I see people make when they start their PhDs is trying to pick a thesis direction way too early. IMO you need to explore for at least a couple of years and try many different things before you find a direction that's a good fit.

Halfway through my own PhD, I'd worked on about a dozen different things without much to show for it, and considered myself a failure. But in retrospect this turned out to be exactly the right approach. Here's why.

Research projects succeed when at least 4 things happen:

1. The problem is important and *recognized* as important

One might get a paper out of solving a made-up problem, but is that really success? Or one might be ahead of one's time, like Mendel with genetics, and be ignored because the significance is not recognized by the community.

2. The problem is amenable to existing research methods

It's often hard to know whether the current state of the field is adequate to tackle a problem, or whether it will take decades to get there.

3. The researcher is passionate about the research area

This one goes without saying — without passion it will be hard to sustain the motivation needed to work consistently until you see the light at the end of the tunnel.

4. The researcher has the skills to solve the problem

This must be true not just in an absolute sense but also in a relative sense. If the problem is known to be important (see #1 above) there will be a whole bunch of smart people working on it, all trying to be the first to solve it.

In addition, you need a dose of luck. In research that's a given.

How often do all these things happen at once?

This is why the early stages of a PhD often feel like flailing. The best thing an advisor can do is to help the researcher come to terms with the fact that this is completely normal and even necessary, and give them the freedom to explore by providing a safety net in terms of funding etc.

LinkedIn · 2024-04-16

My proudest moment as an advisor

My proudest moment as an advisor is when an advisee tells me they're rejecting my advice, and gives me a convincing reason why. The sooner a junior scholar gets to the point where they have the judgment, self-knowledge, and confidence to do this, the better for their career.

Another milestone is no longer needing the advisor to suggest research problems. In my field, most researchers enter PhD programs already having a pretty good idea of how to *solve* research problems, but learning to *ask* good research questions is... basically the entire PhD.

I do sometimes encounter trainees who reject perfectly good advice out of overconfidence, but the far more common problem seems to be giving too much deference to the advisor. Never forget that advisors are just making it up as we go along!

x.com · 2022-08-03

Cringing at your own old papers is normal

When I started out as a researcher I had so much to learn about my topics of interest that I'd feel embarrassed whenever I looked at a paper I'd written a year or two previously.

Now, as a tenured professor, it's exactly the same. What's changed is I've realized that's normal!

It no longer surprises me when I tell a famous scholar how much I enjoyed a seminal paper they wrote and they react by physically recoiling.

There's an ironic selection effect where seminal papers are *more* likely to embarrass the author in later years. That's because if the work is impactful they're likely to continue working on the topic for years and their understanding will improve a lot.

3. PICKING PROBLEMS: TASTE AND RISK

How to pick your first research project

At Princeton I get to advise many gifted graduate and undergraduate students in doing research. Combining my experience as a mentor with reflecting on my experience as a student, here's some guidance on how to pick your first research project.

I'm writing this because selecting a research problem to work on is significantly harder than actually solving it. I mean the previous sentence quite literally and without exaggeration. As an undergraduate and early-stage graduate researcher, I repeatedly spent months at a time working on research problems only to have to abandon my efforts because I found out I was barking up the wrong tree. Scientific research, it turns out, is largely about learning to ask the right questions.

The good news is that three simple criteria will help you avoid most of the common pitfalls.

1. Novelty

Original research is supposed to be, well, original. There are two components to novelty. The first is to make sure the problem you're trying to solve hasn't already been solved. This is way trickier than it seems -- you might miss previous research because you're using different names for concepts compared to the standard terminology. But the issue is deeper: two ideas may be equivalent without sounding at all the same at a superficial level. Your advisor's help will be crucial here.

The other aspect to novelty is that you should have a convincing answer to the question "why has this problem not been solved yet?" Often this might involve a dataset that only recently became available, or some clever insight you've come up with that you suspect others missed. In practice, one often has an insight and then looks for a problem to apply it to. This means you have to put in a good bit of creative thinking even to pick a research question, and you must be able to estimate the difficulty level of solving it.

If your answer to the question is, "because the others who tried it weren't smart enough," you should probably think twice. It may not be prudent to have the success of your first project ride on your intellectual abilities being truly superlative.

2. Relevance

You must try to ensure that you select a problem that matters, one whose solution will impact the world directly or indirectly (and hopefully for the better). Again, your advisor's help will be essential. (That said, professional researchers do produce massive volumes of research papers that no one cares about.) I encourage my students to pick subproblems of my ongoing long-term research projects. This is a safe way to pick a problem that's relevant.

3. Measurable results

This one becomes automatic as you get experienced, but for beginning researchers it can be confusing. The output of your research should be measurable and reproducible; ideally you should be able to formulate your goals as a testable hypothesis. Measurability means that many interesting projects that are novel and make the world better are nevertheless unsuitable for research. (They may be ideal for a startup or a hobby project instead.) "Build a website for illiterate kids in poor countries to learn effectively" is an example of a task that's hard to frame as a research question.

Irrelevant criteria

Let me also point out what's not on this list. First, the general life advice you often hear, to do something you're passionate about, is unfortunately a terrible way to pick a research problem. If you start from something you're passionate about, the chance that it will meet the three criteria above is pretty slim. Often one has to consider a dozen or more research ideas before settling on one to work on.

You should definitely pick a research *area* you're passionate about. But getting emotionally invested in a specific *idea* or research problem before you've done the due diligence is a classic mistake, one that I made a lot as a student.

Second, the scope or importance of the problem is another criterion you shouldn't fret much about for your first project. Your goal is as much to learn the process of research as to produce results. You probably have a limited amount of time in which you want to evaluate if

this whole research thing is the right fit for you. While you should definitely pick a useful and relevant research task, it should be something that you have a reasonable chance of carrying to fruition. Don't worry about curing cancer just yet.

Note that the last point is at odds with advice given to more experienced researchers. Richard Hamming, in a famous talk titled "You and your research", advised researchers to pick the most important problem that they have a shot at solving. I've written a version of the current post for those who're in it for the long haul, and my advice there is to embrace risk and go for the big hits.

How to pick your first research project · 2013-11-01

The importance of research taste

An early-career researcher asked me which topics in my field are under-explored. An important question! But not all researchers in a community will agree on the answers. If they did, those topics won't stay under-explored for long. So how to pick problems?

It's helpful for researchers to develop a "taste" for problems based on their specific skills, preferences, and hypotheses about systemic biases in the research community that create blind spots. I shared two of my hypotheses with the student, but we must each develop our own.

Hypothesis 1: interdisciplinary topics are under-explored because it requires researchers to leave their comfort zones. But collaboration is a learnable skill, so if one can get better at it and find suitable collaborators, rich and important research directions await.

Hypothesis 2 (for computer science research): work that brings accountability to the tech industry's claims is pursued less than research that responds to the industry's needs. The reasons are, of course, funding, prestige, and other such incentives.

These hypotheses have guided me for most of my career, not just in picking research topics but in the way I conduct research — e.g. nurturing collaborations; learning from civil society organizations and investigative journalists who do critical tech accountability work.

Some researchers stay within a topic and develop deep expertise over decades. That's a perfectly valid way to do things. I've instead found it fruitful to evolve a strong taste and research style while migrating relatively frequently between topics. This way is worth considering!

x.com · 2021-06-21

Spotting promising research ideas

A vital skill for a researcher is to be able to take a step back when you have an idea for solving an important problem that no one seems to have worked on, and figure out if it's because you know something that others don't or because they all know something that you don't.

1. Not all knowledge in a community is recorded formally. Networking fills the gap. (But the need for networking contributes to systemic biases and we must work against them.)

2. Spend time earlier in the research pipeline and be prepared to drop ideas.

3. Work with a mentor. Over time, people build up a repository of ideas and can tell which ones are truly new. As an advisor, much of what I offer is helping my mentees filter and prioritize their ideas.

4. Try many things, because any single idea is always risky.

5. Just because an idea has been tried many times doesn't mean it isn't worth another go. Ideas may finally succeed because the *environment* has changed — new tools and techniques may be available, or the community's biases may have changed to be more accepting of the idea.

x.com · 2020-07-21

Contracting the boundaries of knowledge

One way to make your mark as a researcher is to discover new things. But even more valuable is to question existing knowledge. Throughout my career I've been surprised by the speed with which castles of traditional wisdom tend to come crashing down if you poke them the right way. Do we really know what we think we know, or are we falling for hype and groupthink?

In fact, my favorite kind of research is that which shows that a widespread, unquestioned assumption in a research field is completely wrong.

Scholars aspire to expand the boundaries of knowledge but I must admit I've really enjoyed contracting the boundaries of knowledge.

x.com · 2020-11-10 x.com · 2017-09-25 x.com · 2024-04-17

Speculate about the world 5–10 years out

When I was doing startups I picked up the habit of speculating about what the world will look like in 5-10 years. This has turned out to be extremely useful in my research career, too. Research ideas have a long arc, so we need to learn to hit moving targets from a long range.

I made many predictions about tech during my PhD. The majority were wrong. Much of my success today stems from one that I got right: tech will play a much bigger role in our lives, the industry will lack accountability, and computer science research can help address this need.

If this seems obvious today, it's because correct predictions tend to look obvious in retrospect. Many predictions that I got spectacularly wrong — I thought print books would disappear by 2020 — seemed just as obvious to me back in the day.

x.com · 2020-08-28

Some ideas fail because the world isn't ready

Some ideas don't succeed at first because the world isn't yet ready for them. After realizing how common this is, I've been learning to control my frequent urge to shout "this isn't new — it was tried 20 years ago!" and instead ask myself if the idea is worth another chance.

One way to try to quantify this is to look at the citation patterns of research papers. A 2015 study found that many papers are "sleeping beauties" — they gather dust for decades and later become influential, likely because they were ahead of their time.

x.com · 2021-01-13

Why does your field exist?

"Why does your field exist?" is a question every scholar should be prepared to answer. Even if you're never asked that question, it's a useful kind of reflection to engage in once in a while.

In 2012 I was on the academic job market. The majority of professors I met asked why I chose to work on privacy since—as we all know—privacy is dead because of the Internet and it's pointless to fight it. (Computer scientists tend to be technological determinists, who knew?!)

At first I didn't expect that "why does your research field exist?" would be a serious, recurring question. Gradually I came up with a pitch that at least got interviewers to briefly suspend privacy skepticism and hear about my research.

I'm glad I persevered through my peers' skepticism and continued working on privacy. I couldn't have known that the pendulum would start to swing back the very next year after my job interviews due to the Snowden leaks (and, later, Cambridge Analytica).

x.com · 2021-06-30

Everyone should spend time on the philosophy of their own field

In my dream version of the scientific enterprise, everyone who works on X would be required to spend some percentage of their time learning and contributing to the philosophy of X. There is too much focus on the "how?" and too little focus on the "why?" and the "what are we even...?".

Junior scholars entering a field naturally tend to ask critical questions as they aren't yet inculcated into the field's dogmas. But the academic treadmill leaves them little time to voice concerns and their lack of status means that even when they do, they aren't taken seriously.

One possible intervention is for journals and conferences to devote some fraction of their pages / slots to self-critical inquiry, and for dissertation committees to make clear that they will value this type of scholarship just as much as "normal" science.

In other words, we don't even need to create new incentives to encourage scholars to grapple with meta-questions. We just need to remove the disincentives that are currently in place.

x.com · 2021-07-19

Gravitate to the simplest questions, not the hardest

This is the story of how reading “The Selfish Gene” when I was around 15 changed my career decades later. It’s a terrific book. But beyond its substance, it changed my view of what science can be. It showed me that there are simple but profound ideas waiting to be discovered. I’d thought of the frontier of science as necessarily esoteric, but the book proved otherwise. Richard Dawkins’s writing also showed me that it’s possible to explain novel and profound ideas in a way that even a child (me) could understand them.

When I grew up and became a researcher, I never stopped thinking about this. I began to gravitate toward the simplest questions within my areas of expertise, rather than the hardest, contrary to the norm in science. And I taught myself how to communicate my ideas to as broad an audience as possible. Unfortunately, the peer review process heavily penalizes this approach, because the value system prioritizes abstruseness, when ideally it should be the opposite. But no matter — I found that simpler ideas, when they do get published, are much more widely read, which made it all worthwhile. Besides, pushing to make ideas as simple and as simply communicated as possible often made them *better ideas*, more robust and widely applicable than initially anticipated.

Aspirations should be balanced with an awareness of one's limitations. Not everyone can be as successful as Dawkins; I realized that I couldn't count on my ideas being so powerful that they would spread on their own (fittingly, the term "meme" was coined in The Selfish Gene!) So I've tried to put as much effort into spreading ideas as I do into generating and explaining them. That's a topic I've written about elsewhere.



Richard Dawkins in 1976, around the time he published his first best-selling book. Credit: Terry Smith/The LIFE Images Collection/Getty

Books about science tend to fall into two categories: those that explain it to lay people in the hope of cultivating a wide readership, and those that try to persuade fellow scientists to support a new theory, usually with equations. Books that achieve both — changing science and reaching the public — are rare. Charles Darwin's *On the Origin of Species* (1859) was one. *The Selfish Gene* by Richard Dawkins is another. From the moment of its publication 40 years ago, it has been a sparkling best-seller and a scientific game-changer.

x.com · 2025-12-19

4. FAILURE, QUITTING, AND GETTING SCOOPED

Embracing failure: How research projects are like startups

As an academic who's spent time in the startup world, I see strong similarities between the nature of a scientific research project and the nature of a startup. This boils down the fact that most research projects fail (in a sense that I'll describe), and even among the successful

projects the variance is extremely high -- most of the impact is concentrated in a few big winners.

Of course, research projects are clearly unlike startups in some important ways: in research you don't get to capture the economic benefit of your work; your personal gain from success is not money but academic reputation (unless you commercialize your research and start an actual startup, but that's not what this post is about at all.) The potential personal downside is also lower for various reasons. But while the differences are obvious, the similarities call for some analysis.

I hope this post is useful to grad students in particular in acquiring a long-term vision for how to approach their research and how to maximize the odds of success. But perhaps others including non-researchers will also find something useful here. There are many aspects of research that may appear confusing or pathological, and at least some of them can be better understood by focusing on the high variance in research impact.

1. Most research projects fail.

To me, publication alone does not constitute success; rather, the goal of a research project is to impact the world, either directly or by influencing future research. Under this definition, the vast majority of research ideas, even if published, are forgotten in a few years. Citation counts estimate impact more accurately, but I think they still significantly underestimate the skew.

The fact that most research projects don't make a meaningful lasting impact is OK -- just as the fact that most startups fail is not an indictment of entrepreneurship.

2. Grad students: diversify your portfolio!

Given that failure is likely (and for reasons you can't necessarily control), spending your whole Ph.D. trying to crack one hard problem is a highly risky strategy. Instead, you should work on multiple projects during your Ph.D., at least at the beginning. This can be either sequential or parallel; the former is more similar to the startup paradigm ("fail-fast").

I achieved diversity by accident. Halfway through my Ph.D. there were at least half a dozen disparate research topics where I'd made some headway (some publications, some works in progress, some promising ideas). Although I felt I was directionless, this turned out to be the

right approach in retrospect. I caught a lucky break on one of them, and from then on I doubled down to focus on it. This breadth-then-depth approach paid off.

3. Go for the big hits.

Paul Graham's fascinating essay Black Swan Farming is about how skewed the returns are in early-stage startup investing. Just two of the several hundred companies that YCombinator had funded were responsible for 75% of the returns, and in each batch one company outshined all the rest.

The returns from research aren't quite as skewed, but they're skewed enough to be highly counterintuitive. This means researchers must explicitly account for the skew in selecting problems to work on. Following one's intuition and/or the crowd is likely to lead to a mediocre career filled with incremental, marginally publishable results. The goal is to do something that's not just new and interesting, but which people will remember in ten years, and the latter can't necessarily be predicted based on the amount of buzz a problem is generating in the community right now. Breakthroughs often come from unsexy problems (more on that below).

There's a bit of a tension between going for the hits and diversifying your portfolio. If you work on too few projects, you incur the risk that none of them will pan out. If you work on too many, you spread yourself too thin, the quality of each one suffers, and lowers the chance that at least one of them will be a big hit. Everyone must find their own sweet spot. One piece of advice given to junior professors is to "learn to say no."

4. Find good ideas that look like bad ideas.

How do you predict if an idea you have is likely to lead to success, especially a big one? Again let's turn to Paul Graham in Black Swan Farming:

"the best startup ideas seem at first like bad ideas. ... if a good idea were obviously good, someone else would already have done it. So the most successful founders tend to work on ideas that few beside them realize are good."

Something very similar is true in research. There are some problems that everyone realizes are important. If you want to solve such a problem, you have to be smarter than most others working on it and be at least a little bit lucky.

Then there are research problems that are analogous to Graham's good ideas that initially look bad. These fall into two categories: 1. research problems that no one has realized are important 2. problems that everyone considers prohibitively difficult but which turn out to have a back door.

If you feel you are in a position to take on obviously important problems, more power to you. I try to work on problems that everyone seems to think are bad ideas (either unimportant or too difficult), but where I have some "unfair advantage" that leads me to think otherwise.

In summary, if you're a researcher, think carefully about which projects to work on and what the individual and overall chances of success are.

Embracing failure: How research projects are like startups · 2013-01-02 x.com · 2013-01-04

The idea funnel

In a past life I took stand-up comedy lessons and even performed a little. I learnt that for every joke in your set, there were 10 that didn't land well when you tried them, 100 that you wrote down but weren't funny enough to try, and 1,000 you thought of but didn't write down.

Turns out this also applies to research. You'll have many "aha moments", but most of those ideas either won't work or will have been done before. And of the ones you publish, only a few will prove as significant as you'd hoped. This is normal—don't let it faze you!

x.com · 2018-02-11

The flip side of passion and perseverance

I've learnt the hard way that in research we need to be passionate about our research topics while putting emotional distance from specific ideas. Only about 10% of ideas actually work, so if you get too emotionally invested too early, you won't know when it's time to quit.

Often your peers will tell you your idea isn't good, or simply ignore it. Nine out of ten times they're right. Occasionally they're wrong, and need to be convinced with better evidence. How can you tell the difference? Only by being relatively dispassionate.

We've all heard inspiring stories of researchers who persisted for years or decades while the world laughed at them, and finally succeeded. I love such stories—perseverance is a great quality—but so is knowing when to quit. I often see misplaced perseverance in young researchers.

It's hard to tell an inspiring story about a researcher who took a cold, hard look at the data, saw the writing on the wall, submitted it to the Journal of Negative Results, wrote a two-page "lessons learned" memo to the group, and moved on quickly. And yet that's a vital skill.

x.com · 2019-07-30

Getting scooped feels like a gut punch. It doesn't have to.

Getting scooped is a fact of life for every researcher. It feels like being punched in the gut. After decades of being terrified, I've learned that there are many things we can do to reduce the risk. More importantly, getting scooped is not nearly as big a deal as I thought.

Looking back, 3 of my 5 most impactful/most cited papers were actually scooped before we published them! In none of those cases did my fears come true. Being scooped didn't seem to negatively affect those papers at all. There's research that backs this up.

If you get scooped, the thing to do is pivot. A paper is a complex and multifaceted exploration of an idea, so it's exceedingly unlikely that two papers will have exactly the same set of insights. In most cases you can reframe your paper to emphasize what's distinct about it.

I've found three good ways to reduce the risk of being scooped. The first is to work on something that's far from the mainstream of your community, like solving a problem that your community hasn't even recognized as a problem worth solving.

Obviously, this has its own downsides. You need to be sure you know something that others don't, and not vice versa. And when you're done, you'll need to work extra hard to convince your community of the paper's importance. It's best not to work only on this type of paper.

The second strategy is to complete papers faster. This doesn't mean doing shoddy work. Sometimes we slow down at the end because we run out of steam, or we can't bring ourselves to call it done and submit it because of perfectionism. We can train ourselves to avoid those traps.

I've noticed sometimes I try to stuff too much into one paper. By recognizing that what you thought was a paper is actually a series of papers, you can get the first paper out sooner. And please, release a preprint. It *decreases* the risk of being scooped.

The third way is to network and be better connected in your community (which is good for many reasons). In my experience, researchers who trust & respect each other, if they realize they are working on the same thing, will *usually* decide to cooperate rather than compete.

x.com · 2022-06-27

There is no shame in being wrong in public

There is absolutely no shame in making mistakes, as long as we learn from them. Budding scientists seem to get socialized to live in constant fear of saying something publicly that might turn out to be wrong. It took me decades to let go of this. We need to fix the culture.

x.com · 2022-07-16 x.com · 2014-04-04

5. RESEARCH STRATEGY: PROJECTS, PORTFOLIOS, AND DISTRIBUTION

How the "marketplace of ideas" actually works

Back in grad school, when I realized how the “marketplace of ideas” actually works, it felt like I’d found the cheat codes to a research career. Today, this is the most important stuff I teach students, more than anything related to the substance of our research.

A quick preface: when I talk about research success I don’t mean publishing lots of papers. Most published papers gather dust because there is too much research in any field for people to pay attention to. And especially given the ease of putting out pre-prints, research doesn’t need to be officially published in order to be successful. So while publications may be a prerequisite for career advancement, they shouldn’t be the goal. To me, research success is authorship of ideas that influence your peers and make the world a better place.

So the basic insight is that there are too many ideas entering the marketplace of ideas, and we need to understand which ones end up being influential. The good news is that quality matters — other things being equal, better research will be more successful. The bad news is that quality is only weakly correlated with success, and there are many other factors that matter.

First, give yourself multiple shots on goal. The role of luck is a regular theme of my career advice. It’s true that luck matters a lot in determining which papers are successful, but that doesn’t mean resigning yourself to it. You can increase your “luck surface area”.

For example, if you always put out preprints, you get multiple chances for your work to be noticed: once with the preprint and once with the publication (plus if you're in a field with big publication lags, you can make sure the research isn't scooped or irrelevant by the time it comes out).

More generally, treat research projects like startups — accept that there is a very high variance in outcomes, with some projects being 10x or 100x more successful than others. This means trying lots of different things, taking big swings, being willing to pursue what your peers consider to be bad ideas, but with some idea of why you might potentially succeed where others before you failed. Do you know something that others don't, or do they know something that you don't? And if you find out it's the latter, you need to be willing to quit the project quickly, without falling prey to the sunk cost fallacy.

To be clear, success is not all down to luck — quality and depth matter a lot. And it takes a few years of research to go deep into a topic. But spending a few years researching a topic before you publish anything is extremely risky, especially early in your career. The solution is simple: pursue projects, not problems.

Projects are long-term research agendas that last 3-5 years or more. A productive project could easily produce a dozen or more papers (depending on the field). Why pick projects instead of problems? If your method is to jump from problem to problem, the resulting papers are likely to be somewhat superficial and may not have much impact. And secondly, if you're already known for papers on a particular topic, people are more likely to pay attention to your future papers on that topic. (Yes, author reputation matters a lot. Any egalitarian notion of how people pick what to read is a myth.)

To recap, I usually work on 2-3 long-term projects at a time, and within each project there are many problems being investigated and many papers being produced at various stages of the pipeline.

The hardest part is knowing when to end a project. At the moment you're considering a new project, you're comparing something that will take a few years to really come to fruition with a topic where you're already highly productive. But you have to end something to make room for something new. **Quitting at the right time always feels like quitting too early.** If you go with your gut, you will stay in the same research area for far too long.

Finally, build your own distribution. In the past, the official publication of a paper served two purposes: to give it the credibility that comes from peer review, and to distribute the paper to your peers. Now those two functions have gotten completely severed. Publication still brings credibility, but distribution is almost entirely up to you!

This is why social media matters so much. Unfortunately social media introduces unhealthy incentives to exaggerate your findings, so I find blogs/newsletters and long-form videos to be much better channels. We are in a second golden age of blogging and there is an extreme dearth of people who can explain cutting-edge research from their disciplines in an accessible way but without dumbing it down like in press releases or news articles. It's never too early — I started a blog during my PhD and it played a big role in spreading my doctoral research, both within my research community and outside it.

Summary

- Research success doesn't just mean publication
- The marketplace of ideas is saturated
- Give yourself multiple shots on goal
- Pick projects, not problems
- Treat projects like startups
- Build your own distribution

x.com · 2025-07-21

Pick projects, not problems

One question I'm sometimes asked is how my research group picks problems. Do I come up with most of the ideas for new papers, or do the students? Neither!

I strongly believe that research is more effective if we pick projects, not problems. What's the difference?

- Projects are long-term research agendas that last 3-5 years or more. A productive project could easily produce a dozen or more papers (depends on the field, of course — in some fields papers represent a lot more work than in others).

- Projects are defined not by a research question but by a change we want to see in the world. For example, the goal of a current project in my group is to make AI more reliable. We may or may not succeed, but the point is that this is a much more ambitious scope than can be tackled in a single paper.
  (Some fields have a norm that their job is only to describe the world, not change it. This is culturally jarring to me but even in that case I think projects are better defined in terms of a change you want to see in the research community, if not the external world.)
- Projects are best executed by a core team that stays together and provides intellectual continuity but with a diverse and varying set of collaborators for individual papers which helps constantly bring in new perspectives.

Why pick projects instead of problems? If your method is to jump from problem to problem, you face a tradeoff. You could pick small problems that you can tackle in a month or two, but in that case the resulting papers may not have much impact. Or you can go deep into a topic for many years (essentially what I've described as a project, but structured as a single paper), but that's extremely risky.

In my experience, once a research team is committed to a project, generating the research questions that individual papers in the project will tackle is fairly straightforward. Each paper in the project naturally generates a bunch of new questions and directions for future work. So generating new ideas is not the hard part, rather it is the profusion of ideas. How to select among them? Ideally some combination of intellectual curiosity and whatever best furthers the project's overall goals and vision.

x.com · 2025-03-21

Time allocation in research

This is why most research is so ineffective and how you can make yours more effective.

(While I pulled the numbers out of a hat to illustrate a point, I do think the right hand column is a helpful rough-and-ready guide for time allocation.)

Component of research	Fraction of time spent by average researchers	Fraction of time spent by great researchers
Discussing, sifting and selecting research ideas	0%	25%
Carefully formulating the problem	5%	25%
Solving the problem	90%	25%
Communicating the findings effectively	5%	25%

x.com · 2019-11-16

Deadline-driven versus value-driven research

At the start of my research career I operated in a deadline-driven mode because that's what most researchers seemed to do. Gradually I discovered the value-driven way of working. I'm glad I had a supportive advisor who didn't make me chase deadlines. It took me 20 years to fully embrace the switch — it requires developing a long-term vision, willpower to create structure without deadline pressure, a theory of value, project management skills, good taste, the willingness to turn projects down, brutal honesty about whether our work is any good (even if it gets published), and a lot more. But there is no going back!

Two ways of getting research done

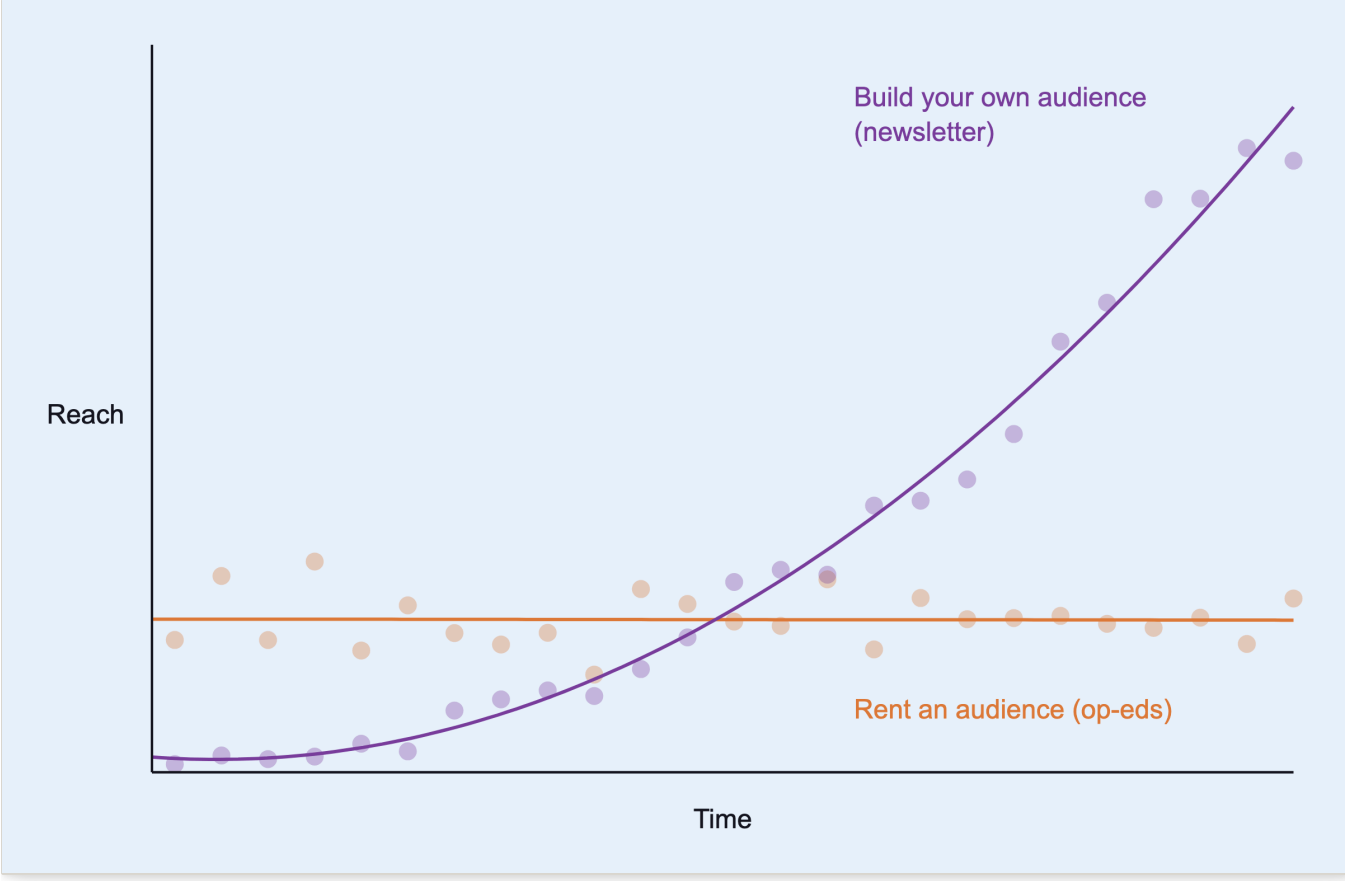
	Deadline driven	Value driven
Unit of work	Papers	Projects (long-term agendas that last 3-5 years)
Source of discipline	External	Internal—requires willpower, creating structure for oneself
What is prioritized	Whatever is close to deadlines	What's most valuable; requires a theory of value
Coordination mechanism	Deadline as a forcing function	Good project management skills
Quality bar	Whatever gets past peer review	Independent assessment of quality
Motivation	Brownie points from getting papers in	Making the world better
Project selection	Say yes promiscuously; drop things that don't get done for the deadline or if reviews are terrible	Highly selective; explicit go/no-go decision; commit to finishing things that are started
Stakes	Low — affects only the research community	High — affects real people
Endpoint	Paper gets accepted/published	The paper is just documentation, not the output, much less the outcome. Projects last years
Feedback signal	Meeting deadlines + paper acceptance. Fast, binary, objective (but ultimately meaningless).	"Did it make the world better?" Fuzzy; requires investing in measurement and being honest with oneself.

The saddest thing about deadline culture is what it does to researchers over time. Tenure is supposed to free us from publish-or-perish incentives. But by the time they get there, most researchers have been on the deadline treadmill for so long they've forgotten any other way to do things. The amount of wasted potential in academia is shocking and sad.

x.com · 2026-07-02

Build your own distribution

Some people have asked me about the decision to start a newsletter rather than using other channels like op-eds. There's one key thing I've learned, and I made a stylized chart to explain it. If you build your own audience, your reach grows over time with each piece you write. But it takes a *long* time — years in most cases — before that audience grows big enough that it all starts to feel worthwhile. Start a newsletter only if you know you're in it for the long haul. (I started my newsletter with Sayash Kapoor when we started writing our book together, so we knew we would be writing about AI for many years.)



There are many other advantages to controlling your own channel, including not having to go through the crapshoot of pitching op-eds or worrying about word limits. In the long run, it is likely to be the better choice. But it's a lonely journey and there is no guarantee of success.

x.com · 2025-08-12

Put it on arXiv already

I'm often told I'm "brave" for putting my papers on arXiv before they are accepted for publication. How silly! It's 100x more likely for academic papers to languish in obscurity than for the ideas in them to be so good that someone would want to steal them.

When I ask those paranoid researchers if they've ever had an idea stolen, it always turns out to be something that happened to a friend of a friend of a friend.

The reason why ideas don't get stolen is not that academia is a paragon of virtue. It's more that someone plagiarizing from arXiv has a high risk of getting caught. Besides, ideas are a dime a dozen and everyone else is too obsessed with their own ideas to pay attention to yours.

An arguably good reason to not put your paper on arXiv is that it might compromise anonymous review. If your community has such a norm, please do respect that. But that's totally different from suppressing your papers because of imagined risks.

If your research is funded by taxpayers, you have a moral duty to make it publicly available in a timely way. Don't put half-baked work on arXiv, but also don't sit on a draft for years while it's trudging through the broken peer review system.

Often an idea doesn't work out, and journals don't accept the paper due to negative results (which is terrible, but that's a different rant.) So the researcher just buries the draft, leaving others to rediscover the same flawed idea and repeat the work. What a tragic waste!

In my experience, for selfish reasons alone, releasing a draft as soon as the paper is ready is absolutely the optimal strategy. Much of my best known work was public for *years* before it was official. Not-yet-accepted work was a big part of my job talk and even my tenure case.

x.com · 2019-11-01

Taking advantage of arbitrage in academia

Academics tend to follow the herd. This yields many "arbitrage" opportunities. Here's one: some ideas deserve to be books; some deserve to be papers. But most academic fields stick to one or the other. So if you're in a "book field", write papers. In a paper field, write books.

In computer science there's an overproduction of papers to an absurd degree while the demand for books is pitifully unmet. I've been trying to shift about half of my writing to the form of books.

A useful piece of research meta-advice that I've tried to follow: identify what seem to be unwritten rules and norms in your community. Search hard for valid reasons behind those norms (such as ethics). Absent such reasons, systematically violate those norms in your work.

x.com · 2019-01-11 x.com · 2019-01-11

Brilliance is vastly overrated. Learn basic productivity techniques instead.

Academics would double our productivity if we learnt some basic project management skills that are bog standard in the industry. We have this myth that scholarly success is all about brilliance and creativity, but in fact 90% of it is getting sh*t done, same as any other job.

Like most scholars, I used to start too many projects and inevitably let half of them wither and die slowly due to inattention. But a few years ago I began making go/no-go decisions after the preliminary stage of a project. Never mind productivity, it has let me regain my sanity!

A 1-hour conversation with your collaborators after having worked on a project a bit, where you honestly examine your time availability and discuss whether the idea is as promising as you thought at first will save you so much wasted work and awkwardness down the line.

Hardly anyone in academia does a retrospective after every paper/project, or even regularly admits failures. Tenure-track professors are evaluated very infrequently, so it's easy for us to never confront things that aren't going well and learn from them.

I really enjoy collaborations with people who've been outside the bubble, because they'll say something like "here are last week's notes in advance of today's meeting" and I'll be like "oh good, you took notes, I didn't think to do that and I've already forgotten everything".

To be clear, we shouldn't fall into the efficiency trap, which is feeling pressure to optimize every minute. The kind of productivity I'm talking about is about doing more of the work we find meaningful over the long run. In my experience, project management helps.

Finally, time management helps creativity, not hurt it. If you want to do deep work you need to bunch your meetings and have as many uninterrupted stretches as possible. That's my superpower.

x.com · 2022-06-02

How to keep up with research (answer: don't!)

Were you told that successful researchers must constantly "keep up" with research in their fields? In my experience, that's both stressful and inefficient to the point of being useless. New papers may be released every day but actual knowledge doesn't accumulate at that rate.

Paying too much attention to the so-called cutting edge of research leads to a loss of perspective and results in incremental work more often than not. If you want to do foundational research, it probably doesn't rely on a preprint that was released last week.

Here's the process I've used for well over a decade. When I see a new paper, I put it in a big list organized by topic. I don't read it right away. Once in a while, I notice that a collection of papers on a topic have resulted in meaningful progress, and I read the papers together.

By reading papers in groups, it's easier to tell which are truly significant. Flaws/limitations become clear. Terminology gets refined over time. The best description of a paper may be in follow-up work.

I've found that this process is more rewarding and builds deeper understanding.

x.com · 2021-02-17

6. PEER REVIEW, PUBLISHING, AND INCENTIVES

Peer review is like democracy

Peer review is like democracy: the worst system except for all the ones we've tried before. We can't throw peer review out yet, but we should be trying our hardest to figure out what comes next. Unlike democracy, we can easily experiment with scientific publishing.

The problem isn't the idea of peer review—how can giving feedback on each others' work be bad?!—but the institutionalization. If quality control is treated as binary, rather than iterative, imperfect, and needing the reader to be responsible, it will fail no matter how it's done.

A silver lining of the flood of AI papers is that it might stretch the system beyond breaking point and force us to come up with something better in its place.

x.com · 2023-07-18

What peer review was like in the 1860s

I often find myself re-reading this short piece about what peer review was like in the 1860s. A reviewer was someone who helped improve a paper through a collegial, interactive process rather than rejecting it with a withering, anonymous comment.

The great benefit of the more formalized system we have today is that it is more impartial, and has helped turn science into less of an old boys' network. But it is also clear that something has been lost.

The problem with reducing bias by formalizing the review process is that it pushes the bias to other parts of the publication pipeline where it is less observable and harder to mitigate.

Better connected authors can:
–Avail of informal, constructive peer review *before* formal review
–Better jump through the hoops required by the ritualized (sometimes years-long) review process
–Better publicize the paper while under review, increasing its impact once published.

I don't think we should go back to the 1860s model. But we should make peer review more constructive. We could also build in an explicit discussion of many types of reviewer biases as part of the review process itself instead of assuming that anonymity solves the problem.

x.com · 2021-06-20

Selfish reasons to do peer review

Peer review is often considered a burden that one grudgingly accepts in order to keep the system working. But in my experience, especially for a junior researcher, the effort is well worth the time.

The most obvious benefit is that it forces you to read papers deeply, spending hours at a time on a single one. If you're the type that never needs external motivation to get things accomplished, this might not matter to you.

But there are also insights you get that you can't reproduce by having perfect self-discipline. Peer reviewing gives you a raw, unfiltered look into the research that people have chosen to work on. This is a 6-month-or-so head start for getting on top of emerging trends compared to only reading published papers. You also get a better idea of common pitfalls to avoid.

Finally, peer review is one of the rare opportunities to read papers critically (it is harder with published work because it doesn't have as many loopholes). This is not a natural skill for most people — our cognitive biases predispose us to confuse good rhetoric with sound logic.

Selfish Reasons to do Peer Review, and Other Observations · 2012-05-02

How to constructively review a research paper

Any piece of research can be evaluated on three axes:

- Correctness/validity -- are the claims justified by evidence?
- Impact/significance -- how will the findings affect the research field (and the world)?
- Novelty/originality -- how big a leap are the ideas, especially the methods, compared to what was already known?

There are additional considerations such as the clarity of the presentation and appropriate citations of prior work, but in this post I'll focus on the three primary criteria above. How should reviewers weigh these three components relative to each other? There's no single right answer, but I'll lay out some suggestions.

First, note that the three criteria differ greatly in terms of reviewers' ability to judge them:

- Correctness can be evaluated at review time, at least in principle.

- Impact can at best be predicted at review time. In retrospect (say, 10 years after publication), informed peers will probably agree with each other about a paper's impact.
- Novelty, in contrast to the other two criteria, seems to be a fundamentally subjective notion.

We can all agree that incorrect papers should not be accepted. Peer review would lose its meaning without that requirement. In practice, there are complications ranging from the difficulty of verifying mathematical proofs to the statistical nature of research claims; the latter has led to replication crises in many fields. But as a principle, it's clear that reviewers shouldn't compromise on correctness.

Should reviewers even care about impact or novelty?

It's less obvious why peer review should uphold standards of (predicted) impact or (perceived) novelty. If papers weren't filtered for impact, presumably it would burden readers by making it harder to figure out which papers to pay attention to. So peer reviewers perform a service to readers by rejecting low-impact papers, but this type of gatekeeping does collateral damage: many world-changing discoveries were initially rejected as insignificant.

The argument for novelty of ideas and methods as a review criterion is different: we want to encourage papers that make contributions beyond their immediate findings, that is, papers that introduce methods that will allow other researchers to make new discoveries in the future.

In practice, novelty is often a euphemism for cleverness, which is a perversion of the intent. Readers aren't served by needlessly clever papers. Who cares about cleverness? People who are evaluating researchers: hiring and promotion committees. Thus, publishing in a venue that emphasizes novelty becomes a badge of merit for researchers to highlight in their CVs. In turn, forums that publish such papers are seen as prestigious.

Because of this self-serving aspect, today's peer review over-emphasizes novelty. Sure, we need occasional breakthroughs, but mostly science progresses in a careful, methodical way, and papers that do this important work are undervalued. In many fields of study, publishing is at risk of devolving into a contest where academics impress each other with their cleverness.

There is at least one prominent journal, PLoS One, whose peer reviewers are tasked with checking only correctness, with impact and novelty being left to be sorted out post-publication. But for most journals and peer-reviewed conferences, the limited number of publication slots means that there will inevitably be gatekeeping based on impact and/or novelty.

Suggestions for reviewers

Given this reality, here are four suggestions for reviewers. This list is far from comprehensive, and narrowly focused on the question of weighing the three criteria.

1. Be explicit about how you rate the paper on correctness, impact, and novelty (and any other factors such as clarity of the writing). Ideally, review forms should insist on separate ratings for the criteria. This makes your review much more actionable for the authors: should they address flaws in the work, try harder to convince the world of its importance, or abandon it entirely?

2. Learn to recognize your own biases in assessing impact and novelty, and accept that these assessments might be wrong or subjective. Be open to a discussion with other reviewers that might change your mind.

3. Not every paper needs to maximize all three criteria. Consider accepting papers with important results even if they aren't highly novel, and conversely, papers that are judged to be innovative even if the potential impact isn't immediately clear. But don't reward cleverness for the sake of cleverness; that's not what novelty is supposed to be about.

4. Above all, be supportive of authors. If you rated a paper low on impact or novelty, do your best to explain why.

Conclusion

Over the last 150 years, peer review has evolved to be more and more of a competition. There are some advantages to this model, but it makes it easy for reviewers to lose touch with the purpose of peer review and basic norms of civility. Once in a while, we need to ask ourselves critical questions about what we're doing and how best to do it. I hope this post was useful for such a reflection.

Thanks to Ed Felten and Marshini Chetty for feedback on a draft.

The cleverness tax in academia

Academia forces you to pay a "cleverness tax" if you want to succeed—it's a tax on your time that goes towards constantly convincing others that your work is clever enough for publication, getting a PhD, tenure, and promotion. It's one of the things that pushes people out.

Reviewer 3: I see you've solved global hunger, but it was always obvious that you could do that by working really hard, so we haven't learned anything from your paper. Perhaps you could try solving global hunger using only purple foods? That would be novel.

The cleverness tax is higher for scholars whose work doesn't fit their discipline's stereotyped notions of what clever work is supposed to look like. You're often forced to pick between having a real impact on the world and just staying in the game.

I don't have a satisfying resolution to the dilemma, but it's helpful that communities and institutions differ a lot in terms of how much they reward meaningful work versus clever work. You'll be happier doing more meaningful work even if it's at a less prestigious institution.

x.com · 2020-07-27

Overproduction is the #1 problem in research

AI being able to produce papers that aren't easily distinguishable from human-authored ones is potentially an extremely positive development. Even before AI, in every field I'm familiar with, the amount of published stuff exceeds the community's collective bandwidth to absorb and build upon ideas by a factor of 100x or more. Inevitably, the vast majority of what's published makes zero impact. Yet we pretend that publication itself has some value. It doesn't.

Overproduction is the #1 problem in research. Dropping the cost of production to zero will force us to stop attaching value to publication, and look for ways to identify actual intellectual value.

x.com · 2025-02-05

Meritocracy theater

I often see this pattern: candidates are asked to go through an elaborate and formal process that gives the appearance of meritocracy, but is so complex and laborious that it's impossible to navigate without inside connections. That's not meritocracy, it's meritocracy theater.

It's not just in hiring that meritocracy theater operates: academic peer review is often like this and essentially denies the ability to publish to those outside the academy or established research institutions, or even to those just trying to switch disciplines.

In many organizations, the process is nonexistent (the people in power make decisions on an ad-hoc basis) or not publicly documented, which are also recipes for bias. But meritocracy theater is arguably worse, because in the first two scenarios the bias is much more obvious.

It's not that any institution sets out with a devious plan to design a fake meritocracy. More likely, the process gradually got more complex over time. But it does usually involve insiders' willful failure to recognize that the system favors people who are just like them.

x.com · 2022-08-08

7. WRITING AND SPEAKING

Most academic writing is optimized to be bad

In most fields—say political theory or quantum physics—scholars write for their peers because the work isn't directly useful to the public.

But in some fields the public is eager to hear what's new. Yet academic norms ensure that scholarship is impenetrable. That's a damn shame.

Most academic writing is *optimized* to be bad — good writing is all about making ideas as simple as possible but career incentives push authors to make their ideas seem more profound and complicated than they are.

On the occasions when I've managed to make my work broadly accessible, it led to more people *within* the community reading and citing it. Turns out most jargon doesn't help us better communicate with our peers. We do it simply because we're socialized by academic convention.

x.com · 2022-05-13 x.com · 2025-06-06

The limiting step in research might be communication

Once upon a time, watching my toddler who had a rich imagination and a fierce desire to communicate but a limited vocabulary, I had an epiphany. It made me realize that the same is true in research — we are limited not by the ideas we can think of but by what we can communicate.

If we think of communication and comprehension, rather than discovery, as the limiting step in research, it helps explain many puzzling phenomena, such as great ideas being initially ignored before being rediscovered or resurrected.

x.com · 2020-12-28

Writing is like singing — or any other skill

Writing is an essential skill for scholars. It took me years to realize that one doesn't get better simply by writing more, just as you can't become a singer by singing in the bathroom. You have to take time to learn the principles and practice without the pressure of deadlines.

There are many excellent books and guides on writing but my favorite is The Sense of Style. The best thing about it is that the author takes his own advice, and it's a pleasure to read.

x.com · 2019-12-22

Write the way you talk

In 2015 I coauthored a textbook. It was successful, and readers especially liked the writing style. Our secret? We didn't mean to write a book! We made lecture videos, and later decided to edit the transcripts into a book. The lesson: write the way you talk. It reads much better.

x.com · 2020-12-19 x.com · 2015-09-02

Edit from the end to the beginning

15 years ago my PhD advisor taught me One Weird Trick for editing your own writing. Edit **back to front** — start with the last paragraph, then the penultimate, and end with the first paragraph. I still use it and it still surprises me how well it works. When I get my students to do it, it often blows their minds. Try it!

Most writing is way too complicated because we imagine our readers as machines who progress linearly through it, maintaining a perfect memory and understanding of everything they've read so far. Back-to-front editing helps us see how jarring the text is to a human reader.

And let's face it—reading your own writing is terminally boring (when it's not nails-on-chalkboard painful). So when we edit front to back, our brains process relatively little of the text. Back-to-front editing forces the brain to work more, so we notice more. And it's more fun!

The best way to edit your writing is to have someone else do it. The second best way is to put it away for a few weeks before editing it, so the text isn't fresh in your mind.

Back-to-front editing is always worthwhile, but especially when the first two options aren't available.

x.com · 2022-10-16

No one outside academia cares about novelty

Francois Chollet writes:

> People with expert knowledge are often reticent to talk/blog about it, either because they think it's obvious stuff (but for most people it isn't!) or because they think it's boring (but many people would love to hear about it!). Share your knowledge!

I've been posting online for over a decade yet I have to remind myself of this every day and I still can't bring myself to do it half the time. In academia there's so much pressure to only say new things that it's easy to forget no one else cares about novelty.

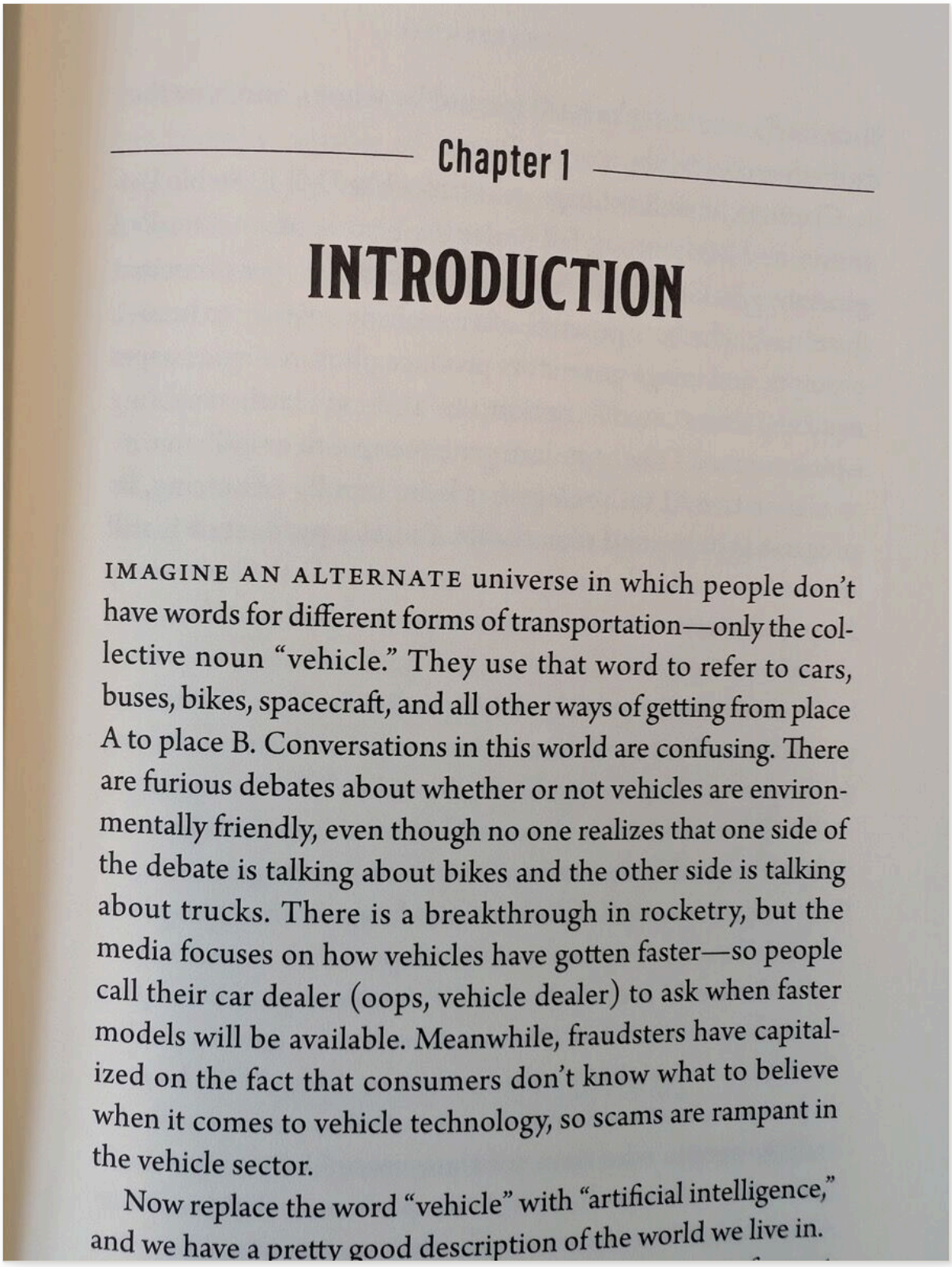
Often the most obvious things are the most important to say out loud. Our inability to say them is one reason why academia so often fails to serve the public.

x.com · 2022-06-24

Publishing is a shot in the dark

Many people have told us that this is one of the most memorable opening paragraphs they've come across in a book. 🙏 The funny thing is that the analogy wasn't a hit with reviewers, so we were going to take it out, but had a last minute change of heart because our

editor liked it. I've learned that it's very hard to predict which parts will resonate with readers, whether in a book, online essay, or any other format. Publishing is always a nerve-wracking shot in the dark.



Chapter 1

INTRODUCTION

IMAGINE AN ALTERNATE universe in which people don't have words for different forms of transportation—only the collective noun "vehicle." They use that word to refer to cars, buses, bikes, spacecraft, and all other ways of getting from place A to place B. Conversations in this world are confusing. There are furious debates about whether or not vehicles are environmentally friendly, even though no one realizes that one side of the debate is talking about bikes and the other side is talking about trucks. There is a breakthrough in rocketry, but the media focuses on how vehicles have gotten faster—so people call their car dealer (oops, vehicle dealer) to ask when faster models will be available. Meanwhile, fraudsters have capitalized on the fact that consumers don't know what to believe when it comes to vehicle technology, so scams are rampant in the vehicle sector.

Now replace the word "vehicle" with "artificial intelligence," and we have a pretty good description of the world we live in.

x.com · 2025-07-30

What it takes to finish a book project

Writing a book is psychologically so different from writing a paper or any other shorter-form piece. For a prolonged period you have to wake up every day truly believing that this is the most important thing you could be working on. Otherwise the project will keep getting sidelined in favor of others with more immediate deadlines, and never get done.

x.com · 2024-11-26

How to prepare a technical talk

I used to suck at giving technical talks. I would usually confuse my audience, and often confuse myself. By the time I became a professor, I sucked a lot less. So something had changed; I'd developed a process. It turned out to be surprisingly formulaic, like an algorithm.

There are three steps. They're simple but not easy.

1. Identify the atomic concepts
2. Identify the dependencies
3. Sort the contents accordingly

Identify atomic concepts

The key word here is atomic. The idea is to introduce only one key concept at one time and give the audience time to internalize the concept before moving on to the next one.

This is hard for two reasons. First, concepts that seem atomic to an expert are often an amalgam of different concepts. Second, it's audience-specific. You have to have a good mental model of which concepts are already familiar to your audience.

Identify the dependencies

This is a tricky step because it's easy to miss dependencies. When the topic I'm teaching is the design of a technical system, I ask myself questions like, "what could go wrong in this component?" and "why wasn't this alternative design used?" This helps me flesh out the internal logic of the system in the form of a graph.

Sort the contents

Finally, we want to order the concepts so that each concept only depends on the ones already introduced. Sometimes this is straightforward, but sometimes there are interlinked concepts.

Among the technical topics I've taught, cryptocurrency seems especially difficult in this regard. Each concept is bootstrapped off of the others, but somehow the system magically works when you put everything together. What I do in these cases is introduce intermediate steps that don't exist in the actual design I'm teaching, and remove them later.

Think of a technical topic as a skyscraper. When it's presented in a paper, it's analogous to unveiling a finished building. The audience can admire it and check that it's stable/correct (say, by verifying theorems or other technical arguments.) But just as staring at a building doesn't help you learn how to build one, the presentation in a typical paper is all but useless for pedagogical purposes. Having dependencies between concepts is perfectly acceptable in papers, because papers are not meant to be read in a single pass.

The instructor's role, then, is to reverse engineer how the final concept might plausibly be built up step by step. This is analogous to showing the scaffolding of the building and explaining each step in its construction. Talks and lectures, unlike papers, must necessarily have this linear form because the audience can't keep state in their heads.

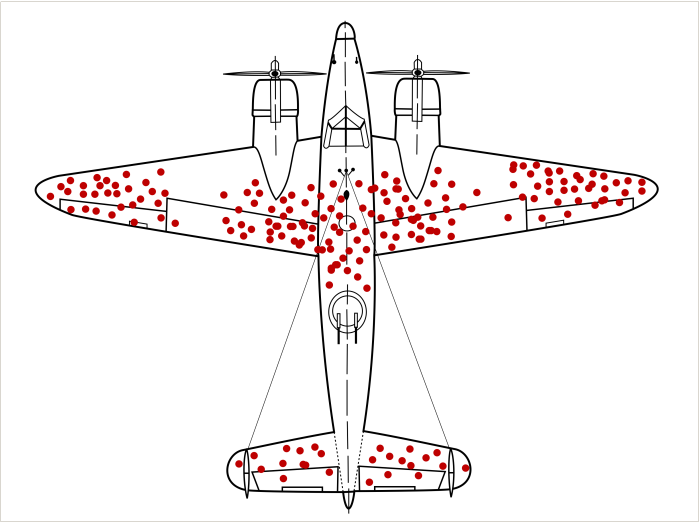
How to prepare a technical talk · 2013-11-26 x.com · 2013-10-09

8. BECOMING AND BEING A PROFESSOR

The role of luck in academia

Professors at top universities are lottery winners, but rarely acknowledge the role of luck in their success. Be skeptical when they give you advice suggesting that the path they took is a repeatable one. If you aspire to an academic research career, have a backup plan.

Deciding to go into academia because all your professors said it worked out pretty well for them is the statistical fallacy of survivor bias.



When I review pre-doctoral, doctoral, post-doctoral, or faculty applications I'm amazed by how much more these candidates have accomplished than I had at the corresponding stage of my career, and how many more qualified candidates there are compared to available positions.

This post is about one of the many improbable things that happened to me, without which I wouldn't be where I am.

x.com · 2021-01-11

The job talk is a performance

Going on the academic job market can be nerve-wracking. After years of hard work, your ability as a researcher will be judged primarily based on a 1-hour talk. Of course, you'll (hopefully) give your talk at many places, but it's going to be the same talk!

The reason I say this is not to stress you out even more. Rather, if at any point the level of preparedness that I suggest seems excessive or disproportionate, remember the stakes.

Public speaking is a performance

I enjoyed the book Confessions of a Public Speaker. There are many useful little details in the book, but it has one central idea that can be boiled down to the phrase "public speaking is a performance." Job talks are even more of a performance than public speaking in general, since the audience is specifically there to judge you. (I should add that teaching is very different from public speaking and is emphatically not a performance.)

This is a generative metaphor -- it allows you to predict things about your job talk based on what you know about performing. Fully appreciating the metaphor will require reading the book, but here are two such predictions that might otherwise be surprising.

Your first priority is to entertain

Certainly you must both entertain and inform, but the point is that you don't really have a shot at the latter if you fail at the former. Sitting in a lecture, as everyone who remembers their student days is surely aware, can be excruciating. Anything you can do to break the routine, linear, hyper-boring pattern of a lecture will help jolt the audience out of their stupor. Otherwise they won't be excited about you nor remember much after the talk.

Rehearse, rehearse, rehearse

You may have heard "practice, practice, practice." I'd rather cast it in the language of performance, as there are some subtle differences. For example, when people tell you to practice, they tell you not to overdo it because you'd lose your spontaneity. I disagree. In a rehearsal, everything is practiced down to the last detail.

Rehearsal should include recording yourself giving the talk and watching it. Yes, it's painful, but essential. In addition to all the obvious facets of good presentation style that I won't repeat, one of the subtle but important things you should watch for is nervous tics or other repetitive behaviors -- almost everyone has one or more of those, and they can almost derail your talk by distracting your audience.

The reason rehearsal makes such a huge difference is that when you're delivering a rehearsed talk, your every word and gesture is subconscious, freeing up your mental bandwidth for observing and reacting in real-time to the facial expressions of your audience. There are usually less than 40-50 people at these talks, a small enough number that you can notice if someone looks confused, skeptical, or bored. But this won't be possible if you have to think through your slides instead.

The job talk is a performance · 2013-02-09

Professors get no training for the job they actually do

Academia is a perfect realization of the Peter Principle: "employees are promoted based on their success in previous jobs until they reach a level at which they are no longer competent, as skills in one job do not necessarily translate to another."

A tenure track professor's main roles are mentoring, management, fundraising, teaching, and administrative work — not research — especially in STEM fields where PhD students do the actual research.

"Congrats on your excellent PhD research. You're promoted! Oh, we know you'll suck at it—that's why you have the job for life!"

It's unfortunate that professors in many fields face pressures that lead them to stop doing the actual research themselves. I'm just observing the way things are, not advocating for this model — quite the opposite.

Since professors get no formal training in the skills they actually need, they rely on informal networks. So the system favors those with access to these connections. One of the best things that academia can do for diversity and inclusion is to prepare people for these roles and remove the reliance on the old boys' network.

x.com · 2022-04-12

Thoughts on tenure and productivity

My academic trajectory might look like smooth sailing from the outside, but I experienced it as bumbling and meandering. Much of the time I felt like I didn't know what I was doing. So here are some things about tenure I wish I'd known when I started.

1. A six-year tenure clock isn't that long! It takes a year or two to settle in and become productive in a new research area, especially if you need to build a team. And your tenure application will likely be due at the end of year 5. So you really have something like 3-4 years.

If you feel like you haven't done much at the end of year 1, don't be discouraged. My first high-impact work wasn't until the end of year 2. What's crucial early on is figuring out your research direction(s). If you need to pivot, pivot early — it's much harder to do it midway.

2. Tenure advice doesn't generalize well (including, of course, this post). You'll often find yourself violating other people's heuristics for how to be productive as a professor. Listen to your mentors, but have the confidence to do your own thing if you need to.

For example, at first I followed the usual advice to travel regularly to present my work and to network. Then I realized that my online presence gave me most of those benefits anyway, and I drastically cut back on travel. That's the single best decision I made.

3. Don't become a manager! I thought the secret to productivity was to hire a bunch of grad students and spend all my time advising them. In the short run, this will probably lead to a big increase in output. But the downside is that your skills will rust over time.

Besides, it's far less enjoyable than doing a mix of advising and your own research, in which you're coming up with the ideas or writing the code or whatever. In recent years I've tried to maintain a healthy balance, but it takes conscious effort.

4. It's stressful even if you're doing well. In comparing notes with other professors who were up for tenure, I thought their tenure cases were slam dunks, and I was surprised to learn that they were stressed about it. Turns out they thought the same about me.

The good news is that while an academic career is stressful, it's no more so than being a doctor or lawyer or any other profession. A major cause of stress is randomness. There's a lot of it in every walk of life. We just have to learn to cope with it.

A final thought: I was a reluctant academic. When I decided to take the plunge, I promised myself I'd opt out of the aspects I disliked (e.g. secrecy, publish or perish). I'm fortunate it's worked out. Nonconformism may be riskier, but far more rewarding and worth considering!

x.com · 2018-07-01

Two opportunities for higher-ed reform

I was asked if the crisis created by AI is also an opportunity for fundamental changes to higher ed. Yes! I've been thinking and writing about this since before ChatGPT's release.

I see two big opportunities. The first is to separate essential skills from incidental ones. Here's an analogy. When calculators became available, we didn't stop teaching arithmetic in elementary school. That's an essential skill. But in college calculus courses, we don't make students do tedious calculations by hand; we let them use calculators. In that context, arithmetic is incidental.

What every instructor needs to do — should have been doing for the last few years, really — is to painstakingly go through everything they teach and figure out which skills/knowledge are in the curriculum because they are essential, and which ones are incidental and can be relegated to AI and automation. Lots of student time can be freed up to focus on what's truly

important. A trivial example: students still need to learn to write, but formatting citations correctly is a pointless skill, and it's distressing that many instructors continue to waste students' time by making them learn this.

The second opportunity to reform higher ed may be even bigger. In many ways the education system is designed to inhibit, rather than foster, students' intrinsic motivation to learn. Imagine if an alien showed up to earth. They would probably find it absurd that humans are supposed to absorb knowledge passively for the first 20+ years without any practical context, and then find jobs, with no further need to learn or revisit what they've learned.

The disconnect between education and the real world makes learning feel like an artificial activity in an artificial environment. This reduces the intrinsic motivation to learn. As a result, college educators rely heavily on the extrinsic motivation of grades to incentivize learning. We've all been through this and I think it goes without saying that this is just about the worst way to learn. There's a well known finding in psychology that intrinsic and extrinsic motivation are in tension. So the two goals of classes — learning and testing — conflict with each other, and currently the former is almost entirely subordinate to the latter.

How can we bring intrinsic motivation back in universities? AI (along with other information technologies) is gradually decreasing the barrier of expertise needed to actually *do* something and contribute something to the world. I think educators should try to lean into this and reorient university education around producing new things and making an impact, and minimize the role of toy problems and simulated environments.

In my dream scenario, learning could be entirely decoupled from the assessment, certification, and credentialing function of universities. The system has too much inertia for this to happen, but even a small movement in this direction would make a big difference!

x.com · 2025-03-27

Things that are more important than how much you know

When I was a student I thought professors are people who know lots of stuff. Then they went and made me a professor. After getting over my terror of not knowing stuff, I realized I had it all wrong. Here are a bunch of things that are far more important than how much you know.

- Knowing what you know and what you don't know.
- Being good at teaching what you know.

- Being comfortable with saying you don't know.
- Admitting when you realize you got something wrong.
- Effectively communicating uncertainty when necessary.
- Spotting BS.
- Recognizing others with expertise.
- Recognizing that there are different domains of expertise.
- Recognizing that there are different kinds of expertise including lived experience.
- Drawing from others' expertise without deferring to authority.
- Being pleasantly surprised, instead of unpleasantly surprised, when we encounter information that challenges our beliefs.
- Accepting the unknowable, because there's so much we simply can't know in our present state of understanding.

x.com · 2020-11-25

Research and teaching are symbiotic

When I started as a professor I resisted the idea that research and teaching are symbiotic. But over the years I've gradually rediscovered why this is true. One way to think about it is that research adds height to the tower of knowledge while teaching anchors the foundations.

When I teach, concepts solidify in my mind and the increase in fluency takes me farther in my research. Conversely, staying at the forefront of knowledge through research helps me figure out how to tweak my teaching so that it remains relevant to my students for the long term.

I'm not saying that all good researchers are good teachers. We've all had professors who were famous for their research but sucked in the classroom. But I do think they'd be even better researchers if they took time to improve their teaching, which makes it doubly unfortunate.

x.com · 2020-08-23

9. BREADTH, SWITCHING FIELDS, INTERDISCIPLINARITY

Breaking out of hyper-specialization

Expertise is important for scholars, but after 5-10 years the benefits of continuing to deepen your expertise are tiny compared to broadening it.

Universities are perfectly set up to prevent breadth of expertise by hiring people for life and putting them into siloed departments.

An aphorism someone shared: "The goal of the PhD is to know more and more about less and less, until you reach the state of perfection where you know everything about nothing".

Since academia rewards hyper-specialization, scholars narrow their interests until only a handful of people in the world can understand them.

That means you can make ground-breaking discoveries and contributions if you have even an undergrad-level fluency in two or more fields.

There is nothing wrong with specializing in one discipline, and it is certainly possible to make deep and meaningful contributions that way. But there is so much low-hanging fruit, so much untapped potential for world-changing findings, at the boundaries between disciplines.

The asteroid theory of dinosaur extinction was developed by a physicist-geologist pair, Luis & Walter Alvarez. They were father and son. Academia is so bad at enabling interesting dialog across fields that they are more likely to happen randomly, such as at a family dinner table!

The intellectual superiority of depth over breadth is a pervasive fiction in academia that sustains the culture of fetishizing specialization. I tried to fight this culture early in my career, but realized it was like punching a bag of sand.

Rather than fight the culture, I found a different strategy. Try to do interdisciplinary work that is so impactful that others will have no choice but to acknowledge its value, even if they continue to mumble that it lacks intellectual depth.

x.com · 2019-08-01

Switching fields becomes more daunting as you get more senior

As you grow more senior and more prominent as a scholar, it gets more daunting to switch your area of expertise, because the inevitable early mistakes in a new area feel more embarrassing. Yet, the alternative — staying in your comfort zone throughout your career — is a sure path to irrelevance.

When I learned about the great Indian physicist Subrahmanyan Chandrasekhar, the fact that he changed topics every few years amazed me. I realized that if such a renowned scholar could repeatedly make himself a beginner in a new area, I sure could do that as a relative nobody.

> Chandrasekhar developed a unique style of mastering several fields of physics and astrophysics; consequently, his working life can be divided into distinct periods. He would exhaustively study a specific area, publish several papers in it and then write a book summarizing the major concepts in the field. He would then move on to another field for the next decade and repeat the pattern. Thus he studied stellar structure, including the theory of white dwarfs, during the years 1929 to 1939, and subsequently focused on stellar dynamics, theory of Brownian motion from 1939 to 1943. Next, he concentrated on the theory of radiative transfer and the quantum theory of the negative ion of hydrogen from 1943 to 1950. This was followed by sustained work on turbulence and hydrodynamic and hydromagnetic stability from 1950 to 1961. In the 1960s, he studied both the equilibrium and the stability of ellipsoidal figures of equilibrium, and general relativity. During the period, 1971 to 1983 he studied the mathematical theory of black holes, and, finally, during the late 80s, he worked on the theory of colliding gravitational waves.

But the luxury of repeatedly switching fields is only available in a supportive institution without short-term publishing pressures and the cushion of tenure to sustain someone through a several-year period of decreased productivity as they strive to master a new area.

To me, the best part about switching fields was the discovery that skills I learned in one area were applicable — and sorely needed — in another, despite the subject matter being unrelated.

It turns out that the portability of research skills is actually a common experience, even between pairs of disciplines that are very far apart. For instance, Duncan Watts has spoken about how his training in physics helped him succeed as a sociologist.

x.com · 2022-06-28 x.com · 2022-06-15

A recipe for interdisciplinary research: Culture, then Language, then Substance

One surprise for me from interdisciplinary collaboration is that the hard part isn't so much the substance as simply being able to communicate and be understood. That, in turn, requires some awareness of the other field's culture, such as what kind of scholarship they value. What's valuable to one person may seem irrelevant to another. So the cultural aspects have to be navigated first. Then you can start establishing a common language and vocabulary and finally get to the substance of the collaboration.

x.com · 2012-06-18 x.com · 2022-07-12 x.com · 2020-09-24

Part 2: Work, learning, and AI

10. LEARNING AND PROFESSIONAL GROWTH

My message to Princeton's incoming undergraduate class

In August 2026 I had the honor of speaking to Princeton's entire incoming undergraduate class to address their AI anxieties. I had three messages for them — good news, bad news, and a note of optimism. Here's a condensed version.

The good news

We have enough evidence now to conclude that the shrill predictions of rapid, massive job loss were misplaced. Even in a field like software engineering where AI has been rapidly adopted, its effect has been to shift, not replace the role of the human (see the decide-execute-deliver framework).

Similarly, the panic about what to major in is also misplaced. There will be enduring demand for computer science, philosophy, and just about everything else. (In fact, AI companies hiring philosophers has been a recent trend, which illustrates that humanistic training retains and perhaps even rises in value in the age of AI.)

The bad news

AI seems to help senior people much more than juniors. I can use AI for coding because I spent 25 years learning how to code, which lets me supervise coding agents effectively. (See the entry on the "growth cycle" vs the "dependence spiral").

You are in a bind — you can't offload your skill-building to AI, but you'll graduate into a market where employers will expect you to get work done with AI. We never faced this dilemma. As a result we haven't figured out how to revamp our classes to help you do both. You'll have to help us figure it out. And you'll need to somehow resist the constant temptation to turn to the shortcut machine.

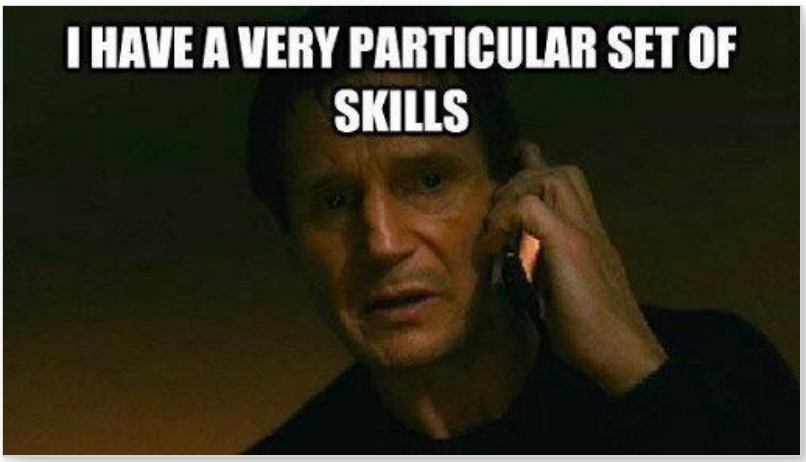
The hope

My point is not that AI is bad for learning. It's an incredibly flexible tool. Is the internet good or bad for learning? Depends — are you using it to find research papers or waste time scrolling? I use AI every day for learning. The key is to use it to increase, not decrease your cognitive load. To learn deeper, not faster. There is no learning without the cognitive sweat. I try to make sure I'm mentally exhausted at the end of the day. I do feel that AI lets me push myself harder than I ever could before, and I have a vision that as AI continues to advance it will enable human-AI "co-superintelligence".

x.com · 2026-08-27 x.com · 2026-08-27

The Liam Neeson principle

Early in my career, I thought the path to success was to be super smart. Pretty soon I realized that the majority of my peers were smarter than me, so that plan went out the window. What saved me was a realization that I'll call the Liam Neeson principle.



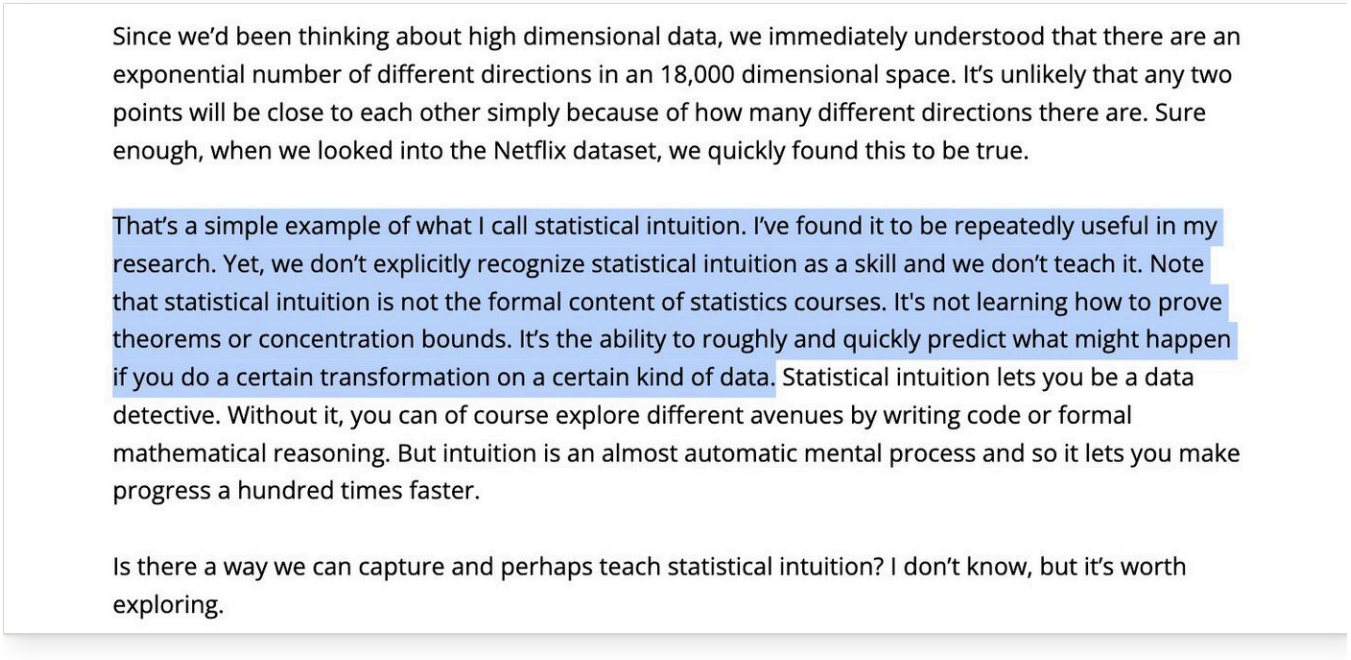
I realized I had a particular combination of skills that most others didn't, even if I wasn't the best at any of those skills. It meant I could solve certain kinds of problems as well as or better than anyone. Once I found my precise niche, I was able to thrive in it.

Once I hit upon this strategy, I gradually started to add to my skillset so that I could expand the niche in which I could be successful. What I like about this approach to career success is that it can work for everyone — it's not a competition.

To be clear, this is not false modesty. I'm smart, but not smarter than my peers. I remember the moment I realized it: in 2006, when I was at Microsoft Research in Mountain View, CA, a lab filled with giants of computer science (which MS later atrociously shut down).

There are a couple of popular books that talk about related ideas. "Range: Why Generalists Triumph in a Specialized World" by David Epstein and Superforecasting by Phil Tetlock.

Figuring out my strengths was surprisingly hard, especially non-standard skills that might not even have a name! One strategy is to reflect on your successes. Here's one example (from a talk in which I discussed my research and drew some lessons).



Since we'd been thinking about high dimensional data, we immediately understood that there are an exponential number of different directions in an 18,000 dimensional space. It's unlikely that any two points will be close to each other simply because of how many different directions there are. Sure enough, when we looked into the Netflix dataset, we quickly found this to be true.

That's a simple example of what I call statistical intuition. I've found it to be repeatedly useful in my research. Yet, we don't explicitly recognize statistical intuition as a skill and we don't teach it. Note that statistical intuition is not the formal content of statistics courses. It's not learning how to prove theorems or concentration bounds. It's the ability to roughly and quickly predict what might happen if you do a certain transformation on a certain kind of data. Statistical intuition lets you be a data detective. Without it, you can of course explore different avenues by writing code or formal mathematical reasoning. But intuition is an almost automatic mental process and so it lets you make progress a hundred times faster.

Is there a way we can capture and perhaps teach statistical intuition? I don't know, but it's worth exploring.

x.com · 2022-10-14

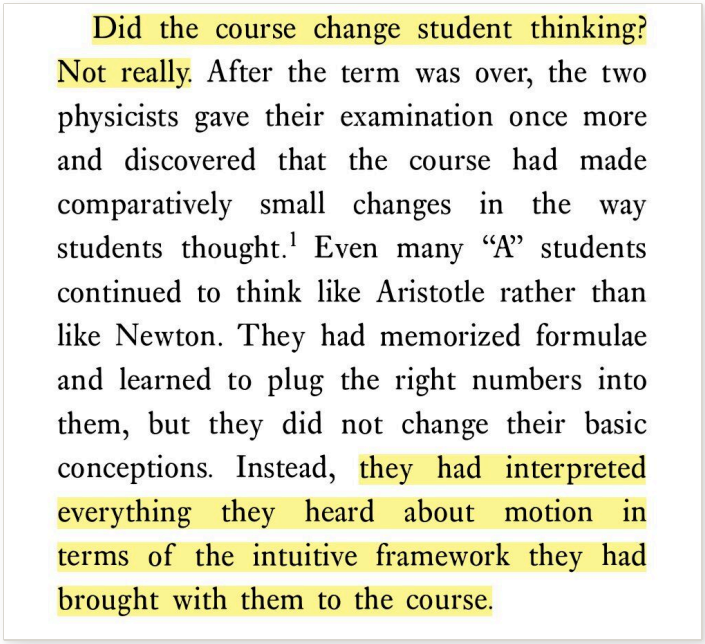
Real learning is supposed to feel uncomfortable

I tell students on the first day of class that if you're truly learning it's supposed to feel uncomfortable. The reason is that real learning is not simply the accumulation of facts; it is deep understanding, building mental models of the world, and other higher-level abilities. The problem is that we already have simple, intuitive, and usually incorrect mental models of most things in the world around us. So real learning usually involves *unlearning*. And that has an extremely high cognitive cost and we're very resistant to doing it, presumably for evolutionary reasons.

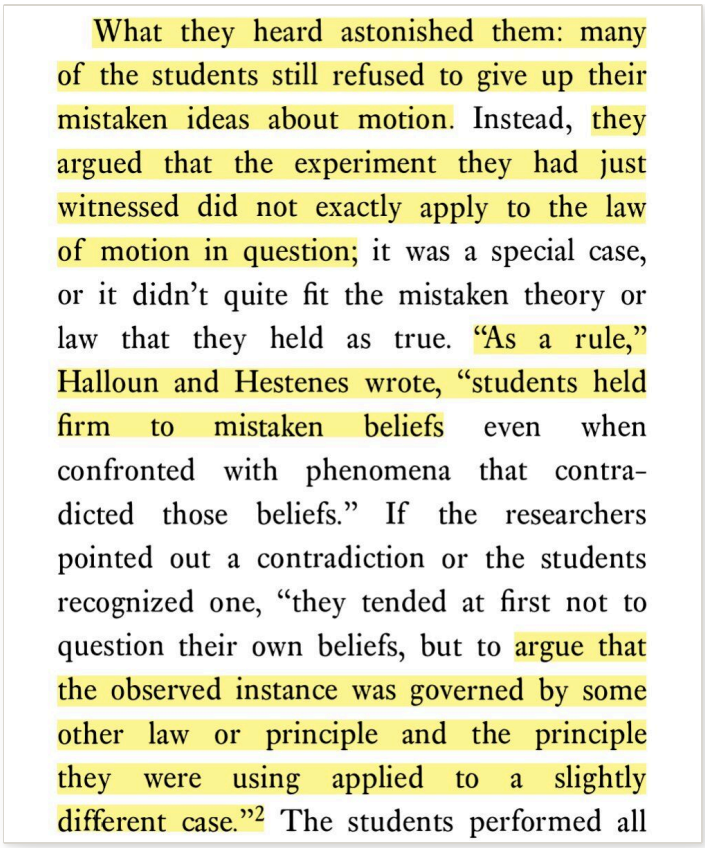
Students go so far as to learn concepts in class but somehow parcel them so that they think the concepts are only applicable to the toy problems on tests but not the world around us! Turns out these mental gymnastics are still easier than actually updating their mental models. (The screenshots are from the book "What the best college teachers do.")

It only gets harder, not easier, to learn as we progress in our careers, because in addition to the cognitive cost of learning, you have to face the prospect of admitting that you were wrong in front of subordinates, if not in public, and admitting to yourself that you've been making suboptimal decisions all along.

I've found that the only way to continue to learn is to develop a kind of masochism where you learn to enjoy, or at least love-hate, the feeling of having been wrong. It's not easy but I think it's necessary!



Did the course change student thinking? Not really. After the term was over, the two physicists gave their examination once more and discovered that the course had made comparatively small changes in the way students thought.[1] Even many "A" students continued to think like Aristotle rather than like Newton. They had memorized formulae and learned to plug the right numbers into them, but they did not change their basic conceptions. Instead, they had interpreted everything they heard about motion in terms of the intuitive framework they had brought with them to the course.



What they heard astonished them: many of the students still refused to give up their mistaken ideas about motion. Instead, they argued that the experiment they had just witnessed did not exactly apply to the law of motion in question; it was a special case, or it didn't quite fit the mistaken theory or law that they held as true. "As a rule," Halloun and Hestenes wrote, "students held firm to mistaken beliefs even when confronted with phenomena that contradicted those beliefs." If the researchers pointed out a contradiction or the students recognized one, "they tended at first not to question their own beliefs, but to argue that the observed instance was governed by some other law or principle and the principle they were using applied to a slightly different case."[2] The students performed all

x.com · 2025-05-02

Systems thinking applies to almost everything

One of the most interesting things I've ever read is the short paper "How Complex Systems Fail". Like many others encountering it, I felt it was the most insightful four pages ever written about software engineering.

Observations like "Complex systems run in degraded mode" and "Root-cause attribution is fundamentally wrong" are both profound yet totally relatable if you've encountered production software. It was an "oh, this explains everything!" moment for me.

Yet I was confused that the paper didn't refer specifically to software engineering, or even engineering. Then I learnt that the paper was written by an MD, and he actually had healthcare systems in mind. Wait, what?!

Apparently complex systems in every domain have striking similarities. And there is a whole field of study of "systems" whose insights apply equally well to ecological systems and medical systems and software and traffic and many, many other things.

Then I read "Thinking in Systems: a Primer", a classic of the field. It's a whole new way of seeing the world and everything in it. I'm sad this wasn't in my high school/college syllabus. If you hadn't heard of systems thinking either, it's not too late!

I've made more highlights in "Thinking in systems" than any other book I've read, in part because it's so widely applicable. Here's an almost throwaway passage brilliantly explaining why the news is not a good way to understand the world.

> Systems fool us by presenting themselves—or we fool ourselves by seeing the world—as a series of events. The daily news tells of elections, battles, political agreements, disasters, stock market booms or busts. Much of our ordinary conversation is about specific happenings at specific times and places. A team wins. A river floods. The Dow Jones Industrial Average hits 10,000. Oil is discovered. A forest is cut. Events are the outputs, moment by moment, from the black box of the system.
>
> Events can be spectacular: crashes, assassinations, great victories, terrible tragedies. They hook our emotions. Although we've seen many thousands of them on our TV screens or the front page of the paper, each one is different enough from the last to keep us fascinated (just as we never lose our fascination with the chaotic twists and turns of the weather). It's endlessly engrossing to take in the world as a series of events, and constantly surprising, because that way of seeing the world has almost no predictive or explanatory value. Like the tip of an iceberg rising above the water, events are the most visible aspect of a larger complex—but not always the most important.

x.com · 2019-09-01 x.com · 2014-12-18

How technical do you need to be?

A question I've gotten for decades from non-technical folks who want to work in tech (say, tech policy) is how technical they should be. I've never had a satisfying answer, falling back on "it depends" — until 2026.

I think most people in this position should be technical enough to use agents to get work done. You can do so even without any technical knowledge, but a little bit of investment in understanding how the tech works will go a long way toward getting more out of them and, more importantly, avoiding the ever-present pitfalls (verification, overreliance, skill erosion). It is also a great way to get an intuition for where the frontier of AI capabilities lies.

Acquiring this just-enough-to-be-useful level of technical knowledge has gotten easier than ever because agents themselves are good teaching tools. Of course, you still have to put in the work to learn, and maintain a critical mindset. But the "where do I even start?" problem

has gone away, you don't need a human tutor, and don't have to worry about hitting some technical snag (setting up a compiler or whatever) that's going to block your progress.

Usually, people asking how technical they should be want to know if they should learn to code. My answer used to be "probably", but not anymore. The alpha of learning to code has gone down a lot. It used to be a good way to understand what computers can and can't do, but AI has shifted that boundary. And non-technical folks who learn to code typically aren't planning to write production software but rather do simple things like web scraping — tasks that can be delegated to agents now.

Finally, there is enduring value in learning basic computer science concepts, and AI is unlikely to undermine that.

x.com · 2026-08-20

The growth cycle versus the dependence spiral

There's a big, under-appreciated reason why people may have very different experiences and opinions about using AI for work — are they using it for tasks they're already an expert at, or tasks they can't do themselves? The former leads to a *growth cycle* and the latter leads to a *dependence spiral*.

When I use AI to do something I'm an expert at, like coding, I treat it as a tool. I can build quickly, maintaining an understanding of the code, knowing that if necessary, I can fix the code myself. It feels empowering. It frees up my time to think about the complex, judgment-oriented parts of software engineering that I can't or won't delegate to AI. That means my own skills improve rapidly, and I get to climb the ladder of complexity and develop higher-level skills, much more so than when I write the code myself. I feel in control. I can lock in and achieve a flow state — when AI is working, I'm reviewing, building understanding, and planning the next steps. I never get the feeling that the tool is about to replace me. This is the growth cycle.

(Of course, the growth cycle is not automatic. I still need to exercise agency to use AI responsibly. But it's the same challenge with any productivity-enhancing technology, and those who've navigated such transitions before are well-equipped to navigate it with AI as well.)

On the other hand, if I use it for tasks I don't understand and haven't learned to perform myself, I have no choice but to treat it as a superintelligence. If something breaks, the best I can do is ask AI to fix it and hope for the best. I generally can't evaluate the quality of the output myself. The only way to find out if it's any good is if and when the work is ultimately reviewed by an actual expert. The experience is confusing, unsettling and disempowering. And forget about flow state. By over-relying on AI, I risk losing whatever skill I had at the task in the first place, even if it boosts productivity in the short term. This is the dependence spiral.

It's no wonder that entry-level workers and students preparing to enter the workforce find themselves in a bind. To compete with the AI-enabled productivity of more seasoned workers, they must adopt AI themselves, but doing so risks the dependence spiral.

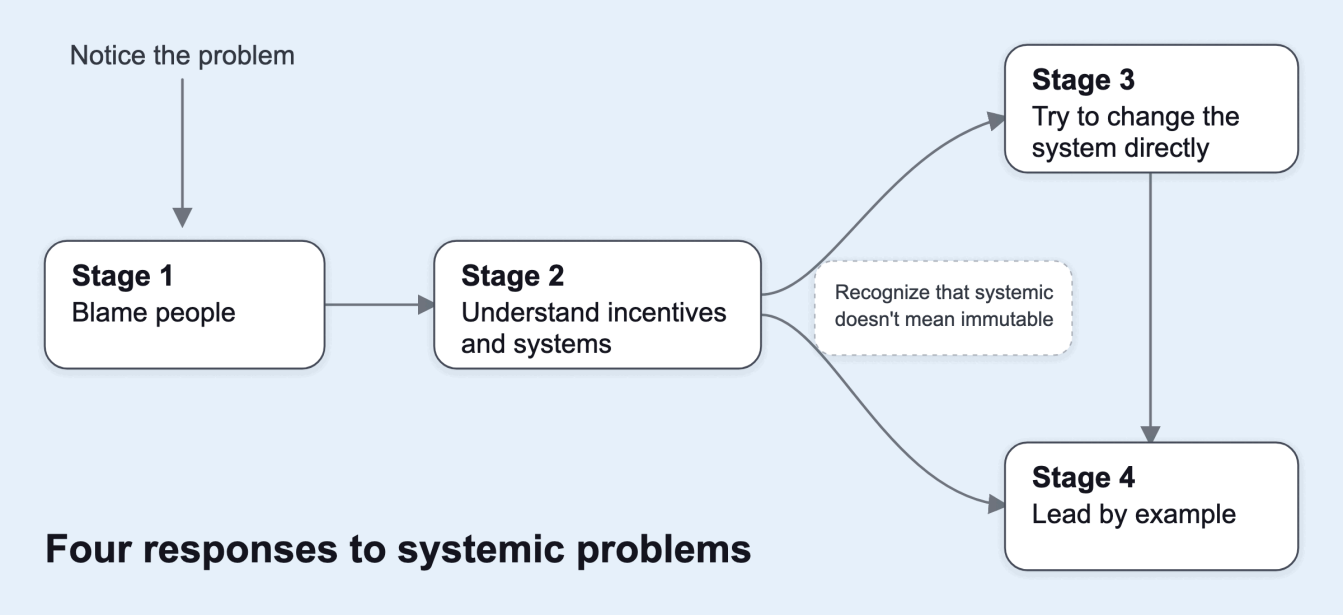
x.com · 2026-06-12

Four responses to systemic problems

There are many problems in our world and most of them are systemic. A pattern I've observed is that people go through 4 stages in how they respond to the problems they notice, gradually shifting their approach as they gain more life or work experience. I think it's worth sharing.

Stage 0 is not noticing the problem. Most people don't notice most problems, and that's okay. Our time and attention are limited. The key question is how we conceptualize the problems we do notice and what we do about them.

Stage 1 is viewing it as a matter of individual incompetence or bad actors. This is the most intuitive, initial reaction that we tend to have. In rare cases this perception is accurate, but most big problems in society are systemic.



Stage 2 is the recognition of the systemic nature. As Steve Jobs put it:

“When you're young, you look at television and think, There's a conspiracy. The networks have conspired to dumb us down. But when you get a little older, you realize that's not true. The networks are in business to give people exactly what they want. That's a far more depressing thought. Conspiracy is optimistic! You can shoot the bastards! We can have a revolution! But the networks are really in business to give people what they want.”

People who notice a problem tend to eventually graduate from Stage 1 to Stage 2, but stop there — with resigned acceptance and a feeling of powerlessness.

In contrast, stages 3 and 4 involve deciding to take some action despite recognizing how hard the problem is. Systemic does not mean immutable. Institutions are maintained by norms, incentives, and coordinated behavior that can sometimes be changed.

Stage 3 is trying to change the system directly through activism or reform. Many junior people in a field want to be reformers. They’re in the sweet spot — when they go from outsider to insider and have a little bit of power and leverage, but they haven’t yet adapted to the system and become vested in its preservation.

Media reform through collective action has happened, though nothing that successfully tackled the “networks are dumbing things down because that’s what audiences want” problem.

Trying to directly change any entrenched system requires dedication for a sustained period, and even then has only a low likelihood of success. Most people decide sooner or later that that path isn't for them.

But that's not the end of the road, because in many cases, change can happen incrementally.

Stage 4 is realizing that systemic doesn't mean universal. Systems exert pressure toward an equilibrium, but there are niches where enterprising individuals or teams can buck the trend. They can benefit — financially or reputationally — from doing things differently. In other words, when systemic failure is severe enough, the unmet demand it generates can become an opportunity. A successful exception may remain an exception, but sometimes it attracts imitators by demonstrating the feasibility of an alternative path.

Coming back to television as our case study, one of the best examples of this is The Wire, a famously high-quality, cerebral and accurate show. It was not what mainstream TV economics ordinarily rewarded, but this niche was so underserved that there was an opportunity for the show to survive for many seasons on HBO. As audiences gradually shifted over time, demand for this kind of show increased in later decades, and DVDs and streaming made it much easier to cater to those audiences, The Wire became an important precedent and gained enormous cultural prestige.

To be clear, Stage 4 is not always better than Stage 3. Sometimes we really need to reform or even tear down the system. But more often, we can contribute to change through leading by example, without having to sacrifice our career to pursue reform — and in fact benefit from the leadership opportunity.

In my own small way, I've tried to put this into practice in my career. For example, early on I noticed that academic writing is often jargon-filled and unreadable. I assumed this was because there's something wrong with the sort of people who become academics (Stage 1). Pretty soon I realized the incentives are messed up — scholars write to impress reviewers and advance their careers, more so than to inform the ultimate reader (Stage 2). Early in my career I made some feeble efforts toward changing peer review (Stage 3), but gave up pretty soon because of the obvious difficulties.

My breakthrough was realizing that I should simply write papers the way I wanted to (Stage 4). Even though this resulted in a peer review penalty, when I did publish papers (or even put preprints online) their influence was amplified because more people read them. Over the years I've heard from many junior scholars that reading my writing helped them realize that

it’s possible to have a career writing plainly. This has been incredibly gratifying to hear, and is a small but personally meaningful change — much more than I would have ever been able to accomplish by tilting my lance at the windmill of peer review.

x.com · 2026-09-22

11. GETTING THINGS DONE

Thinker mode vs. manager mode

Something I've observed in academia but I suspect is true in industry as well: as you become more senior, you'll find that you can be much more productive if you spend all your time being a manager and letting junior people do all the real work. But after a few years of this, you're likely to lose your skill at doing the work yourself, the opportunity to practice that skill, and the gratification of doing so that drew you to your chosen profession in the first place. And you'll probably also get worse at being a manager as you lose touch with the actual work.

Avoiding managerdom is hard. You have to give up substantial short-term productivity gains. If you're in academia in a tenure track, you risk not getting tenure by being less productive. And personally I find that switching between thinker mode (deep work with no distractions for a whole day) and manager mode (a thousand quick but urgent tasks) to be highly unpleasant. When I'm in thinker mode, it kills me that people have to wait a day (often a lot more) for a 1-minute response from me that could unblock them on what they're stuck on and save them hours of work. When I'm in manager mode, it kills me that I have all these creative ideas sloshing around in my brain that I'm not able to execute on.

So it's extremely tempting to perpetually be in just one mode or the other. When you have a team, not managing them is not an option, but being a pure manager is an option. That makes it a constant struggle to carve out time for your own work and not give in to the temptation.

Note: I think this is only tangentially related to Founder Mode vs Manager Mode. It's much more related to another Paul Graham essay, Maker's Schedule vs Manager's Schedule. And Deep Work by Cal Newport is very relevant.

This interesting essay argues that you can't do both simultaneously, and yet need to do each of them to be good at the other, so the solution is to be on a perpetual 2-3 year pendulum where you go from an individual contributor to manager and vice versa.

I don't think you can do it in an academic research career — once you recruit PhD researchers you're committed to mentoring them for 5 or more years — but it seems like a great strategy in industry.

x.com · 2025-01-09 x.com · 2025-01-10

Is cold outreach dead?

AI agents seem to be killing the ability to do cold outreach. Look at it from my perspective — someone who gets dozens of unsolicited invitations/requests in my email from strangers every day. In the past there used to be a sharp distinction between unpersonalized mass emails and those that were meant specifically for me. My spam filter could mostly automate this classification, and even when it didn't, it would take me only a few seconds to tell if an email was worth reading/responding to. Now the distinction is gone because it is trivial to use AI to send out mass emails that appear deeply knowledgeable about each recipient's work and expertise. I guess we have to go back to the bad old days of needing to be "introduced" to someone before communicating with them.

As I write this in September 2026, we're months away from the Fall *2027* PhD admissions cycle and I've gotten about 75 inquiries. I expect this to increase exponentially until December. And this is just one of 10-15 categories of unsolicited email I get. Sadly I'm long past the point of being able to open all mail.

One question I received — what's wrong with needing introductions for outreach? Why am I sad about the death of cold emails? It doesn't affect me much, because I'm well established in my career. But once upon a time I was the one writing cold emails. Growing up in India, I didn't have many connections to rely on for reaching out to prominent researchers. Without cold emails, I wouldn't have ended up where I am today.

Of course this is far from a catastrophe, but there was something magical about the egalitarian potential of the internet, and we lose something when we go back to clubby hierarchies.

In response to some comments — yes, it's possible that a few people including non-native English speakers used AI for understandable reasons. But the massive flood combined with the fact that AI makes it trivial to fake signals of genuine interest means that the cold email channel is effectively dead. It's not like I can discern and adjudicate individual motivations. It's a tragedy of the commons.



x.com · 2026-03-26 x.com · 2026-09-13

What's in it for them?

I've seen hundreds of bits of professional advice. Here's the one that had the biggest impact on me.

When inviting someone to do something, be explicit about what's in it for them.

If there's nothing in it for them, stop. Inviting them is disrespectful and wastes their time.

Of course, the answer to "what's in it for me" doesn't have to be money; it can be a platform for your ideas, or an intellectually interesting project, or a learning opportunity, or many other things.

Whatever the benefit to the invitee is, make it explicit and up front. Don't assume that it will be obvious — it rarely is.

Sometimes, talking about why your invitation is a cool opportunity might feel like tooting your own horn. Well, there's nothing wrong with that!

I wish this practice were more widely followed. When I look at the invites that land in my inbox, less than 10% of them are explicit about how I might benefit if I said yes. Look at your own inbox — I suspect you'll be surprised.

I've learned the hard way that many emails superficially look like they're opportunities but are in fact requests for unpaid labor.

If you're invited to do something and it isn't clear what's in it for you, remember that saying 'no' is always an option. No one is entitled to your unpaid labor.

(Sometimes even saying a polite 'no' can be emotional labor. I don't have a good solution to this.)

x.com · 2022-08-02 x.com · 2020-02-23

The sunk cost fallacy fallacy

The sunk cost fallacy fallacy is when you put a lot of time into a project, find your motivation running out right before you get to the finish line, and tell yourself that because of the sunk cost fallacy the rational decision is to quit.

There's a different concept from behavioral economics that explains why we so often quit just before completing a project: hyperbolic discounting. We focus too much on the short term pain of finishing and too little on the sustained long term gain from releasing the project.

x.com · 2020-10-23

The imagined-dependency problem

A project management pitfall that's super common in my experience but I've never heard anyone mention is the "imagined dependency" problem. We assume or imagine that task A is a dependency for B, when in fact it isn't. Usually this causes no problems, but sometimes A gets delayed, and we fail to start on B, unnecessarily delaying project completion.

Why does this happen? I don't know but I have a few guesses. Maybe initially we came up with a sequence A, B, C because it is often best to focus our energies on one thing at a time (totally reasonable) but over time we incorrectly start to see this arbitrary sequence as a dependency chain A → B → C because we have so often envisioned starting B once A gets done and C when B gets done.

Another possibility is that it could be related to a famous cognitive bias called the disjunction effect. In one experiment, students said they would buy a vacation to Hawaii if they passed an impending exam (to celebrate) and also if they failed the exam (to recover). But they would refuse to actually make the decision until they learned the result of the exam.

How can we avoid imagining dependencies? Being explicit about what's a dependency and why is important. And project management tools like Kanban can help. But the most important thing is simply being aware of our tendency to do this — I'm no exception!

x.com · 2026-04-30

Working around the limits of our memory

This observation about people forgetting to take life-saving pills is a good reminder to go easy on ourselves and on others when we forget things, and start using checklists, reminders, and other memory aids instead. From the book *Scarcity* by Mullainathan and Shafir:

> Some may find this hard to believe: how can you forget something so important? But memory doesn't work that way. You don't remember as a function of long-term value. Certainly no one forgets to take painkillers: the pain is a constant reminder. Diseases such as diabetes, though, are "silent"; their consequences are not immediately felt. There is nothing to remind a person with an overburdened bandwidth to take those medications.

At least once a day, I serendipitously come across an idea that's relevant to a problem I'm thinking about. Over the years, I've learned the importance of interrupting what I'm doing to record the idea in my project notes.

I used to think I'd be able to remember those ideas when I resumed working on the project. But no — if I don't capture that fleeting moment when my unconscious brain makes a connection, it's gone. Accepting the limits of my memory has made me both happier and more productive.

A corollary is that it helps to have well-organized and easily accessible project notes for problems that you're not actually working on but just thinking about maybe working on someday.

It may feel slightly awkward to interrupt a free flowing social conversation but it's always ok to say, "excuse me, do you mind if I make a note of what you just said?" In fact, people usually appreciate it.

And don't forget your shower thoughts! I'm grateful that my wife knows and accepts that it's often necessary for me to run to my laptop for a few minutes right after I shower. Learning to use tricks like memory palace has also helped a bit.

x.com · 2019-08-06

Mornings for thinking, afternoons for meetings

I have a simple heuristic for productivity: mornings are for thinking and writing; afternoons are for meetings, collaboration, administrivia.

This only works if you're a morning person, but if you're not, you can find your own heuristic.

x.com · 2017-10-11

Meeting friction is a feature, not a bug

Every few years there's hype about automating meeting scheduling. The most recent round is because of AI agents. My advice: if you value your time, stay away from such tech like it's radioactive waste. Making scheduling painless is how you get "meetings that should have been emails". Scheduling friction makes it easier to be thoughtful about how you're spending your time. Not all types of friction are bad — I think the majority serve some useful purpose, albeit accidentally in some cases.

x.com · 2022-09-24 x.com · 2026-01-23

Parenthood and prioritization

I used to have trouble saying "no" to unnecessary work stuff. Then I became a father. The clarity you get from the voice in your head constantly going "would you really rather do this than spend time with the kid?" is magical.

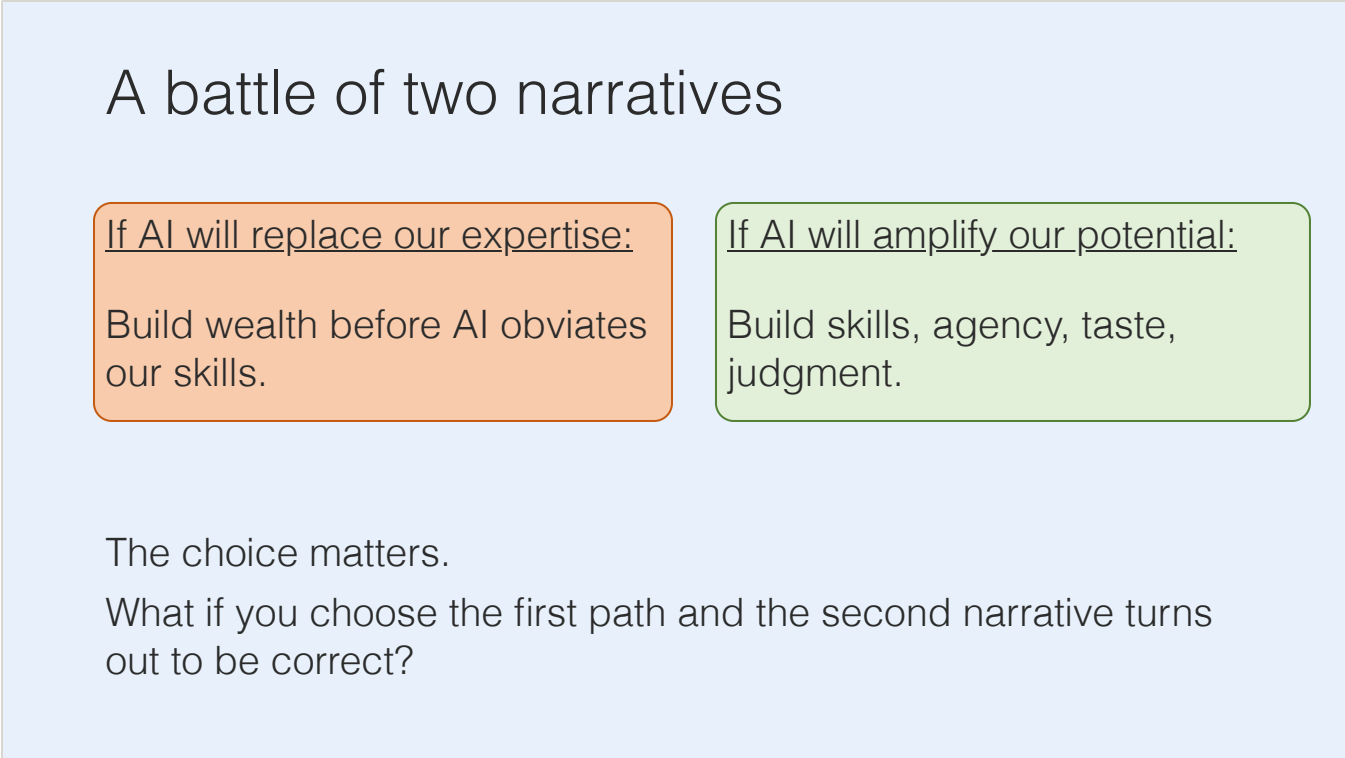
By forcing me to figure out what's most important, having kids has surprisingly helped my productivity.

x.com · 2019-12-19

12. AI AND THE FUTURE OF JOBS

In my **ICML 2026 keynote** titled "What will be left for us to work on?", I presented a framework for thinking about AI and the future of work. Many of the entries below are based on the talk.

Will AI replace or complement human abilities? ICML



From the beginning of AI, historically there have been these two battling narratives. In the past, the distinction was academic and philosophical, but now it has become an acutely practical question. Each one of us has to decide which camp we're in, or where on this spectrum we're in, because the practical consequences of believing in one versus the other are very, very different.

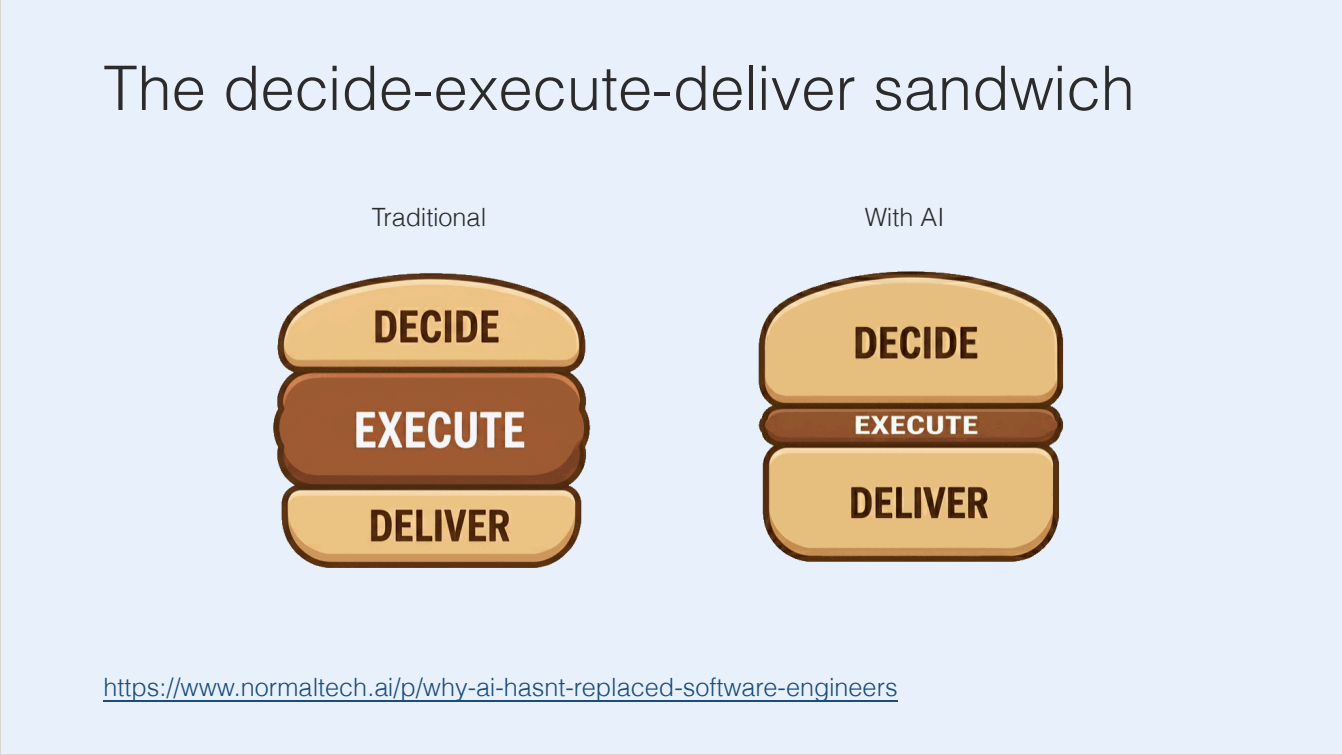
If you think this is a technology which in a few years is going to be able to replace everything we do today, then perhaps the correct response is to build wealth as quickly as possible before our skills become irrelevant. And this is the path that many have chosen in Silicon Valley. You may have heard of the "permanent underclass" meme.

On the other hand, if you believe, as I do, that this is a technology that will greatly amplify our potential, then now is the best time to build skills — especially the skills that are going to be complementary to what AI is doing and is going to be able to do — as well as to build all the things around it, such as agency and taste and judgment.

If you choose the first path, and it turns out that AI actually ends up being an amplifying technology as opposed to a replacing technology, then I would argue that over the next few years you've perhaps lost the best time in history to build these skills that will give us superpowers. That's why we all need to think about this question, even if we won't all land in the same place.

slide 4 · ICML 2026 · 2026-07-09

Why AI hasn't replaced software engineers: the decide-execute-deliver sandwich ICML



As of 2026, the best data suggests that software engineer employment in the U.S. is still *growing*, though slightly more slowly compared to a no-AI counterfactual world. But even this doesn't account for increased entry into entrepreneurship.

Why haven't AI agents replaced software engineers? Many kinds of knowledge work, including software development, can be seen as a "decide-execute-deliver sandwich". AI compresses the middle but the other two layers resist automation in a way that will not be overcome by capability improvements alone. In 2026, the middle has already been squeezed and there isn't much room for further compression, so there isn't a future capability leap that will cause discontinuous impacts, either. This framework is our answer to the question.

The decide layer: understanding customer requirements, developing the specification, planning, etc. That is not getting compressed by AI.

The execute layer: the actual coding and debugging. This is getting compressed, but it was only maybe one-third of the work to begin with.

The deliver layer: Understanding your code deeply enough to be accountable for what you release; carrying out integration into customer systems, maintenance, testing, etc. This layer is not getting compressed either.

Read the full essay.

slide 21 · ICML 2026 · 2026-07-09

Radiology, and the limits of the task model

I find the story of AI and radiology fascinating. Of course, Hinton's famous 2016 prediction about radiologists going obsolete was wrong. Tech advances don't automatically and straightforwardly cause job replacement. But that's not the interesting part.

Radiology has embraced AI enthusiastically, and the labor force is growing nevertheless. The augmentation-not-automation effect of AI is despite the fact that AFAICT there is no identified "task" at which human radiologists beat AI. So maybe the "jobs are bundles of tasks" model in labor economics is incomplete. Paraphrasing something Melanie Mitchell pointed out to me, if you define jobs in terms of tasks maybe you're actually defining away the most nuanced and hardest-to-automate aspects of jobs, which are at the boundaries between tasks.

Can you break up your own job into a set of well-defined tasks such that if each of them is automated, your job as a whole can be automated? I suspect most people will say no. But when we think about *other people's jobs* that we don't understand as well as our own, the task model seems plausible because we don't appreciate all the nuances.

If this is correct, it is irrelevant how good AI gets at task-based capability benchmarks. If you need to specify things precisely enough to be amenable to benchmarking, you will necessarily miss the fact that the lack of precise specification is often what makes jobs messy and complex in the first place. So benchmarks can tell us very little about automation vs augmentation.

x.com · 2025-06-19

Humans versus AI is the wrong comparison

If we compared AI capabilities against humans with no access to tools, such as the internet, we would probably find that AI already outperformed humans at many or most cognitive tasks we perform at work. But of course this is not a helpful comparison and doesn't tell us much about AI's economic impacts. We are nothing without our tools.

And yet, many predictions about the impact of "AGI" are based on hypothetical human-AI comparisons in which the humans have internet access but no AI access. This kind of comparison is equally irrelevant.

The real question is humans + AI vs AI alone. In such a comparison, AI isn't going to outperform human-AI pairs, except in narrow, computationally heavy domains like games where speed is paramount and having a human in the picture only slows things down.

So whether or not AI will replace humans comes down to factors beyond accuracy — things like accountability, the ability to handle unknown unknowns, and potential preferences of customers and other workers to interact with a human, all weighed against the cost of employing a human.

This is not to say that AI won't displace jobs. But looking at capability benchmarks and going straight to claims about job loss is completely naive.

x.com · 2025-07-18

Telling young people to pick blue-collar work is exactly backwards ICML

The “lump of labor” fallacy

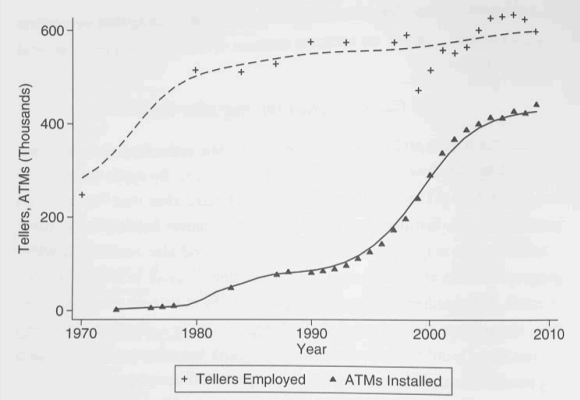


Figure 7.1. Adoption of automated teller machines did not reduce teller jobs. (Ruggles et al., Integrated Public Use Microdata Series: Version 5.0; Bureau of Labor Statistics, Occupational Employment Survey, http://www.bls.gov/oes/; Bank for International Settlements, Committee on Payment and Settlement Systems, various publications [see p. 243, note 9]).

In most jobs, the total amount of work is not fixed.

As ATMs were introduced, teller jobs grew.

Based on the work of economist James Bessen.

ATMs made it economically feasible for banks to open lots of regional branches. And those branches still needed human tellers to handle the things ATMs couldn't. So, paradoxically, employment actually grew. Economists have found this type of thing repeatedly. You might have heard the term Jevons' paradox; I like the term "lump-of-labor fallacy".

On the other hand, U.S. agriculture employment share dropped 50x because of mechanization. This is very different from the effect that automation has on white collar work. The reason is simple. Blue collar work is real work. There is a fixed, finite amount of it out there. So the effect of automation tends to be substitution.

Robotics could have this effect on many present-day jobs. Take warehouses. There are credible forecasts that by 2030, half of new warehouses will be built for primarily autonomous operation (much easier than replacing workers in existing warehouses with humanoid robots). After that, it takes 1-2 decades for the economics to force most existing warehouses to switch over. That's millions of jobs in the US alone.

White collar jobs are extremely malleable in their definition and variable in demand. When AI starts to do what we used to do, we just switch to doing new stuff and redefine the job. And we produce lots more units of work. There's no real ceiling to the amount of software that

can be written or legal work that can be done.

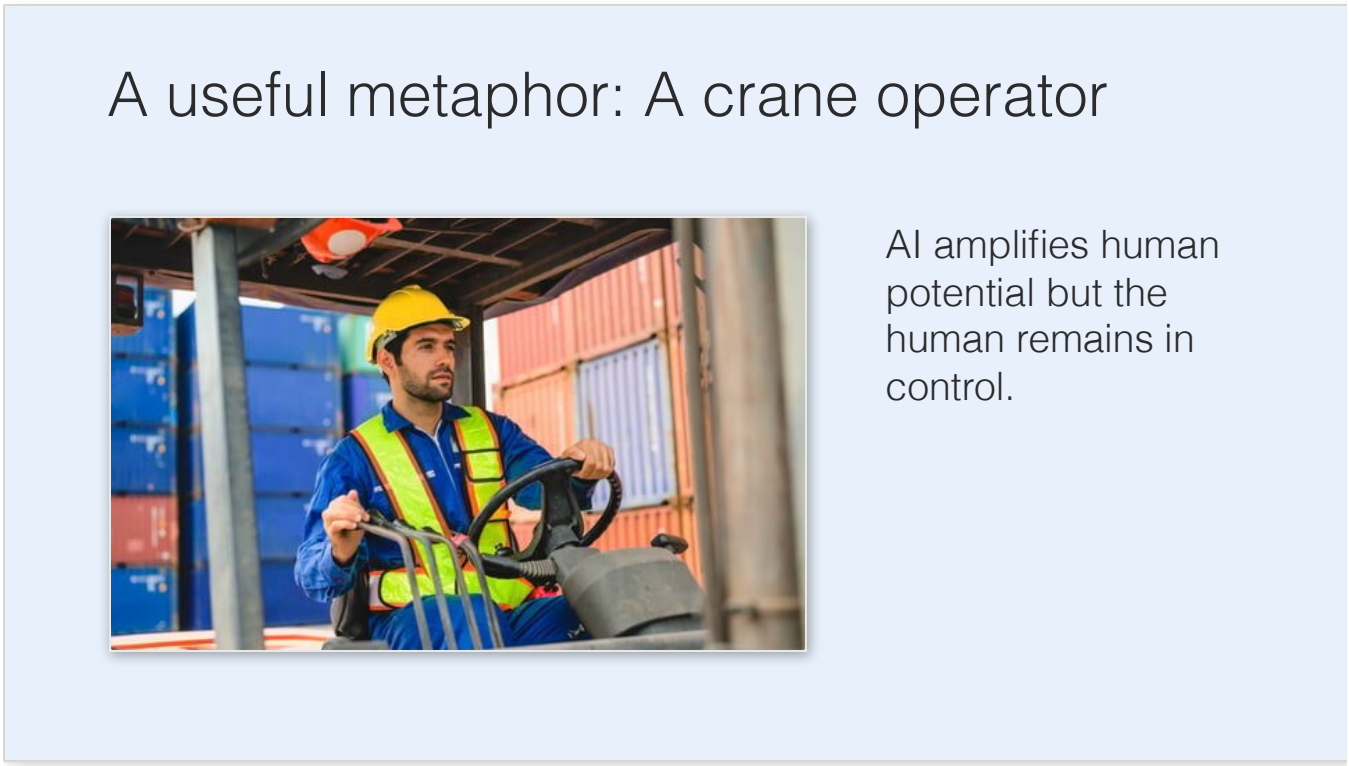
I can't believe some "AI experts" tell young people to pursue blue collar occupations because AI will eat the white collar work. I think this will turn out to be exactly ass-backwards.

Two interesting readings on this:

- What will be scarce? by Alex Imas
- So You Think You Could Be An Electrician? by Jesse Smith

slide 24 · ICML 2026 · 2026-07-09 x.com · 2026-09-01

How to think about cognitive jobs of the future ICML



I think it is already the case in software engineering, and will increasingly be the case in many professions, that we can think of knowledge workers as similar to a crane operator or a forklift operator. The machine greatly amplifies the human potential to do physical work — it's doing all the heavy lifting, but the person still remains in control. And I think this is what is happening with cognitive work. Machines are going to increasingly do the cognitive heavy lifting, but the person still remains in control.

The entire job gets reconceptualized as being about operating the machine, understanding the machine, and controlling the machine, as opposed to doing the cognitive work ourselves.

ICML 2026 keynote, slide 22 · 2026-07-09

Purely technical skills get devalued due to AI ICML

AI isn't replacing programmers, but it *is* making it harder to survive as a programmer with purely technical skills and no interest or expertise in how those skills translate to business or societal value. Funny thing is, this has always been true—it's just being accelerated a bit due to AI. There's a famous essay by Patrick McKenzie from 15 years ago called "Don't Call Yourself A Programmer, And Other Career Advice".

Technical skills tend to be verifiable tasks, and AI will continue to get better at them. Starting in the 2000s, there started to be a stark difference in the labor demand for programming jobs versus software engineering jobs. Programming jobs are conceived narrowly around the technical skills of coding and debugging. Software engineering jobs are responsible for all three layers of the decide, execute, deliver sandwich that I talked about — figuring out what even needs to be built, understanding customers, that sort of thing. It requires domain knowledge, judgment, and more.

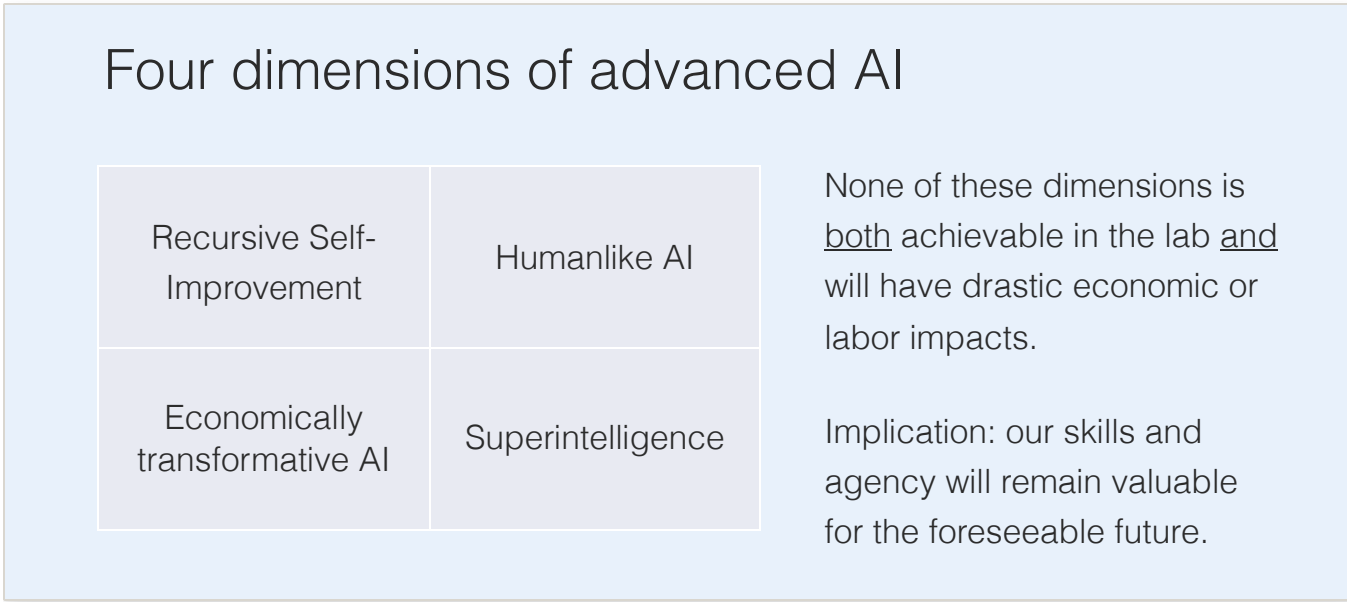


Chart by the Washington Post

I predict that the devaluation of purely technical skills will happen in more and more fields over time, not just software engineering.

ICML 2026 keynote, slide 44 · 2026-07-09

No milestone achieved in an AI lab will suddenly put people out of



For example, Recursive Self Improvement (RSI) may be achievable in the lab, but it won't immediately put people out of work. That's because while certain attributes of AI like speed and creativity are subject to recursive self-improvement, many of the bottlenecks to learning economically useful skills is external, such as medical research.

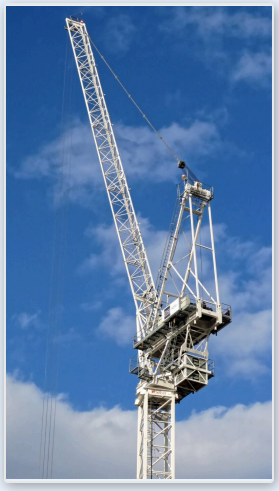
On the other hand, I view it as essentially certain that AI will be economically transformative, but it's not because of the next model release — it's because of things that will gradually happen over the next couple of decades, even if current capabilities were to remain frozen in time.

slide 42 · ICML 2026 · 2026-07-09

The vision: co-superintelligence ICML

The vision: co-superintelligence

We can ensure that the superintelligences of the future are AI-augmented humans rather than AI systems acting alone.



A ~~bicycle~~ crane

for the mind

Computers have often been called bicycles for the mind. I think AI can be more than that. I think AI can be a crane for the mind, if you will indulge my metaphor, in the sense that it can amplify our potential to previously unimaginable heights. Getting there can seem daunting. It has an incredible learning curve. I feel like I'm on a treadmill all the time, but I'm very excited about it. I think it's a fun challenge, and I think it's worth fighting this fight. Compared to five years ago, in a way, I feel ... maybe superintelligent is not the right word, but I feel like I have superpowers, given the extent to which AI allows me to take on new ambitious things that were not possible before, and push myself harder than was possible before.

And I think we can ensure that this remains the case even as AI capabilities advance, for the foreseeable future. It might be that in some distant future this becomes impossible to do, but it is very premature to give up the fight now. I certainly plan to continue this fight, and I hope to work towards this vision of co-superintelligence, and I hope you will join me.

slide 55 · ICML 2026 · 2026-07-09

13. WORKING WITH AI — TIPS AND STRATEGIES

Personal reflections on adapting to AI ICML

Given that AI capabilities are rapidly improving we're all struggling with the challenge of adapting our workflows. I think everyone should choose their own path, but here are some reflections that are hopefully interesting to consider.

The floor keeps moving up but we can move the ceiling too

By floor I mean what AI can do on its own. By ceiling I mean what AI allows us to do by augmenting our capabilities — the ability to take on new and ambitious projects that were not possible before AI. But the ceiling is not going up automatically. It will only happen if we work on pushing the ceiling up.

The optimal balance between growth and productivity has shifted

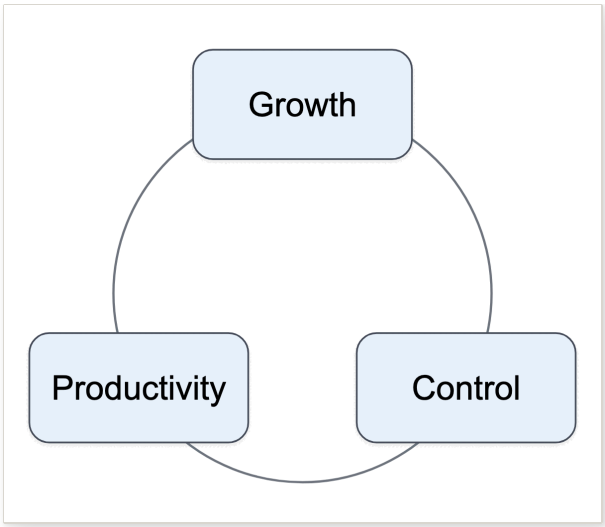
I find that I'm spending something like 10 hours per week just learning and experimenting with new workflows, as well as learning new topics. So one way to think about it is that AI enables big productivity improvements, and I've been trying to take that time saved and "reinvest" it into long-term growth and picking up new and complementary skills. Managing, controlling, evaluating, and directing highly capable AI agents requires very different skills than doing the cognitive heavy lifting ourselves. And as AI enables us to take on ever more ambitious projects, we will need a broader base of substantive knowledge and skills, which demands continual learning.

Embracing cognitive sweat

I've learned that if I don't feel exhausted at the end of the day, I've done something wrong. I've offloaded too much to AI. I've sacrificed too much of my long-term growth in the pursuit of short-term productivity.

Staying in control

Growth and productivity are two legs of a three-legged stool that we need to learn how to balance. And the third leg of that stool is staying in control.



Here are two heuristics that I've tried to use in order to do that. The first is **resisting the black-box temptation**. Companies want us to use agents as black boxes — just prompt it and it will go off and do its thing. I think that's a trap. I think that's very dangerous, and it will lead us to gradually give up control over time.

Second, and related, is **avoiding the dependence spiral**. It's very tempting to use AI for tasks that I am myself not yet an expert at, because learning new things is of course very hard. But that leads to my losing whatever little skills I have in that task over time. It's much better in the long run if I first put in the time to master the task myself before I use AI to augment my productivity.

ICML 2026 keynote, slide 54 · 2026-07-09 x.com · 2026-07-19

Two contrasting patterns of working with agents: delegation and collaboration

Two contrasting patterns of working with agents are emerging: delegation and collaboration. Delegation makes sense when it's a long-horizon task that you want the agent to tackle asynchronously, your intent and specifications are clear to the agent, and it's easy to verify the output at the end even if you didn't stay in the loop. Effectively delegatable tasks are rarer than the hype would suggest, because it’s limited by what you can cheaply verify, not what the model can do.

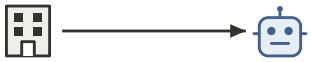
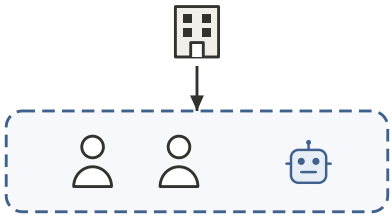
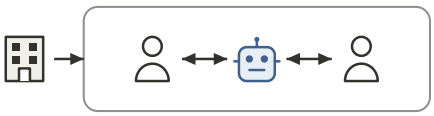
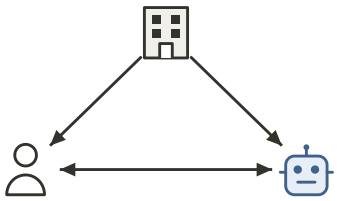
Collaboration makes sense when the task is hard to fully specify a priori and you want to be in the loop to iteratively figure out what you want, stay in control, recover from mistakes, sharpen your own skills through collaborative task performance, and have fun.

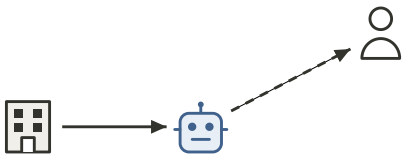
The design criteria for automation/delegation agents and collaboration agents are very different. If you're going to delegate a big task, accuracy and reliability are what matter the most. You want a frontier system that will do the best possible job. It doesn't matter if the task will take minutes or hours. If you're going to collaborate with an agent, the criteria are more multifaceted: latency (even at the expense of accuracy), transparency/controllability, creativity, and more. The agent should allow the user to stay in a "flow state" instead of having to delegate a task and come back later. It should promote the user's agency and be fun to work with.

We're at the very early stages of an emerging bifurcation between these two types of agents. (The conceptual distinction is ancient, but I'm talking about the product design of LLM-based agents + the practice of how people work with them.) I predict that the distinction will sharpen in the coming months. What's less clear is whether the specialization will happen at the level of companies, with some making a bet on automation and others on collaboration / human amplification, or at the level of products, with many companies going after both markets.

I've been going on for a while about how automation and collaboration agents need to be designed differently, but that's just the start. Here are 11 different configurations of firm/worker/agent that I think will require different agent designs. The industry's current focus is on automation / delegation, overindexing on metrics like time horizon while ignoring others like collaboration skill, and treating the human as the bottleneck. So there's a huge unmet market need. This is also a great opportunity for research, provided we treat harness design not as an afterthought but as a set of open problems that will require expertise not just in machine learning but also human-computer interaction and organizational behavior.

Configurations of firm, worker(s), and agent

#	Configuration	Description
1		**Delegation.** The worker hands a task to an agent she directs, and checks the result.
2		**Automation of existing work.** The firm runs the agent directly, on work people already do.
3		**Automation of new work.** The firm runs the agent directly, on work no human was doing.
4		**Augmentation of existing work.** The worker and an agent do a task together, in the loop.
5		**Augmentation into new work.** The same, on work no human was doing.
6		**Portable human capital.** The worker uses an agent for upskilling, with no employer in the picture.
7		**Ambient agent.** The firm deploys the agent across the whole workspace. It automatically provides help and context by observing what the workers are doing, without needing to be explicitly invoked .
8		**Mediation / collective intelligence.** The agent helps the team collaborate better with each other.
9		**Teammate.** The manager directs the worker and the agent separately; the two then work as peers. The agent's line to management is its own, not the worker's.

#	Configuration	Description
10		**Algorithmic management.** The agent stands between the firm and the worker, directing and evaluating the work.
11		**Exception handling.** The agent does the routine work and escalates what it cannot handle.

x.com · 2026-08-12 x.com · 2026-08-25

Even flawed AI is helpful for error checking

Error checking is a great application of generative AI capabilities and there are low-hanging fruits in just about every domain:

- Software: automatic detection of security vulnerabilities
- Writing: identifying logical gaps, unclear structure, and weak arguments (can be seen as a step up from spellcheck and grammar checking)
- Science: detecting errors in calculations, bugs in code, citations to papers that have been retracted, etc.
- Legal contracts: flagging missing boilerplate clauses, contradictory terms, loopholes, etc.
- Finance: fraud detection, errors in financial reports, etc.

Reasons why error checking is such a valuable application:

- It can be done automatically and unobtrusively, just like spellcheck.
- Even a 50% false positive rate is usually fantastic because a human can quickly review and accept/reject error reports, whereas in most other generative AI applications a high error rate would be fatal.
- In many cases, it automates drudgery rather than creative work.

The risk, of course, is that people become dependent on AI assistance and compensate by becoming more sloppy in their work. While it's important to think about and mitigate this risk, it is not a reason to avoid AI, any more than it is a reason to turn off spellcheck.

One example from my own workflow: I use LLMs / agents to review almost all my writing before posting, from social media posts to papers and everything in between. AI suggestions tend to vary in usefulness — some are straightforwardly good, others aren't useful by themselves but get me thinking in a productive way, yet others are things that I already considered and rejected, and some are totally off-base.

Overall, even though only a fraction of the suggestions end up being useful, the fact that the feedback is near-instant makes it quite worthwhile!

I make sure to do this after I've done as much human review (by myself or others) as I otherwise would. That way it augments my thinking instead of replacing it and decreases the risk of developing an over-reliance on AI feedback.

x.com · 2025-06-18

The pile of ideas grows ever bigger, and that's okay

I use AI heavily for validating ideas and prototyping things, but I'm much more cautious about AI use for anything that I need to ultimately publish and stand behind. One unfortunate consequence of speeding up prototyping more than execution is that I now have a much bigger pile of interesting, validated ideas that I simply don't have time to properly get started on. This has been bothering me but I've learned to come to terms with it. What would be even worse is starting tons of projects with AI, making partial progress, but never completing and publishing them. Project management skills like ranking projects by priority and making explicit go/no-go decisions were always a good idea, but have now become essential in order to avoid losing our sanity.

x.com · 2026-08-31

On using AI for writing

The real sign of AI writing is not superficial stuff like "It's not X—it's Y". It's the hollowness. Polished writing but relatively mundane ideas. The giveaway is that you're less impressed when you read it the second time. With good writing, it should be the other way around.

I'm not sure this is inherently about AI. It's more about the fact that people tend to turn to AI when they don't have much to say.

Reading text that has the syntactic smell of AI is mildly annoying, but when I read hollow writing I feel the writer is wasting my time, which is much more frustrating. So don’t do it. People are unlikely to respond to your email or subscribe to your newsletter or whatever you’re trying to get them to do. And they’ll probably remember that you betrayed their trust as a reader.

Once in a while I read something that has the syntactic smell of AI all over it, but then I do my habitual "second read" and it turns out to be actually deep. It's a rare treat when this happens. Like it says "It's not X—it's Y" but then brings the receipts to show that X is widely believed but Y is actually true.

It's even rarer when a writer is able to consistently deliver AI-assisted writing that has this quality. I've had the privilege of having a few incredible students in my classes who were able to use AI this way (all of whom are non-native English speakers). I don't think AI assisted writing is inherently hopeless and I'm hopeful that more people will learn to use AI effectively and responsibly, even for tasks like writing where the default way of using it is deeply problematic.

x.com · 2026-03-06

Part 3: My story

A story about techno-optimism

A small story. When I was a kid growing up in India, the K-12 school system mostly sucked and there were no public libraries. So it was a life-changing experience when my parents got me a computer with a CD-ROM encyclopedia. It was over a year's worth of their savings. 🙏

Sidenote: it turned out to be a slightly insane encyclopedia called Grolier that peddled an Americentric worldview. It taught me that the US won the Vietnam war. In any case, the computer was my best friend (of a total of two friends). I spent all my time on it.

In college I had terrible grades because I spent all my time programming for fun. I invented some algorithms to speed up the search for large prime numbers with special properties, collaborated with someone I met on the Internet, and got some new primes into "top ten" lists.

It was a totally useless hobby, but it amazed me that computers allowed me to compete on a level playing field worldwide. It seemed crazy, because as a child I was constantly aware that the technologies I got to interact with were decades behind the state of the art.

I saw computers and the Internet as a tool of both social and economic liberation, especially for the developing world, and an unalloyed force for good. I decided that computer science was what I wanted to do with my life.

So it was a crushing disappointment to grow up and realize that digital technology simply magnifies the best and the worst in us and that developing new tech isn't enough or automatically good. That realization, more than anything else, shapes my research agenda today.

One question I received — how did I get to where I am with terrible grades? It's a crazy story, but basically someone once took a chance on me. I often think back to that moment, especially since my current position regularly puts me on the other side of that process.

x.com · 2019-01-03

Grad school applications: my stroke of luck and what I learned from it

I was rejected by 8 of 9 schools I applied to, including what I thought would be a "safe" school (based on their ranking, but I would have been thrilled to go there). My backup job offer was a software engineering position that involved copy-pasting Java code all day.

After I got 8 rejections I emailed the 9th school to ask if they had any news. They said they had no record of my application. (Back then you had to mail in your materials.)

I panicked so hard. I reached out to a friend who was a grad student there. He went down to the admin office, and it turned out they'd put my envelope in some other pile by accident. I couldn't believe what I was hearing and was even more amazed when they actually accepted me.

If it weren't for the friend who was willing to help locate my application, it would have been a wipeout. That stroke of luck separated me from a life of staring at FactoryFactoryFactory.

When I look back at my career so far I see so many occasions when I was extraordinarily lucky. Hard work is important, but don't forget that we're in a universe ruled by randomness.

There's a great book about this. In every walk of life, we dramatically underestimate the role of randomness in how things turn out, and overestimate the effect of intentional actions.

The lesson I take from this that we should design our institutions to give people multiple chances in their careers and in life. And when evaluating candidates we should keep in mind that predicting success is extraordinarily hard, if not impossible.

We can't eliminate the impact of randomness on our lives, but we can manage it. For example, in my research I try to maintain 2–3 independent research agendas, because whether not not each of those will succeed is determined primarily by factors outside my control.

x.com · 2018-01-09

A story about my name

I want to share a story about my name, why I once decided to quit academia, why I came back, what I learnt from it, and why I'm grateful to have an audience on social media.

Near the end of my PhD, I had three main publications. I ran into someone at a conference whom I knew from previous events. He asked me about my work, and I briefly mentioned my papers. His reaction was something I didn't expect...

He'd seen and liked my papers, but the name hadn't registered. He didn't realize the 3 papers had the same first author and that it was someone he knew. He was embarrassed and apologetic. If he'd known, he said, he'd have advocated for me at his school for a faculty position(!)

I was naive back then, but I wasn't an idiot. I knew that my name tends to register as "generic Indian name with too many A's". I was used to jokes and many, many misspelled name tags (it's like banana, you never know when to stop!). But this seemed more serious.



In academia, your name is your brand. In a game of reputation, that's everything. Having your name be known in the community is how your get invited to talks, how people look you up to cite your papers, and how they decide whether your job application is worth a serious look.

At first I wondered if I was overreacting. But you know how when you learn a new word you suddenly start seeing it everywhere? I realized it wasn't just that one guy. People knew my work but didn't recognize my name. Paradoxically, my work was highly visible, but I was not.

That was a major reason I decided academia wasn't for me. I'd have to work way harder due to my invisibility. If you have a name that's easy to remember, you may wonder if I'm exaggerating. If you have a name that people forget, you know exactly what I'm talking about.

The startup world seemed egalitarian in contrast. If people liked your product, they would use it, no matter who built it. But a couple of years was all it took to become disillusioned with Silicon Valley. If you've read my work or seen my tweets, you probably know why that is.

So I was going to try to make a career of academia after all. But I decided I was going to do things differently. I would do everything I could to try to make sure that people who encountered my work would learn something about the author.

I tried various things, but what worked was online outreach. I had a somewhat popular blog that led to "ahh, now I know who you are!" comments when I was on the job market. And it's one reason I got on Twitter and stuck with it for years while it felt like shouting into the void.

Don't get me wrong—outreach is good, important, and rewarding, no matter what your name is. It's worked out well for me in many ways. But it was also a survival strategy long before I had a reputation and the security of tenure. Twitter, for all its faults, let me say: "I exist!"

Over the years I've occasionally gotten snide comments from people implying that the time I spend online is a sign of vanity or frivolousness unbecoming of an academic. Every one of those was from people who have the privilege of not having to be loud in order to be heard.

For anyone who thinks academia is a meritocracy: the consequences of a hard-to-pronounce name were enough to nearly make me quit, but that's a relatively *tiny* handicap. I've never faced sexual harassment or had people question my ability because of how I look.

There's a simple way to fight familiarity bias. When you read a good paper with an author you don't know, especially if they're junior, take a minute to look them up, get to know their work, cite them, and keep them in mind for events you organize. In short, remember the name!

x.com · 2019-11-26

Posts, talks and essays by Arvind Narayanan. Last updated 2026-10-19.