An Unsolicited Guide to Being A Researcher

Eugene Vinitsky

EMERGE Lab - New York University

Overview

- Don't do overview slides, it's bad practice.
- What function do they serve?
- Why does anyone need to know what order you're going to present things in.
- Just tell people the key idea upfront.
- There's no key idea here though.
- Also a lot of bullet points are bad
- Slides should have maybe a sentence of text at most

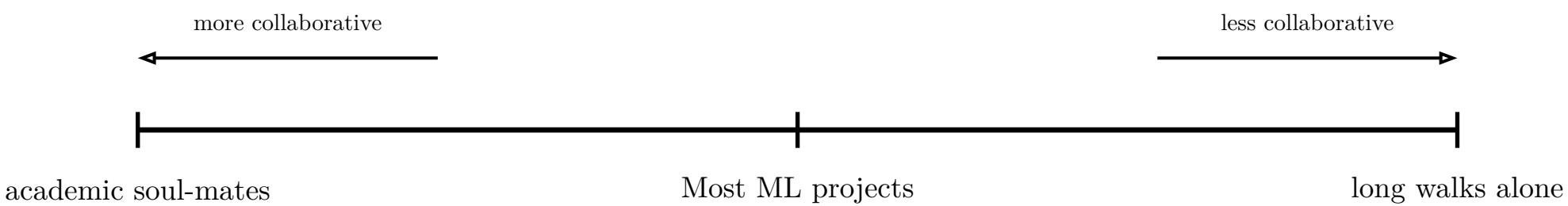
Your goals as a student

There is no one valid goal for a student to have. There are many that are possible. What's important is knowing what your goals are.

Your goals as a student

- Develop / pursue a unique view of the world that lets you see / do things differently
- Put out good research that you are proud of
- Reach the frontier of your field
- Become part of a community that you enjoy being a part of
- Chase a dream of something you want to build / know
- Put out enough research that you can get the job you want
- Have a good time
- Avoid entering the real world (not a good goal)
- Develop deep collaborations
- Become the best at a particular skill

Collaboration



Collaboration

- Most ML projects have some engineering component
- It is hard to do engineering alone
- It is also (for many people) boring to work alone too much
- Research can be inherently isolating which can be hard for people
- Therefore, many ML projects wind up being at least 2 people, often more.

Collaboration

- How does one become a good collaborator? The golden rule: **Do not block**.
- If your collaborator is waiting on something, they are blocked.
  - They will be annoyed and/or sad
  - *At that moment*, adding you to the project has made it worse rather than better.
- You should always be thinking: how can I make sure I am not blocking my collaborator.

Collaboration

- But research requires deep thought and time alone!
  - You can't spend all day on slack. You shouldn't!
  - You will have other demands on your time besides keeping your collaborators happy.
  - It's okay to have rules about what you will and won't do:
    - Times you don't want to work and want to block out.
    - Vacations, family events, exercise, hobbies
    - Just be communicative about them!

Collaboration

- Balancing multiple constraints with collaboration. Some ideas that might work for you, some won't:
  - **Overcommunicate about your constraints and timelines**
    - Share your calendar
    - Tell folks when things will be done
      - Try to stick to them
      - If this falls through, let them know. This is important, they are relying on you.
  - **Make your work visible**.
    - If they can see your project state, they can plan around you.
    - It can be fine to share a research notebook or research summaries. But it must be **legible**: short and clear.
  - There is no getting around talking.

Collaboration

- Shared project docs can be helpful
  - There's a cost to maintaining them, but they help with coordination.
  - Someone has to eat that cost. It's nice to try to share it around, I find it often doesn't work.

Collaboration

- Presenting to collaborators (or in large group meetings):
  - Your collaborators cannot read your mind
  - Results need to be immediately legible
  - For example:
    - Plots should be reasonably titled
    - Different experiments should be grouped together

Keeping up with the literature

- Note, this is mostly geared for folks who are or will be PhD students
- There's so much literature!! How are you supposed to stay on top of it?!

Keeping up with the literature

- Note, this is mostly geared for folks who are or will be PhD students
- There's so much literature!! How are you supposed to stay on top of it?!
- **You can't**. The faster you accept this, the better. What you can do:
  - Have a good mathematical toolbox
  - Have a broad sense of trends in your field
  - Be widely read in your subfield and associated areas
  - Be mostly on top of your current niche (if small enough)
  - Have a network that fills in gaps / provides pointers as needed

Keeping up with the literature

- Have a good mathematical toolbox : *Courses, self-study as needed*
- Have a broad sense of trends in your field : *Conferences, workshops, online seminars, talking to folks, social media (sparingly)*
- Be widely read in your subfield and associated areas : *Occasionally read papers from nearby areas (not too often)*
- Be mostly on top of your current niche (if small enough) :
  - *alerts from small set of author/keywords, talking to collaborators.*
  - *Tools like Scholar Inbox or Semantic Scholar.*
  - *Becomes easier as you acquire expertise, can often skim.*
- Have a network that fills in gaps / provides pointers as needed

Keeping up with the literature

- The underlying rule: **you want a reading list that is tailored to you, not to the mean researcher**.
- Develop a list of "must read" authors. Folks whose work you love and trust and usually learn something from. There are researchers who only write bangers. This is easy to tell because when you read a banger you know it.
- Tools like scholar inbox, which you can tune specifically to **your** preferences are good.
- Use survey papers to get a sense of the landscape.
  - **Good survey papers are opinionated on what you should read**
  - Bad survey papers are giant undifferentiated lists
- You can sometimes find lists of the "critical" works in a subfield
- LLMs are a component of this but can be overly homogeneous.
- Ask me!
- Ask your lab-mates!

Keeping up with the literature

Do not let social media source most of your reading. **This is a way to be behind the puck / wind up chasing the same trend. Virality based social media is inherently homogenizing.**

It is hard to contribute unique research if you're reading the same thing as everyone else. Become a weird butterfly.

Keeping up with the literature

- Be ruthless and honest with your reading list.
  - Most papers are worth at best a skim
  - Some small set of papers will be worth a full read
  - Some even smaller set will be worth hours of your time.
  - If you're never going to read a paper, don't bother saving it (except perhaps for a related work section).
- You will get faster at reading as you read. I can get the gist of most papers in < 10 minutes now because I just need to categorize them as something I already know with a little twist on top.
- Generally for papers go: abstract, intro, conclusion. Then decide if it's worth actually reading the details.

Keeping track of your research

- You must keep a research notebook. **The sooner you accept this, and invest the 1-2 needed hours a week, the faster you become a good researcher.**
- Some students struggle to do this, they don't see the immediate payoff.
- If you are not using a notebook, you are missing the single greatest cognitive tool developed by humanity in our lifetimes.
- You are not as smart as you think you are: you will forget things you have done.
- Corollary: if you can remember everything you've done, you're not running enough experiments.
- It's less work than you think and more valuable than you know.

What goes in a notebook?

- Every experiment that you run or idea you work through.
  - Details about what it was testing, why you ran it, overview of the results.
  - If it wasn't worth writing up the outcome, why did you even run the experiment?
- You cannot be too detailed here.
  - Invest the time in making it easy and develop a system that works for you.
  - You will probably change the system over time.

What goes in a notebook?

12_13 Learning Rate Sweep

Properties

date	2023-12-13 12:18
tags	stratego-experiment
analyzed	✅

+ Add property

Purpose

We wanted to see if there were better learning rates than the default we were using

Outcome

Ran this script to eval

python round_robin.py --model_paths /scratch/ev2237/log_dir/12_13_23/lr1e-4_train_per_log10_log_per_save10_ent_coef0.005/model4010.pthw,/scratch/ev2237/log_dir/12_13_23/lr1e-5_train_per_log10_log_per_save10_ent_coef0.005/model4010.pthw,/scratch/ev2237/log_dir/12_13_23/lr2e-5_train_per_log10_log_per_save10_ent_coef0.005/model4010.pthw,/scratch/ev2237/log_dir/12_13_23/lr5e-5_train_per_log10_log_per_save10_ent_coef0.005/model4010.pthw --num_envs 300

```
[[ 0.    0.   -0.25 -0.21]
 [ 0.    0.   -0.23 -0.15]
 [ 0.    0.    0.    0.02]
 [ 0.    0.    0.    0.  ]]
```

It seems that 2e-5 is the best one as I thought!

Next Steps

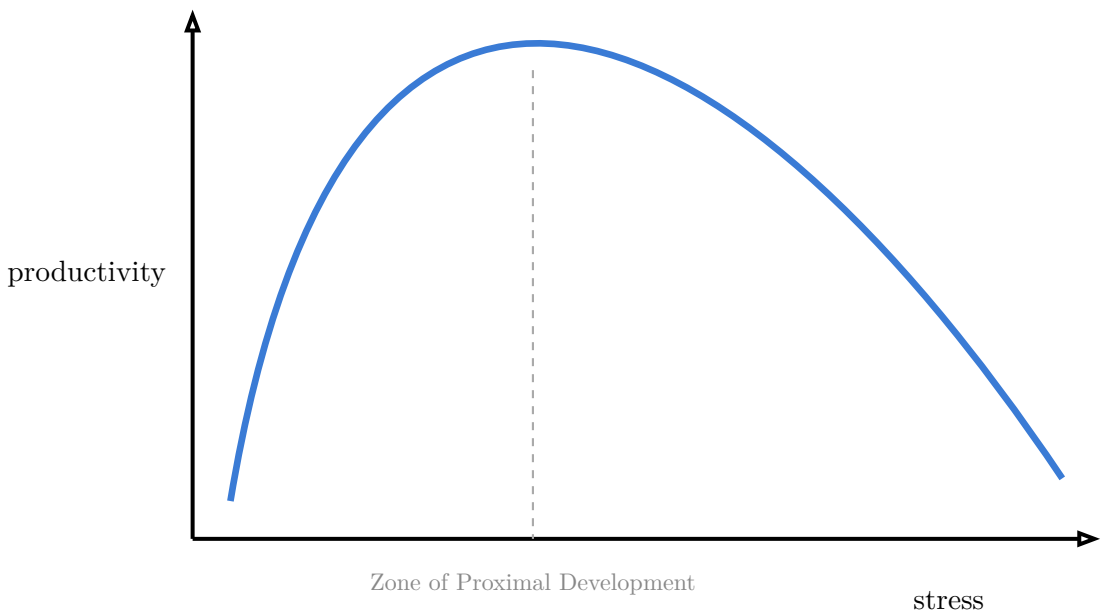
What goes in a notebook?

- Weekly write-ups:
  - What did I do this week? What did I learn? What remains unresolved?
  - This can usually be done in 15 minutes before you go home on Friday.
  - Value:
    - Can quickly scan over it to see what hypotheses you've already run through. I promise you will forget.
    - Your progress and effort becomes tangible. A source of emotional support in tough times.
  - Overviews of tested / untested hypotheses
  - This can also be the thing your collaborators look at.

What goes in a notebook

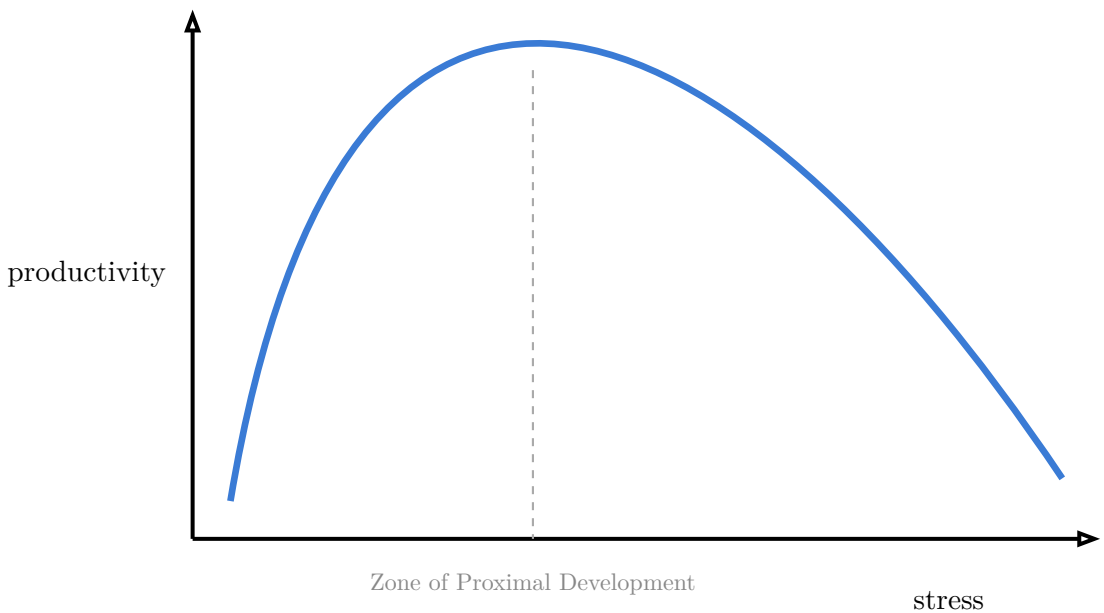
- Ideas you've had
- Links to experiments
- Papers you've read and your thoughts on them

Stress and Productivity



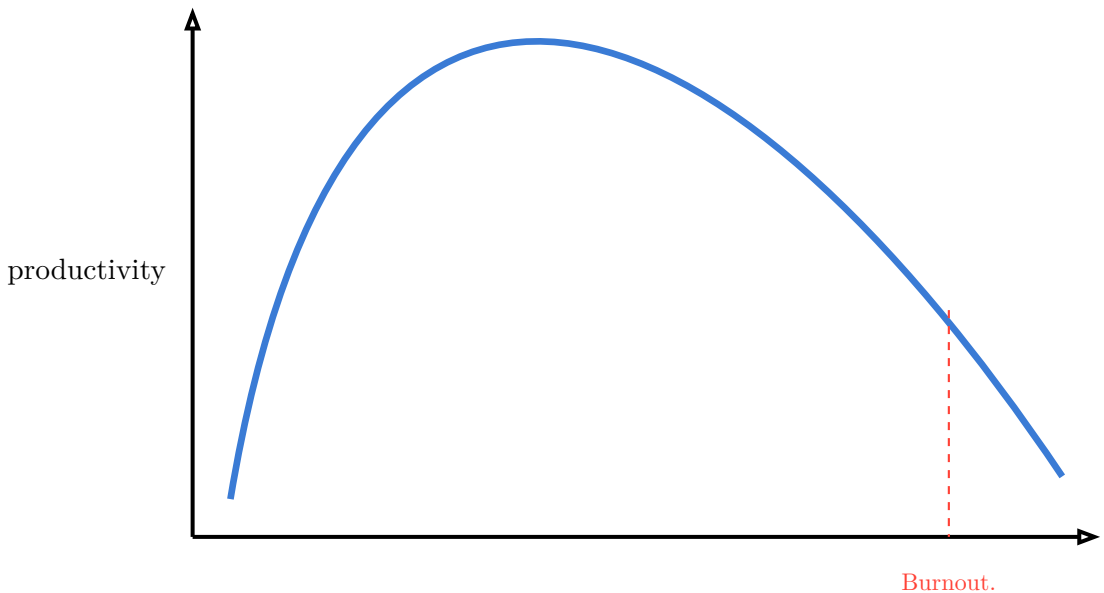
- We want you here.
- If your goal is to be a better researcher, you'll have to push yourself a little.
- You need to be honest with yourself about where that line is for you. The peak moves over the course of your life.

Stress and Productivity



- What does this peak feel like?
  - Intense but not overwhelming.
  - Sense of purpose?
  - Should mostly feel like a good time. Not always, but mostly.

Stress and Productivity



- Talk to me before you wind up here.
- You don't want to wind up here.
- We will do anything to keep you out of that zone.

- Golden rule: **never be blocked**.
  - There is always a way around wherever you are stuck. Take this as an item of faith.
    - Ask for help.
    - Try a different technique.
    - Go work on something else for a bit.
- Sometimes being stuck is productive. Banging your head against the wall can be good. **You have to learn for yourself what unproductive stuck vs. productive stuck feels like.** No one can peek into your head and know it for you.
- Ask yourself: "Am I blocked?" Refine your sense of this.
- Never be blocked != always be working.

Every experiment tests a single thing. It makes one change.

Tips for research productivity

- **Every experiment tests one thing. It makes one change.** If you make two changes, how do you know which one caused the difference?
- Every experiment has a clear way to measure the effect of that change.
- Every experiment is the minimal, testable next step.
- If you learn nothing else from me, this is the most important thing. Going too fast makes you slow.

Tips for research productivity

- Almost every paper you read will claim a positive result.
- Many of these positive results will turn out not to be positive results.
- How can you tell? You'll develop a sense of it, this is part of the expertise.
- Frequently, discovering **why** the result wasn't what it said it was will lead somewhere interesting.
- But only if the original result was interesting enough. It's not worth building on little benchmarkmaxxing results unless you really care about solving the benchmark.

Research productivity

- Generally, you are looking to put out 1-2 papers a year.
- Why?
  - Keeps you honest. Prevents you from being really stuck but not noticing it.
  - The best practice for research is doing research
- There are caveats and times where you may not care about this. At the beginning your career, you should perhaps care more.

Research productivity

- If the cluster is not on fire, you are not using it enough.
- Most types of ML are empirical science, more akin to biology than math.
- Of course, it can be about ideas too, but it still benefits from experimental feedback to guide your ideas.

Developing taste

- Researchers develop in stages:
  - First, they are mostly responding directly to other people's work. These can be small, incremental projects that try to close small gaps. This is important for your development!
  - As you close those gaps, you start to observe larger methodological and conceptual issues.
  - As you close those gaps, and you have mastery of your field, you start to notice totally missing directions or ideas.
- Each of these stages are important and equally valuable. It's hard to skip one stage and jump right to the next.

Developing Taste

Naomi Saphra @nsaphra.bsky.social · 15h

When shaping your research agenda, your objective is to find the weirdest niche possible that still has the potential to change everything.

Closing remarks

- We are literally being paid to play and learn.
- Research is a beautiful responsibility.
- Make the most of it and rise to the level of that responsibility!