Lessons from My First Two Years of AI Research

By Tom Silver

A friend of mine who is about to start a career in artificial intelligence research recently asked what I wish I had known when I started two years ago. Below are some lessons I have learned so far. They range from general life lessons to relatively specific tricks of the AI trade. I hope others find them useful.

Starting out

Find someone who you feel comfortable asking "dumb" questions

I was initially very intimidated by my colleagues and hesitant to ask basic questions that might betray my lack of expertise. It was many months before I felt comfortable enough with a few colleagues to ask questions, and still my questions were carefully formulated. Now I have three or four go-to people. I wish I had found them sooner! Before I was drowning in a backlog of terms to Google after work. Now I immediately ask a question when it comes and my confusion is resolved before it compounds.

Search for research inspiration in different places

Deciding what to work on can be the hardest part of research. Some general strategies that I have seen employed by researchers with long track records:

1. **Talk to a researcher in a different field.** Ask what problem they are excited about and try to restate the problem in computational terms. Ask if they have any datasets that they want to analyze, for which existing techniques seem insufficient. A lot of the most impactful work in machine learning results from collisions with bio/chem/physics, social sciences, or pure math. I am thinking, for example, of this NIPS 2016 paper by Matthew Johnson et al., which was motivated by a dataset of mouse behavior, or this ICML 2017 paper by Justin Gilmer et al. with applications to quantum chemistry.

2. **Code a simple baseline to get a feel for a problem.** For example, try to write some carefully calibrated code for controlling an inverted pendulum, or try see how far you can push a bag-of-words model on a natural language dataset. I often run into unanticipated situations when writing baselines -- bugs in my mental model or code. By the time my baseline is working, I usually have a handful of other ideas to try and a deeper understanding of the problem.

3. **Extend the experiments section of a paper you like.** Read the methods and results carefully. Try to find the duct tape. Consider the simplest extensions first and ask whether the paper's method would suffice. Think about baseline methods not discussed and imagine where those might fall short.

Invest in visualization tools and skills

The strategy I have come to adopt for writing research code is to start by creating visualization scripts. When the rest of the code has been written, running the visualization scripts will allow me to quickly verify whether my code matches my mental model. Even more importantly, good visualizations will often make bugs in my thinking or code far more obvious and interpretable than they would be otherwise. There is also something to be said here for self-motivation: when I finish this code, I will have a pretty figure or video to show people!

Coming up with the right visualization for the problem at hand can be tricky. If you are iteratively optimizing a model (e.g. deep learning), plotting loss curves is always a good place to start. There are also many techniques for visualizing and interpreting learned weights of (especially convolutional) neural networks such as guided backpropagation (e.g. guided backpropagation). In reinforcement learning and planning, the obvious thing to visualize is the agent acting in its environment, be it an Atari game, a robotic task, or a simple grid world (e.g. the environments in OpenAI Gym). Depending on the setup, it may also be possible to visualize the value function and how it changes over the course of training (shown below), or the tree of explored states. When dealing with graphical models, visualizing the distribution of a one or two dimensional variable as it changes over inference can be highly informative (shown below). One barometer for the effectiveness of a visualization technique is to estimate the amount of information that you have to hold in your head everytime you analyze a visualization. A bad visualization will require recalling in detail the code you wrote to produce it; a good one will scream an obvious conclusion.

Tensorboard is a popular GUI for visualizing Tensorflow deep learning models.

Plotting a distribution as evidence accumulates can make it a lot easier to debug a graphical model (from Wikimedia).

A value function being learned with Q-learning can be visualized on the grid world that it represents (by Andy Zeng).

Identify the fundamental motivations of researchers and papers

Researchers publishing in the same conferences, using the same technical jargon, and calling their field Artificial Intelligence can have polar opposite research motivations. Some folks have even suggested different names for the field in an effort to clear things up (e.g. Michael Jordan, in an excellent recent blog post). There are at least three main clusters that might be called the "math", "engineering", and "cognitive" motivations.

- "Math" motivation: what are the fundamental properties and limits of an intelligent system?
- "Engineering" motivation: how can we develop intelligent systems that solve real problems better than alternative approaches?
- "Cognitive" motivation: how can we model natural intelligence like that found in humans and other animals?

These motivations can be harmonious and many AI papers are interesting from multiple perspectives. Moreover, individual researchers are often driven by more than one of these motivations, which helps glue the field of AI together.

However, the motivations can also be at odds. I have some friends and colleagues who are distinctly of the "Engineering" bent and others who are primarily interested in "Biology." A paper showing that some clever combination of existing techniques is sufficient to break the state of the art on a benchmark will pique the interest of the engineers but might earn yawns or even scorn from the cognitive scientists. The reverse will happen towards a paper with only theoretical or toy results but claims of biological plausibility or cognitive connections.

Good papers and researchers will state at the outset their motivation, but often the fundamental impetus is buried. I have found it useful to consider papers through each lens one at a time in case the motivation is not obvious.

Drinking from the research community firehose

Finding papers

Papers in AI are fairly accessible and often published on arXiv. The sheer number of papers coming out right now is exciting and overwhelming. A number of folks in the community have made it easier to sort signal from noise. Andrej Karpathy hosts the arXiv sanity preserver with some helpful sorting, searching, and filtering features. Miles Brundage used to tweet a lightly curated list of arXiv papers each night; this duty has largely been assumed by the Brundage Bot. Many other tweeters share interesting references from time to time -- I recommend following your favorite researchers on Twitter (here are the people I follow). If Reddit is your thing, r/MachineLearning is pretty good, but the posts are often geared more towards ML practioners than academic researchers. Jack Clark publishes a weekly community newsletter called "Import AI" and Denny Britz has one called "The Wild Week in AI."

Scrolling through conference proceedings when they are published can also be worthwhile. The big three conferences are NIPS, ICML, and ICLR. Other reputable general-audience conferences include AAAI, IJCAI, and UAI. Each subdiscipline has more specific conferences too. For computer vision, there is CVPR, ECCV, and ICCV; for natural language, there is ACL, EMNLP, and NAACL; for robotics, there is CoRL (for learning), ICAPS (for planning, including but not limited to robotics), ICRA, IROS, and RSS; for more theoretical work, there is AISTATS, COLT, and KDD. Conferences are by far the dominant venue for publication, but there are journals as well. JAIR and JMLR are the two most prominent journals specific to the field. Occasionally high profile papers will also come out in general scientific journals like Nature and Science.

It is equally important but often much harder to find older papers. Those considered "classic" will often turn up from following reference trails, or from browsing the reading lists of graduate courses. Another way to discover older papers is to start with a senior professor in the field and find their earlier works, i.e. the research that paved the path to their professorship. Also feel free to email those professors to ask for additional references (though don't take offense if they are too busy to reply). I don't know of a consistent way to find older papers that are lesser known or overlooked beyond searching for keywords in Google scholar.

How much time should be spent reading papers?

I have heard two common pieces of advice regarding the amount of time one should spend with prior work. First, when just starting out, read all of the papers! People often say that the first semester or year of graduate school should be nothing but paper reading. Second, presumably beyond this initial ramp-up period, do not spend too much time reading papers! The rationale for the latter being that it is easier to creatively pose and solve problems if one is not biased towards previous approaches.

Personally, I agree with the first bit of advice and disagree with the second. I think one should read as many papers as possible always, so long as there is still time left over for original research. The notion that I will be better equipped to come up with a novel, superior approach to a hard problem if I am unfamiliar with what others have tried seems unlikely at best and arrogant at worst. Yes, a fresh perspective on a problem can be key, and yes, stories of amateurs solving longstanding challenges because of their outside-the-box thinking are inspiring (e.g. George Dantzig showing up late to lecture). But a career researcher cannot really depend on these fortunate jumps to sections of solution space not yet considered. The vast majority of time is spent patiently following the gradient, chipping away at a problem slowly and methodically. Reading relevant papers then is just a far more efficient way to figure out where we are and what to try next. (See also Julian Togelius on "tinkering versus research.")

With regard to reading as many papers as possible, there is one important caveat: taking time to digest a paper is just as important as reading it. It is better to spend a day with a handful of papers, taking careful notes and reflecting on each, than it is to devour paper after paper in succession. Read all of the papers that you can, but no more than that.

Conversations >> videos > papers > conference talks

Papers are definitely the most accessible source for understanding an unfamiliar research idea. But what path is most efficient? Different people may answer this question differently. For me, I have found that having a conversation (ideally with folks who already understand the idea in question) is by far the quickest and most effective path to understanding. In the case that such people are unavailable, videos about the subject, e.g. the author of the paper giving an invited talk, can provide very good insight. When the presenter is addressing a live audience, they tend to prioritize clarity more than concision. The priorities are swapped in most paper writing, where word count is king and background explanations may even be viewed as evidence of an author's unfamiliarity with the field. Finally, short conference talks are often more of a formality than an educational opportunity. Of course, a conversation with the presenter afterwards could be invaluable.

Beware the hype

Successful AI research solicits public attention, which brings more people into the field, which leads to more successful AI research. This cycle is mostly virtuous, but one pernicious side effect is hype. Journalists trying to get clicks, companies vying for investors and recruits, and researchers aiming for high profile publications and citations are all guilty of inflating the hype bubble. It is important to remain mindful of these various motives when assessing a headline or press release or paper.

At NIPS 2017, during the Q&A portion of a paper talk in a room with several hundred audience members, a prominent professor took the microphone ("on behalf of the hype police") and admonished the authors for using the word "imagination" in their paper title. I have mixed feelings about these sorts of public confrontations and I happen to have liked the particular paper in question. But I completely sympathized with the professor's frustration. One of the most common and aggravating manifestations of hype in AI research is the renaming of old ideas with flashy new terms. Beware of these buzzwords -- judge a paper based primarily on its experiments and results.

Running the research marathon

Always be making measurable progress

When searching for research projects early on, I spent hours and hours brainstorming. Brainstorming, for me at the time, meant putting my head down at my desk and hoping that some vague intuitions would coalesce into a concrete insight. At the end of a day of brainstorming, I would often feel tired and discouraged. Was this research, I wondered?

There is, of course, no recipe for research progress, and fumbling around in the dark is part of (most of) the process. However, I now find it much easier and more fulfilling to structure my work around measurable objectives. If I have very little idea what I'm doing next, the objective can be: write down a vague idea in the greatest detail available; if, in the course of writing the idea I rule it out, write down the reason for ruling it out (rather than scrapping the whole thing and losing the measure of progress). In the absence of any ideas, progress can take the form of papers read or conversations with colleagues had. By the end of each day, I now try to have some tangible evidence of my work. Even if the ideas are never used, my morale is much improved, and I need not worry about wasting future cycles on the same ideas that I ruled out that day.

Learn to recognize and backtrack from dead-ends

Strong researchers spend more time on good ideas because they spend less time on bad ideas. Being able to sort the good from the bad seems to be largely a function of experience. Nonetheless, researchers at any level constantly encounter the following decision. My research idea is flawed or inconclusive. Should I A) try to salvage or support the idea further, or B) try to justify abandoning the idea completely? I personally regret spending more time doing A) when I should have done B). Especially early on, I became stuck several times in what I now recognize as dead-ends and remained there for too long. My reluctance to leave was likely rooted in the sunk cost fallacy -- in backtracking from the dead-end, I would be sacrificing the time that I had already expended.

I still feel a twinge of disappointment when I leave research dead-ends. What I am now trying to internalize is that backtracking is forward progress, counterintuitively enough. The cost was well spent, not sunk. If I hadn't explored this dead-end today, I might have considered it tomorrow. Dead-ends are not the end. Also they're a healthy part of life. Hopefully one of these mantras will stick. If not, there's also a Feynman quote.

Write!

I once had an occasion to ask a very prominent AI researcher for early career tips. His advice was simple: write! Write blog posts and papers of course, but even more importantly, write down your thoughts throughout the day. Since he said that, I have noticed an obvious difference in progress that I make when I am actively writing versus simply thinking.

Mental and physical health are prerequisites for research

There is the dangerous trope of the academic researcher who forgoes sleep and self-care in an obsessive pursuit of scientific discovery. I have often been guilty of putting such behavior on a pedestal and striving towards it myself. I now understand (at a rational level, at least) that exercise and mental breaks are investments, not distractions. If I spend 8 hours sleeping and 4 hours working, I am immensely more productive than having spent 4 hours sleeping and 8 hours working, saying nothing of the downstream effects.

It can be very difficult to stop working in the middle of a tough problem. I still have the tendency to grind away at something even when I have passed the point of exhaustion or frustration and have no real chance of progress without a break. When I am able to step away and take a long breath, I am always happy that I did so. I hope to continue internalizing this fact as I move on to the next phase of my research career.

Acknowledgements

Many thanks to Hannah Blumberg, Kate Donahue, Mihail Eric, Ankit Gupta, Eric Purdy, Austin Stone, and Scott Swingle for reading and providing excellent feedback on an early version of this post.