Jeremy Karnowski   (Follow)

AI Lead at Insight

Aug 9 · 6 min read

Transitioning from Academic Machine Learning to AI in Industry

Jeremy Karnowski and Emmanuel Ameisen, Insight AI

Richard Socher (Chief Scientist at Salesforce) chatting with Insight AI Fellows about the future of NLP research.

Beyond Implementing Papers

It requires more than just taking online courses or being able to implement papers to get a job in the modern AI industry. After speaking with over 50 top Applied AI teams all over the Bay Area and New York, who come to Insight to find Applied AI practitioners, we have distilled our conversations into a set of actionable items outlined below. If you want to make yourself competitive and break into AI, not only do you have to understand the fundamentals of ML and statistics, but you must push yourself to restructure your ML workflow and leverage best software engineering practices. This means you need to be comfortable with system design, ML module implementation, software testing, integration with data infrastructure, and model serving.

Why do people implement papers?

Frequent advice for people trying to break into ML or deep learning roles is to pick up the required skills by taking online courses which provide some of the basic elements (e.g. Google's Deep Learning course, Fei-Fei Li's Computer Vision course, Richard Socher's NLP

course). Many of these courses teach participants in the same way as students learn—through lectures and structured coursework.

While these core concepts of machine learning and deep learning are essential for Applied AI roles in industry, the experience of grappling with a real, messy problem is a critical piece required for someone seeking an industry role in this space. Because of this, Andrew Ng recommends that people to prepare for this transition by implementing research papers. This forces one to work through some of the same issues that Applied AI professionals must face when moving research to production. Through this process, newcomers to the field also learn the many tips and tricks that allows researchers to debug algorithmic issues and iterate more rapidly for better performance.

However, through our many conversations with 50+ teams across industry, we consistently hear that **just learning concepts or implementing papers is not enough.**

.   .   .

Top 5 Skills you need to acquire before transitioning to Applied AI

1—System Design

When machine learning and deep learning is employed to solve business problems, you must design systems that consider the overall business operations. The system's components should be architected in a modular way, operate under a solid logic, and have extensive documentation for others.

Questions to ask for each project:

- How do you efficiently train your model without impacting day to day production?

- Where do you store and backup models?

- How do you run quick inference?

- What are concrete metrics you can relate to your model?

- If needed, how do you integrate a human feedback loop?

- Do you need deep learning to solve the problem, and if so, why?

2—Structured ML Modules

Jupyter notebooks, while wildly popular for rapidly prototyping deep learning models, are not meant to be deployed in production. For this reason, academics should push themselves to build structured ML modules that both use best practices and demonstrate you can build solutions that others can use.

Action Items:

- Take a look at this GitHub repo from an Insight AI Fellow that, in addition to having some exploratory notebooks, converts these ideas into a well structured repo that can be called from the command line.

- Read up on Tensor2Tensor (repo) and work to expose your model's training and inference through an elegant API.

3— Software Testing

Academics often run code to find and eliminate errors in an ad hoc manner, but building AI products requires a shift towards using a testing framework to systematically check if systems are functioning correctly. Using services like Travis CI or Jenkins for automatic code testing is a great first step to showing you can work in a company's production environment.

Action Items:

- Check out a good starter blog post on testing by Alex Gude (Insight alumnus, now Staff Data Scientist at Intuit).

- Read Thoughtful Machine Learning, which goes more in depth on how to test machine learning systems.

- Read this paper on the tests and monitoring that companies care about for production-ready ML

- Work through how you would test machine learning algorithms. For example, design tests that ensure a piece of your ML system

is modifying data in the way you assumed (e.g. correctly preprocessing image data by making it the correct size for the model to use).

- Check out testing options in Python.

- Read this article on the differences between different continuous integration services. We recommend trying out Travis CI as a quick intro into this industry-standard practice.

4—Integrating with data infrastructure

No matter what company you join, you will have to access their often large data stores to provide the training and testing data you need for your experiments and model building. To show that you would be able to contribute on day one, demonstrate that you are able to interface with structured data records.

Academics typically experience a world where all the data they use can be stored locally, which is often atypical in industry. Similarly, many competitions and research problems are structured in a way that academics only need to use a folder of images.

To demonstrate industry know-how, academics should show that they can (1) query from large datasets and (2) construct more efficient datasets for deep learning training.

Action Items:

- Check out the latest information on the current data engineering ecosystem and understand what large data stores exist, how you would query them, and how ML systems leverage distributed data.

- Read some of the early chapters in Designing Data-Intensive Applications

- Know how to do some basic queries in SQL, PostgreSQL, MongoDB, Hadoop, Spark, and other proprietary systems like Amazon Redshift and Google BigQuery.

- Read how different deep learning frameworks conceptualize how to load large datasets in a fast and efficient manner and organize

your data in this way (e.g. HDF5 for Keras, TFRecords in TensorFlow (another site), MXNet).

5—Model Serving

It's one thing to have built a solid ML or deep learning model that has excellent accuracy. It's another thing to turn that model into a package that can be incorporated into products and services. While many academics using ML are very familiar with model metrics (e.g. accuracy, precision, recall, F1 scores, etc), they need to become familiar with metrics that companies care about when it comes to fast, reliable, and robust ML services.

Action Items:

- Take one of your models and turn it into a service that others, including yourself, could use in other projects. As a first step, try out TensorFlow Serving (one example from Zendesk), FloydHub, and AWS Lambda & MXNet

- Read up about the metrics that companies care about when they are putting a model into production (e.g. here and here), which must be able to robustly handle requests. In addition, take into account users desires and preferences when looking to use ML APIs.

Parting Advice

AI is an exciting, ever-changing field. The demand for Machine Learning Engineers is strong, and it is easy to get overwhelmed with the amount of news surrounding the topic. We recommend following a few serious sources and newsletters, to be able to separate PR and abstract research from innovations that are immediately relevant to the field. Here are some sources to help out:

- Hacker News: Hacker News is a social news website focusing on computer science, data science, and entrepreneurship. It is run by Y Combinator, a well-known startup incubator. Don't be thrown off by the name! The original usage of the term "hacker" has nothing to do with cyber criminals, but rather someone who comes up with clever solutions to problems through their programming skills.

- Import AI: Great newsletter by Jack Clark of OpenAI, that stays on top of most important advances in the field.

- Insight Blog: We maintain a pretty active pace. Some posts in our AI blog talk about different past projects, and can serve as good inspiration for interesting problems to tackle. We also regularly send AI content to our mailing list, sign up here.

.   .   .

Want to learn applied Artificial Intelligence from top professionals in Silicon Valley or New York? Learn more about the Artificial Intelligence program.