

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
▼<rss xmlns:atom="http://www.w3.org/2005/Atom" version="2.0">
  ▼<channel>
    <title>Kyle Roth</title>
    <link>https://kylrth.com/</link>
    <description>Recent content on Kyle Roth</description>
    <generator>Hugo -- gohugo.io</generator>
    <language>en-us</language>
    <lastBuildDate>Mon, 18 Apr 2022 23:08:52 -0400</lastBuildDate>
    <atom:link href="https://kylrth.com/index.xml" rel="self" type="application/rss+xml"/>
  ▼<item>
    <title>WordBurner beta</title>
    <link>https://kylrth.com/post/wordburner/</link>
    <pubDate>Mon, 18 Apr 2022 23:08:52 -0400</pubDate>
    <guid>https://kylrth.com/post/wordburner/</guid>
    <description>Update 2022-04-27: The beta is over, but the apk is still installable with the instructions below and any feedback sent from inside the app will be received by me. I&rsquo;m going to be working on this more over the summer, and eventually publishing it on the app store. :) Ever since learning Spanish, it has been a dream of mine to create a vocabulary study app that meets my needs. Duolingo won&rsquo;t cover advanced vocabulary, Anki requires manually-generated decks, and other apps have expensive subscription plans.</description>
  </item>
  ▼<item>
    <title>PaLM</title>
    <link>https://kylrth.com/paper/palm/</link>
    <pubDate>Mon, 11 Apr 2022 12:17:25 -0400</pubDate>
    <guid>https://kylrth.com/paper/palm/</guid>
    <description>This was a paper I presented about in Bang Liu&rsquo;s research group meeting on 2022-04-11. You can view the slides I used here.</description>
  </item>
  ▼<item>
    <title>QA-GNN: reasoning with language models and knowledge graphs for question answering</title>
    <link>https://kylrth.com/paper/qa-gnn/</link>
    <pubDate>Tue, 05 Apr 2022 22:54:43 -0400</pubDate>
    <guid>https://kylrth.com/paper/qa-gnn/</guid>
    <description>This post was created as an assignment in Bang Liu&rsquo;s IFT6289 course in winter 2022. The structure of the post follows the structure of the assignment: summarization followed by my own comments. The authors create a novel system for combining an LM and a knowledge graph by performing reasoning over a joint graph produced by the LM and the KG, thus solving the problem of irrelevant entities appearing in the knowledge graph and unifying the representations across the LM and KG.</description>
  </item>
  ▼<item>
    <title>ethics drift within bubbles</title>
    <link>https://kylrth.com/post/ethics-drift/</link>
    <pubDate>Fri, 01 Apr 2022 08:35:54 -0400</pubDate>
    <guid>https://kylrth.com/post/ethics-drift/</guid>
    <description>Here are some snippets from a Lex Fridman interview with John Abramson, outspoken critic of Big Pharma. Lex: Are people corrupt? Are people malevolent? Are people ignorant that work at the low level and at the high level, at Pfizer for example? How is this possible? I believe that most people are good, and I actually believe if you join Big Pharma your life trajectory often involves dreaming, wanting, and enjoying helping people.</description>
  </item>
</rss>
```

- ▼<item>
 - <title>notes about neuroscience</title>
 - <link><https://kylrth.com/post/neuroscience/></link>
 - <pubDate>Thu, 31 Mar 2022 10:02:44 -0400</pubDate>
 - <guid><https://kylrth.com/post/neuroscience/></guid>
 - <description>How much of brain structure is coded for in the genome? For example, the hippocampus is generally thought to be responsible for consolidating long-term memories. Is the specialization of this region an epigenetic phenomenon due to optimization in the environment, or is it coded more directly? Will we eventually see these structures emerge in artificial networks with sufficient scale and good optimization, or will we need to code it more directly?</description></item>
- ▼<item>
 - <title>Experienced well-being rises with income, even above \$75,000 per year</title>
 - <link><https://kylrth.com/paper/experienced-well-being/></link>
 - <pubDate>Wed, 30 Mar 2022 13:34:53 -0400</pubDate>
 - <guid><https://kylrth.com/paper/experienced-well-being/></guid>
 - <description>Turns out that money does buy happiness. You may have heard that people’s average happiness stops improving once you make more than \$75,000/year? Researchers did a better survey with more data and found that that was not the case. The researchers cited 5 methodological improvements over the old research that suggested that it didn’t matter after \$75,000: They measured people’s happiness in real time, instead of having people try to remember past happiness levels.</description></item>
- ▼<item>
 - <title>Neural message passing for quantum chemistry</title>
 - <link><https://kylrth.com/paper/neural-message-passing/></link>
 - <pubDate>Fri, 25 Mar 2022 14:46:11 -0400</pubDate>
 - <guid><https://kylrth.com/paper/neural-message-passing/></guid>
 - <description>This post was created as an assignment in Bang Liu’s IFT6289 course in winter 2022. The structure of the post follows the structure of the assignment: summarization followed by my own comments. To summarize, the authors create a unifying framework for describing message-passing neural networks, which they apply to the problem of predicting the structural properties of chemical compounds in the QM9 dataset. paper summarization The authors first demonstrate that many of the recent works applying neural nets to this problem can fit into a message-passing neural network (MPNN) framework.</description></item>
- ▼<item>
 - <title>keep your tasks in the heap</title>
 - <link><https://kylrth.com/post/tasks-stack-heap/></link>
 - <pubDate>Tue, 22 Mar 2022 11:23:20 -0400</pubDate>
 - <guid><https://kylrth.com/post/tasks-stack-heap/></guid>
 - <description>Often when someone (usually a professor) is sharing their screen I see that their browser has so many tabs open that the descriptions are lost: That was my best impersonation as a Firefox user. Chrome will let you go a lot further (like ~113 tabs) before starting to provide a dropdown to show you the list of open tabs: Besides the obvious fact that this makes it hard to find a tab you’re looking for, you also waste computer memory and add to your cognitive load while you’re working.</description></item>
- ▼<item>
 - <title>The effect of model size on worst-group generalization</title>
 - <link><https://kylrth.com/paper/effect-of-model-size-on-worst-group-generalization/></link>
 - <pubDate>Thu, 17 Mar 2022 14:34:33 -0400</pubDate>
 - <guid><https://kylrth.com/paper/effect-of-model-size-on-worst-group-generalization/></guid>
 - <description>This was a paper we presented about in Irina Rish’s neural scaling laws course (IFT6167) in winter 2022. You can view the slides we used here.</item>

```
</description>
</item>
▼<item>
  <title>quotes from a Lex Fridman interview with Philip Goff</title>
  <link>https://kylrth.com/post/philip-goff/</link>
  <pubDate>Wed, 09 Mar 2022 09:39:16 -0500</pubDate>
  <guid>https://kylrth.com/post/philip-goff/</guid>
  <description>Here are some snippets from a Lex Fridman interview with Philip Goff, a panpsychist. The Enlightenment ideal is to follow the evidence and the arguments where they lead, but it's very hard for human beings to do that. I think we get stuck in some conception of how we think science ought to look. People talk about religion as a crutch, but I think a certain kind of scientism, a certain conception of how science is supposed to be, gets into people's identity and their sense of themselves and their security.</description>
</item>
▼<item>
  <title>hosting a Tor onion service</title>
  <link>https://kylrth.com/post/tor-onion-service/</link>
  <pubDate>Thu, 24 Feb 2022 18:28:49 -0500</pubDate>
  <guid>https://kylrth.com/post/tor-onion-service/</guid>
  <description>Tor onions are a way to host secure services that protect the anonymity of you and your clients. It also removes load from Tor exit nodes. If you open this page in the Tor browser it will redirect you to the following address: http://kylrthhj7mpvktolz7u6fnudt3hpdvjw4hzquanjpepgsf5vcq5divad.onion/post/tor-onion-service/ which can only be opened from inside the Tor network. getting started To host an onion service, we'll have a Docker container running Tor that decodes requests and forwards them to another container hosting the service.</description>
</item>
▼<item>
  <title>Scaling laws for the few-shot adaptation of pre-trained image classifiers</title>
  <link>https://kylrth.com/paper/scaling-laws-few-shot-image-classifiers/</link>
  <pubDate>Tue, 22 Feb 2022 13:19:12 -0500</pubDate>
  <guid>https://kylrth.com/paper/scaling-laws-few-shot-image-classifiers/</guid>
  <description>The unsurprising result here is that few-shot performance scales predictably with pre-training dataset size under traditional fine-tuning, matching network, and prototypical network approaches. The interesting result is that the exponents of these three approaches were substantially different (see Table 1 in the paper), which says to me that the few-shot inference approach matters a lot. The surprising result was that while more training on the "non-natural" Omniglot dataset did not improve few-shot accuracy on other datasets, training on "natural" datasets did improve accuracy on few-shot Omniglot.
  </description>
</item>
▼<item>
  <title>Learning explanations that are hard to vary</title>
  <link>https://kylrth.com/paper/learning-explanations-hard-to-vary/</link>
  <pubDate>Tue, 22 Feb 2022 12:29:17 -0500</pubDate>
  <guid>https://kylrth.com/paper/learning-explanations-hard-to-vary/</guid>
  <description>The big idea here is to use the geometric mean instead of the arithmetic mean across samples in the batch when computing the gradient for SGD. This overcomes the situation where averaging produces optima that are not actually optimal for any individual samples, as demonstrated in their toy example below: In practice, the method the authors test is not exactly the geometric mean for numerical and performance reasons, but effectively accomplishes the same thing by avoiding optima that are "inconsistent" (meaning that gradients from relatively few samples actually point in that direction).</description>
</item>
▼<item>
  <title>In search of robust measures of generalization</title>
  <link>https://kylrth.com/paper/robust-measures-of-generalization/</link>
```

- [<pubDate>Mon, 21 Feb 2022 15:33:22 -0500</pubDate>](#)
[<guid>https://kylrth.com/paper/robust-measures-of-generalization/</guid>](#)
[<description>These authors define robust error as the least upper bound on the expected loss over a family of environmental settings \(including dataset, model architecture, learning algorithm, etc.\): \$\mathbb{E}_{\omega \in P^e} \left\[\ell\(\phi, \omega\) \right\]\$ The fact that this is an upper bound and not an average is very important and is what makes this work unique from previous work in this direction. Indeed, what we should be concerned about is not how poorly a model performs on the average sample but on the worst-case sample.</description>](#)
</item>
- ▼ [<item>](#)
[<title>It's not just size that matters: small language models are also few-shot learners</title>](#)
[<link>https://kylrth.com/paper/not-just-size-that-matters/</link>](#)
[<pubDate>Fri, 18 Feb 2022 13:13:54 -0500</pubDate>](#)
[<guid>https://kylrth.com/paper/not-just-size-that-matters/</guid>](#)
[<description>We presented this paper as a mini-lecture in Bang Liu's IFT6289 course in winter 2022. You can view the slides we used here.</description>](#)
</item>
- ▼ [<item>](#)
[<title>Educated</title>](#)
[<link>https://kylrth.com/book/educated/</link>](#)
[<pubDate>Thu, 17 Feb 2022 09:51:11 -0500</pubDate>](#)
[<guid>https://kylrth.com/book/educated/</guid>](#)
[<description>thinking for yourself I recognized myself a little in this book, not in the events, severity, or locations but in the path to being "educated" in the sense that Westover intends. I'll try to convey what that sense is with some quotes from the book. The first moment is after she takes a class on American history at BYU. She returns home and gets her face dirty while working, and her brother calls her a "r", a joke he had made many times before.</description>](#)
</item>
- ▼ [<item>](#)
[<title>Scaling laws for transfer</title>](#)
[<link>https://kylrth.com/paper/scaling-laws-for-transfer/</link>](#)
[<pubDate>Wed, 16 Feb 2022 14:12:26 -0500</pubDate>](#)
[<guid>https://kylrth.com/paper/scaling-laws-for-transfer/</guid>](#)
[<description>This post was created as an assignment in Irina Rish's neural scaling laws course \(IFT6167\) in winter 2022. The post contains no summarization, only questions and thoughts. Sometimes these scaling laws can feel like pseudoscience because they're a post hoc attempt to place a trend line on data. How can we be confident that the trends we observe actually reflect the scaling laws that we're after? In the limitations section they mention that they didn't tune hyperparameters for fine-tuning or for the code data distribution.</description>](#)
</item>
- ▼ [<item>](#)
[<title>Deep learning scaling is predictable, empirically</title>](#)
[<link>https://kylrth.com/paper/scaling-predictable-empirically/</link>](#)
[<pubDate>Mon, 14 Feb 2022 10:38:11 -0500</pubDate>](#)
[<guid>https://kylrth.com/paper/scaling-predictable-empirically/</guid>](#)
[<description>This was a paper we presented about in Irina Rish's neural scaling laws course \(IFT6167\) in winter 2022. You can view the slides we used here. It's important to note that in the results for NMT \(Figure 1\) we would expect the lines in the graph on the left to curve as the capacity of the individual models is exhausted. That's why the authors fit the curves with an extra constant added.</description>](#)
</item>
- ▼ [<item>](#)
[<title>Masked autoencoders are scalable vision learners</title>](#)
[<link>https://kylrth.com/paper/masked-autoencoders-are-scalable-vision-learners/</link>](#)
[<pubDate>Fri, 11 Feb 2022 14:18:30 -0500</pubDate>](#)

<guid><https://kylrth.com/paper/masked-autoencoders-are-scalable-vision-learners/></guid>

<description>This post was created as an assignment in Irina Rish's neural scaling laws course (IFT6167) in winter 2022. The post contains no summarization, only questions and thoughts. In this paper they mention that the mask vector is learned, and it sounds like the positional embeddings are also learned. I remember in Attention is all you need they found that cosine positional embeddings worked better than learned ones, especially for sequences of longer length.</description>

</item>

▼<item>

<title>Data scaling laws in NMT: the effect of noise and architecture</title>

<link><https://kylrth.com/paper/data-scaling-laws-nmt/></link>

<pubDate>Wed, 09 Feb 2022 20:47:59 -0500</pubDate>

<guid><https://kylrth.com/paper/data-scaling-laws-nmt/></guid>

<description>This paper is all about trying a bunch of different changes to the training setup to see what affects the power law exponent over dataset size. Here are some of the answers: encoder-decoder size asymmetry: exponent not affected, but effective model capacity affected architecture (LSTM vs. Transformer): exponent not affected, but effective model capacity affected dataset quality (filtered vs. not): exponent and effective model capacity not effected, losses on smaller datasets affected dataset source (ParaCrawl vs.</description>

</item>

▼<item>

<title>Parallel training of deep networks with local updates</title>

<link><https://kylrth.com/paper/parallel-training-with-local-updates/></link>

<pubDate>Wed, 09 Feb 2022 10:50:21 -0500</pubDate>

<guid><https://kylrth.com/paper/parallel-training-with-local-updates/></guid>

<description>This post was created as an assignment in Irina Rish's neural scaling laws course (IFT6167) in winter 2022. The post contains no summarization, only questions and thoughts. Once I learned how the loss functions worked for each chunk, my first question was whether the earlier chunks were going to be able to learn the low-level features that later chunks would need. Figure 7 seems to show that they do, although their quality apparently decreases with increasingly local updates.</description>

</item>

▼<item>

<title>A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification</title>

<link><https://kylrth.com/paper/cnn-sentence/></link>

<pubDate>Wed, 02 Feb 2022 15:35:00 -0500</pubDate>

<guid><https://kylrth.com/paper/cnn-sentence/></guid>

<description>This post was created as an assignment in Bang Liu's IFT6289 course in winter 2022. The structure of the post follows the structure of the assignment: summarization followed by my own comments. paper summarization Word embeddings have gotten so good that state-of-the-art sentence classification can often be achieved with just a one-layer convolutional network on top of those embeddings. This paper dials in on the specifics of training that convolutional layer for this downstream sentence classification task.</description>

</item>

▼<item>

<title>Learning transferable visual models from natural language supervision (CLIP)</title>

<link><https://kylrth.com/paper/clip/></link>

<pubDate>Wed, 02 Feb 2022 12:35:03 -0500</pubDate>

<guid><https://kylrth.com/paper/clip/></guid>

<description>This post was created as an assignment in Irina Rish's neural scaling laws course (IFT6167) in winter 2022. The post contains no summarization, only questions and thoughts. This concept of wide vs. narrow supervision (rather than binary "supervised" and "unsupervised") is an interesting and flexible way to think about the way these training schemes leverage data. The zero-shot CLIP matches the performance of 4-shot CLIP, which is a surprising result. What

do the authors mean when they make this guess about zero-shot's advantage:

</description>

</item>

▼<item>

<title>Distributed representations of words and phrases and their compositionality</title>

<link><https://kylrth.com/paper/distributed-representations/></link>

<pubDate>Tue, 01 Feb 2022 16:09:19 -0500</pubDate>

<guid><https://kylrth.com/paper/distributed-representations/></guid>

<description>This post was created as an assignment in Bang Liu's IFT6289 course in winter 2022. The structure of the post follows the structure of the assignment: summarization followed by my own comments. paper summarization This paper describes multiple improvements that are made to the original Skip-gram model: Decreasing the rate of exposure to common words improves the training speed and increases the model's accuracy on infrequent words. A new training target they call "negative sampling" improves the training speed and the model's accuracy on frequent words.</description>

</item>

▼<item>

<title>meaning-making in the post-modern world</title>

<link><https://kylrth.com/post/meaning-making/></link>

<pubDate>Tue, 01 Feb 2022 14:35:37 -0500</pubDate>

<guid><https://kylrth.com/post/meaning-making/></guid>

<description>Here are some snippets from a Lex Fridman interview with Peter Wang, co-founder and CEO of Anaconda: For a lot of human history, there wasn't so much a meaning crisis as just a food and not getting eaten by bears crisis. Once you get to a point where you can make food there was a not getting killed by other humans crisis. Sitting around wondering what it's all about is a relatively recent luxury.</description>

</item>

▼<item>

<title>Deep learning</title>

<link><https://kylrth.com/paper/deep-learning/></link>

<pubDate>Thu, 20 Jan 2022 15:11:00 -0500</pubDate>

<guid><https://kylrth.com/paper/deep-learning/></guid>

<description>This post was created as an assignment in Bang Liu's IFT6289 course in winter 2022. The structure of the post follows the structure of the assignment: summarization followed by my own comments. paper summarization The authors use the example of distinguishing between a Samoyed and a white wolf to talk about the importance of learning to rely on very small variations while ignoring others. While shallow classifiers must rely on human-crafted features which are expensive to build and always imperfect, deep classifiers are expected to learn their own features by applying a "general-purpose learning procedure" to learn the features and the classification layer from the data simultaneously.</description>

</item>

▼<item>

<title>Cratylus</title>

<link><https://kylrth.com/book/cratylus/></link>

<pubDate>Sun, 16 Jan 2022 15:11:14 -0500</pubDate>

<guid><https://kylrth.com/book/cratylus/></guid>

<description>In this dialog Hermogenes comes to Socrates to discuss Cratylus's view of the nature of names, whether they are true to the objects they represent or are just conventional. Hermogenes believes that names are purely conventional, while Cratylus believes the opposite. Socrates falls somewhere in the middle: I quite agree with you that words should as far as possible resemble things; but I fear that this dragging in of resemblance, as Hermogenes says, is a shabby thing, which has to be supplemented by the mechanical aid of convention with a view to correctness; for I believe that if we could always, or almost always, use likenesses, which are perfectly appropriate, this would be the most perfect state of language; as the opposite is the most imperfect.</description>

</item>

- ▼<item>
 - <title>Crito</title>
 - <link><https://kylrth.com/book/crito/></link>
 - <pubDate>Mon, 27 Dec 2021 14:05:02 -0700</pubDate>
 - <guid><https://kylrth.com/book/crito/></guid>
 - <description>In this dialogue Crito comes to Socrates who is in prison waiting to be executed by the state. Crito has come to convince Socrates to come and escape with him. Crito's escape plan will not cause great inconvenience for any of Socrates's friends, and he would be able to live well in Thessaly. Socrates ends up convincing Crito that it would be wrong for him to escape. "the opinion of the many" CRITO: But you see, Socrates, that the opinion of the many must be regarded, for what is now happening shows that they can do the greatest evil to anyone who has lost their good opinion.</description>
- ▼<item>
 - <title>Apology of Socrates</title>
 - <link><https://kylrth.com/book/apology/></link>
 - <pubDate>Sun, 26 Dec 2021 13:40:45 -0700</pubDate>
 - <guid><https://kylrth.com/book/apology/></guid>
 - <description>I'm starting a course of foundational texts in philosophy with a friend of mine, and this is the first one we've read. Socrates is often considered a founder of Western philosophy, and it was easy for me to see in the text some common philosophical themes I've been exposed to growing up in the West. the fear of death is irrational Socrates argues that the fear of death is irrational from two perspectives: one, that what happens after death cannot be bad; and two, that a righteous person needs to be more concerned with whether he is doing right or wrong than whether death occurs or not.</description>
- ▼<item>
 - <title>avatarify</title>
 - <link><https://kylrth.com/post/avatarify/></link>
 - <pubDate>Wed, 24 Nov 2021 11:58:34 -0500</pubDate>
 - <guid><https://kylrth.com/post/avatarify/></guid>
 - <description>Avatarify is a cool project that lets you create a relatively realistic avatar that you can use during video meetings. It works by creating a fake video input device and passing your video input through a neural network in PyTorch. My laptop doesn't have a GPU, so I used the server/client setup. setting up the server Be sure you've installed the Nvidia Docker runtime so that the Docker container can use the GPU.</description>
- ▼<item>
 - <title>hosting my own web services</title>
 - <link><https://kylrth.com/post/self-hosting/></link>
 - <pubDate>Tue, 23 Nov 2021 07:57:00 -0500</pubDate>
 - <guid><https://kylrth.com/post/self-hosting/></guid>
 - <description>I host several services on an Alienware gaming computer I keep at my apartment. (We call it the spaceship.) I originally got the computer so I could have a computer with a GPU for machine learning projects, but I've since started using this computer to host a bunch of different services. Here I've documented how I set up the server. operating system To keep things simple I use Ubuntu 20.04 LTS.</description>
- ▼<item>
 - <title>moving to Québec</title>
 - <link><https://kylrth.com/post/qu%C3%A9bec/></link>
 - <pubDate>Fri, 29 Oct 2021 11:55:00 -0400</pubDate>
 - <guid><https://kylrth.com/post/qu%C3%A9bec/></guid>
 - <description>We just moved our family from Utah, USA, to Montréal, Québec, Canada. I entered Canada on August 18, 2021 by car, and my wife and daughter entered a few days later by air. The process actually began on April 27 when I got my acceptance letter to the Université de Montréal as a master's student in the Département

d’informatique et de recherche opérationnelle. After a few days of scrambling to find out if I would be able to study there without knowing French (turns out you can as a grad student at DIRO!</description>

</item>

▼<item>

<title>Jupyter Lab Hub in Docker with Nvidia GPU support</title>

<link><https://kylrth.com/post/jupyter-lab/></link>

<pubDate>Tue, 19 Oct 2021 20:28:25 -0400</pubDate>

<guid><https://kylrth.com/post/jupyter-lab/></guid>

<description>This is how I set up my headless home server with a Jupyter Lab Docker container with an Nvidia GPU runtime. Login is handled by a GitHub OAuth application. Nvidia drivers and the container runtime First, check here (replacing the CUDA version in the URL with your own) to see which Nvidia drivers you need for the CUDA toolkit version you want. I’m using CUDA 11.4.2, which means I need at least driver version 470.</description>

</item>

▼<item>

<title>Minecraft in Docker</title>

<link><https://kylrth.com/post/minecraft/></link>

<pubDate>Mon, 02 Aug 2021 13:25:28 -0600</pubDate>

<guid><https://kylrth.com/post/minecraft/></guid>

<description>This guide shows how to host multiple Minecraft servers on a single machine with docker-compose. mkdir minecraft_server cd minecraft_server mkdir data/ wget <https://kylrth.com/post/minecraft/docker-compose.yml> \ -O docker-compose.yml This docker-compose setup uses itzg’s Dockerimage, which you see further documentation for here. If you’re moving from a vanilla Minecraft world, do the following to get the different world directories in the right position: cp -r \${OLD}/world data/server/world mkdir data/server/world_{nether,the_end} mv data/server/world/DIM-1 data/server/world_nether/DIM-1 mv data/server/world/DIM1 data/server/world_the_end/DIM1 Here’s the map from vanilla Minecraft directories to Spigot directories (which is what itzg’s container uses):</description>

</item>

▼<item>

<title>Matrix setup with Synapse, Postgres, Maubot, and matrix-registration</title>

<link><https://kylrth.com/post/matrix-setup/></link>

<pubDate>Mon, 02 Aug 2021 10:30:00 -0600</pubDate>

<guid><https://kylrth.com/post/matrix-setup/></guid>

<description>This is how I set up my own Matrix server on a Raspberry Pi with Docker. Unfortunately, the Matrix community has stopped releasing ARM images, so the latest version that will work on ARM is v1.26.0. These instructions will work the same for x86_64 systems, except you’ll be able to use the default x86_64 images in the docker-compose file. This installation comes with Maubot and matrix-registration containers too. If you don’t want to use those features, leave out those sections of the docker-compose config and don’t follow the instructions in the corresponding sections.</description>

</item>

▼<item>

<title>I really just want to edit and compile my LaTeX files in VS Code</title>

<link><https://kylrth.com/post/latex-vscode/></link>

<pubDate>Mon, 12 Jul 2021 21:43:47 -0600</pubDate>

<guid><https://kylrth.com/post/latex-vscode/></guid>

<description>LaTeX has a ton of different flavors, releases, and installations: MacTeX, MiKTeX, TeXworks, XeTeX, pdfTeX, LuaTeX… If you’re using Linux and just want to edit LaTeX files in Visual Studio Code and have them automatically rendered as PDFs, follow these instructions: On Arch-based distros, install the packages listed here. On Debian-based systems, sudo apt install texlive. Install some Perl dependencies: sudo cpan Log::Log4perl Log::LogDispatch Log::Dispatch::File YAML::Tiny File::HomeDir If you want to use FontAwesome on Arch-based systems, install the oft-font-awesome package and then do the following (source):</description>

</description>


```
</item>
▼<item>
  <title>The last speakers: the quest to save the world's most endangered
  languages</title>
  <link>https://kylrth.com/book/last_speakers/</link>
  <pubDate>Mon, 31 May 2021 08:29:00 -0600</pubDate>
  <guid>https://kylrth.com/book/last_speakers/</guid>
  <description>This book argues that language loss is always bad, but that we can do
  something to save it. While the stories in the book leave me feeling like every
  language lost is a terrible cost, I think it's inevitable as our species merges
  into a global society due to technology. I think we ought to prioritize the proper
  treatment and respect of marginalized and alternative cultures, including their
  languages and how these cultures want to maintain them.</description>
</item>
▼<item>
  <title>The seven principles for making marriage work: a practical guide from the
  country's foremost relationship expert</title>
  <link>https://kylrth.com/book/seven-principles-for-marriage/</link>
  <pubDate>Wed, 26 May 2021 06:52:44 -0600</pubDate>
  <guid>https://kylrth.com/book/seven-principles-for-marriage/</guid>
  <description>Better communication doesn't really solve marriage problems. It
  has a low success rate, and that makes sense because there are plenty of marriages
  that yell and dispute. Disputation is not a sign of an unhealthy marriage.
  You'd have to be really magnanimous to take criticism about you, even if
  presented as softly as possible. Personality does not make a marriage incompatible.
  People can be friends but have very distinct personalities. Handle each other's
  strange side with caring and respect, as you would a friend.</description>
</item>
▼<item>
  <title>Harry Potter and the methods of rationality</title>
  <link>https://kylrth.com/book/hpmor/</link>
  <pubDate>Thu, 20 May 2021 07:25:00 -0600</pubDate>
  <guid>https://kylrth.com/book/hpmor/</guid>
  <description>Spoiler warning: no plot held back in this review. science is at least
  as beautiful as magic In chapter 7 Harry introduces Draco to the beauty of scientific
  advancement, and it actually moved me to tears. You should read the whole thing, but
  here are some of the best quotes: "Anyway," Harry said, "I'm
  saying that you don't seem to have been paying much attention to what goes on
  in the Muggle world.</description>
</item>
▼<item>
  <title>Planted: belief and belonging in an age of doubt</title>
  <link>https://kylrth.com/book/planted/</link>
  <pubDate>Sun, 18 Apr 2021 10:21:00 -0600</pubDate>
  <guid>https://kylrth.com/book/planted/</guid>
  <description>(My own thoughts appear as sidenotes or in italics, to distinguish from
  the author's thoughts.) Richard Bushman categorizes those who leave the church
  into two broad categories: those who feel "switched off", and those who
  feel "squeezed out". Mason summarizes the switched-off group as those who
  encounter troubling information about church history or doctrine, and as they
  discover more information they become jaded by it until they can no longer see the
  good the church does for them or for others.</description>
</item>
▼<item>
  <title>using GPG to prove you wrote your code</title>
  <link>https://kylrth.com/post/gpg/</link>
  <pubDate>Mon, 12 Apr 2021 07:13:54 -0600</pubDate>
  <guid>https://kylrth.com/post/gpg/</guid>
  <description>GPG is cool. You can use GPG to send encrypted messages, sign files to
  prove you generated them, and sign git commits to prove you committed them. You can
  get my key here. DigitalOcean has a neat guide to getting started with GPG. It
```

explains asymmetric encryption, key generation and revocation, and key signing and maintenance. Git commit authorship can be modified by anyone, as demonstrated by this tool. But by uploading your GPG public key to GitHub, you allow anyone who trusts GitHub to be sure that commits marked “verified” were actually created by you.</description>

</item>

▼<item>

<title>favorite art</title>

<link>https://kylrth.com/post/art/</link>

<pubDate>Sat, 03 Apr 2021 22:16:52 -0600</pubDate>

<guid>https://kylrth.com/post/art/</guid>

<description>Here’s some of my favorite art. Edvard Munch, The Scream, 1893 (source) Ben Shahn, All That Is Beautiful, 1966 (source) Peter Doig, Architect’s Home in the Ravine, 1991 (source)</description>

</item>

▼<item>

<title>The life-changing magic of tidying up: the Japanese art of decluttering and organizing</title>

<link>https://kylrth.com/book/life-changing-magic/</link>

<pubDate>Wed, 31 Mar 2021 21:21:04 -0700</pubDate>

<guid>https://kylrth.com/book/life-changing-magic/</guid>

<description>This book was my first real exposure to minimalism, and it completely changed how I feel about the possession of objects. It was super fortunate that my wife and I listened to it together on a road trip, and became equally enthralled with the idea of dumping all of our excess clutter. allat once We have excess clutter because of a fundamental problem with the way we deal with possessions.</description>

</item>

▼<item>

<title>The smartest kids in the world</title>

<link>https://kylrth.com/book/smartest-kids/</link>

<pubDate>Tue, 30 Mar 2021 07:00:00 -0700</pubDate>

<guid>https://kylrth.com/book/smartest-kids/</guid>

<description>The PISA test tests common senses reasoning. The countries that did best on the test were a surprise to everyone. Finland, South Korea, and Poland were all standouts in their own ways, and Ripley compares the policies and learning environments in these countries with those of the US to determine why the US is falling behind, especially in math and science. We talk a lot about parent involvement in the US, but the US actually has above average parental involvement.</description>

</item>

</item>

▼<item>

<title>The infinite Atonement</title>

<link>https://kylrth.com/book/infinite-atonement/</link>

<pubDate>Tue, 30 Mar 2021 06:30:00 -0700</pubDate>

<guid>https://kylrth.com/book/infinite-atonement/</guid>

<description>These notes are made while reading this with a Mormon theological background, so I skip noting some of the basic Mormon doctrines about the Atonement that he teaches. The Atonement is the central doctrine of Christianity. All scripture should be at least partially focused on it, and we’re invited to “speak of the atonement of Christ, and attain to a perfect knowledge of him” (Jacob 4:12). What is the significance of the Atonement?</description>

</item>

▼<item>

<title>How not to diet: the groundbreaking science of healthy, permanent weight loss</title>

<link>https://kylrth.com/book/how-not-to-diet/</link>

<pubDate>Wed, 03 Mar 2021 07:00:37 -0700</pubDate>

<guid>https://kylrth.com/book/how-not-to-diet/</guid>

<description>I read this book with Irresistible and the Social Dilemma on my mind, so I have a lot of notes here about addiction and big business. Just like everything else, capitalism has screwed over our diets by giving companies the incentive to put

shareholders above customers. Food companies employ lobbyists to keep subsidies on sugar/corn syrup/meat, and keep a stranglehold on public organizations. They buy billions of dollars of ads to communicate the message that it’s laziness that has caused the obesity epidemic and to push their products that appeal to the unconscious desires of our brains to produce artificial hunger.</description>

</item>

▼<item>

<title>The gene: an intimate history</title>

<link><https://kylrth.com/book/the-gene/></link>

<pubDate>Fri, 22 Jan 2021 06:14:58 -0700</pubDate>

<guid><https://kylrth.com/book/the-gene/></guid>

<description>These are notes I made after finishing the book, so they’ll be more heavily weighted toward concepts discussed near the end. The first half of the book was primarily dedicated to a history of genetic research, which I think helped the reader understand the issues discussed in the latter half. playing God It seems like our identity derives from a complicated combination of genes and chance environmental effects. Part of our strength as a species has been our natural variation, and to begin editing the genome is to assume that we can do it better than evolution has done up until this point.</description>

</item>

▼<item>

<title>Faith is not blind</title>

<link><https://kylrth.com/book/faith-is-not-blind/></link>

<pubDate>Mon, 18 Jan 2021 08:11:23 -0700</pubDate>

<guid><https://kylrth.com/book/faith-is-not-blind/></guid>

<description>Elder Hafen struggled as a missionary with the concept of knowing versus believing: he felt he believed it was true, but not that he knew it. On the mission he felt pressure to bear testimony with the word “know”, but he chafed at that. In this book, Elder Hafen hopes to discuss the complex boundaries between believing and knowing, Richard Bushman, a prominent LDS historian, found himself in a similar situation.</description>

</item>

▼<item>

<title>Cross-lingual alignment of contextual word embeddings, with applications to zero-shot dependency parsing</title>

<link><https://kylrth.com/paper/cross-lingual-alignment-contextual/></link>

<pubDate>Fri, 11 Dec 2020 06:30:43 -0700</pubDate>

<guid><https://kylrth.com/paper/cross-lingual-alignment-contextual/></guid>

<description>Recent contextual word embeddings (e.g. ELMo) have shown to be much better than “static” embeddings (where there’s a one-to-one mapping from token to representation). This paper is exciting because they were able to create a multi-lingual embedding space that used contextual word embeddings. Each token will have a “point cloud” of embedding values, one point for each context containing the token. They define the embedding anchor as the average of all those points for a particular token.</description>

</item>

▼<item>

<title>Inductive biases for deep learning of higher-level cognition</title>

<link><https://kylrth.com/paper/inductive-biases-higher-cognition/></link>

<pubDate>Tue, 08 Dec 2020 06:40:48 -0700</pubDate>

<guid><https://kylrth.com/paper/inductive-biases-higher-cognition/></guid>

<description>This is a long paper, so a lot of my writing here is an attempt to condense the discussion. I’ve taken the liberty to pull exact phrases and structure from the paper without explicitly using quotes. Our main hypothesis is that deep learning succeeded in part because of a set of inductive biases, but that additional ones should be added in order to go from good in-distribution generalization in highly supervised learning tasks (or where strong and dense rewards are available), such as object recognition in images, to strong out-of-distribution generalization and transfer learning to new tasks with low sample complexity.

</description>

</item>

- ▼<item>
 - <title>SpanBERT: improving pre-training by representing and predicting spans</title>
 - <link><https://kylrth.com/paper/spanbert/></link>
 - <pubDate>Sat, 05 Dec 2020 16:08:03 -0700</pubDate>
 - <guid><https://kylrth.com/paper/spanbert/></guid>
 - <description>BERT optimizes the Masked Language Model (MLM) objective by masking word pieces uniformly at random in its training data and attempting to predict the masked values. With SpanBERT, spans of tokens are masked and the model is expected to predict the text in the spans from the representations of the words on the boundary. Span lengths follow a geometric distribution, and span start points are uniformly random. To predict each individual masked token, a two-layer feedforward network was provided with the boundary token representations plus the position embedding of the target token, and the output vector representation was used to predict the masked token and compute cross-entropy loss exactly as in standard MLM.</description></item>
- ▼<item>
 - <title>Tools and weapons: the promise and peril of the digital age</title>
 - <link><https://kylrth.com/book/tools-and-weapons/></link>
 - <pubDate>Thu, 03 Dec 2020 20:58:02 -0700</pubDate>
 - <guid><https://kylrth.com/book/tools-and-weapons/></guid>
 - <description>I started taking notes later in the book. There were lots of good insights in the first half. Sorry! broadband access Getting the internet to rural communities is a big deal for the rural economy. Just like electricity, it’s something that needs government support because there isn’t the economic incentive for ISPs to reach some of these locations.ethical AI The focus on AI now is not just a fad, but a convergence of several trends that have made AI the next logical step: the increased computational resources, flexible access to compute through the cloud, etc.</description></item>
- ▼<item>
 - <title>Deep contextualized word representations</title>
 - <link><https://kylrth.com/paper/deep-contextualized-word-representations/></link>
 - <pubDate>Thu, 03 Dec 2020 12:01:43 -0700</pubDate>
 - <guid><https://kylrth.com/paper/deep-contextualized-word-representations/></guid>
 - <description>This is the original paper introducing Embeddings from Language Models (ELMo). Unlike most widely used word embeddings, ELMo word representations are functions of the entire input sentence. That’s what makes ELMo great: they’re contextualized word representations, meaning that they can express multiple possible senses of the same word. Specifically, ELMo representations are a learned linear combination of all layers of an LSTM encoding. The LSTM undergoes general semi-supervised pretraining, but the linear combination is learned specific to the task.</description></item>
- ▼<item>
 - <title>Removing a keyword from git history</title>
 - <link><https://kylrth.com/post/removing-keyword-from-git-history/></link>
 - <pubDate>Wed, 02 Dec 2020 11:34:25 -0700</pubDate>
 - <guid><https://kylrth.com/post/removing-keyword-from-git-history/></guid>
 - <description>I recently had to remove a keyword from the git history of a project I was working on. This meant not just removing a file but modifying commits where the keyword was added, commits where the keyword was removed, and even commits with the keyword in the commit message. I eventually came to the right solution through a mix of blog posts and the documentation for git rebase. For this example, assume the keyword is “matrix”.</description></item>
- ▼<item>
 - <title>Blink: the power of thinking without thinking</title>
 - <link><https://kylrth.com/book/blink/></link>
 - <pubDate>Tue, 17 Nov 2020 20:44:48 -0700</pubDate>
 - <guid><https://kylrth.com/book/blink/></guid>

<description>Our subconscious not only manages bodily systems but also performs processing of features in our experience that our conscious does not have time to process. This has been proven in lots of experiments where people have been given subconscious cues to help them solve problems, but the people are unaware of this and make up answers when asked to explain how they came to conclusions. It's important to trust these judgments that seem to come out of nowhere, but if we try to explain them we'll start trying to provide rational answers, which can be totally false or misleading.**</description>**

</item>

▼**<item>**

<title>A short history of nearly everything**</title>**

<link><https://kylrth.com/book/short-history-nearly-everything/>**</link>**

<pubDate>Wed, 07 Oct 2020 11:19:03 -0600**</pubDate>**

<guid><https://kylrth.com/book/short-history-nearly-everything/>**</guid>**

<description>We are extremely lucky to be here, and even more lucky to be able to appreciate it. Let's not waste it.**</description>**

</item>

▼**<item>**

<title>Overcoming catastrophic forgetting in neural networks**</title>**

<link><https://kylrth.com/paper/overcoming-catastrophic-forgetting/>**</link>**

<pubDate>Thu, 01 Oct 2020 10:47:28 -0600**</pubDate>**

<guid><https://kylrth.com/paper/overcoming-catastrophic-forgetting/>**</guid>**

<description>In the paper they use Bayes's rule to show that the contribution of the first of two tasks is contained in the posterior distribution of model parameters over the first dataset. This is important because it means we can estimate that posterior to try to get a sense for which model parameters were most important for that first task. In this paper, they perform that estimation using a multivariate Gaussian distribution.**</description>**

</item>

▼**<item>**

<title>Learning neural causal models from unknown interventions**</title>**

<link><https://kylrth.com/paper/neural-causal-models/>**</link>**

<pubDate>Tue, 22 Sep 2020 10:39:54 -0600**</pubDate>**

<guid><https://kylrth.com/paper/neural-causal-models/>**</guid>**

<description>This is a follow-on to A meta-transfer objective for learning to disentangle causal mechanisms Here we describe an algorithm for predicting the causal graph structure of a set of visible random variables, each possibly causally dependent on any of the other variables. the algorithm There are two sets of parameters, the structural parameters and the functional parameters. The structural parameters compose a matrix where (σ_{ij}) represents the belief that variable (X_j) is a direct cause of (X_i) .**</description>**

</item>

▼**<item>**

<title>A meta-transfer objective for learning to disentangle causal mechanisms**</title>**

<link><https://kylrth.com/paper/meta-transfer-objective-for-causal-mechanisms/>**</link>**

<pubDate>Mon, 21 Sep 2020 08:46:30 -0600**</pubDate>**

<guid><https://kylrth.com/paper/meta-transfer-objective-for-causal-mechanisms/>**</guid>**

<description>Theoretically, models should be able to predict on out-of-distribution data if their understanding of causal relationships is correct. The toy problem they use in this paper is that of predicting temperature from altitude. If a model is trained on data from Switzerland, the model should ideally be able to correctly predict on data from the Netherlands, even though it hasn't seen elevations that low before. The main contribution of this paper is that they've found that models tend to transfer faster to a new distribution when they learn the correct causal relationships, and when those relationships are sparsely represented, meaning they are represented by relatively few nodes in the network.**</description>**

</item>

▼**<item>**

<title>Deep learning generalizes because the parameter-function map is biased towards simple functions**</title>**

<https://kylrth.com/paper/parameter-function-map-biased-to-simple/>
<pubDate>Tue, 08 Sep 2020 07:29:09 -0600</pubDate>
<guid>https://kylrth.com/paper/parameter-function-map-biased-to-simple/</guid>
<description>The theoretical value in talking about the parameter-function map is that this map lets us talk about sets of parameters that produce the same function. In this paper they used some recently proven stuff from algorithmic information theory (AIT) to show that for neural networks the parameter-function map is biased toward functions with low Kolmogorov complexity, meaning that simple functions are more likely to appear given random choice of parameters. Since real world problems are also biased toward simple functions, this could explain the generalization/memorization results found by Zhang et al.</description>

</item>

▼<item>

<title>The moment of lift: how empowering women changes the world</title>
<link>https://kylrth.com/book/moment-of-lift/</link>
<pubDate>Tue, 01 Sep 2020 05:25:38 -0600</pubDate>
<guid>https://kylrth.com/book/moment-of-lift/</guid>
<description>This book is about empowering women by giving them the freedom to make their own choices and speak for themselves. She said some important things about stigma in society. She talked specifically about the stigma of not talking about birth control, but she made general statements too. It's each person's responsibility to work against stigma and stop the human tendency to cast out others. I need to spend more time thinking about my own stigmas and biases, so that I can help those who are marginalized.</description>

</item>

▼<item>

<title>A closer look at memorization in deep networks</title>
<link>https://kylrth.com/paper/closer-look-at-memorization/</link>
<pubDate>Mon, 31 Aug 2020 11:52:35 -0600</pubDate>
<guid>https://kylrth.com/paper/closer-look-at-memorization/</guid>
<description>This paper builds on what we learned in "Understanding deep learning requires rethinking generalization". In that paper they showed that DNNs are able to fit pure noise in the same amount of time as it can fit real data, which means that our optimization algorithm (SGD, Adam, etc.) is not what's keeping DNNs from overfitting. experiments for detecting easy/hard samples It looks like there are qualitative differences between a DNN that has memorized some data and a DNN that has seen real data.</description>

</item>

▼<item>

<title>Naked economics: undressing the dismal science</title>
<link>https://kylrth.com/book/naked-economics/</link>
<pubDate>Sun, 30 Aug 2020 06:46:49 -0600</pubDate>
<guid>https://kylrth.com/book/naked-economics/</guid>
<description>An important question is how much we need to fight income inequality. Is it fair to have 35% growth in the upper class and 3% growth in the lower class? Where is a good balance? We have grown a lot richer since the Industrial Revolution, because we've become more productive. Wealth is not a zero-sum game. Globalization is good because it allows us to buy cheaper, better products. We can offset short-run job loss by paying or giving human capital to those who lose their jobs to globalization Policies often don't do what we intend them to do, because they change people's decisions for the involved choice.</description>

</item>

▼<item>

<title>The faith of a scientist</title>
<link>https://kylrth.com/book/faith-of-a-scientist/</link>
<pubDate>Sun, 30 Aug 2020 06:46:49 -0600</pubDate>
<guid>https://kylrth.com/book/faith-of-a-scientist/</guid>
<description>Scientific thinking and religion go hand in hand, and help refine and give purpose to each other. Descartes' approach wasn't as good as Newton's. Descartes relied on the soundness of his own reasoning. "The erroneous conception that revelation ended with the apostles promotes the

misconception among sectarian religions that the Gospel is complete and that with a liberal admixture of human wisdom, all will be crystal clear.” God places messages in everything.</description>

</item>

▼<item>

<title>Weapons of math destruction: how big data increases inequality and threatens democracy</title>

<link><https://kylrth.com/book/weapons-of-math-destruction/></link>

<pubDate>Sun, 30 Aug 2020 06:46:49 -0600</pubDate>

<guid><https://kylrth.com/book/weapons-of-math-destruction/></guid>

<description>In fact, I saw all kinds of parallels between finance and Big Data. Both industries gobble up the same pool of talent, much of it from elite universities like MIT, Princeton, or Stanford. These new hires are ravenous for success and have been focused on external metrics–like SAT scores and college admissions–their entire lives. Whether in finance or tech, the message they’ve received is that they will be rich, that they will run the world.</description>

</item>

▼<item>

<title>A disciplined approach to neural network hyperparameters: part 1</title>

<link><https://kylrth.com/paper/disciplined-approach-to-hyperparameters/></link>

<pubDate>Fri, 28 Aug 2020 14:16:29 -0600</pubDate>

<guid><https://kylrth.com/paper/disciplined-approach-to-hyperparameters/></guid>

<description>The goal of hyperparameter tuning is to reach the point where test loss is horizontal on the graph over model complexity. Underfitting can be observed with a small learning rate, simple architecture, or complex data distribution. You can observe underfitting decrease by seeing more drastic results at the outset, followed by a more horizontal line further into training. You can use the LR range test to find a good learning rate range, and then use a cyclical learning rate to move up and down within that range.</description>

</item>

▼<item>

<title>Forward and reverse gradient-based hyperparameter optimization</title>

<link><https://kylrth.com/paper/gradient-based-hyperparameter-optimization/></link>

<pubDate>Fri, 28 Aug 2020 14:16:29 -0600</pubDate>

<guid><https://kylrth.com/paper/gradient-based-hyperparameter-optimization/></guid>

<description>In the area of hyperparameter optimization (HO), the goal is to optimize a response function of the hyperparameters. The response function is usually the average loss on a validation set. Gradient-based HO refers to iteratively finding the optimal hyperparameters using gradient updates, just as we do with neural network training itself. The gradient of the response function with respect to the hyperparameters is called the hypergradient. One of the great things about this work is that their framework allows for all kinds of hyperparameters.</description>

</item>

▼<item>

<title>Understanding deep learning requires rethinking generalization</title>

<link><https://kylrth.com/paper/understanding-requires-rethinking-generalization/></link>

<pubDate>Wed, 26 Aug 2020 08:42:58 -0600</pubDate>

<guid><https://kylrth.com/paper/understanding-requires-rethinking-generalization/></guid>

<description>It turns out that neural networks can reach training loss of 0 even on randomly labeled data, even when the data itself is random. It was previously thought that some implicit bias in the model architecture prevented (or regularized the model away from) overfitting to specific training examples, but that’s obviously not true. They showed this empirically as just described, and also theoretically constructed a two-layer ReLU network with $(p=2n+d)$ parameters to express any labeling of any sample of size (n) in (d) dimensions.</description>

</item>

▼<item>

<title>Why does unsupervised pre-training help deep learning?</title>

<link><https://kylrth.com/paper/why-unsupervised-helps/></link>

<pubDate>Mon, 24 Aug 2020 11:40:00 -0600</pubDate>
<guid>https://kylrth.com/paper/why-unsupervised-helps/</guid>
<description>They’re pretty sure that it performs regularization by starting off the supervised training in a good spot, instead of by somehow improving the optimization path.</description>
</item>

▼<item>
<title>Essentialism</title>
<link>https://kylrth.com/book/essentialism/</link>
<pubDate>Sun, 23 Aug 2020 07:47:12 -0600</pubDate>
<guid>https://kylrth.com/book/essentialism/</guid>
<description>The main character of the first story slowly changed his attitude toward demands on his resources. “Can I actually fulfill this request, given the time and resources I have?” “Is this the very most important thing I should be doing with my time and resources right now?” “Just because I was invited didn’t seem a good enough reason to attend.” “It’s important to pursue “less but better” in a disciplined way.</description>
</item>

▼<item>
<title>The consciousness prior</title>
<link>https://kylrth.com/paper/consciousness-prior/</link>
<pubDate>Fri, 14 Aug 2020 09:05:56 -0700</pubDate>
<guid>https://kylrth.com/paper/consciousness-prior/</guid>
<description>System 1 cognitive abilities are about low-level perception and intuitive knowledge. System 2 cognitive abilities can be described verbally, and include things like reasoning, planning, and imagination. In cognitive neuroscience, the “Global Workspace Theory” says that at each moment specific pieces of information become a part of working memory and become globally available to other unconscious computational processes. Relative to the unconscious state, the conscious state is low-dimensional, focusing on a few things.</description>
</item>

▼<item>
<title>Troubling trends in machine learning scholarship</title>
<link>https://kylrth.com/paper/troubling-trends-in-ml/</link>
<pubDate>Thu, 13 Aug 2020 10:36:05 -0700</pubDate>
<guid>https://kylrth.com/paper/troubling-trends-in-ml/</guid>
<description>The authors discuss four trends in AI research that have negative consequences for the community. problems explanation vs. speculation It’s important to allow researchers to include speculation, because speculation is what allows ideas to form. But the paper has to carefully couch speculation inside a “Motivations” section or other verbage to ensure the reader understands its place. It’s extremely important to define concepts before using them. Terms like internal covariate shift or coverage sound like definitions without actually being such.</description>
</item>

▼<item>
<title>Attention is all you need</title>
<link>https://kylrth.com/paper/attention-all-you-need/</link>
<pubDate>Wed, 05 Aug 2020 12:37:42 -0700</pubDate>
<guid>https://kylrth.com/paper/attention-all-you-need/</guid>
<description>I also referred to this implementation to understand some of the details. This is the paper describing the Transformer, a sequence-to-sequence model based entirely on attention. I think it’s best described with pictures. model overview From this picture, I think the following things need explaining: embeddings these are learned embeddings that convert the input and output tokens to vectors of the model dimension. In this paper, they actually used the same weight matrix for input embedding, output embedding, and the final linear layer before the final softmax.</description>
</item>

▼<item>

`<title>BERT: pre-training of deep bidirectional transformers for language understanding</title>`

`<link>https://kylrth.com/paper/bert/</link>`

`<pubDate>Tue, 04 Aug 2020 08:57:44 -0700</pubDate>`

`<guid>https://kylrth.com/paper/bert/</guid>`

`<description>The B is for bidirectional, and that's a big deal. It makes it possible to do well on sentence-level (NLI, question answering) and token-level tasks (NER, POS tagging). In a unidirectional model, the word "bank" in a sentence like "I made a bank deposit." has only "I made a" as its context, keeping useful information from the model. Another cool thing is masked language model training (MLM). They train the model by blanking certain words in the sentence and asking the model to guess the missing word.</description>`

`</item>`

▼`<item>`

`<title>Compositional generalization by factorizing alignment and translation</title>`

`<link>https://kylrth.com/paper/factorizing-alignment-and-translation/</link>`

`<pubDate>Mon, 27 Jul 2020 09:11:16 -0700</pubDate>`

`<guid>https://kylrth.com/paper/factorizing-alignment-and-translation/</guid>`

`<description>They had a biRNN with attention for alignment encoding, and then a single linear function of each one-hot encoded word for encoding that single word. Their reasoning was that by separating the alignment from the meaning of individual words the model could more easily generalize to unseen words.</description>`

`</item>`

▼`<item>`

`<title>Semi-supervised training for automatic speech recognition</title>`

`<link>https://kylrth.com/paper/semi-supervised-for-asr/</link>`

`<pubDate>Tue, 14 Jul 2020 08:06:00 -0600</pubDate>`

`<guid>https://kylrth.com/paper/semi-supervised-for-asr/</guid>`

`<description>This was Manohar's PhD dissertation at JHU. Chapter 2 provides a relatively clear overview of how chain and non-chain models work in Kaldi. In chapter 3 he tried using negative conditional entropy as the loss function for the unsupervised data, and it helped a bit. In chapter 4 Manohar uses [CTC loss]/paper/ctc/. In chapter 5, he discusses ways to do semi-supervised model training. It's nice when you have parallel data in different domains, because then you can do a student-teacher model.</description>`

`</item>`

▼`<item>`

`<title>Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks</title>`

`<link>https://kylrth.com/paper/ctc/</link>`

`<pubDate>Fri, 10 Jul 2020 09:14:59 -0600</pubDate>`

`<guid>https://kylrth.com/paper/ctc/</guid>`

`<description>RNNs generally require pre-segmented training data, but this avoids that need. Basically, you have the RNN output probabilities for each label (or a blank) for every frame, and then you find the most likely path across that lattice of probabilities. The section explaining the loss function was kind of complicated. They used their forward-backward algorithm (sort of like Viterbi) to get the probability of all paths corresponding to the output that go through each symbol at each time, and then they differentiated that to get the derivatives with respect to the outputs.</description>`

`</item>`

▼`<item>`

`<title>Google's neural machine translation system: bridging the gap between human and machine translation</title>`

`<link>https://kylrth.com/paper/google-nmt-2016/</link>`

`<pubDate>Tue, 30 Jun 2020 08:22:30 -0600</pubDate>`

`<guid>https://kylrth.com/paper/google-nmt-2016/</guid>`

`<description>This model was superseded by this one. They did some careful things with residual connections to make sure it was very parallelizable. They put each LSTM layer on a separate GPU. They quantized the models such that they could train using`

full floating-point computations with a couple restrictions and then convert the models to quantized versions.</description>

</item>

▼<item>

<title>Google's multilingual neural machine translation system</title>

<link><https://kylrth.com/paper/google-zero-shot/></link>

<pubDate>Fri, 26 Jun 2020 08:02:12 -0600</pubDate>

<guid><https://kylrth.com/paper/google-zero-shot/></guid>

<description>They use the word-piece model from “Japanese and Korean Voice Search”, with 32,000 word pieces. (This is a lot less than the 200,000 used in that paper.) They state in the paper that the shared word-piece model is very similar to Byte-Pair-Encoding, which was used for NMT in this paper by researchers at U of Edinburgh. The model and training process are exactly as in Google’s earlier paper. It takes 3 weeks on 100 GPUs to train, even after increasing batch size and learning rate.</description>

</item>

▼<item>

<title>Japanese and Korean voice search</title>

<link><https://kylrth.com/paper/word-piece-model/></link>

<pubDate>Wed, 24 Jun 2020 14:44:02 -0600</pubDate>

<guid><https://kylrth.com/paper/word-piece-model/></guid>

<description>This was mentioned in the paper on Google’s Multilingual Neural Machine Translation System. It’s regarded as the original paper to use the word-piece model, which is the focus of my notes here. the WordPieceModel Here’s the WordPieceModel algorithm: func WordPieceModel(D, chars, n, threshold) -> inventory: # D: training data # n: user-specified number of word units (often 200k) # chars: unicode characters used in the language (e.g. Kanji, Hiragana, Katakana, ASCII for Japanese) # threshold: stopping criterion for likelihood increase # inventory: the set of word units created by the model inventory := chars likelihood := +INF while len(inventory) < n && likelihood >= threshold: lm := LM(inventory, D) inventory += argmax_{combined word unit}(lm.</description>

</item>

▼<item>

<title>Towards a multi-view language representation</title>

<link><https://kylrth.com/paper/multi-view-language-representation/></link>

<pubDate>Tue, 23 Jun 2020 08:40:04 -0600</pubDate>

<guid><https://kylrth.com/paper/multi-view-language-representation/></guid>

<description>They used a technique called CCA to combine hand-made features with NN representations. It didn’t do great on typological feature prediction, but it did do well with predicting a phylogenetic tree for Indo-European languages.</description>

</item>

▼<item>

<title>Universal phone recognition with a multilingual allophone system</title>

<link><https://kylrth.com/paper/universal-phone-recognition/></link>

<pubDate>Tue, 23 Jun 2020 08:33:48 -0600</pubDate>

<guid><https://kylrth.com/paper/universal-phone-recognition/></guid>

<description>These guys made sure to model allophones. They had an encoder that produced a universal phone set, and then language-specific decoders. This meant they could use data from various languages to train the system. The decoder has an allophone layer, followed by other dense trainable layers. The allophone layer is a single trainable dense layer, but was initialized by a bunch of linguists who sat down and described the phone sets belonging to each phoneme in each language present in the training set.</description>

</item>

▼<item>

<title>using Matrix</title>

<link><https://kylrth.com/post/matrix-registration/></link>

<pubDate>Sun, 02 Feb 2020 12:34:56 -0700</pubDate>

<guid><https://kylrth.com/post/matrix-registration/></guid>

`<description>Matrix is a federated, open source chat system. By federated, we mean that people can communicate across different servers, like in the image below. In that way, it works sort of like email: even though you may use you@gmail.com and I might use me@kylrth.com, we can still write each other emails. In our case, I host the server at matrix.kylrth.com, and you and I can connect to it with various clients.</description>`

`</item>`

`▼<item>`

`<title>Increase in learning: spiritual patterns for obtaining your own answers</title>`

`<link>https://kylrth.com/book/increase-in-learning/</link>`

`<pubDate>Wed, 20 Nov 2019 10:14:43 -0700</pubDate>`

`<guid>https://kylrth.com/book/increase-in-learning/</guid>`

`<description>Chapter 1 We are give the opportunity to have the Spirit as a constant companion! To take advantage, we need to sincerely desire it, invite it through action, and be worthy of it through obedience. Chapter 2 Knowledge is the accumulation of facts. Understanding comes when we apply our hearts to knowledge, which lets the Holy Ghost testify to us of the truthfulness of it. Understanding comes by revelation. “Intelligence is the righteous application of knowledge and understanding in action and judgment.</description>`

`</item>`

`</channel>`

`</rss>`