**Proposal:**

• What problem you’re planning to solve

• How you’re planning to evaluate / analyze your solution

• What dataset you’re going to use

• What computing resources you need to train your model

Resnet Satori:

<https://satori-portal.mit.edu:8443/auth/realms/satori-portal-dev/protocol/openid-connect/auth?response_type=code&scope=openid&client_id=satori-portal.mit.edu&state=TXqmsHniaG-qfaq6o20yWfxBWvg&redirect_uri=https%3A%2F%2Fsatori-portal.mit.edu%2Foidc&nonce=i1Ok96MjFqYeRNOR0EpVafaSoZst7-tTUuYrXdfdQ2s>

Sample project ideas: <https://canvas.mit.edu/courses/7503/modules/items/316174>

**SOME IDEAS**

1. Handwriting to text (Hebrew).
2. Sentiment analysis (Hebrew).
   1. Write about the weirdness of Hebrew, how words have gender, how some words have multiple meanings, etc.
   2. <https://github.com/NLPH/NLPH_Resources>
   3. <https://github.com/galhev/Neural-Sentiment-Analyzer-for-Modern-Hebrew>
   4. https://github.com/avichaychriqui/HeBERT
3. Paper summarization for academic papers.
   1. <https://github.com/jananiarunachalam/Research-Paper-Summarization>
4. Translating biblical hebrew to modern hebrew
   1. <http://www.yeshua.co.il/bibleapp/?book=1>
   2. Shakespear English to Modern English can be a good   
      <https://github.com/tokestermw/tensorflow-shakespeare>   
      <https://github.com/ashleynicole472/Shakespeare-Translator>  
      https://github.com/parthpatel20010/Shakespeare-to-Modern-English-Translator
   3. <https://www.haaretz.co.il/misc/1.1347443> our guy
5. Add nekud to existing text
   1. Remove nekud from text to create a dataset then create a model to try to add the nekud

Problem: Translate biblical hebrew to modern hebrew.

1. The data will be line by line translations from here: <http://www.yeshua.co.il/bibleapp/index.php>
2. Concern: Do we have enough computing resources for this problem?
3. Model seq2seq

<https://www.chabad.org/library/bible_cdo/aid/8165#lt=both>

----------------------------------------------------------------------------------------------------------------------

Remove nekud from text: https://stackoverflow.com/questions/50828136/remove-hebrew-vowels-nikkud-from-selected-unicode-hebrew-tex

Proposal:

**Ancient Hebrew to Modern Hebrew Translator**

Doron Hazan (doronh), Zachary Metzman (zmetzman), Dana Rosenfarb (danarose), Yorai Shaoul (yorai)

We propose creating a machine translation model between ancient Hebrew and modern Hebrew. We motivate our choice by recalling how, as younger students who were required to study the Hebrew Bible in ancient Hebrew, we faced difficulties in understanding the written language. These hardships may have arisen because Biblical Hebrew tends to be more formal than its contemporary counterpart and also because its structure tends to differ from modern Hebrew. To the untrained reader, Biblical Hebrew seems to introduce ideas in an unexpected manner. To the benefit of current students, historians, and researchers, we wish to make ancient languages more approachable to modern-day readers.

We propose designing and implementing a machine translation module that would map ancient-language texts to more easily understood corpora. To this end, we choose to train a deep-learning sequential model on excerpts from the Old Testament for producing modern Hebrew translations of the input text. We choose the Hebrew Bible as our training corpus as it is a widely available source of ancient text. Our main goal is to implement and train a sequence-to-sequence architecture to perform machine translation. We plan on comparing the efficacy of our model to word-to-word translation modules (that are local in nature and do not take sentence structure into account) and also to other methods.

Finally, we note that by "translating" ancient Hebrew texts to modern Hebrew, we gain the option of translating them to any other modern language by exploiting other translation tools. The difficulty, in our eyes, lies in the challenge of translating ancient Hebrew to a modern language, and not necessarily to modern Hebrew.

We plan to do the following:

* Generate an ancient Hebrew to modern Hebrew sentence dataset. Mapping sentences between language versions.
* Implement a sequence-to-sequence model and train it on our dataset.
* Evaluate our trained model on language of similar nature (Biblical excerpts not seen during training).
* Evaluate our model on modern-day Hebrew - results should remain understandable to modern readers with similar contents to the inputs.
* Evaluate other methods for this translation task (word-to-word, HMM).
* Potentially try ancient-Hebrew to English with the same architecture, if time allows.

-------------------------------------------------------------------------------------------------------------------------------

Tentative division of work:

1. Data processing + "other method" #1.
2. Creation of Seq-to-seq framework and training.  
   <https://github.com/Developer-Zer0/Get-To-The-Point-Summarization-with-Pointer-Generator-Networks>
3. "Other method" #2 and evaluation on modern-day Hebrew.
4. Creation of Seq-to-seq #2 (second seq-to-seq) model and training. (This might not be done)(work allocation is subject to change :) ).