

An aerial photograph of Singapore at dusk. The sun is low on the left, casting a warm glow over the city. The Marina Bay Sands hotel is prominent on the left, with its three towers and skybridge. The Esplanade - Theatres on the Bay is in the foreground, showing its two large, dome-shaped structures. The city skyline is filled with illuminated skyscrapers, and the water reflects the lights. The text "SGnlp Workshop 2022" is overlaid in white on the right side of the image.

SGnlp Workshop 2022

Agenda

- ❖ Introduction to SGnlp
- ❖ Why SGnlp?
- ❖ Use Cases
- ❖ Contribution guide
- ❖ Hands-on session
- ❖ Q&A

An aerial photograph of Singapore's skyline at sunset. The image shows the Marina Bay Sands hotel on the left, the Esplanade - Theatres on the Bay in the foreground, and the dense financial district with numerous skyscrapers on the right. The sun is low on the horizon, casting a warm glow over the city and reflecting on the water in the bay.

Introduction to SGnlp

Overview of AI Singapore Activities



AI RESEARCH

Invent next generation AI techniques/ algorithms



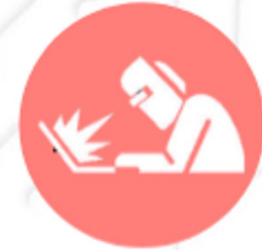
AI TECHNOLOGY

Create significant economic and social impact through solving national challenges



AI INNOVATION

Catalysing AI Innovation in the Industry through projects and talents pipeline



AI MAKERSPACE

SME and Start-ups goAI through platform and suite of AI bricks, open datasets, training, and consulting services



AI SINGAPORE

SGnlp is an initiative by AI Singapore's NLP-Hub

Our Goals

- To bridge the gap between research and industry
- To promote translational research
- To encourage adoption of NLP techniques by the industry

Key Features

SGnlp

By AI Singapore

AI SINGAPORE

Docs

Community

Contributing

🗨️ Conversational Tools

✍️ Grammar Correction

🔍 Knowledge Mining

🐦 Social Media Analysis

📄 Document Analysis

😊 Sentiment Analysis

Welcome to the SGnlp Demo page!



SGnlp is an initiative by AI Singapore's NLP Hub. We aim to bridge the gap between research and industry, promote translational research, and encourage adoption of NLP techniques in the industry.

This website hosts various models developed by Singapore's NLP research community. You can navigate the menu on the left to find the different models available.

User statistics

Total users	2,637	
Views	10,639	

Demo website traffic between
27 Aug 2021 - 23 Sep 2022



An aerial photograph of Singapore at sunset. The image shows the city's skyline with numerous skyscrapers, including the Marina Bay Sands hotel on the left. The sun is low on the horizon, creating a warm, orange and yellow glow. The water of the bay is visible, reflecting the city lights and the sunset. The text "Why SGnlp?" is overlaid in the center in a large, white, sans-serif font.

Why SGnlp?

SGnlp

A consistent &
maintainable
codebase

Converts codes written in **different** styles and structures for new models **towards** a more **consistent & maintainable** style for industry and NLP practitioners **to use**.

SGnlp

Covers a wide range
of NLP models

- **Conversational Tools**
 - ☐ Emotion Entailment
 - ☐ Causal Span Extraction
 - ☐ Identify Follow up question
- **Grammar Correction**
 - ☐ Grammatical Error Correction
- **Knowledge Mining**
 - ☐ Relation Extraction
- **Social Media Analysis**
 - ☐ Rumor Detection
- **Document Analysis**
 - ☐ Discourse Parsing
- **Sentiment Analysis**
 - ☐ Aspect-Based Sentiment Analysis
 - ☐ Cross Lingual Cross Domain

SGnlp

User-friendly

Installation

```
pip install sgnlp
```

Prediction with pretrained model

```
from sgnlp.models.ufd import UFDModelBuilder, UFDPreprocessor

# Instantiate model builder and preprocessor
model_builder = UFDModelBuilder(
    source_domains=['books'],
    target_languages=['de'],
    target_domains=['dvd'])
preprocessor = UFDPreprocessor()

# Build pretrained model groups
model_groups = model_builder.build_model_group()

# Model predict ('books_de_dvd' model example)
instance = """Wolverine is BACK Der Film ist im Grunde wie alle Teile der X-Men f
              Hugh Jackman spielt seine Rolle wie immer so gut was ich von den ei
              sagen kann. Story und Action sind aber genug Gründe um sich die Blu

instance_features = preprocessor([instance])
output = model_groups['books_de_dvd'](**instance_features)
```

An aerial photograph of Singapore's skyline at sunset. The image shows the Marina Bay Sands hotel on the left, the Esplanade - Theatres on the Bay in the foreground, and a dense cluster of skyscrapers in the background. The sky is a mix of orange, pink, and blue, and the water reflects the city lights and the sky. The text "Use Cases" is centered over the image in a large, white, sans-serif font.

Use Cases

Fake News Detector



- **Goal:** Apply the algorithms to determine objectivity and truthfulness of posts on web
- **How:** Segment texts into elementary components according to basic segmentation principles,
- **Application:** Rhetorical Structure Theory (RST) as the first cut to various discourse components, allowing human to review in a more effective and efficient way

Contract Analytics



- **Goal:** Support the comprehensive contract analytics.
- **How:** Segment texts into elementary components according to basic segmentation principles,
- **Various Applications:** Automated Clause Analysis, Intelligent Contract Comparisons, Document Classification and Lineage etc.

Optimising Customer Service



- **Goal:** Support customer support services
- **How:** Process conversation threads and reviews
- **Applications:** Automate recognizing the cause behind emotions in conversations/text, aspect based sentiment analysis, etc.

Conversational Tools > Emotion Entailment

Example out of SG-NLP Demo:

Utterances

David , why didn ' t you clean the room ?

I ' m not in the mood .

Why are you feeling depressed ?

I was told my girlfriend was speaking ill of me .

CAUSAL UTTERANCE FOR SADNESS

That ' s a real let-down .

TARGET UTTERANCE

Amazon Product Review:

Utterances

A: I cannot understand the positive reviews for this monocular.

B: Tell me about it. The magnification is so minimal, it's worthless.

A:

The slick ads totally lied about this monocular.

CAUSAL UTTERANCE FOR ANGER

B:

Complete garbage. I'm sending it back ASAP.

CAUSAL UTTERANCE FOR ANGER

TARGET UTTERANCE

Adoption by Industry

- ❖ Live Commerce Platform

- Uses an SGnlp hosted model to identify sentiments in real time

- ❖ And more to come

- Ongoing discussions with other interested parties



Contribute (resources)

- ❖ 10-step Contribution Guide ([Demo site](#))
- ❖ Extended Contribution Guide ([Github repo](#))
- ❖ The Docs
- ❖ Reach out to us!



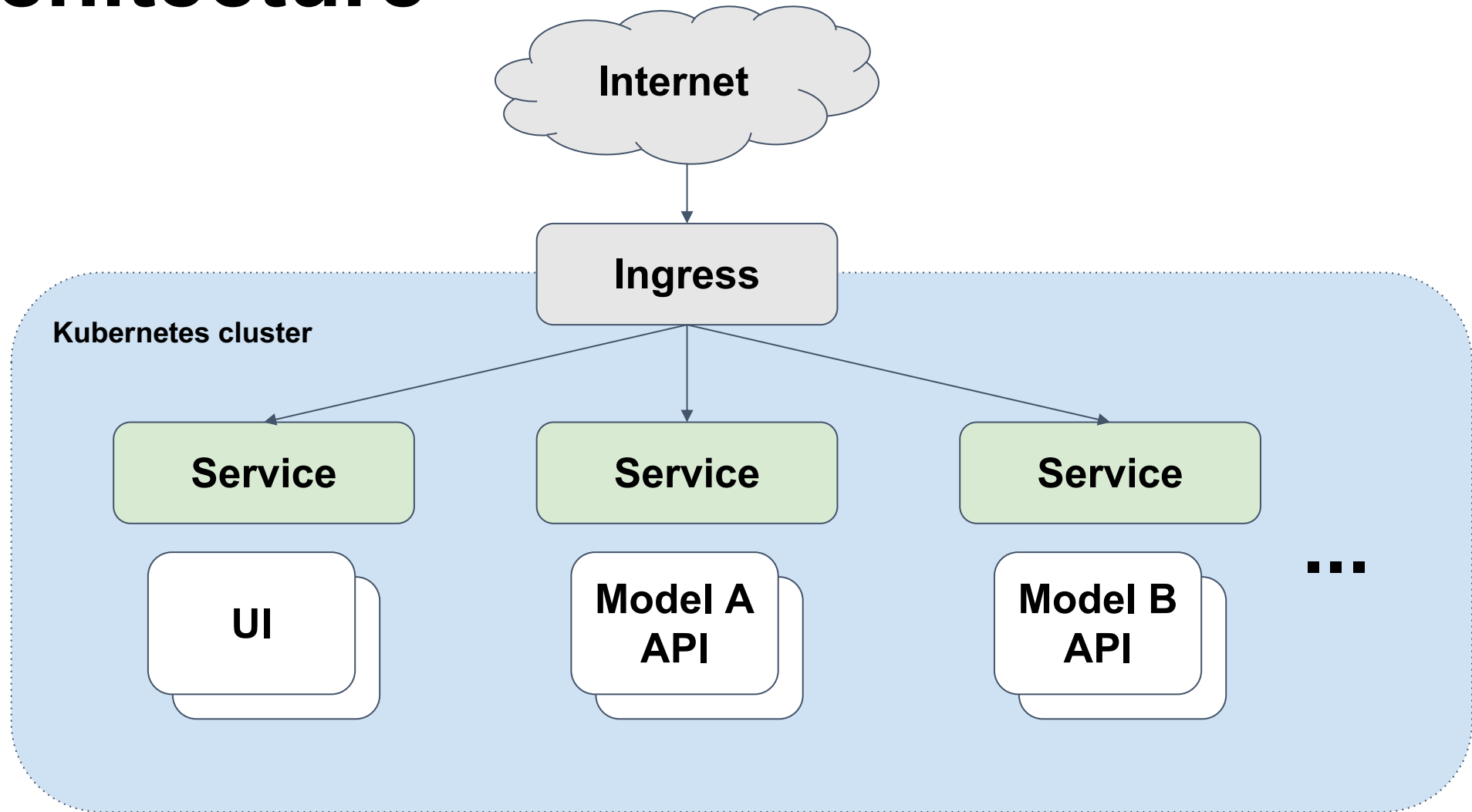
An aerial photograph of Singapore at sunset. The image shows the Marina Bay Sands hotel on the left, the Esplanade - Theatres on the Bay in the foreground, and the dense financial district with numerous skyscrapers on the right. The sun is low on the horizon, casting a warm glow over the city and reflecting on the water in the bay.

Technical Sharing

Overview

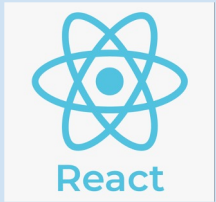
- Python package design
- Tech stack
- Architecture
- Deployment
- Development experiences

Architecture



Tech stack

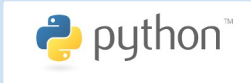
UI



Deployment



Backend



Others



An aerial photograph of Singapore at sunset. The Marina Bay Sands hotel is prominent on the left, with its three towers and skybridge. The Esplanade - Theatres on the Bay is in the foreground. The city skyline is visible in the background, with many skyscrapers. The sky is a mix of orange, pink, and blue. The water of the bay is calm, reflecting the city lights and the sky. The overall scene is a beautiful representation of Singapore's modern architecture and coastal location.

Hands-on Session

General model classes

PreTrainedConfig

- is_encoder_decoder
- output_attentions
- from_pretrained()
- save_pretrained()

PreTrainedModel

- config
- from_pretrained()
- save_pretrained()
- init_weights()
- generate()

Each model specific config inherits all generic attributes and functions from **PreTrainedConfig** and adds model specific attributes.

Each model specific **PreTrainedModel** has access to **self.config**, inherits all functions from **PreTrainedModel** and provides a model specific initialization method used by **init_weights()**

PreTrainedConfig.from_pretrained(dir):

Load serialized configuration object from dir (locally or from url). File is expected to be found as dir/config.json.

PreTrainedConfig.save_pretrained(dir): Serializes configuration instance as dir/config.json.

PreTrainedModel.from_pretrained(dir): 1st calls **init_weights()** to initialize all weights. Then loads serialized model from dir (locally or from url) and overwrites all initialized weights found in dir/pytorch_model.bin with loaded values. Also calls **PreTrainedConfig.from_pretrained(dir)** and sets to **self.config**.

PreTrainedModel.save_pretrained(dir): Serializes model instance as dir/pytorch_model.bin.

PreTrainedModel.generate(): Generates output_ids given input_ids. Calls **PreTrainedModel.forward()** in a while loop

Specific to each model

BrandNewBertConfig

inherit from **PreTrainedConfig**

- vocab_size
- num_hidden_layers
- num_attention_heads

BrandNewBertConfig is the model specific configuration and contains all parameter that are required to instantiate a **BrandNewBertModel**, such as *num_hidden_layers*, etc. Every parameter should have a default value that corresponds to the main model checkpoint to be added

BrandNewBertPreTrainedModel

inherit from **PreTrainedModel**

- **base_class_prefix** = "brand_new_bert"
- **_init_weights()**

BrandNewBertPreTrainedModel is the model specific **PreTrainedModel** class. It is important to define the initialization strategy and common class attributes, such as *base_class_prefix* for all **BrandNewBert** implementations.

Both the "base" model without any head and models with a specific head inherit from a model specific **PreTrainedModel**

BrandNewBertModel

inherit from **BrandNewBertPretrainedModel**

- brand_new_bert_encoder
- **forward()**

BrandNewBertModel is the "base" model that has torch.nn.Modules, such as *brand_new_bert_encoder* as attributes and outputs just the encoded hidden states.

BrandNewBertModel.forward(): defines the inference function for models. It is called when running *model(input_ids)*. The goal for every new model should be that **BrandNewBertModel.forward()** is a self-contained function of the model file so that the reader does not have to look into other files to under it

BrandNewBertForMaskedLM

inherit from **BrandNewBertPretrainedModel**

- brand_new_bert
- masked_lm_head
- **forward()**

BrandNewBertForMaskedLM is a head-specific model that has **BrandNewBertModel** and a specific head layer, called *masked_lm_head* as attributes.

BrandNewBertForMaskedLM.forward(): calls *brand_new_bert.forward()* and *masked_lm_head.__call__()* to generated the head specific outputs, e.g. the logits in this case. The function is usually very short

https://huggingface.co/docs/transformers/add_new_model#overview-of-models

Why contribute?

- **Learning** a consistent way of organizing code
- **Using** a tried and tested interface
- **Increase** exposure for your paper/model
- **Make a difference!**
Models being tested and used in industry use-cases.

Workshop's Github Repository

<https://github.com/aisingapore/sgnlp-workshop>



Q & A



Our Team @ NLP-Hub



William Tjhi



Jayden Tan



Alon Tenzer



Jonathan Heng



Weiqi Leong



Thanh Ngan Nguyen



Raymond Ng



Sherry Zhu

Thank you!

Please visit us at
<https://sgnlp.aisingapore.net/>



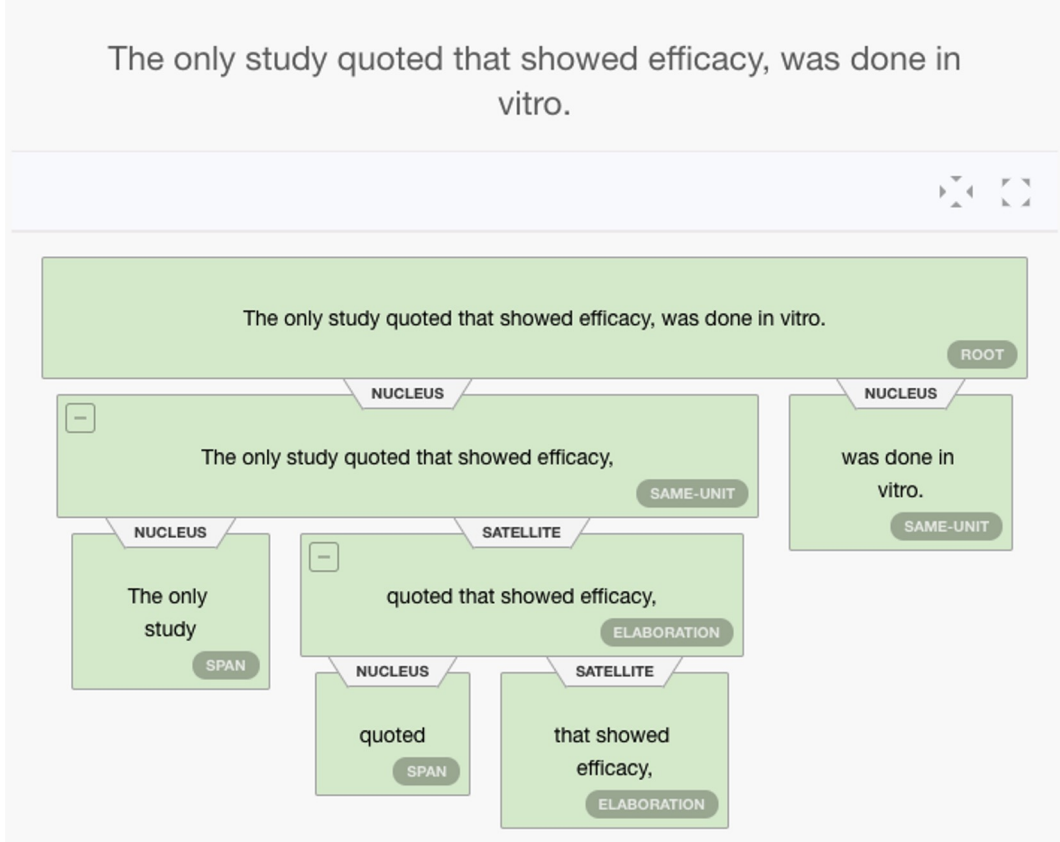
An aerial photograph of Singapore's skyline at sunset. The image captures the Marina Bay Sands hotel on the left, the Esplanade - Theatres on the Bay in the foreground, and the dense cluster of skyscrapers in the Financial District on the right. The sky is a mix of orange, pink, and blue, with the sun low on the horizon. The water of Marina Bay reflects the city lights and the colorful sky.

Appendix

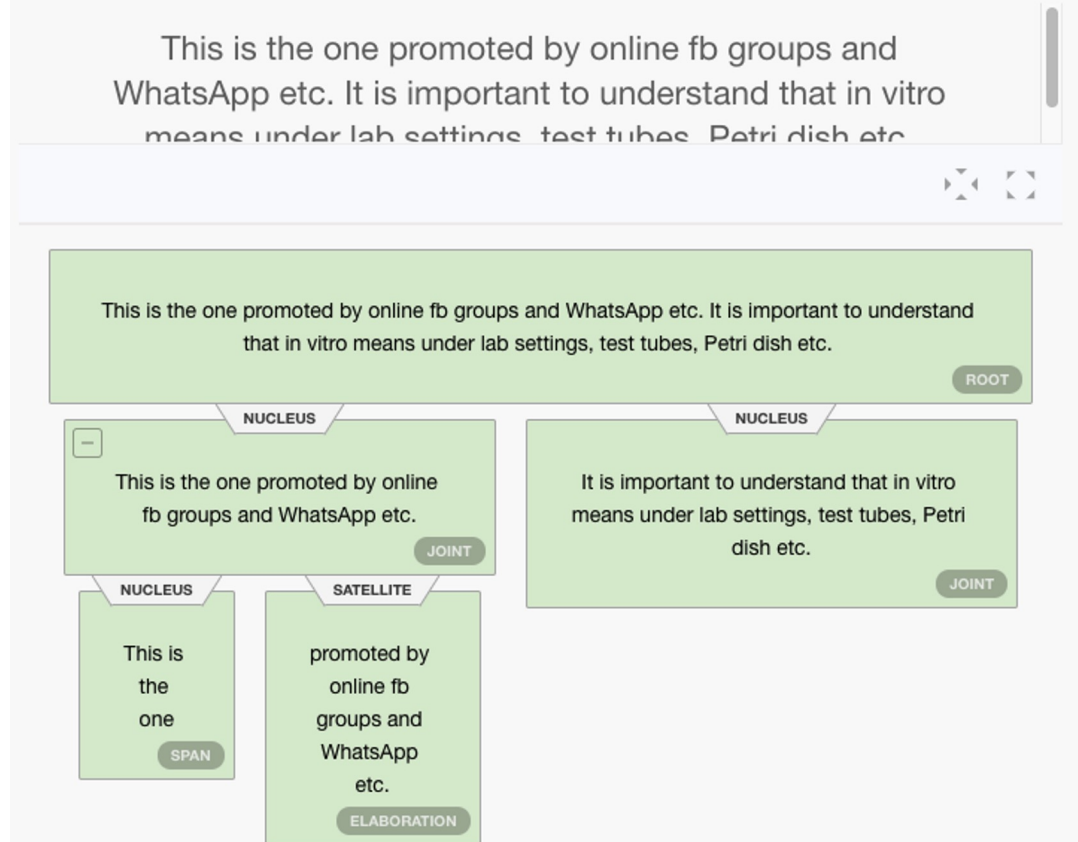
Example 1 (High Quality)

The only study quoted that showed efficacy, was done in vitro. This is the one promoted by online fb groups and WhatsApp etc. It is important to understand that in vitro means under lab settings, test tubes, Petri dish etc. The human body is infinitely more complicated than that. When the same experiment was done on human trials, there's no evidence it worked. It's possible for many things to work in vitro. But we should only trust results from human trials.

Model Results



Model Results



Example 1

Model Results

The human body is infinitely more complicated than that.

The human body is infinitely more complicated than that.

ROOT

Model Results

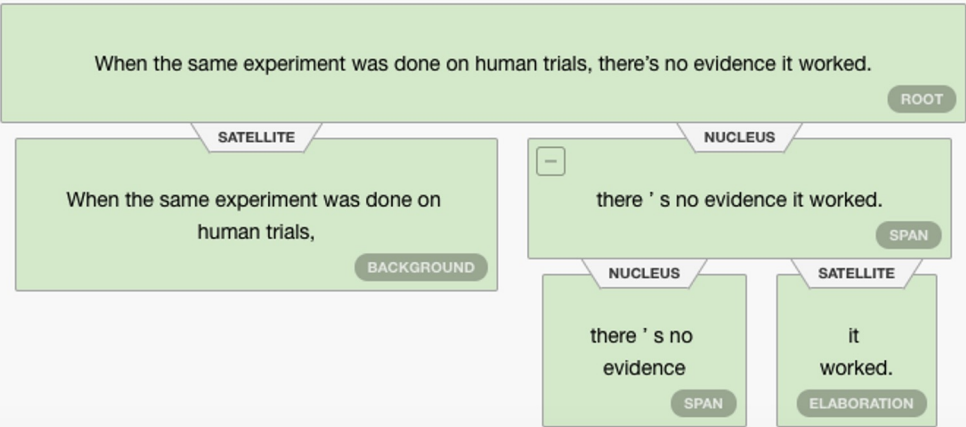
It's possible for many things to work in vitro.

It's possible for many things to work in vitro.

ROOT

Model Results

When the same experiment was done on human trials, there's no evidence it worked.



Model Results

But we should only trust results from human trials.

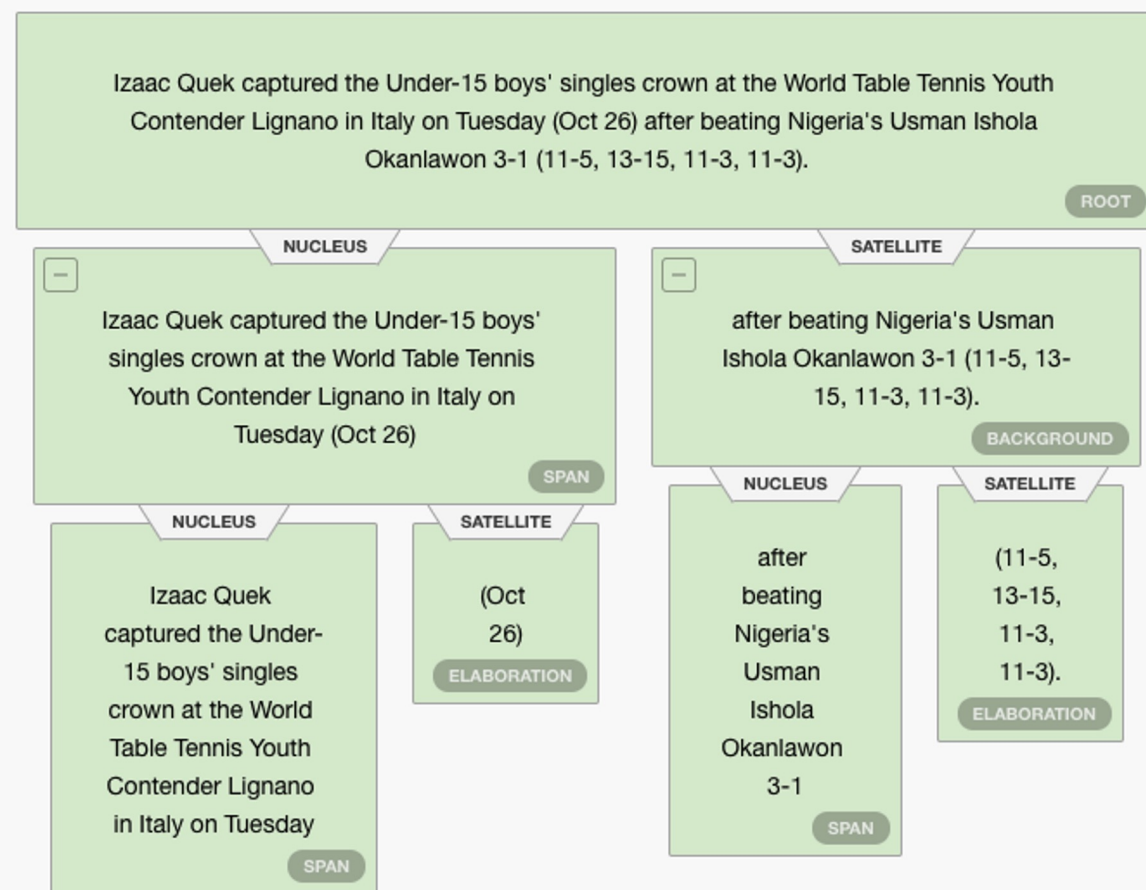
But we should only trust results from human trials.

ROOT

Izaak Quek captured the Under-15 boys' singles crown at the World Table Tennis Youth Contender Lignano in Italy on Tuesday (Oct 26) after beating Nigeria's Usman Ishola Okanlawon 3-1 (11-5, 13-15, 11-3, 11-3).

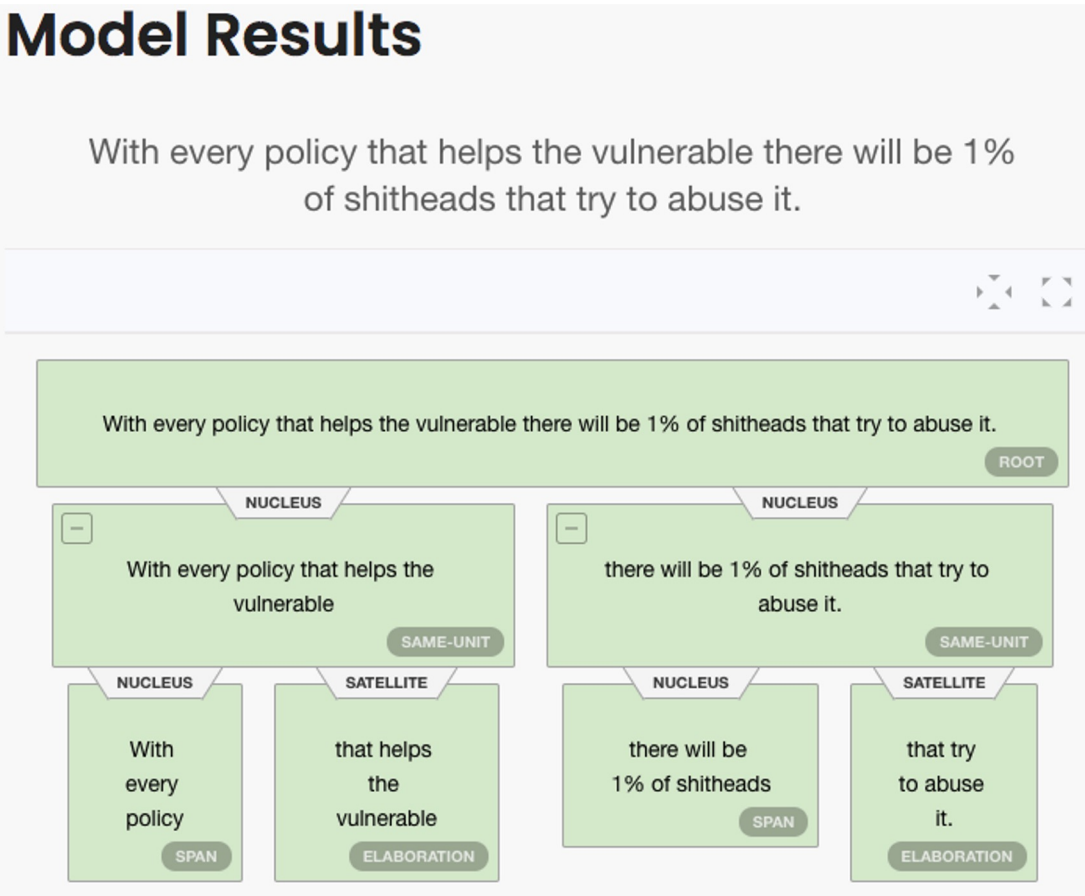
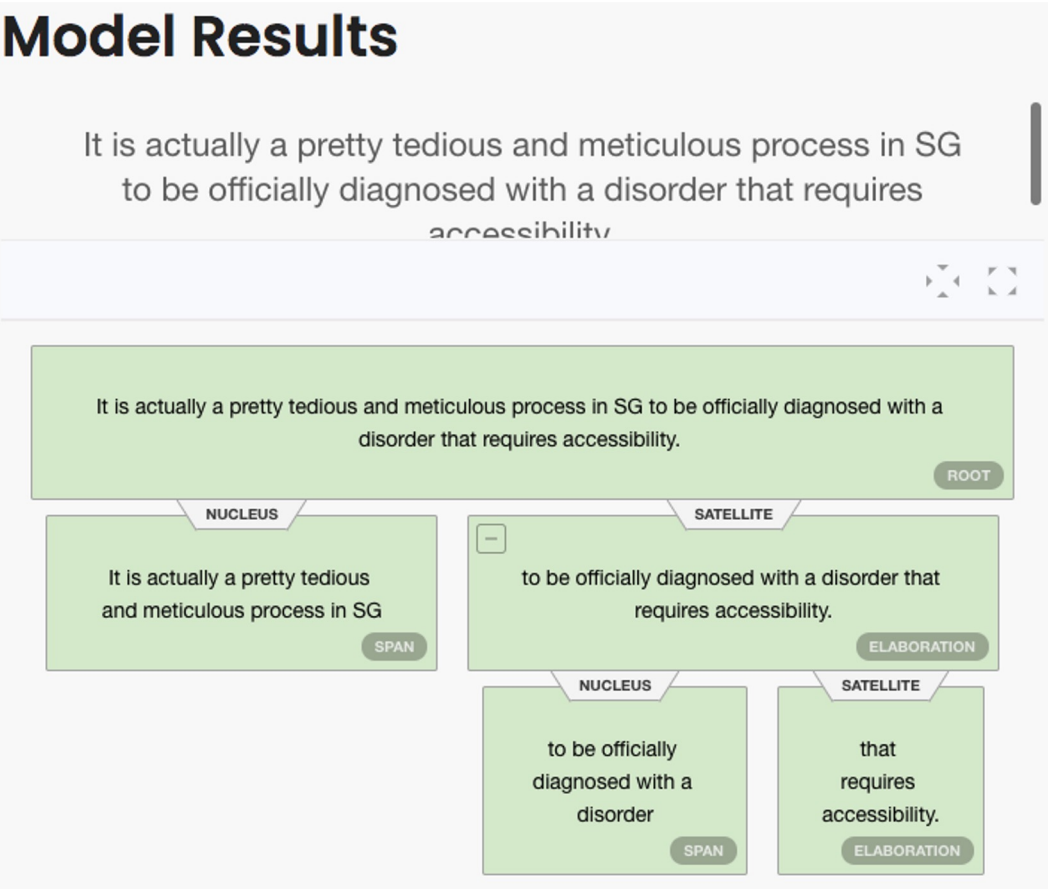
Izaak Quek captured the Under-15 boys' singles crown at the World Table Tennis Youth Contender Lignano in Italy on Tuesday (Oct 26) after beating Nigeria's Usman Ishola Okanlawon 3-1 (11-5, 13-15, 11-3, 11-3).

Izaak Quek captured the Under-15 boys' singles crown at the World Table Tennis Youth Contender Lignano in Italy on Tuesday (Oct 26) after beating Nigeria's Usman Ishola.



Example 2 (Subjective/ Biased)

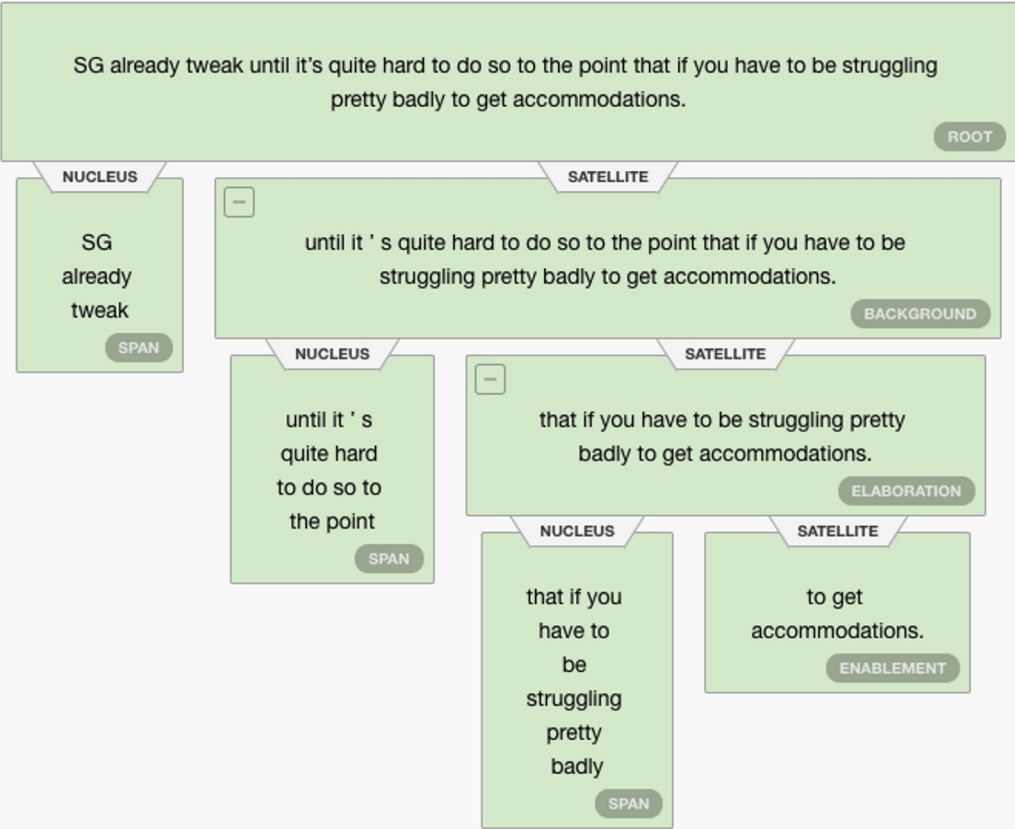
It is actually a pretty tedious and meticulous process in SG to be officially diagnosed with a disorder that requires accessibility. With every policy that helps the vulnerable there will be 1% of shitheads that try to abuse it. SG already tweak until it's quite hard to do so to the point that if you have to be struggling pretty badly to get accommodations. Just accept the 1%, it will always exist with every good thing. Be careful you don't smear the 99% who need it when you judge the 1%.



Example 2

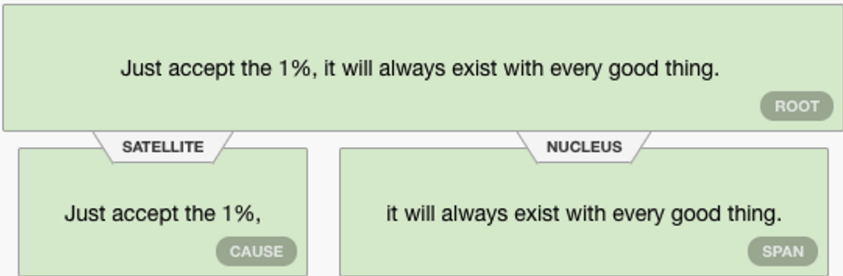
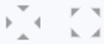
Model Results

SG already tweak until it's quite hard to do so to the point
that if you have to be struggling pretty badly to get
accommodations



Model Results

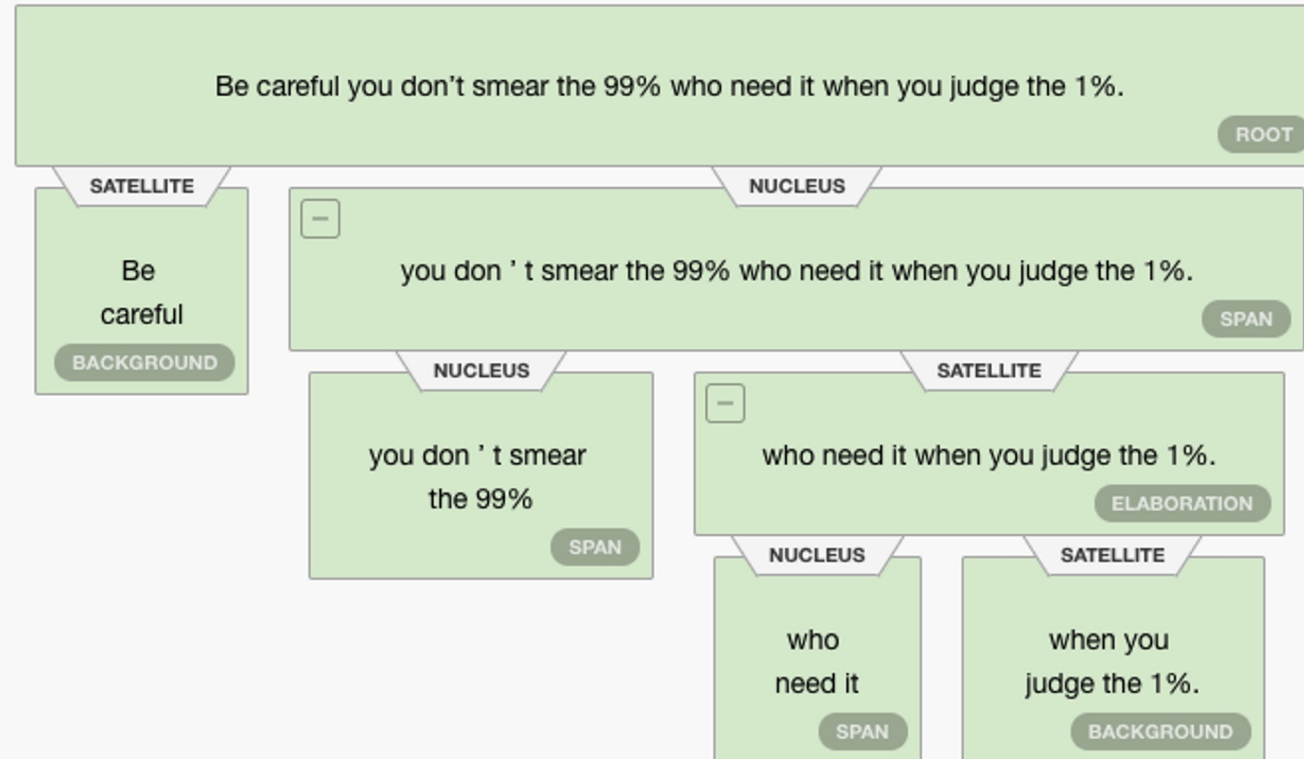
Just accept the 1%, it will always exist with every good
thing.



Example 2

Model Results

Be careful you don't smear the 99% who need it when you judge the 1%.



Example 4 (Subjective/Biased)

As an Australian I am deeply disappointed and frustrated with how they've treated their citizens, permanent residents as well as all visa holders. Imagine being locked in your country, or being locked out, it's madness. Then they made it such that if you were Australian but resident if another country, and you visited, you had to get permission to exit the country. Now this arbitrary ruling about citizenship for entry.

Model Results

As an Australian I am deeply disappointed and frustrated with how they've treated their citizens, permanent residents as well as all visa holders

As an Australian I am deeply disappointed and frustrated with how they've treated their citizens, permanent residents as well as all visa holders.

ROOT

Model Results

Imagine being locked in your country, or being locked out, it's madness.

Imagine being locked in your country, or being locked out, it's madness.

ROOT

SATELLITE

Imagine being locked in your country, or being locked out,

ENABLEMENT

NUCLEUS

it ' s
madness.

SPAN

NUCLEUS

Imagine being locked in your
country,

JOINT

NUCLEUS

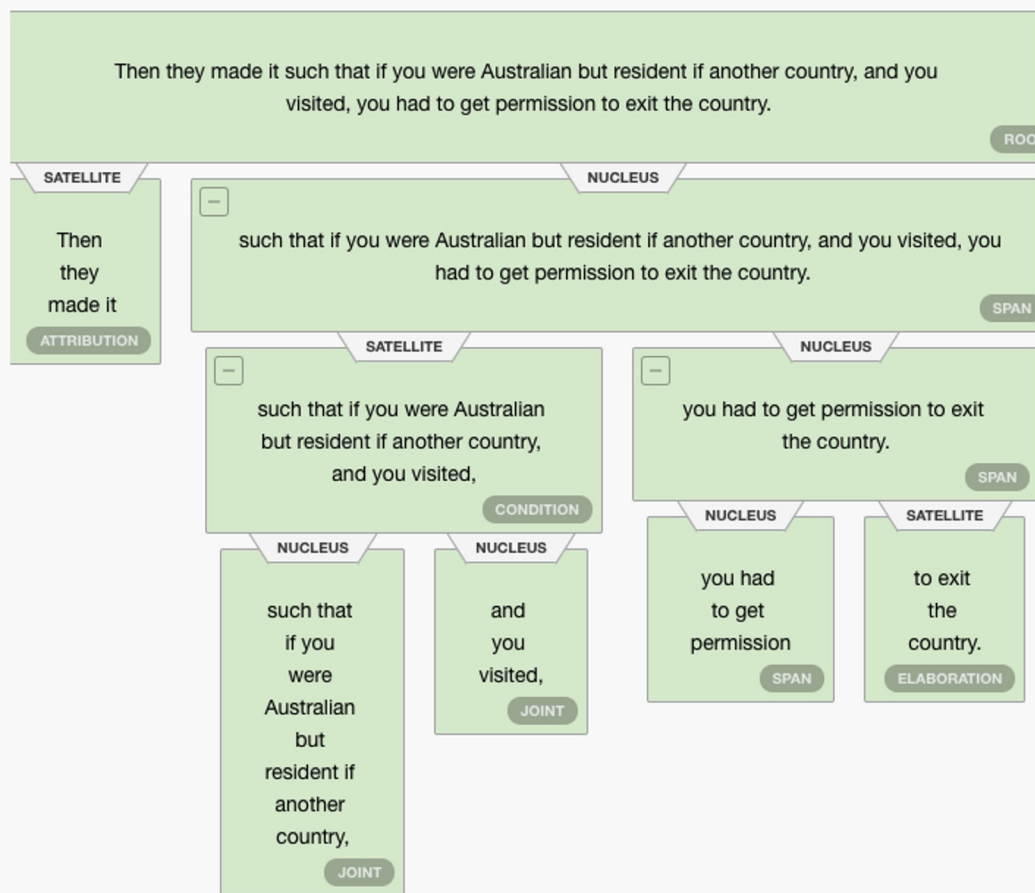
or being locked
out,

JOINT

Example 4

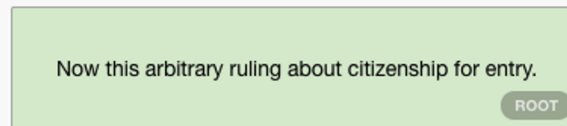
Model Results

Then they made it such that if you were Australian but resident in another country, and you visited, you had to get permission to exit the country.



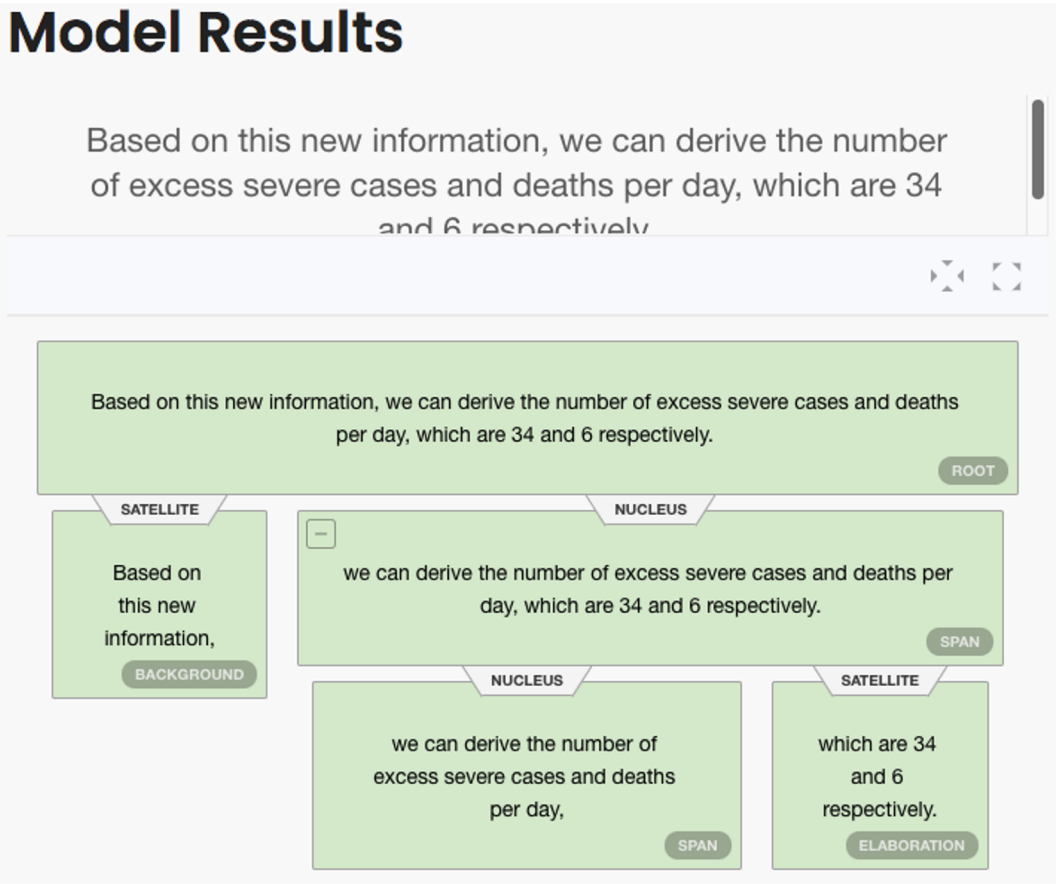
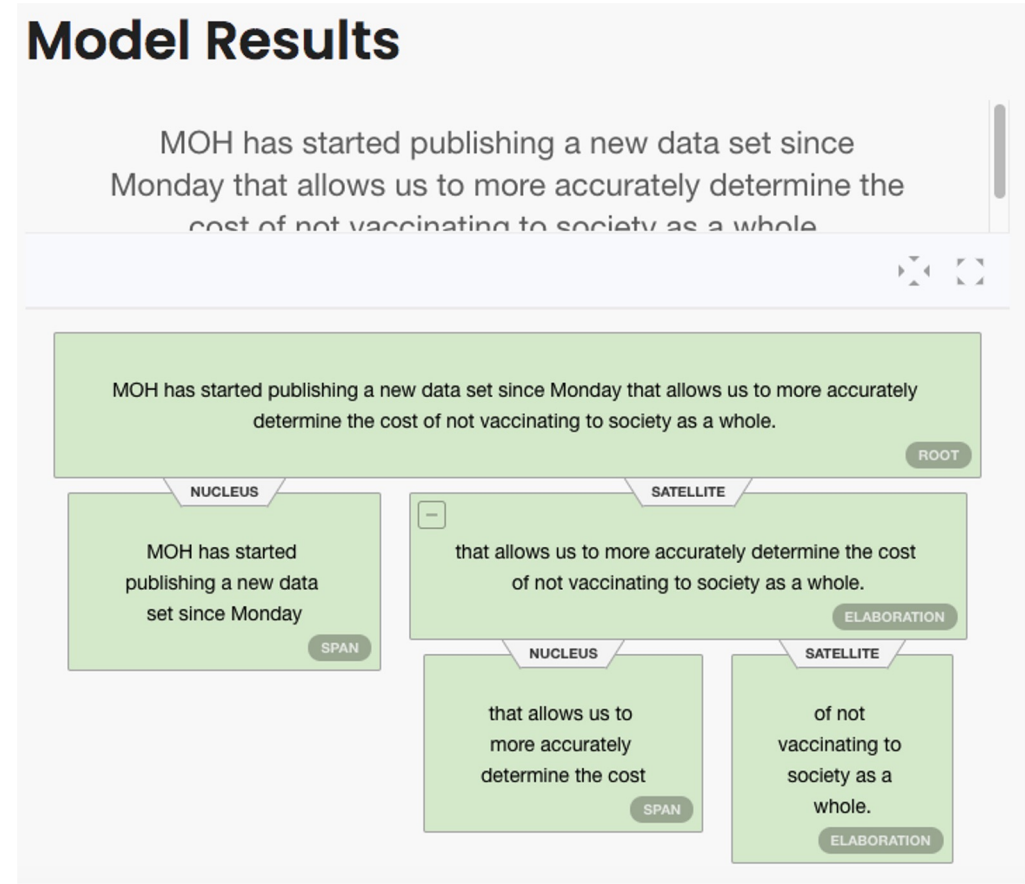
Model Results

Now this arbitrary ruling about citizenship for entry.



Example 5 (High Quality)

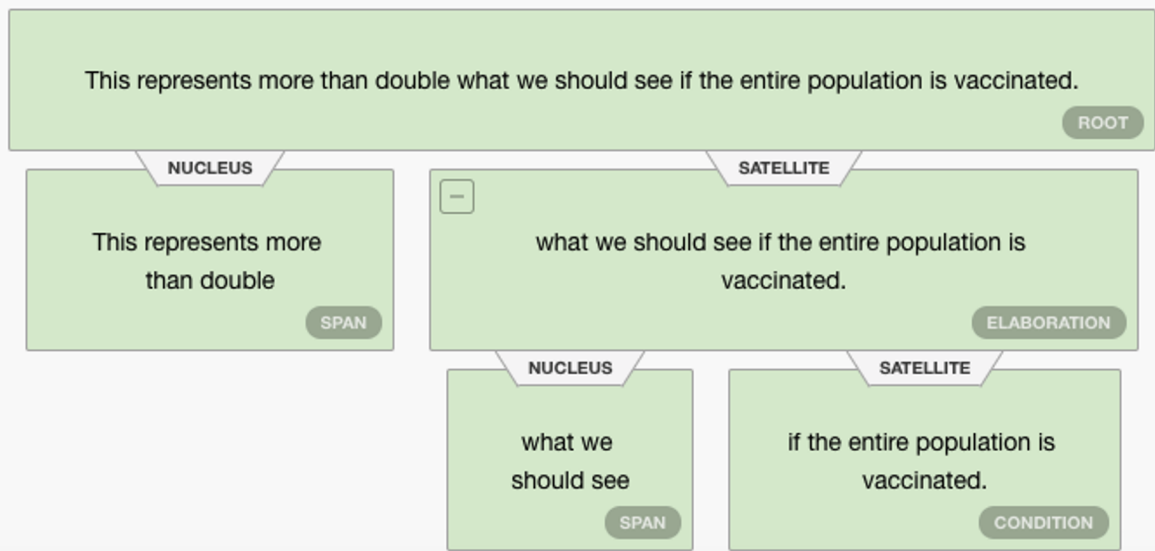
MOH has started publishing a new data set since Monday that allows us to more accurately determine the cost of not vaccinating to society as a whole. Based on this new information, we can derive the number of excess severe cases and deaths per day, which are 34 and 6 respectively. This represents more than double what we should see if the entire population is vaccinated. In other words, if everyone were vaccinated, our ICU problem could have been half of what it is now. Interestingly, because nearly all deaths are of people 60 or older, mandating vaccines on those 60 or older would suffice to achieve the same impact as it relates to death. To achieve the same impact as it relates to all severe illness cases, we would need to include those 40 or older.



Example 5

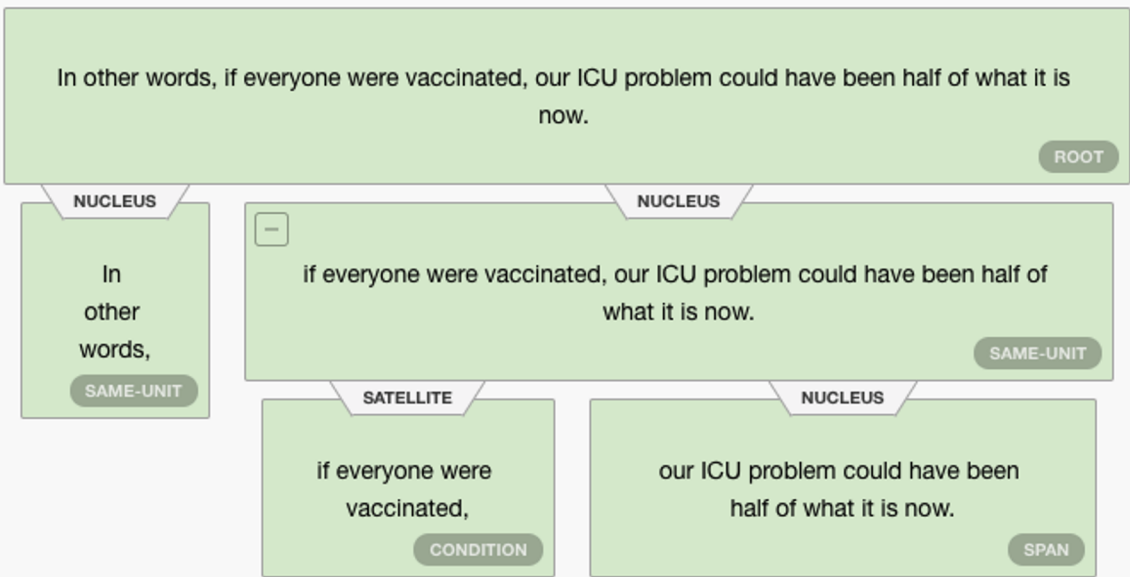
Model Results

This represents more than double what we should see if the entire population is vaccinated.



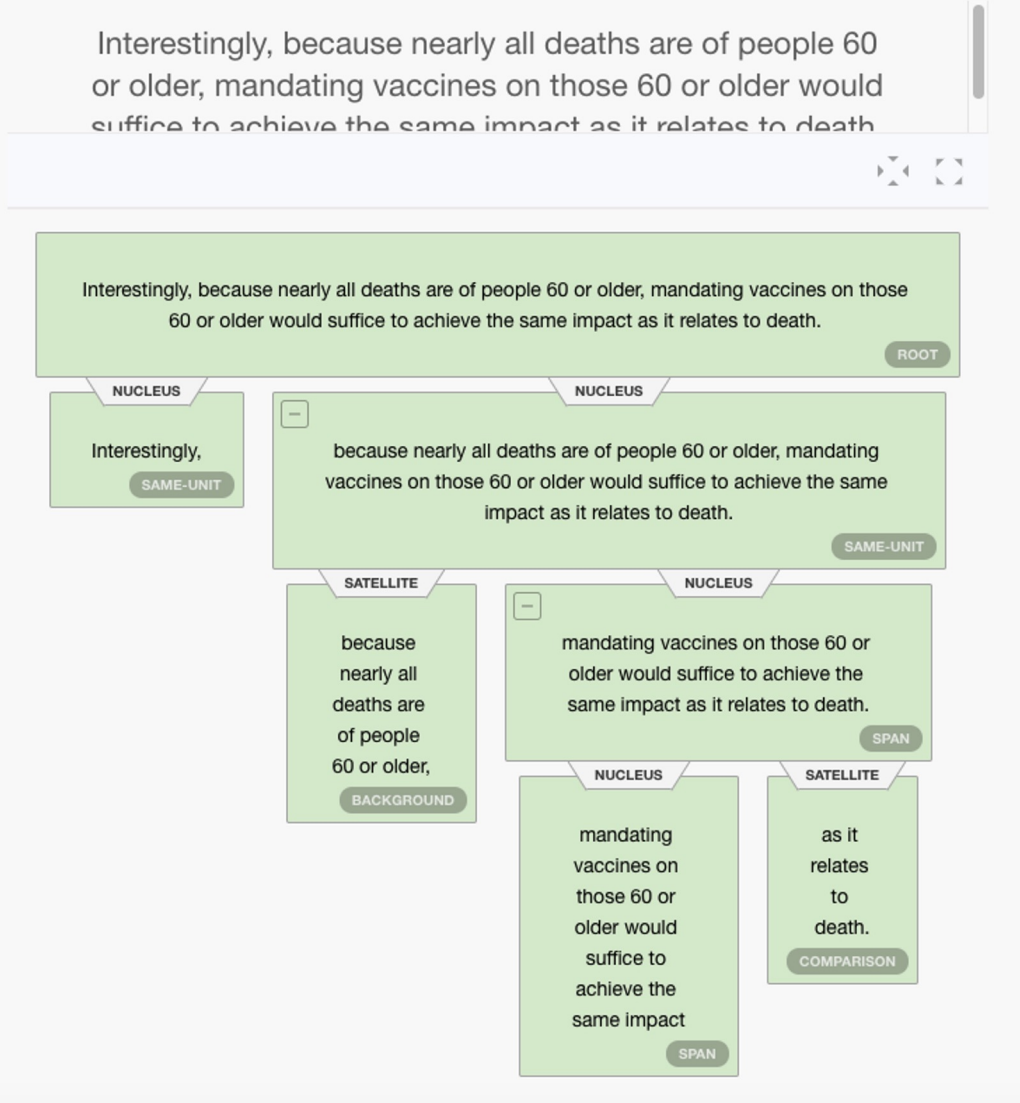
Model Results

In other words, if everyone were vaccinated, our ICU problem could have been half of what it is now.



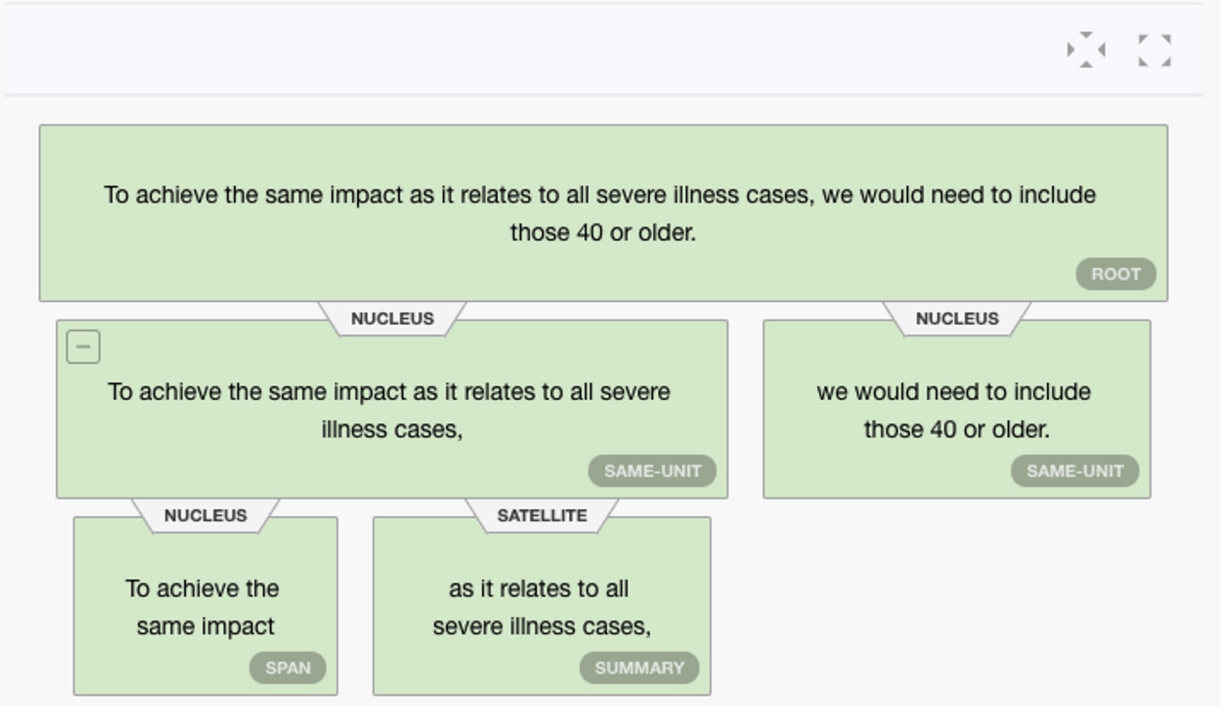
Example 5

Model Results



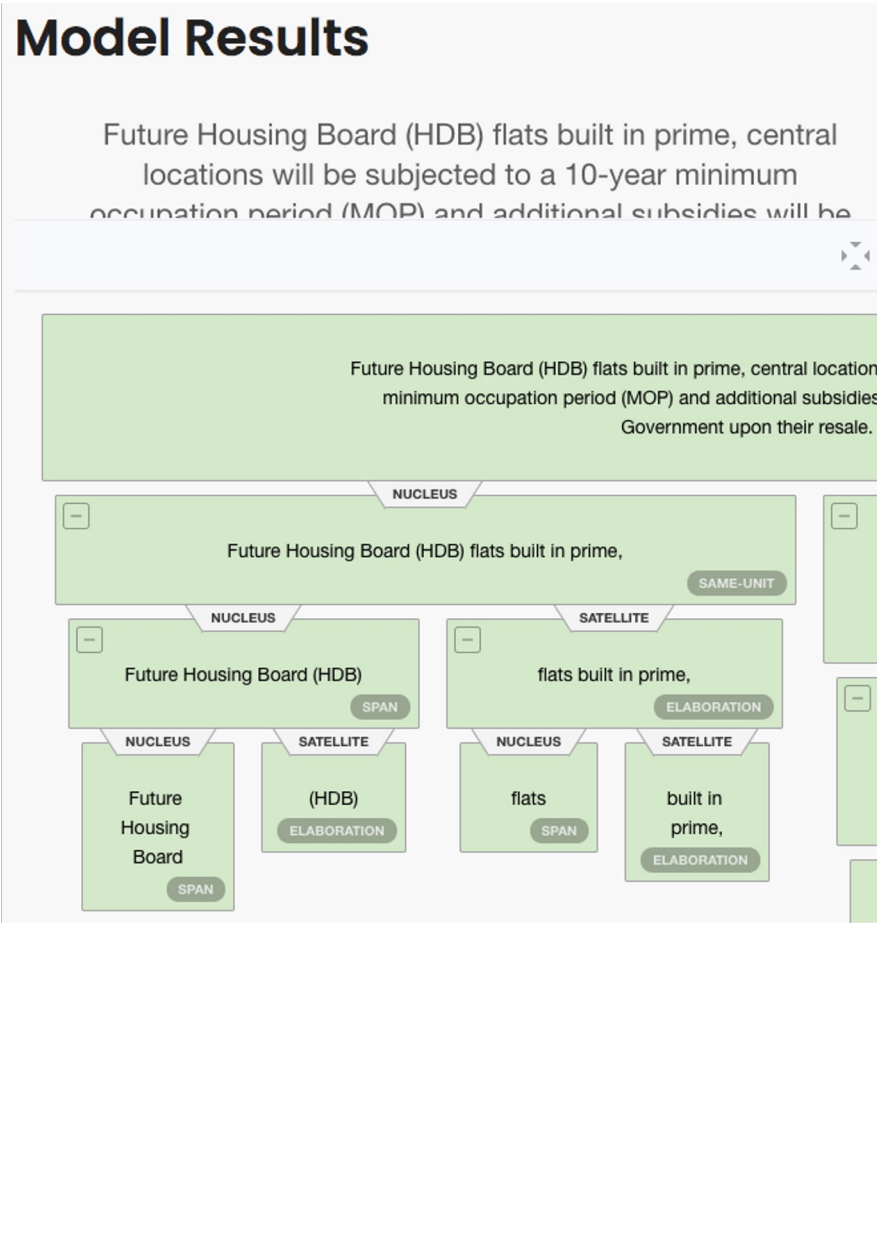
Model Results

To achieve the same impact as it relates to all severe illness cases, we would need to include those 40 or older.



Example 7 (High Quality)

Future Housing Board (HDB) flats built in prime, central locations will be subjected to a 10-year minimum occupation period (MOP) and additional subsidies will be clawed back by the Government upon their resale.

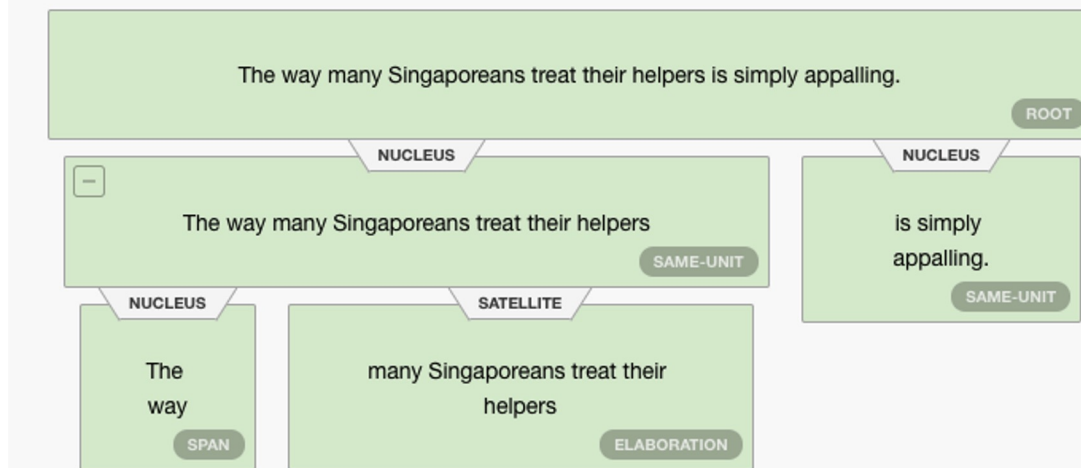


Example 8 (Subjective/Biased)

The way many Singaporeans treat their helpers is simply appalling. Beyond the obvious maid abuse cases that make the news, if you have a look at some local helper hiring Facebook groups you can see the difference between foreign family job scope and a Singaporean family job scope. The foreign family post will look something like, Monday's to Saturday's till dinner time with evenings off, all sundays off, PH off, transport ez link card given, paid OT if necessary, \$800/month. A Singaporean post is along the lines of 2 off days a month, 7am-10pm work, mop floor twice a day, oh and no food sharing, helper only gets cheap Maggie or canned food, \$450/month.

Model Results

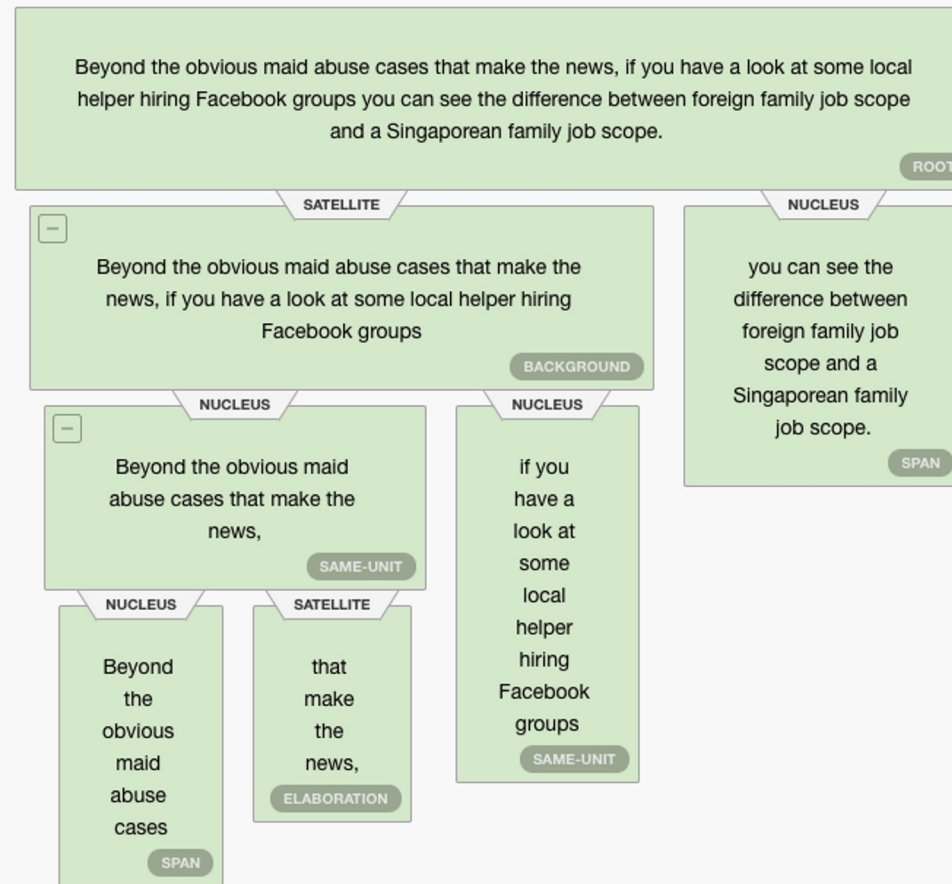
The way many Singaporeans treat their helpers is simply appalling.



Example 8

Model Results

Beyond the obvious maid abuse cases that make the news, if you have a look at some local helper hiring Facebook groups you can see the difference between



Example 8

Model Results

The foreign family post will look something like, Monday's to Saturday's till dinner time with evenings off, all sundays off, PH off, transport ez link card given, paid OT if



The foreign family post will look something like, Monday's to Saturday's till dinner time with evenings off, all sundays off, PH off, transport ez link card given, paid OT if necessary, \$800/month.

ROOT

NUCLEUS

The foreign family post will look something like, Monday ' s to Saturday ' s till dinner time with evenings off, all sundays off, PH off, transport ez link card

SPAN

SATELLITE

given, paid OT if necessary, \$800/month.

ELABORATION

NUCLEUS

given,

SPAN

SATELLITE

paid OT if necessary, \$800/month.

BACKGROUND

Model Results

A Singaporean post is along the lines of 2 off days a month, 7am-10pm work, mop floor twice a day, oh and no food sharing, helper only gets cheap Maggie or canned



A Singaporean post is along the lines of 2 off days a month, 7am-10pm work, mop floor twice a day, oh and no food sharing, helper only gets cheap Maggie or canned food, \$450/month.

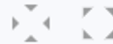
ROOT

Example 10 (Subjective/Biased)

It just seems like in a fit of anger and a bit of a 'diva' moment, he really tried his best to ruin this student's life. If he did try to destroy her over nothing, he's nothing but a scumbag. Like it or not, when you're a 'public' figure representing a minority group you have to conduct yourself in a responsible manner. Doing things like this negatively affects his industry and community, so shame on him.

Model Results

It just seems like in a fit of anger and a bit of a 'diva' moment, he really tried his best to ruin this student's life.



It just seems like in a fit of anger and a bit of a 'diva' moment, he really tried his best to ruin this student's life.

ROOT

Model Results

If he did try to destroy her over nothing, he's nothing but a scumbag.



If he did try to destroy her over nothing, he's nothing but a scumbag.

ROOT

SATELLITE

NUCLEUS

If he did try to destroy her over nothing,

CONDITION

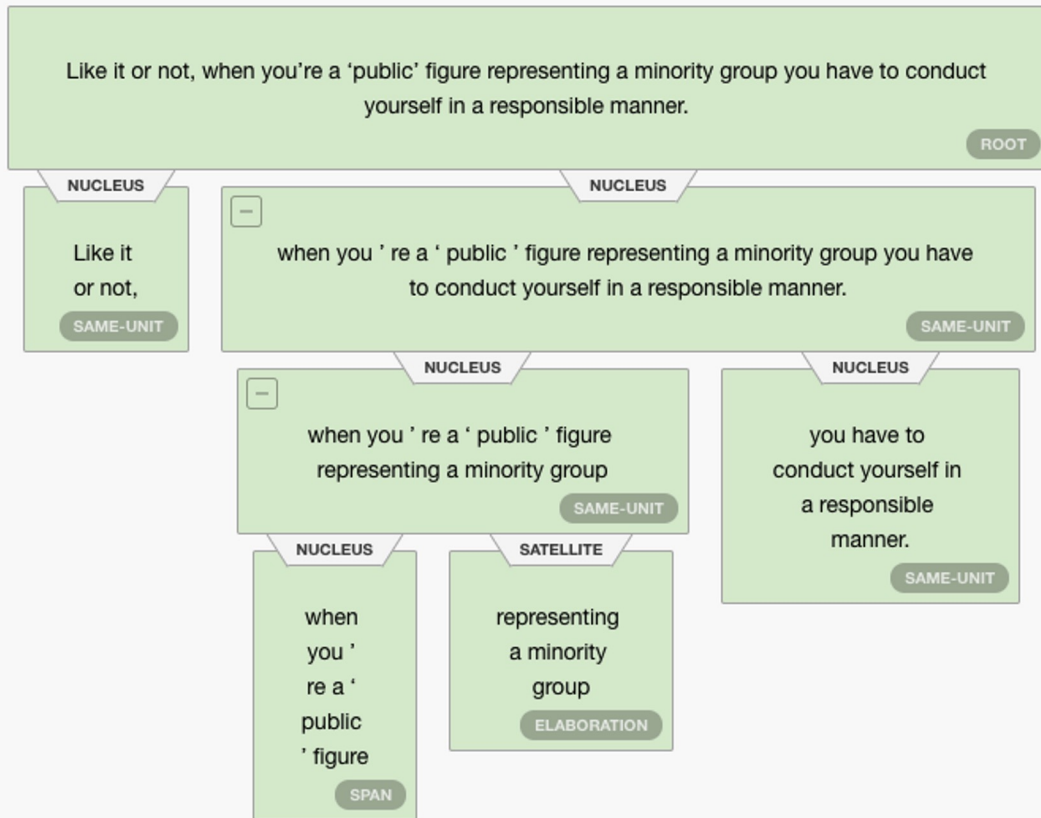
he ' s nothing but a scumbag.

SPAN

Example 10

Model Results

Like it or not, when you're a 'public' figure representing a minority group you have to conduct yourself in a responsible manner



Model Results

Doing things like this negatively affects his industry and community, so shame on him.



Doing things like this negatively affects his industry and community, so shame on him.

ROOT