Homework 2

Diya Benjamin
diyabenjamin@gatech.edu

1 QUESTION 1:

1.1 Incremental Concept Learning:

Incremental Concept Learning is a process where The AI agent learns from one example at a time. Positive example generalizes the concept definition while negative example specializes it.

Table 1. Positive and Negative examples of sandwich

A sandwich	Not a sandwich
BLT on white bread, hamburger, turkey and swiss on potato roll, Meatball sub, tuna salad on brioche, chip butty, ice cream sandwich, grilled cheese, turkey hero, vada pav, veggie burger, egg & cheese biscuit, buttered biscuit, patty melt, sloppy joe	Chicken wrap, burrito, ice cream taco, toast, cheese quesadilla, toaster strudel, Klondike bar, gyro, sushi rolls, calzone

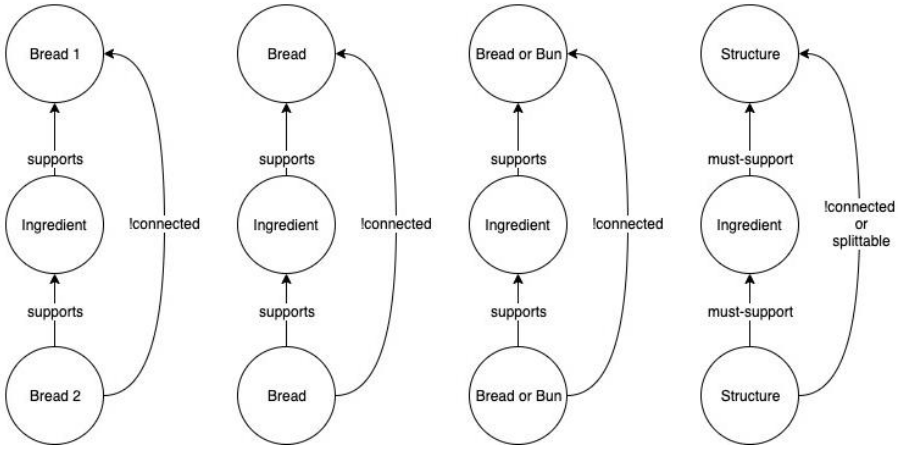

Figure 1. a) Initial Concept Definition of Sandwich. b) Variablization c) Enlarge-set heuristic d) Final Concept Definition of Sandwich.

To begin with, the 1st positive example is **BLT on white bread** and a concept definition has been built using this example is as shown in Figure 1.a above. We will use **variabilization** to generalize bread 1 and bread 2 as bread. The order or placement doesn't matter as long as relationship remains same. The 2nd positive example is **Vada pav** which consists of a bun that is splittable and has a potato filling

inside it. Hence, we use **enlarge-set heuristic** to add this feature of splittable to the link between the breads. The background knowledge can be generalized to replace bread or bun to container. This is called **climb tree heuristic** which is as shown in Figure 2.b below where a structure can be a bread or bun or biscuit.

The 3rd negative example is **Burrito** which is a wrap consisting of meat, beans and veggies wrapped inside a tortilla. This example does not fit the current concept definition and hence there will be no modification i.e. **do nothing**. The 4th negative example is **Toast**. It consists of a toasted bread with no filling. Hence, we will use the **require-link heuristic** to specialize the 'contains' link to 'must-contain' link in order to specify it as a required feature**.**

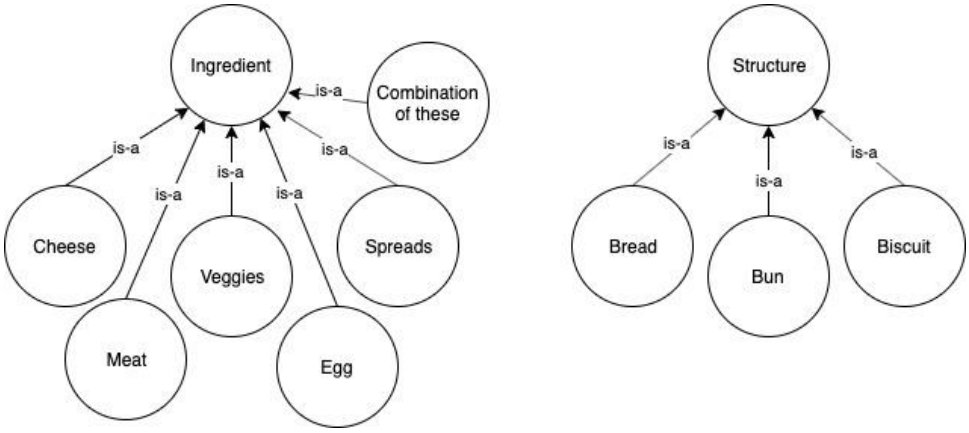

Figure 2. a) Climb-tree heuristic for Ingredient. b) Climb-tree heuristic for Structure.

Calzone is an example which, if considered with above examples, would have made some significant difference to concept definition model of sandwich. Given Calzone as a negative example, it requires the filling to be placed inside dough and then cook (or bake) them just like pizza. This would require changes to model to consider both the container and fillings as being already cooked. A club sandwich or a 3 layered sandwich is a positive example which would require a **drop link heuristic** in order to generalize it to ignore the extra set. This would allow the new concept definition to cover this example as well as all previous examples.

1.2 Classification:

The sandwich can also be defined using a different approach known as classification where the complex problem is divided into number of smaller problems. I have defined four different percepts or parameters for a sandwich and the values for these parameters for 6 sandwiches as provided in Table 2 below. The

abstracted classification tree-based structure of a sandwich is based on these values and are as shown in Figure 3 below.

Table 2. Values of parameters for sandwiches based on classification

	structure made of dough?	**structure has 2 or more slices of bread or bun?**	**structure is splittable?**	**has ingredient in the middle?**
Hamburger	Yes	Yes	Yes	Yes
Tuna salad on brioche	Yes	No	Yes	Yes
Grilled cheese	Yes	Yes	Yes	Yes
Chicken wrap	Yes	No	No	Yes
Toaster strudel	Yes	No	No	Yes
Sushi rolls	No	No	No	Yes

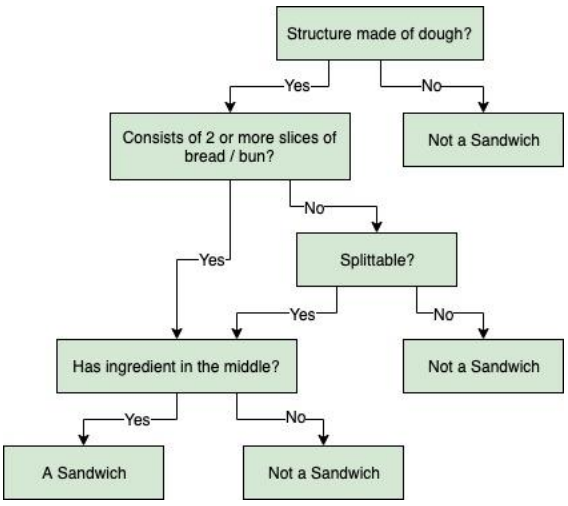

Figure 3. Abstracted classification of what a sandwich is.

1.3 Is hot dog a sandwich?

Using the model in figure 1 above for incremental concept learning, the hot dog is a positive example of sandwich. It consists of a bun which is splittable and a hotdog (or sausage) inside the bun. Based on classifier, the parameters and their values are as shown below. This shows that hot dog is a sandwich.

- Structure made of dough? Yes
- Structure has 2 or more slices of bread or bun? Yes
- Structure is splittable? Yes

- Ingredient is between or within the structure? Yes

Based on case-based reasoning, we will retrieve a previous example 'Meatball Sub' that would be most similar to hotdog. Both consists of bun as structure which is splittable, with some meat and other ingredients in between. The only difference is meatball is replaced by hotdog which is minor tweak required for case adaptation. Thus, evaluation of this adaptation provides a positive outcome i.e. a hot dog is a sandwich.

2 QUESTION 2:

We will see here how an AI agent understands the meaning of the sentence '**I never said Amy planted that seed**'. The AI agent will perform lexical, syntactic and semantic analysis on this sentence to understand and figure out its meaning.

Lexical Analysis of this sentence categorizes each of the words into various lexical categories as shown below:

- I, Amy, seed are nouns
- said, planted are verbs
- never is an adverb

Syntactic Analysis shows that 'I', 'Amy' are noun phrase, while 'never said', 'planted that seed' are verb phrase.

These 2 analyses can be used to form the **Semantic Analysis**. The AI agent uses this semantic analysis to form a thematic role frame representation of the sentence to understand its meaning. In the above frame representation, the thematic role represents the relationship of the other words in the sentence with the action 'say'.

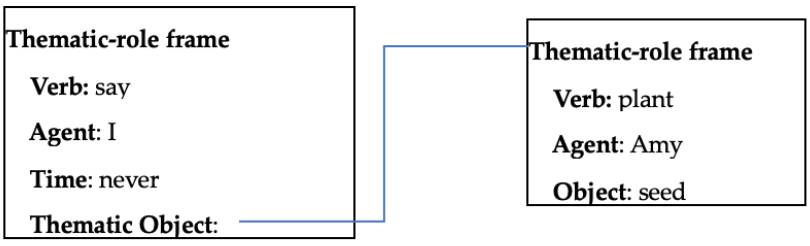

Figure 4. Thematic role frame representation of the sentence

Emphasizing any individual word changes the implication of the sentence. Here, we will emphasize on the words 'I' and 'Amy' and see how AI agent understands how different emphases changes the meaning of the sentence. Here, the additional knowledge provided is emphasis on specific words.

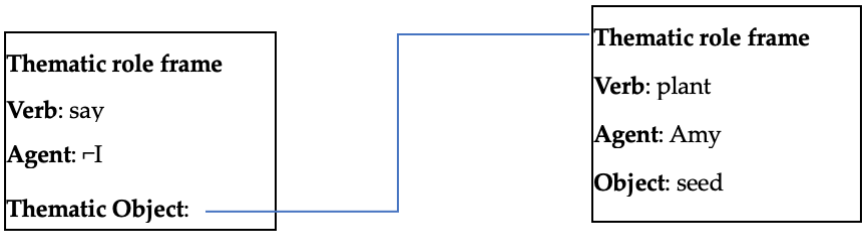

Figure 5. Thematic role frame representation of the sentence with emphasis on 'I'

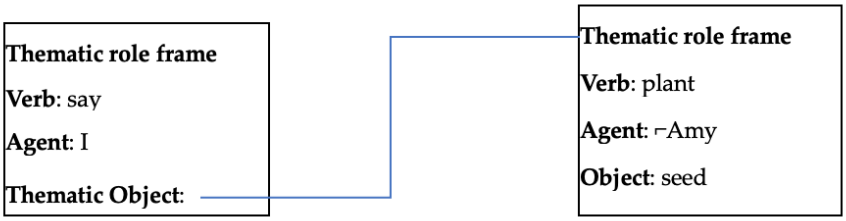

Figure 6. Thematic role frame representation of the sentence with emphasis on 'Amy'

The above two frame representations show that emphasizing different words in the same sentence provides different interpretations of the sentence. If the emphasis is on 'I', it results in negation of the agent (⌐I) in the 'say' action frame. This helps the AI agent to understand the meaning of the sentence as: someone said that Amy planted the seed, but that someone is not 'I' and 'Amy planted that seed' is true. Similarly, if the emphasis is on 'Amy', it results in negation of the agent (⌐Amy) in the 'plant' action frame. Thus, AI agent can interpret the sentence as: agent 'I' said that someone planted that seed, but that someone is not 'Amy' and 'Amy planted that seed' is not true.

It is difficult for an AI agent to decide if 'planted seed' in the sentence is to be taken literally or figuratively wherein seed represents an idea. This is not specific form the given sentence. If the AI agent is provided with a preceding sentence in that story, this ambiguity could be cleared. Currently, we can say that there are two particular interpretations for this sentence.

3 QUESTION 3:

3.1 Toronto Declaration:

The **Preamble** to the Toronto Declaration (TD) states that even though machine learning has made great advancements in their capability and usage, we still need to consider both positive and negative implications of these technologies. The main focus is on how these technologies will affect human rights. TD aims to:

- Emphasize on the importance of human rights law and standards.
- Build on existing discussions, principles and papers exploring the discriminatory risks of machine learning systems.

The systems employing machine learning technologies can change the power structures or inequalities that are risk to human rights. It also emphasizes on promoting the positive right to the enjoyment of benefits of these technologies and its progress in the field of ML and AI. Misuse of this technology will affect the **rights to equality and non-discrimination** which is the main focus of TD, in addition to other areas.

Using the framework of international human rights law section discusses about the obligations of public and private sectors to protect and respect human rights. To **prevent discrimination**, the private sectors need to check for compliance with the international laws and address various concerns such as systemic bias. If these rights are violated or denied, the states must ensure access to effective remedy for all affected individuals. **Protecting the rights of all individuals and groups and promoting diversity and inclusion diversity** are key components to ensure that machine learning systems respect non-discrimination, equality and other human rights. More work is needed to ensure that all human rights are protected as artificial intelligence touches nearly all aspects of modern life.

Duties of states: human rights obligations section mentions that States bear the main responsibility of promoting and protecting human rights and non-discrimination. This requires: i. identifying risks through regular impact assessments, ii. ensuring transparency and accountability by enabling independent analysis and avoiding 'black box systems' and iii. enforcing oversight which includes adopting diverse hiring, human rights training, independent oversight and conforming to international standards. This section also states the responsibility of government in **promoting equality** and non-discrimination.

Three steps to the process of human rights due diligence:

i. Identify potential discriminatory outcomes
ii. Take effective action to prevent and mitigate discrimination and track responses
iii. Be transparent about efforts to identify, prevent and mitigate against discrimination in machine learning systems."

The right to an effective remedy section mentions the responsibility of the States to provide effective remedy to victims of discrimination that includes reparation, sanctions against those responsible and guarantees of non-repetition.

In my analysis, I believe there are not many tradeoffs implied by the TD. In following the declaration, it provides social benefits such as promoting equality, transparency, avoiding bias, etc. In my opinion, it fails to provide innovations in artificial intelligence especially for the image recognition systems. Providing data to these systems by avoiding bias brings in a lot of challenges thus increasing production cost. However, if the declaration is discarded, it can cause serious threat to human rights by replicating biases, hindering due process and undermining the laws of war ("Toronto Declaration: Protecting the rights to equality and non-discrimination in machine learning systems," 2018). While it may make things easier for the private sectors and states, the risks outweigh the benefits. The private sectors will anyway introduce these rules themselves, in order to amass their customers.

I think the Toronto Declaration is a useful document and it is essential to understand the need for it. It will help in the legal design and implementation of various ML and AI-based technologies. It is important to emphasize on the human rights to equality and non-discrimination, which is effectively done in TD. There could be a couple of changes to include people in the development process from a different race, culture, gender, and socio-economic background, which will make sure there is a proper inclusion of each of the culture to avoid bias. This will be extremely useful in image recognition systems that is used by autonomous cars in detecting pedestrians. There should also be inclusion of laws to control the development of autonomous weapons that cause harm to humans.

4 QUESTION 4:

4.1 New article on AI development – Positive light:

I will be discussing on the recent news article from techxplore.com that discusses about how "Artificial intelligence improves biomedical imaging". The researchers from ETH Zurich used AI to improve the quality of images produced by optoacoustic imaging. This technique can be used to visualize blood vessels, to study brain activity, to diagnose breast cancer. The quality of image produced by this device increases as the number of sensors increases. This is not feasible as more sensors leads to higher device cost. Thus, using AI to improve the quality of image using a smaller number of sensors helps in reducing the device cost.

This development in AI can also be applied to other imaging techniques because it operates on reconstructed images, not the raw recorded data. Thus, the quality of these images can be improved using AI methods for use in better and more accurate diagnosis by physicians. Currently, the scientists have used this technique on small animals and trained the machine learning algorithms with the resultant images. Their next step involves applying this method on human patients ("Artificial intelligence improves biomedical imaging," 2019).

This article clearly portrays this AI development to its audience who are unfamiliar with this field. It begins by explaining about Optoacoustic and then moves on to how to the researchers implemented AI in this method. The researchers initially used a high-end device with 512 sensors to develop superior-quality images. These images were then fed to artificial neural network for learning the features of the high-quality images. Later, the sensors in the device were reduced to 128 or 32 to produce resultant images. These distorted low-quality images were then corrected by the trained neural network to produce images nearly similar to the images produced with 512 sensors.

The development of AI in improving the biomedical imaging is fairly portrayed in this article. The article provides accurate details without trying to oversell its significance. It mentions clearly that though this AI development is a success, it still has few limitations. The optoacoustic imaging uses light waves to produce images which cannot fully penetrate the human body. Hence, this technique can be used for producing images of tissues to a depth of a few centimeters beneath the skin.

4.2 New article on AI development – Negative light:

I was going through The Guardian news and found this article 'The racism of technology and why driverless cars could be the most dangerous example yet'. The researchers at our esteemed Georgia Tech found that the detection systems and sensors used in driverless cars were ineffective in recognizing pedestrians with different skin tones. They were better at spotting people with light skin tone and 5% less accurate on average in detecting people with darker skin. The researchers undertook this study after noticing higher error rates by these systems for certain demographics (Hern, 2019).

Autonomous cars are a huge development in the field of AI and machine learning. The machine learns from the data that we provide and hence it is important to include data from all demographics as training sets. Biasing such data will only pose threat to the life of humans. The article provides this information to the mass audience for them to understand the underlying technology and how it is used. Those unfamiliar with artificial intelligence are able to get a gist of it from this article data.

The article tries to portray the actual development of AI in the field of autonomous vehicles. It speaks about the advanced development of the various image-recognition systems. In addition to this, the article provides examples of other similar systems that failed notoriously to be unbiased, for example, Kodak color film and motion-activated taps and dryers. The article mentions about an initiative to switch to camera-only systems because of the huge cost savings. These systems also need to be properly studied and analyzed if they are safer than human drivers. But, the concern remains the same, if these systems will stay unbiased and safe for pedestrians regardless of their skin tones, race or gender.

5 REFERENCES

1. The Toronto Declaration: Protecting the rights to equality and non-discrimination in machine learning systems. (2018, July 10). Retrieved from https://www.hrw.org/news/2018/07/03/toronto-declaration-protecting-rights-equality-and-non-discrimination-machine

2. Artificial intelligence improves biomedical imaging. (2019, September 30). Retrieved from https://techxplore.com/news/2019-09-artificial-intelligence-biomedical-imaging.html
3. Hern, A. (2019, March 13). The racism of technology - and why driverless cars could be the most dangerous example yet. Retrieved from https://www.theguardian.com/technology/shortcuts/2019/mar/13/driverless-cars-racist
4. Self-driving cars are be more likely to drive into black people, study claims. (2019, March 6). Retrieved from https://www.independent.co.uk/life-style/gadgets-and-tech/news/self-driving-car-crash-racial-bias-black-people-study-a8810031.html