CS 7637: Homework 2

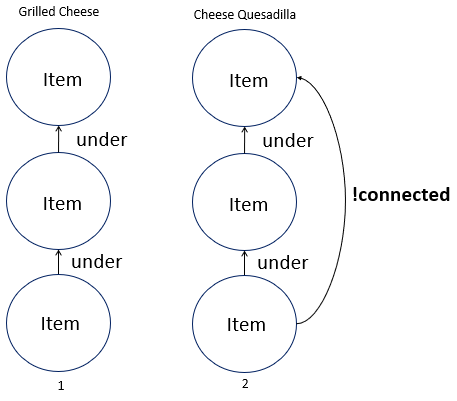
Zachary Todd  
ztodd3@gatech.edu

***Abstract—***This paper contains responses to the questions posed for homework 2 in CS 7637.

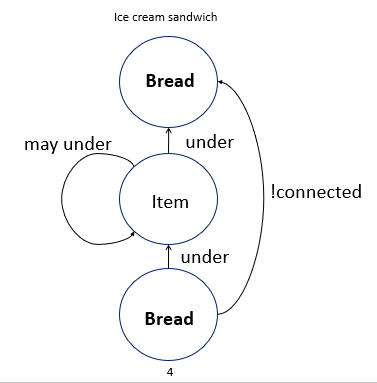
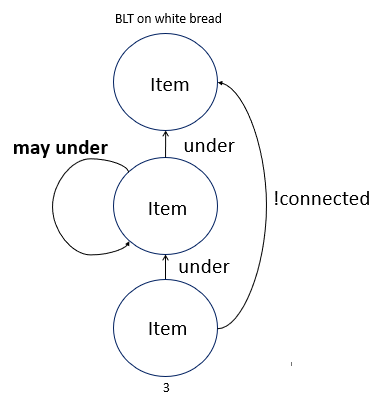
# 1 To be, or to not be a sandwich

The following food items **are** sandwiches: BLT on white bread; hamburger; turkey and swiss on potato roll; tuna salad on brioche; chip butty; grilled cheese; turkey hero; veggie burger; egg & cheese biscuit; patty melt; sloppy joe

The following food items are **not** sandwiches: meatball sub; chicken wrap; burrito; ice cream sandwich; ice cream taco; vada pav; toast; cheese quesadilla; toaster strudel; Klondike bar; buttered biscuit; gyro; sushi rolls; calzone;



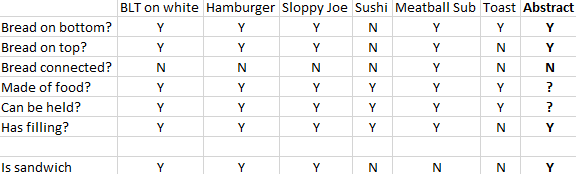
***Figure 1—***(1) – the original model created from presentation of a grilled cheese (sandwich). (2) – new model generated by the forbid-link heuristic when presented with a cheese quesadilla (not sandwich).



***Figure 2—*** (3) – new model generated by climb-tree heuristic when presented with BLT on white bread (sandwich). (4) new model generated by minimize-set heuristic when presented with Ice cream sandwich (not sandwich).

Figures 1 and 2 show the models generated through incremental concept learning when presented with different examples. The bolded parts are the portion of the model that changed at each iteration. Presenting the model with a sloppy joe, tuna salad on brioche, and buttered biscuit would make significant changes to the model. The first two would make the model take into consideration sandwich fillings which are more homogenous instead of the current structure which expects the insides of the sandwich to be stacked. The buttered biscuit would change the model by requiring the insides of a sandwich to not be condiments.

***Table 1—*** Table showing parameters to determine whether something is a sandwich utilizing classification. Abstract sandwich classification is shown on the far right.

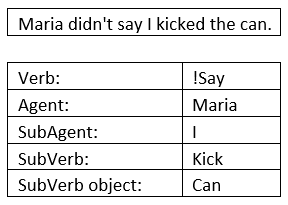


**Neither** of these two models would consider a hot dog a sandwich. The incremental learning model would break under the “!connected” requirement and the classification model would also break under the “bread connected?” requirement. A case-based reasoning model would likely compare a hot dog to meatball sub as both types of bread are long and connected at one end with fillers in the middle. It would thus return the same value it had for the meatball sub: **not** a sandwich.

# 2 Dissecting sentences

For the sentence “Maria didn’t say I kicked the can,” an AI agent would first recognize that the verb “say” will generate a subsentence. The AI would construct a frame for that subsentence and look for the verb there. It can then go back to the main sentence and extract that the main agent is Maria and there is a negation of the verb. The subsentence should be formed into another frame, but for brevity, I included it in the frame shown in Table 2.

***Table 2—*** Table showing frame representation of the given sentence.

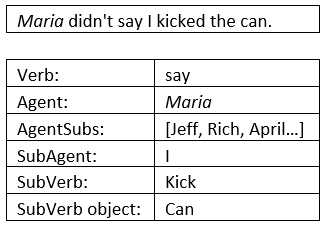
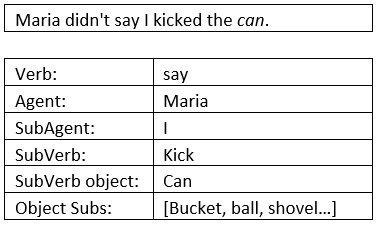


## 2.1 Emphasis

Emphasis on a word in the sentence implies that the word used is not the correct one. For instance, if the emphasis in the above sentence was on “Maria,” it would be implied that someone other than Maria said I kicked the can. The AI agent could attempt to understand this by knowing there is some other agent that it could insert into the sentence to make it correct. When it does this, note that it would also have to negate the verb of the sentence. For instance, if the emphasis of the sentence was on Maria, the agent could reason that the correct sentence may be something like “[Agent] said I kicked the can” where [Agent] is some other agent from the AI’s knowledge base.

This would require the agent to pick up on emphasis (italicized character recognition) as well as being able to look through its memory to find appropriate words to substitute in.

***Table 3—*** (Left) Table showing frame representation with emphasis on “Maria.” AgentSubs is an array of possible agent substitutes for Maria retrieved from memory of agents associated with Maria. (Right) Table showing frame representation with emphasis on “can.” Object Subs is an array of possible object substitutes for can retrieved from memory of objects like can.

The agent could further understand the sentence by distinguishing figurative and literal forms of speech. In this sentence, “kick the can” could be figurative or literal. The agent could use context of the conversation to know which of the two was more appropriate. This would be done by seeing if a literal can had been discussed yet. If not, the figurative form would be the most likely explanation for the word here.

# 3 The toronto declaration

The preamble of the declaration is a basis for the purpose of the document. It talks about how there are certain “universal, binding and actionable laws and standards” that prevent discrimination and how they should be applied to AI systems (Bacciarelli). It also briefly mentions that systems which make decisions or process data can have an impact on human rights through discrimination.

## 3.1 Using the framework

This declaration mirrors the thought process applied to human rights law. This internationally accepted standard requires mechanisms be put in place to protect individuals and that these mechanisms be able to be influenced by “academic, legal, and civil society experts” (Bacciarelli). Both governments and private sector actors must act to prevent discrimination based on human rights. These same expectations carry over to machine learning systems. Furthermore, this section of the article highlights how inclusion, diversity, and equity are essential to non-discrimination.

## 3.2 Duties of states

“States bear the primary duty to promote, protect, respect, and fulfil human rights” (Bacciarelli). The duties of states section focuses on how this concept carries over to machine learning systems. This includes updating existing documents and regulations to apply to more modern technologies. States must identify risks of discrimination, ensure transparency and accountability of public sector use of machine learning, and enforce oversite of these systems. Finally, states must not just take reactionary measures against discrimination, but must promote equality through proactive measures.

## 3.3 Responsibilities of private sector actors

Private sector actors will be held accountable for their actions as they would for violation of human rights through other means. They have a due diligence framework they are responsible for which includes identifying potential discriminatory outcomes, taking action to prevent discrimination, and be transparent about their attempts and processes in ML systems.

## 3.4 Right to an effective remedy

Individuals have a right to justice when they feel they have been discriminated against by the implementation of ML systems. This requires that states and private sector actors must provide an avenue for feedback and redress.

## 3.5 Trade-offs

The Toronto Declaration inherently causes some negative effects. One of which is extra red tape which states and companies much abide by to implement their products. This causes processes to slow and inevitably hurts profits. Additionally, ML algorithms may not be able to pick the most optimal algorithms if they are not allowed to discriminate. For example, a ML system for a bank may recognize that customers from a certain zip code are more likely to pursue high value loans. Maybe this is due to the income from that zip code being a lot higher than surrounding areas. The ML system would then try to push loans to that demographic while failing to do the same to clients from other locations.

However, if the document were to be withdrawn, there would be far worse outcomes. Companies could use the lack of oversite to their advantage and fashion AI agents that adjusted costs of products on their website based on the person who was logged in. Government agencies would also have large incentives to capitalize on the endless possibilities of ML.

## 3.6 My stance

The intent of this document is excellent. It is aimed at protecting individuals’ rights in the modern era. I believe the benefit of this protection far surpasses the negative impact it has on businesses and government agencies. I would keep the entire document and don’t have too much to add. The only thing I can think of adding would be a few more examples of ways both private entities and state agencies that implement AI systems could be transparent as well as offer means for effective feedback and grievances.

# 4 A discussion of brains and bodies

I will assume that the reader of this paper has read about the procedures performed on Ethan, Sofia, and Akhila described in HW 2. I will be using the definition of *identity* as “the distinguishing character or personality of an individual” (Identity). Furthermore, I will be using the definition of *consciousness* as “the state of being characterized by sensation, emotion, volition, and thought” (Consciousness).

## 4.1 Ethan

Ethan had all his body parts, other than his brain, replaced incrementally. At the end of his transformation, Ethan is still Ethan. This is because the brain is the portion of his body which houses his consciousness, the essence of Ethan. His biological body is merely the housing he was born with and his identity does not change when his housing does.

In his new body, Ethan still has the same character and personality. He still interacts with the world in the same way, its just that his means of doing so has changed. Before, he used flesh and bone to interact with the world, but now his brain sends signals to his robot body to interact. This same conclusion would be reached if Ethan replaced his body parts all at once.

## 4.2 Sofia

Sofia had her entire brain replaced incrementally. In contrast to Ethan, along this procedure, Sofia loses her identity and a new individual is born. This new being is a **copy** of Sofia, not Sofia herself. This creation of a new being occurs each time the brain stem, thalamus, or cerebral cortex are replaced.

“The brain stem, thalamus, and cerebral cortex are the necessary components for the generation of consciousness” (Sack). By this reasoning, as soon as one of these organs are replaced, Sofia does not have the same consciousness and thus changes identity. Even though these parts are scanned, and an exact (robotic) copy is substituted, they are two distinct items and thus generate two distinct beings. This is remarkably similar to cloning, where a copy is made from an original, but in the case of Sofia, the original ceases to exist. Sofia would reach the same outcome of change in identity if these changes to her brain were all conducted at once.

## 4.3 Akhila

Akhila had both her body and brain replaced incrementally. By the same reasoning I used for Sofia, replacing Akhila’s brain is what ultimately creates a new individual. Having a new body does not change Akhila’s identity, it just changes how she is able to interact with the world. Replacing Akhila’s body parts and brain all at once would lead to the same outcome.

# 

# 5 REFERENCES

1. Bacciarelli, Anna, and Joe Westby. “The Toronto Declaration.” Toronto Declaration, Amnesty International, 16 May 2018.
2. “Consciousness.” *Merriam-Webster*, Merriam-Webster, www.merriam-webster.com/dictionary/consciousness.
3. “Identity.” *Merriam-Webster*, Merriam-Webster, www.merriam-webster.com/dictionary/identity.
4. Sack, Georgeann. “What Brain Structures Are Required for Consciousness?” *Medium*, Awake & Alive Mind, 20 Aug. 2019, medium.com/awake-alive-mind/what-brain-structures-are-required-for-consciousness-33ba8890977e#:~:text=Therefore%20the%20brain%20stem%2C%20thalamus,system%2Dwide%20failure%20of%20consciousness.