CS 7637 Homework 2:

Abhijeet Gaurav
agaurav@gatech.edu

Question 1

This is a highly debated topic on what is considered a Sandwich. I consider Sandwich to be a portable food consisting any substantial (only butter/spices/sauces not allowed) edible thing wrapped exactly on two sides with two separate distinctly visible starch/carbs based pieces made of grain and can be placed horizontally on a table with both starch/carbs pieces in the same horizontal orientation as the table. Note: The edible thing in between should not be the same as the starch/carbs pieces between which it is wrapped.

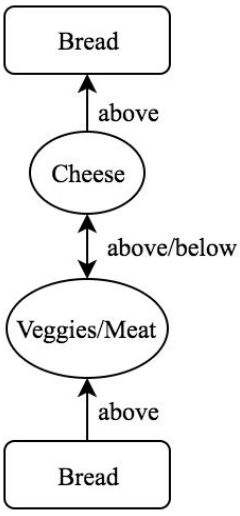
Based on the above/my definition, I would consider the following to be sandwich:

BLT on white bread, Hamburger, Meatball sub, Tuna salad on brioche, Chip butty, Ice cream sandwich, Grilled cheese, Turkey hero, Vada pav, Veggie burgers, Egg & cheese biscuit, Patty melt, Sloppy joe.

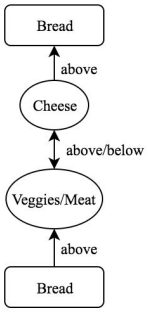
And I consider the following not be in the sandwich category:

Turkey and swiss on potato roll, Chicken wrap, Buttered biscuit, Burrito, Ice cream taco, Toast, Cheese quesadilla, Toaster strudel, Klondike bar, Gyro, Sushi rolls, Calzone.

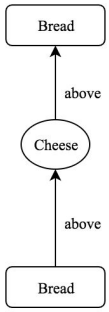
Increment concept learning



For increment learning let's start with this concept of a sandwich (which here is taken from the concept of a Hamburger). In a hamburger, we have Cheese, meat, above a bread and then again another piece of bread encloses the Hamburger on the top.
However, this concept would not represent sandwich such as Grilled cheese, Ice cream sandwich etc. Therefore, we would do incremental learning on this concept in a few iterations. We'll first do two positive examples with principles of *'Generalization to ignore features'* & *'Generalization to abstract features.'* and then will do two negative examples with principles of *'Specialization to Exclude Features'* & *'Specialization to Require Features'*

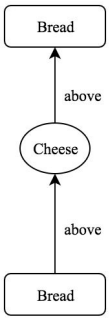


Current concept

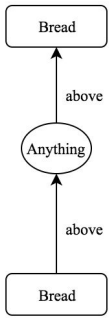


Grilled cheese

The right concept represents the concept of a Grilled cheese (which we consider as a sandwich) in which a cheese is enclosed within two pieces of bread. Since, this new concept will also represent a Hamburger, we'll update the previous concept with the new concept of the Grilled cheese by using *'Generalization to ignore features'*.
(Note that there can be a case where Hamburger might not have cheese in it but for the scope of this assignment, I'll choose Hamburger to have grilled cheese.)

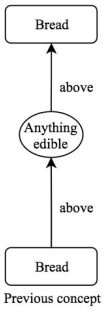


Current concept

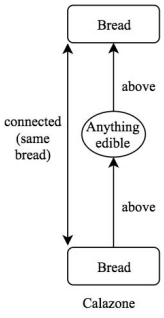


Veggie Sandwich

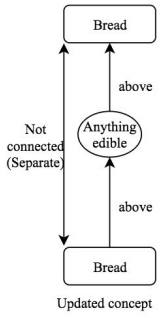
Now, using increment learning, we want to make our concept even more precise. Let's suppose that we have a veggie sandwich (which we consider as a sandwich) but doesn't have cheese in it. In this case our concept will fail. Therefore, we once again would like to generalize our concept from just considering Cheese to **anything edible** that can be put inside the sandwich. Therefore, our updated concept would be the one on the right. Here we generalize by using the concept of *'Generalization to abstract features.'*



Previous concept

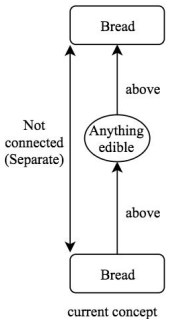


Calazone

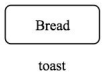


Updated concept

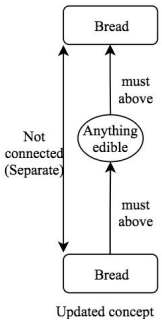
We have a calzone (which is not a sandwich) but the previous concept would mark calzone a sandwich. In calzone we have one bread the wrapping the meat. Therefore, we in the updated concept, we mark the bread to not be connected and the two breads wrapping the sandwich should be separate. Here, we specialize by using the concept of *'Specialization to Exclude Features'*.



current concept



toast



Updated concept

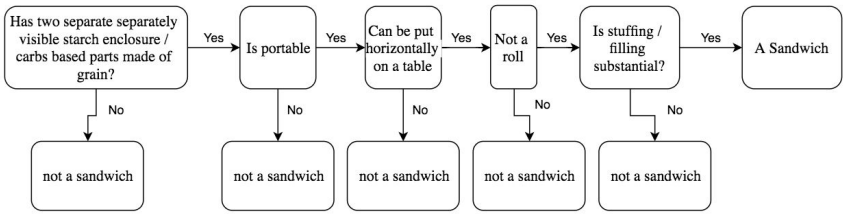
We have toast (which is not a sandwich) but the previous concept would mark it as a sandwich because the we don't a requirement of bread 'above' anything edible 'above' bread. Therefore, we'll specialize the concept by using the principle of *'Specialization to Require Features'* to always have 'must above' instead of 'above'.

The above model is quite robust but would fail on 'ice cream sandwich'. To make the above model work for 'ice cream sandwich', we'll need to update model to use 'starch/carbs pieces' instead of 'bread' by using the principle of *'Generalization to abstract features'*.

Classification approach:
Parameters: 1) Has two separate distinctly visible enclosing starch/carbs based pieces(eg. bread/biscuit etc.)? 2) Can be put horizontally on a table with both starch/carbs pieces in the same horizontal orientation as the table? 3) Is portable? 4) Is stuffing/filling substantial (only butter/spices/sauces not allowed)? 5) Not a roll? 6) Has veggies? 7) Has cheese? 8) Has bread? 9) Has meat? 10) Is spicy? 11) Has chicken? 12) Has tuna? 13) Has meatballs? 14) Has beef? 15) Has pork? 16) Has salami? 17) Has shrimp? 18) Has lamb? 19) Has turkey? 20) Has bananas? 21) Has butter? 22) Is vegan? 23) Is vegetarian? 24) Has jam? 25) Has egg? 26) Has apple etc. and numerous other similar classification questions.
I'll consider following sandwiches for answering the above questions. Due to space constraints, I won't rewrite questions.

a) Ice cream sandwich: 1)Yes 2)Yes 3)Yes 4)Yes 5)Yes 6)No 7)No 8)No 9)No 10)No 11)No 12)No 13)No 14)No 15)No 16)No 17)No 18)No 19)No 20)No 21)No 22)No 23)Yes 24)No 25)Maybe 26)No
b) Grilled cheese: 1)Yes 2)Yes 3)Yes 4)Yes 5)Yes 6)No 7)Yes 8)Yes 9)No 10)No 11)No 12)No 13)No 14)No 15)No 16)No 17)No 18)No 19)No 20)No 21)No 22)Maybe 23)Yes 24)No 25)No 26)No
c) Veggie burgers: 1)Yes 2)Yes 3)Yes 4)Yes 5)Yes 6)Yes 7)Maybe 8)Yes 9)No 10)Maybe 11)No 12)No 13)No 14)No 15)No 16)No 17)No 18)No 19)No 20)Maybe 21)Maybe 22)Maybe 23) Yes 24)Maybe 25)No 26)Maybe
d) Egg & cheese biscuit: 1)Yes 2)Yes 3)Yes 4)Yes 5)Yes 6)No 7)Yes 8)Yes 9)Yes 10)No 11)No 12)No 13)No 14)No 15)Yes 16)No 17)No 18)No 19)No 20)No 21)Yes 22) No 23)No 24)No 25)Yes 16)Maybe
e) Sloppy joe: 1)Yes 2) Yes 3)Yes 4)Yes 5)Yes 6)Yes 7)Yes 8)Yes 9)Yes 10)Maybe 11)No 12)No 13)No 14)Yes 15)Yes 16)Maybe 17)No 18)No 19)No 20)No 21)Maybe 22)No 23)No 24)No 25)No 26)No
f) Vada pav: 1)Yes 2)Yes 3)Yes 4)Yes 6)Yes 7)No 8)Yes 9)No 10)Yes 11)No 12)No 13)No 14)No 15)No 16)No 17)No 18)No 19)No 20)No 21)Maybe 22)Maybe 23)No 24)No 25)No 26)No

Finally, abstracted classification of what a sandwich is given below:



Hot dog is not a sandwich. If we look through incremental learning perspective, I have a condition that breads in the bun should be separated. However, in a hotdog, the bun is joint. If we look through the classifier approach, hot dog won't be considered a sandwich because it can't be put horizontally on a table. If we look into case based reasoning, the closest sandwich to Hot dog would be Sloppy Joe as they are quite similar in terms of the ingredients (meat).

Question 2

'I never said Amy planted that seed.'

The above sentence can be understood by AI agent by semantically analyzing the sentence. We'll first build a frame representation for understanding the sentence.

Agent	I
Verb	Say
Adverb	Never
Co-Agent	Amy
Co-Agent's verb	Plant
Thematic object	Seed

To be precise, this is a complex sentence which can have two (or maybe more if we talk in the figurative sense) meanings. Thus, first building a knowledge representation in terms of frame and may be using some scoring mechanism will let the AI agent give more context on how to tackle the multiple meanings problem.

Frame (Frame I) representation with stress on 'Amy' i.e. I never said *Amy* planted that seed):

Agent	Amy

Verb	Plant
Adverb	Never
Thematic object	Seed
Co-Agent	I
Co-Agent verb (not useful)	Say

Now, let's consider the frame representation for '*I* never said Amy planted that seed':

(Frame II)

Agent	I
Verb	Say
Adverb	Never
Thematic object	'Amy planted the seed'

Note here that thematic object is a whole sentence 'Amy planted the seed'.

The meaning from Frame I is very different from Frame II. From Frame I, AI agent is associanting adverb 'never' with the verb 'plant' and Amy because, in the sentence, the primary focus is on Amy as she is being considered as the primary Agent. So, the entire sentence revolves around her and thus means that the 'speaker (co-agent) said that a seed was planted, but not that Amy (primary-agent) was the one that did it. However, in Frame II, 'I' (current speaker) is the primary agent and adverb 'never' is being associated with the verb 'say' and thus the meaning is different which is 'Amy planted that seed' was spoken previously, but the sentence was not spoken by the current speaker.

If the AI agent is a text based NLP, which accepts text as input, we can maybe build a scoring mechanism where agent in italic font would be given a higher score and thus, the primary agent is the one which is written in italics. After getting the primary agent, our AI agent can thus build Frame I vs Frame II accordingly and thus understand the meaning of the sentence. The other way is to give AI agent more data, like in a form of a second sentence which can help it

in understanding the context. The other way to make this sentence more clear is to add a few words to the same sentence avoid ambiguity.

The AI agent can also interpret this sentence figuratively (as opposed literally as it did in the above example). The key here is to provide more data to the AI agent. As for example, if the sentence was say: 'I never said Amy planted that seed. The plant that has grown from that seed, seems to be quite dangerous.' This sentence truly tells the AI agent that we are talking about a plant which seems to be poisonous has grown from a seed. So, this added extra sentence makes the literal meaning to AI agent very clear as the meaning is revolving around plant which was created by seed and so on and it belongs to the context of plants/trees/seeds/fruits etc. Thus, the AI agent can give the literal meaning of seed a higher score and move forward with the interpretation. Similarly, if we want to talk about the figurative meaning of the sentence, we should feed more information in the form of sentences to the AI agent. One such Eg could be: 'I never said Amy planted that seed. The whole idea of the project, was created by me and Amy had no contribution in that.' This sentence makes it very clear that the number of occurrences for the word 'seed' was low, and the further discussion has been on ideas. Since, seed can be interpreted and related to ideas, the AI agent can quickly understand the notion of the word 'seed' here. Thus, the AI agent can give the figurative meaning of the seed a higher score and move forward with the interpretation.

Question 3

Preamble: The Toronto declaration focuses mainly on safeguarding human rights (especially right to equality) as the machine learning systems advances. Lately, there has been concerns on the issues pertaining to intentional or unintentional biases against certain individuals or groups of people which might happen because of the advancement in machine learning systems. Thus, Toronto declaration mentions that there is a need to address the issues on how these systems will affect human rights and who will account for human rights violations. The Toronto declaration refers to the current international human rights law and standards framework to make sure there is no discrimination (by adopting and implementing necessary safeguards) as it beleives that machine learning systems can quickly reinforce or change power structures on an unprecedented scale.

Using the framework of international human rights law: The Toronto declaration ensures that concrete steps should be taken if potential human rights violations/ harm is created or facilitated because of ML/AI systems. It mentions that government and private/public sector should examine and address risks associated with ML systems. It expects government to take binding and adequate measures to protect and promote the rights and at the same time requires entities to put mechanisms to account those who are responsible for any harms that impact human rights.

Duties of States: States should ensure that the safety measures are continuously updated so that they can take into account the risks and address them accordingly. Also, it is a duty of state to refrain from using/requiring private sector to use tools that leads to discriminatory outcomes. State should regularly monitor risks and identify them. Any tools to be used must be assessed to make sure that it doesn't contain bias. States should also take measures to mitigate the identified risks through impact assessments. The use of any type of ML systems by the states should be transparent and accounted. They should not use any black-box system, and should regularly audit and publicly disclose, places where ML systems are used in the public sphere. State should also enforce oversight by involving diverse hiring, engage in consultations on design/implement/review of ML systems; ensure proper training in human rights and data analysis for officials, ensure mechanisms for independent oversight by judicial authorities

Responsibilities of Private Sector Actors: Private sectors must respect human rights and make sure that deployed ML systems follow human rights framework. The core concepts that should be followed by these actors include: 1) Identify discriminatory outcomes. 2)Take action to mitigate, prevent discrimination and track responses. 3) Always be transparent on the efforts required to identify, prevent and mitigate against bias in ML systems.

Right to an Effective Remedy: Companies and private sector actors should make sure that victim individuals and groups have access to quick remedies in case the ML system violates human rights. Those individuals/private entities who are found to be responsible for violations must be accounted. Since, sometimes because of the opaque nature of ML systems, private entities may not be able to explain the decision making process. However, the individuals should ensure that victims get access to meaningful remedies which is required. It is thus a duty to: a)have a proper standards of due process to be carried out for ML systems

deployed in the public sector. b) Cautious use of ML systems in judiciary because of risks to fair trials. c) Clearly outline individuals who should be legally responsible for decisions made by ML systems.

The Toronto declaration is just a guideline and is not legally binding. Therefore, in the short term, I don't see any tradeoffs. In the long term, assuming it becomes binding, I see many innovations and opportunities lost. The innovations include companies focusing on safe use of AI/ML systems and to make sure that the systems developed are human-centric. The other one is companies and states ensure ML systems being secure and sticking to human rights. This declaration encourages companies to not build any technologies that would go against 'right to equality'. Current some AI agents are significantly biased. Eg: COMPAS algorithm which currently predicts the probability of criminal reoffending. The COMPAS has a racial bias against people of color and marks black defendants as posing higher risk and reverse for white defendants. Another example is: Google Ads has a racial bias and is found to show high income jobs to men more than women. Another future eg: Voice assistants like google now have become very common nowadays. Now, tvoice assistants wouldn't be able to discriminate and talk with people differently by say how rich they are. Another example is say, internet companies build an ML model to start giving a taste of very high speed internet on promotional basis to all customers who they think will be able to afford their new expensive high speed plan. However, this won't be allowed now as the companies won't be able to discriminate against people in any way. Some opportunities lost might be for companies who in future might be very sceptic to new ML systems like superintelligence which are innovative but at the same time might discriminate people in some way. Also, most ML models are trained from collected dataset (which have bias) which in turn makes the ML model bias. So, companies might be sceptical on launching any new innovative ML systems fearing that the bias shown can lead to potential fines which in turn would lead to slower growth in ML adoption.

Toronto declaration is a good step in making sure that we develop AI in a safe way and human centric. I agree with most of stance on Toronto declaration which makes sure that ML systems follow human rights. I mostly agree with cautious use of ML systems in judiciary because of risks to fair trials which is a huge problem that people of color are facing because of COMPAS algorithm. I don't agree with the right to equality in every aspect as it might impact personalization of ML systems to a very great extent. I would keep almost all the

clause of Toronto declaration except would tone down a bit on 'right to equality' to allow better personalization. I would also add a clause that overall 3% of development of ML systems by a company can be biased and should not be regulated. Instead, users should be made aware that they are using a biased ML system and then should be allowed to use such systems if they want to do so. This would allow companies to continue doing cutting edge research and would not slow down advancement in ML/AI. I would also add a clause to ban and regulate the use of ML/AI in weaponry and cyberweapons.

Question 4: AI in positive light: According to Forbes Technology Council, AI will have an enormous positive impact on the human society. AI has enhanced our throughput in the automobile industry. From voice assistants to self driving cars, AI has had a huge way. It's unbelievable that in few years, we won't need to drive our cars at all. Also, humans are not very good at tedious tasks. Since, AI has resulted in a lot of automation, machines have taken most of the manual tedious tasks and has allowed us humans to focus on what we want to do. AI has freed humans to do what they do want. Some people worry about job losses because of AI automation. Although, according to the Forbes, there will be some loss of jobs because of automation, much more number of new jobs will be created. There will be evolution of job in the job market in which people will work better with the help of AI. In fact, AI will enhance our lifestyle. AI will create more efficient businesses. The mundane tasks like answering emails, data entry will be done in future by AI powered intelligent assistants. AI powered smart home will reduce energy consumption and would increase security. Marketing of businesses will become more targeted because of better personalization. Healthcare will become better too with AI powered health diagnosis. Moreover in the sector of Healthcare, AI will supervise learning for telemedicine. Eg: an image based AI diagnosis of medical condition will become the norm and would allow comprehensive development of telemedicine. Overall, AI will elevate the condition of mankind. Since, AI agents are efficient, they can solve more problems, answer more unsolved questions and has ultimate computation power, with an even more improvement in AI, human quality of life will increase. It will benefit multiple industry from facial recognition to digital content analysis and accuracy in identifying patterns which will be used in academic research, technology applications, health sciences etc.

The article does a fantastic job in explaining positives of AI. It tells unfamiliar audience a diverse perspective on the AI by exploring different fields from

automobile industry to healthcare industry to intelligent assistants to smart homes to Telemedicine to Facial recognition etc.

The article portrays the AI actual development and contribution fairly. It doesn't oversell its significance.

AI in negative light: Bernard Marr in his article on AI has talked on a diverse range of negative impacts. The first one he mentioned is the AI bias. AI bias can be created by either intentionally/unintentionally introducing them into the core algorithm or by training the ML model on the biased dataset. This leads to AI agent giving biased results which can have far fetching unintended consequences like Microsoft's Twitter bot that became racist. The second one is the loss of jobs. He believes that these new jobs that will be created because AI will require training and educating the future workforce which he believes is a challenging task. The third one is the shift in human experience. Bernard believes that once AI takes all the menial tasks, humans will significantly reduce the amount of time they spend at work. This will lead to more mental and physical issues in absence of a fulfilling job they had before. The fourth one is lack of consistent global regulation. Bernard believes that in the absence of a standardized and consistent global regulation in AI, the AI development will not be safe. According to him, this is happening today too. In EU countries, there is a robust regulatory approach to consent and transparency in the development of AI which is not present in the US and particularly and other countries. This has allowed some countries to use AI to do hacking. The fifth one is accelerated hacking. With the advancement in AI, there is a likely scenario that accelerated AI hacking acts such as phishing, delivery of viruses to software and taking advantage of AI systems to harm the world may make it a lot difficult for humans to recover from. The sixth one is AI terrorism. Bernard believes that in future, there might be a new AI-enabled form of terrorism to deal with. There will be new problems such as expansion of autonomous drones, introduction of robotic swarms, remote attacks, delivery of disease through nanorobots. The law enforcement would need to be adjusted to tackle these future problems. The article does a fantastic job in explaining negatives of AI. It tells unfamiliar audience a diverse perspective on the AI negatives by exploring different points from AI Bias to Loss of Certain Jobs to A shift in Human Experience to Global Regulations to Accelerated Hacking to AI Terrorism etc. The article portrays the AI actual development and contribution fairly. It doesn't oversell its significance.

REFERENCES

1. The Toronto Declaration: Protecting the rights to equality and non-discrimination in machine learning systems Policy Team - https://www.accessnow.org/the-toronto-declaration-protecting-the-rights-to-equality-and-non-discrimination-in-machine-learning-systems/
2. Discriminating Algorithms: 5 Times Ai Showed Prejudice Daniel Cossins -https://www.newscientist.com/article/2166207-discriminating-algorithms-5-times-ai-showed-prejudice/
3. 14 Ways Ai Will Benefit Or Harm Society Forbes Editors - https://www.forbes.com/sites/forbestechcouncil/2018/03/01/14-ways-ai-will-benefit-or-harm-society/#6ec0430a4ef0
4. What Are The Negative Impacts Of Artificial Intelligence (ai)? https://bernardmarr.com/default.asp?contentID=1827