CS 7637 - KBAI: Homework 1

Sergiy Palguyev

sergiy.palguyev@gatech.edu

Question 1: Start Wars Transport Problem

Semantic Network:

The state representation in the semantic network consists of a planet (oval), a ship (rectangle) and a transport vessel (plus sign shape). An arrow from the transport ship indicates whether the transport ship is docked at the planet or the main ship. Ray, Snoke, Kylo and Leia, are all colored pentagons with the first Initial inside the pentagon. At the state transition arrow, a triangle is drawn pointing either left or right to indicate whether the passengers are being transported to the ship (right) or to the planet (left). If the planet and ship are highlighted orange, that state has previously been encountered in the diagram. If the planet and ship are highlighted red, the state is invalid. If the planet and ship are highlighted green, We have reached the goal state of moving all characters to the main ship.

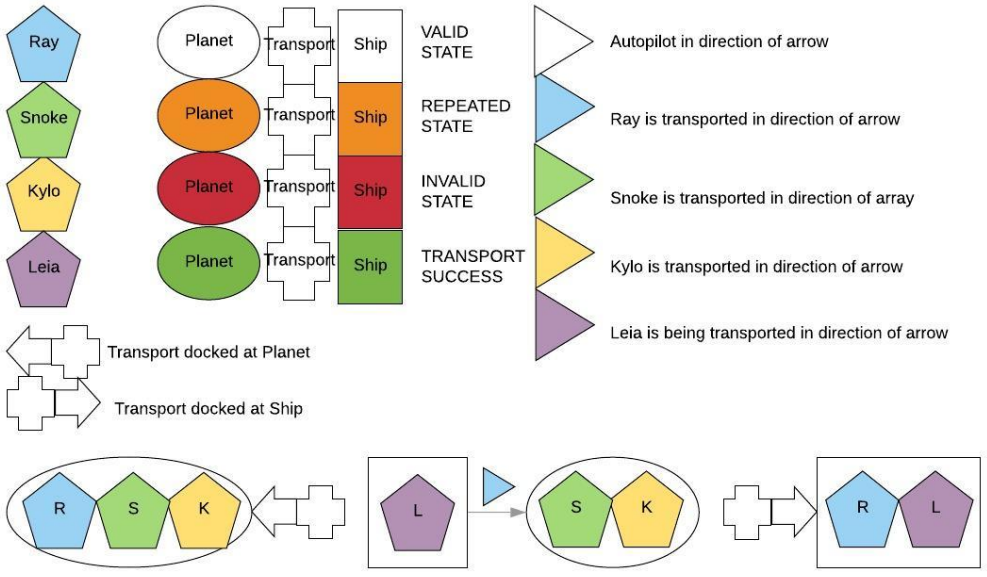

Figure 1: Semantic Network

Generate & Test:

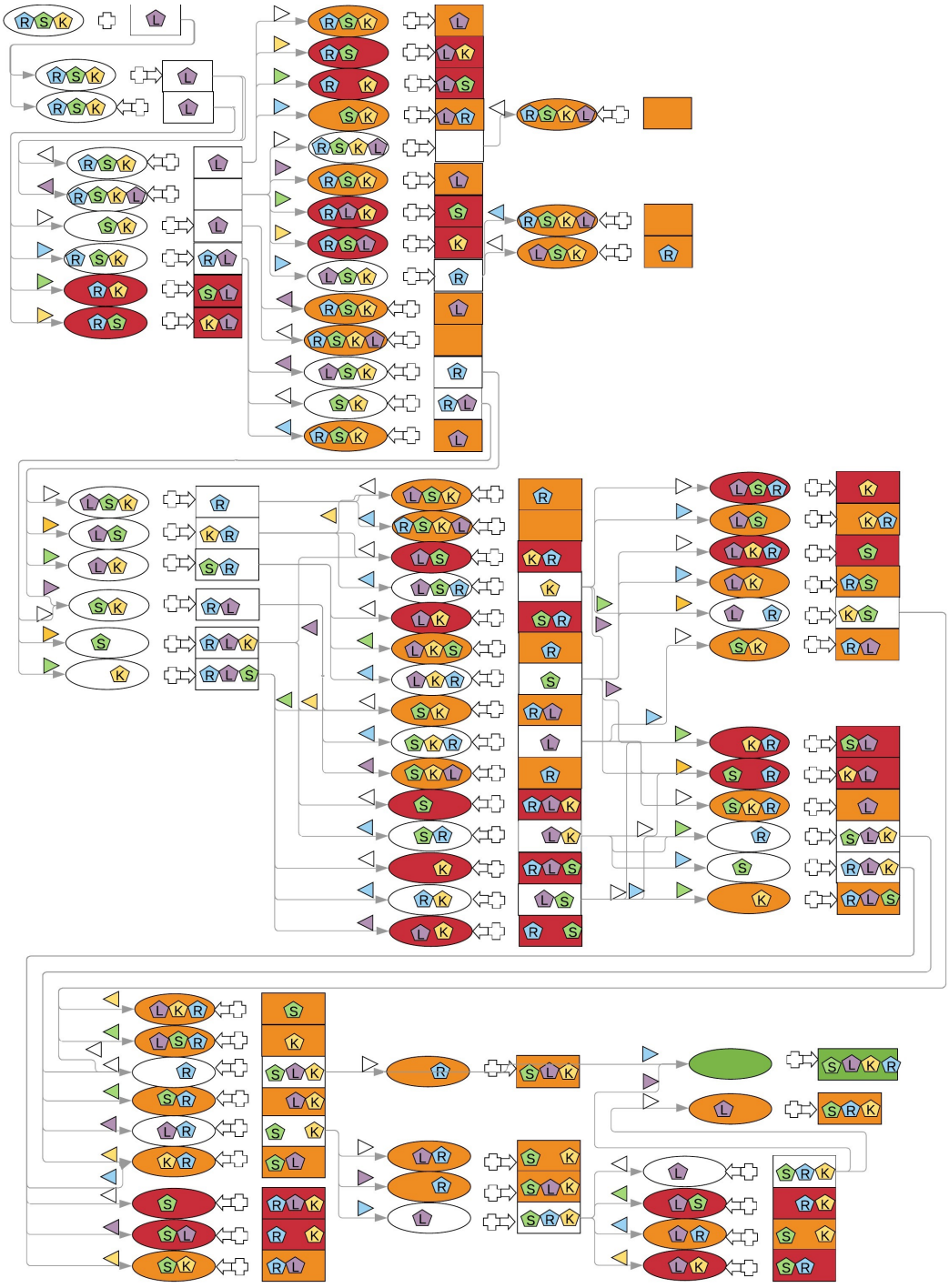

Figure 2: Generate and Test Results

Question 2: Artificial Intelligence Pros and Cons

AI Cynic:

Nick Bostrom is a prominent Philosopher and founding director of the Future of Humanity Institute at Oxford University (Bostrom 2019). Bostrom, although not a hard-lined cynic, engages in thought experiments about the end of the world caused by AI agents. Bostrom makes many assumptions and uses them as a premise to make the case for a dangerous AI to bring about the end of the world. His assumptions can be summarized in three main concessions. First, human intelligence is an accident and any intelligent life that may observe us would be unimaginably far more superior to us. Secondly, computing agents do not need to possess all capabilities of humans (soul, self-awareness, feelings, etc.) in order to surpass our intelligence and become dangerous to human existence. Thirdly, intelligent AI will have available, for it's use, two special advantages: computing overhang and recursive self-improvement. Computing overhand is the availability of cheap source hardware that would allow AI to execute as soon as the algorithms for a fully functioning AI are discovered. Recursive self-improvement is the idea that the only goals an AI will have is to survive, self-improve, and gather resources to perform the first two tasks (Bostrom 2014). It is these three premises that Bostrom uses as foundation for his claim that if left unchecked, AI will recognize humans as an impeding obstacle to one of its goals and must eliminate humanity in order to achieve success. Although skimmed, a subtle but profound notion that Bostrom claims is that AI does not need full human level intelligence in order to be effective in perceiving humanity as a threat and eliminating us. This brings the idea of annihilation by AI anywhere on the AI developmental spectrum, not only at full sentience as most fear.

AI Optimist:

Ray Kurzweil, is a Director of Engineering at Google and one of the most prominent futurists and proponents of Artificial Intelligence (Kurzweil 2019). Kurzweil's main argument is to consider the threats humanity has faced in the past century such as nuclear weaponry, incurable diseases, and bioterrorists. These events have come and gone and humanity is only thriving on the lessons learned from the mistakes made. Kurzweil points out that while nuclear war or

terrorism may be in the hands of the few, AI is in the hands of the billions of people working on developing the technology (Kurzweil 2014). Furthermore, the benefits of AI have been affecting the entirety of the world for years and will continue to do so as humanity pushes into the future. Incidentally, Kurzweil goes so far as to envision a point where man and machine will meld into a human existence augmented by Artificial Intelligence, expanding both longevity and our own intelligence (Kurzweil 2001).

Comparison:

Kurzweil and Bostrom are considering Artificial Intelligence from completely different points of view. Bostrom is a philosopher, contemplating and steering human perception of Artificial Intelligencein such a way so that the creators of this technology would consider all aspects of ethics and ramafications for every line of code. Kurzweil, on the other hand, is one of the prominent creators of the technology. He is a leader in implementing the logic for safety and security into the algorithms developed at Google, and thus has a first-row seat in understanding the capabilities and dangers of Artificial Intelligence.

Discussion of other perspectives:

Jaron Lanier, a prominent Computer Scientist and one of the most famous Virtual Reality developers, falls somewhat in the middle between Kurzweil and Bostrom. As an AI realist, Lanier has not doomsday predictions as Bostrom, but also has no desire to become a cyborg as Kurzweil. Lanier argues that machine intelligence has because a religion for many (Lanier 2019). Lanier argues that when we call a software “intelligent” we trust it and adapt our own habits to that of the machine when it, in fact, acts unintelligibly.

As a student of Machine Learning and Artificial Intelligence, the evidence presented to me so far, has shown unquestionable progress for humanity in the use of Artificial Intelligence. Although there are some inherent dangers in such powerful technology, believing that the future of humanity is an AI augmented world seems the topic of science fiction as well. I believe that the view of Jaron Lanier is best. The questions we ask, in order to solve the problems with Artificial Intelligence, are much more important than hypothesizing all possible future outcomes to irrelavant questions.

Question 3: General Data Protection Regulation

General Data Protection Regulation (GDPR):

The General Data Protection Regulation is a new regulation for data privacy for those living in the European Union. Specifically, GDPR adds the necessity of designer to include provisions to access, delete and make opt-in's more legible and easier to understand when users are navigating their accounts (GDPR 2018). Any Artificial Intelligence that may use personal data to customize experiences, must also include provisions to delete data sets from its database as well as from the final product if the user chooses to erase his/her account.

Company Rooted in Personalization:

Equifax, is a company dedicated to credit reporting services. Considering that an individual credit score cannot exist without personal information, the company's business depends solely on the customer information.

Evaluation:

In order for Equifax to comply with GDPR, certain criteria must be met. Namely, simplified opt-in directions, personal information access, breach notifications, and account deletion capabilities will all have to be implemented. More importantly, any marketing or sale of personal information to other business entities would also have to scrupulously tracked and redacted if any of the information shared is selected to be deleted by the original user. Although these compliance regulations are difficult to implement, executing on them may prove cumbersome as witnessed by the data breach which led to the release of millions of personal accounts information. Under GDPR, Equifax would fail the Regulation standard and could be fined either 20 million Euro, or 4% of annual global turnover (GDPR 2018).

Question 4: Artificial Intelligence in Pop-Culture

Positive Pop-Culture AI: "I, Robot" movie

"I, Robot" is a 2004 film about the future of robotics and Artificial Intelligence. In this futuristic world, man and machine live in harmony and peace with each other. Robots are bound by three simple laws which protect humans from harm, namely "A robot may not injure a human being, A robot must obey orders given to it by humans, and a robot must protect its own existence as long as it does not interfere with the First or Second law". The plot of the movie, and Artificial Intelligence super-computer, realizes that humans are, by nature, self-destructive. As such, the AI decides that in order to protect humanity, it must enact a totalitarian regime and possibly sacrifice some humans, in order to ultimately protect humanity. In a twist to the plot, the robot (Sonny) traveling with Will Smith's character has been given the ability to ignore the three laws, in addition to having dreams and thoughts of his own. This allows Sonny to help save humanity, and in the end become more human himself. As such, I see this film as a positive AI movie where the destruction brought upon the world was caused by humans interfering with natural AI development by constraining it within the three laws.

Negative Pop-Culture AI: "The Terminator" franchise

"The Terminator" is a franchise starting in 1984. Throughout the franchise, both humans and SkyNet (a power Artificial Intelligence) fight a war for survival. SkyNet is an Artificial Intelligence that exists in the cloud, without any one specific location which can be destroyed. As such, it is capable to organizing a nuclear war on all continents of earth as soon as it becomes self-aware and realizes the danger that humans pose to its existence. As part of that war, certain machines, created by SkyNet, are captured by the humans and reprogrammed to serve the humans in the war. Throughout the struggle for survival, Artificial Intelligence is always the enemy as the heroes of the movies are forced to resort to either crude weapons or simply hiding from disaster as the world is taken over by Artificial Intelligence.

Comparison:

Both "The Terminator" and "I, Robot" movies depend heavily on Artificial Intelligence and robotics in order to enact the will of the AI. However, "I, Robot"

AI is a program that is dedicated to fulfilling its pre-programmed goals of ensuring human survival. The misinterpretation by the AI of that goal leads to the conflict depicted in the movie. Subsequently, a free-to-think AI robot is capable to reason on behalf of humanity and save the world.

“The Terminator” on the other hand is completely diabolical from the beginning. Its goal of human destruction stems from the need for self-preservation and survival, evaluating human existence as the most incredible threat.

These two AI agents could not exist in the same world as their drivers for destruction stem from competing schools or thoughts on ethics of Artificial Intelligence.

Bibliography:

1. Bostrom, N. Nick Bostrom's Home Page. Retrieved from https://nickbostrom.com/
2. Bostrom N., Muehlhauser L., (2014). Why We Need Friendly AI.
3. Kurzweil, R. Kurzweil accelerating intelligence. Retrieved from http://www.kurzweilai.net/ray-kurzweil-biography/
4. Kurzweil R., (2014), Don’t Fear Artificial Intelligence
5. Kurzweil R., (2001), The Web Within Us: Minds and Machines Become One
6. Lanier, J., (2019) Mindless Thought Experiments (A Critique of Machine Intelligence)
7. GDPR Key Changes. Retrieved from https://eugdpr.org/the-regulation/