Homework 1
OMS CS7637: Knowledge-Based AI (Fall 2018)

Angela Ambroz
aambroz3@gatech.edu

Question 1

A snapshot of the semantic network representing the Kylo, Snoke, and Rey transport problem is below:

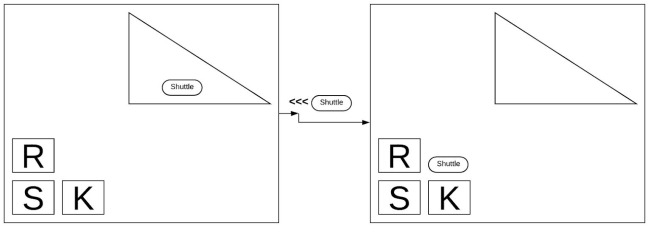


Figure 1. A semantic network, with an example transition between two states. In the left-hand state, Rey, Snoke, and Kylo are on the planet, while the shuttle is on the ship. In the right-hand state, the shuttle has arrived from the ship to the planet.

The full solution to this problem, using a generate and test approach on the above semantic network, is available below (please zoom in) and in Appendix 1 (in a slightly larger version).

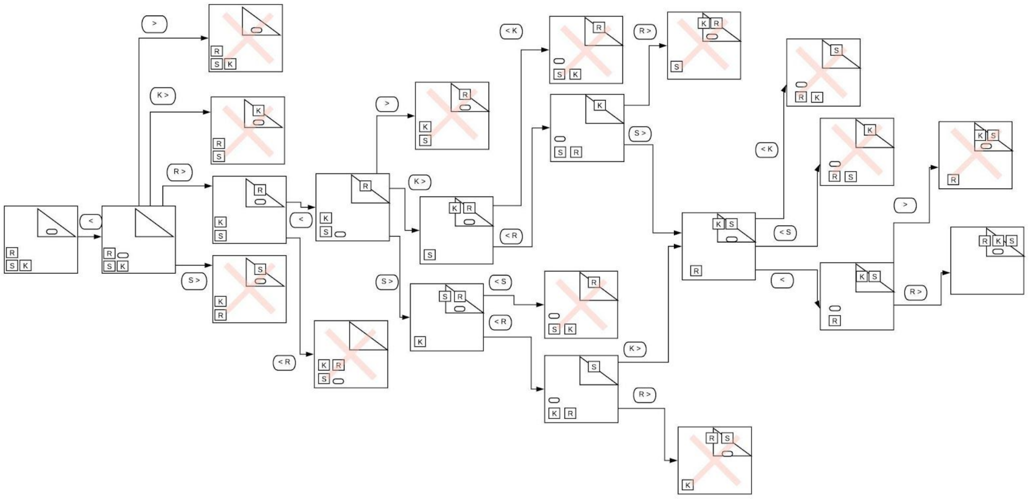


Figure 2. A semantic network, solved using a *generate and test* approach. The network can be read from left to right. Please zoom in; a larger version is available in the Appendix.

Per the instructions, the generator was smart enough to make only valid moves - that is, moves that did not violate the rules (such as the shuttle only traveling with zero or one passenger). However, it was not smart enough to make *strategic* moves, and it was not smart enough to make moves which resulted in valid states.

The tester scanned through each new state and eliminated those where that state had already been seen in the network (for example, moving Rey from the planet (state 1) to the orbiting ship (state 2) and back (state 3) would result in state 3 being eliminated), or states that were invalid by the rules (for example, Kylo and Rey being alone without the shuttle). The solution took eight steps.

Question 2

An AI cynic: Jaron Lanier

Jaron Lanier is a prominent Silicon Valley technologist and the author of several books of generalized “tech skepticism” (Lanier 2010, 2012, 2017, 2018). He has been a voice for caution against the excitement and optimism that has surrounded technologies like the Internet, social media and artificial intelligence. He has taken, in general, a strongly humanist standpoint: arguing against the standardization of human experience when mediated through an imperfectly-designed digital lens (Lanier 2010).

On artificial intelligence, he argues that its promises are overblown and even the debate surrounding it is “itself askew, and confuses us, and does real harm to society and to our skills as engineers and scientists” (Lanier 2014). In particular, rather than treating AI as a technical problem with modest near-term aspirations, Lanier notes that discussions surrounding it involve people “dramatizing their beliefs with an end-of-days scenario” (Khatchadourian 2015). He notes how this thinking is linked to the boom-and-bust cycle of “AI winters” - that is, overblown expectations cycling with disillusionment and existential fear. He also notes the hidden economic costs of big data and crowdsourced human labor which power AI algorithms (Lanier 2014; Intelligence Squared U.S. 2016).

An AI optimist: Ray Kurzweil

An exemplar of Lanier’s “religiously-minded” prognosticators is Ray Kurzweil, a long-time proponent of transhumanism - that is, the blending of human and artificial bodies to achieve longer lives and expanded consciousness (Kurzweil 1999, 2005). Kurzweil has long used the language of myth and religion in his predictions (one book is called *The Age of Spiritual Machines*); and his predictions are large-scale and positive. For example, regarding the use of nanotechnology to enhance consciousness, Kurzweil notes that, by the “2030s [...] our thinking then will be a hybrid of biological and non-biological thinking [...] enabling a qualitative leap in culture and technology” (Kurzweil 2014a). Regarding fears of AI as an existential threat, he notes that “we do have time to devise ethical standards” for managing this (Kurzweil 2014b).

Kurzweil furthermore addresses the more pragmatic economic anxieties, for example around automation and AI displacing human workers. He notes that, “for every job we eliminate, we’re going to create more jobs at the top of the skill ladder” - jobs in “industries and concepts that don’t exist yet” (Lev-Ram 2017).

A comparison and conclusion

Comparing Kurzweil to Lanier's arguments, there is disagreement on both the content (how will AI develop in the near and far future?) as well as the "meta" (how do we think about AI?). In his 2014 TED talk, Kurzweil likens the Singularity - when AI and human consciousnesses meld and expand - to an evolutionary leap akin to the final scene in *2001: A Space Odyssey,* auguring in a new age of "culture and technology" (Kurzweil 2014a). Lanier is more circumspect and less fanciful: his expectations for AI's development are more modest, and he argues that the grandiose way in which AI is discussed actually *slows* real, technological progress (Lanier 2014).

They somewhat talk past each other on the more immediate issue of AI's economic consequences - Kurzweil waves away the risks of economic restructuring ("we're going to create more jobs"), while Lanier notes the ways in which AI facades can mask crowdsourced *human* intelligence (e.g. Google machine translation leveraging large corpora of human-translated texts - see Lanier 2014), and that this system is unsustainable.

My opinion is closer to Lanier's than Kurzweil's - though I found Kurzweil's ideas influential many years ago, when I was beginning to get interested in AI. I think my process from Kurzweil-sympathetic to Lanier-sympathetic is also fairly typical: Kurzweil is often vague on the technical details, but grand in scope - good for a general or beginner audience. Lanier is meticulous in picking apart the realities of the technology, and the socio-economic factors that drive its development - good for the expert. In terms of AI realists, Andrej Karpathy, Director of Artificial Intelligence at Tesla, combines Kurzweil's visionary predictions with Lanier's technical scrutiny - yet he does this without "mythologizing" AI one way or the other. For this reason, his work is sometimes even *more* disturbing to me - since he precisely and rigorously outlines the enormous technological leaps that are imminent, yet leaves it up in the air as to how they could change our world (Karpathy 2017a, 2017b).

Question 3

The General Data Protection Regulation (GDPR) became European law on May 25, 2018 (Burgess 2018). It allows users to have more control over their personally identifiable information (PII) online; specifically, companies which collect online information on individuals must first obtain their consent. This moves from an opt-out to an opt-in system. Furthermore, individuals have a right to access their information, data breaches must be reported to the country's data protection regulator (based on where the company is located), and companies over a certain size must hire a data protection officer to ensure compliance with the regulation (Burgess 2018).

The GDPR has clear implications for companies which sell services personalized by artificial intelligence. For example, for users whose data becomes part of machine learning algorithms or deep learning networks, those users must explicitly consent to having their data collected and used in this way. In the model training process, as much PII as possible should be removed during the analysis - this is in order to protect against unexpected data breaches.

One company for which personalization is the essential service is Facebook. While other social media companies - such as Twitter or LinkedIn - also collect significant PII and offer personalized feeds, they can also offer relatively generic feeds as an alternative. For example, Twitter does not rely on as much PII as Facebook - there is no need to use your real name, and you can select how to tailor your feed by choosing who to follow. LinkedIn requires much more PII (name and employment history), and personalizes its recommendations for who to connect with, but does not use a personalized news feed in the same way as Facebook does to drive engagement.

Facebook, indeed, is unique in this regard. Its business is essentially tied to its use of personal data; both the service it provides its users, and the revenue it receives from advertisers.

For example: Facebook has long held a "real name" policy - criticized by digital civil liberties activists for either excluding vulnerable populations or, if they want to use the service, exposing them to bullying or harassment (Doctorow 2011). Beyond that baseline PII requirement, Facebook actively collects additional PII (date and place of birth, current place of residence, educational and employment history) and passively collects social and behavioral data (who your friends are, which articles you click on and read, how low your phone battery is). It buys and sells user data with third-party data brokers, advertising agencies, and other companies (LaForgia and Dance 2018; Head 2014; Angwin et al. 2016). Indeed, its actual bottom line - its revenue and profitability - comes from its role as one of the world's biggest data collection companies.

Its service to users, however, is in the personalized news feed. In order to continue collecting data which it can then share with third parties, Facebook needs to ensure that users continue engaging with the site. To do this, machine learning algorithms personalize each individual's feed - feed items which drive engagement (clicking the link, sharing the post, spending more time on the site) reward the algorithm. The average user is unlikely to know both how much data is collected by Facebook, and how the "recommended" posts on their feed are selected (Smith 2016).

As such, the GDPR is a direct blow to Facebook's service and bottom line. Given that much of Facebook's AI personalization and data broker revenue derives from the large quantity of PII data it collects about each user, and given the regular public relations scandals surrounding Facebook's data collection and privacy violations (e.g. Cambridge Analytica - see Granville 2018), it is clear that (a) Facebook's business would be enormously slowed if it had to obtain true consent for every piece of data it collects from its 2+ billion monthly active users, and (b) Facebook therefore has an economic incentive to use opt-out-style privacy settings, and to minimize the process of gaining consent. Yet a minimal, hidden consent (for example, buried in Terms & Conditions or hard-to-find privacy settings and written in legalese) is no true consent at all - and flies in the face of the spirit of the GDPR. As such, Facebook would need to make significant changes to adhere properly.

How could it adhere to the GDPR? It depends on how the GDPR would be enforced. On the one hand, Facebook already allows users to download their data archives. On the other, its consent and opt-in process is - in my opinion - insufficient. To truly address the spirit of the GDPR, Facebook would need to frequently alert users to which data is collected and how it is used: for example, whether their data has been used in training machine learning algorithms, how and why their posts appear in other feeds, what other companies their data was shared with, and so on.

Regardless of the spirit of the GDPR, much depends on how it is enforced. As of the summer of 2018, European Economic Area users mostly saw pop-ups explaining, via a few short paragraphs, how their data was used. Most of these pop-ups included an explicit "I Agree" consent-style button; a few allowed an "opt out" (where the service could be used without agreement). Whether these short pop-ups truly reflect the GDPR's ambitions - that is, for users to understand how their data is used, and to knowingly allow this data collection - is unclear. Facebook, as the archetypal violator of user privacy, continues to operate in Europe - demonstrating that the regulators are, for now, taking a pragmatic approach. It remains to be seen whether this will continue: public opinion regarding Facebook may change, or new regulators may enforce the law more stringently.

Question 4

I will compare the following two depictions of AI from popular culture:

- Data, from *Star Trek: The Next Generation*, as an example of a positive representation. Data is an android and Starfleet officer aboard the *USS Enterprise.*
- The Cylons, from the mid-2000s *Battlestar Galactica* remake, as an example of a negative representation. The Cylons were robots who were previously enslaved by humans; they then rebelled and upgraded into more human-like android bodies.

In both television shows, AI - coupled with advanced robotics - have made "artificial people" indistinguishable from humans. This raises challenging questions regarding selfhood and identity: what makes a person a person? Do artificial intelligences have "souls"? What are the moral rights and obligations of artificial persons? Can they be enslaved as property? Do they suffer, and should our compassion extend to them?

In the case of Data, and in keeping with *Star Trek*'s utopian philosophy, these challenges are addressed with rationality and humanism. Indeed, "humanism" is extended early on to include Data. For example, in the season 2 episode, "The Measure of a Man", Data's personhood is put on trial - literally - when scientists argue that Data is Starfleet "property" and can be dismantled for research and study. Arguing on his behalf, Data's colleagues come to his defense with a legalistic argument about what makes Data sentient. They argue that, given that his artificial intelligence is so advanced, Data meets two important criteria for qualifying as sentient life (and therefore the right to self-determination): he is intelligent and he is self-aware. Other episodes throughout the series explore Data's attempts to deepen and enrich this personhood: he is particularly challenged by understanding biological persons' emotional responses. Indeed, it seems that Data's AI - as embodied in his "positronic brain" - is a little closer to "optimal AI" on the "optimal versus human-like" AI spectrum. Data is often unable to understand emotional responses which he perceives to be irrational or self-defeating, though he is anxious to experience these responses as well. Data fulfills the same narrative role that Spock fulfilled from the original series; however, instead of a cultural conditioning which leads a character to suppress their emotional responses in the pursuit of hyper-rationality, Data is physically incapable of experiencing these responses - and views this as a flaw. In a way, *Star Trek* is recognizing the value of both optimal rationality and human-like emotional response.

The Cylons from *Battlestar Galactica* face many of the same challenges as Data: their advanced AI blurs the lines between biological and artificial personhood, and this causes tension between themselves and humans. Yet this tension is not resolved in a thoughtful, one-off court case; rather, it turns violent and lasts many years. Much of the Cylons' actions are motivated by a religious belief in their own personhood - the conflict then stems largely from humans' inability to accept the Cylons as deserving of the same rights and privileges as biological persons. First enslaved (again, similar to Data being considered Starfleet "property"), the Cylons violently rebel, killing a majority of the human population, and pursuing the last surviving military ship - the Battlestar Galactica - in an effort to achieve total genocide. In the first episode of the series, a human representative waits to meet a Cylon representative in neutral territory; when the Cylon arrives, she asks the human: "Are you alive? Prove it." This essential question - how can any individual, human or Cylon, prove their sentience and their personhood? - drives the rest of the series.

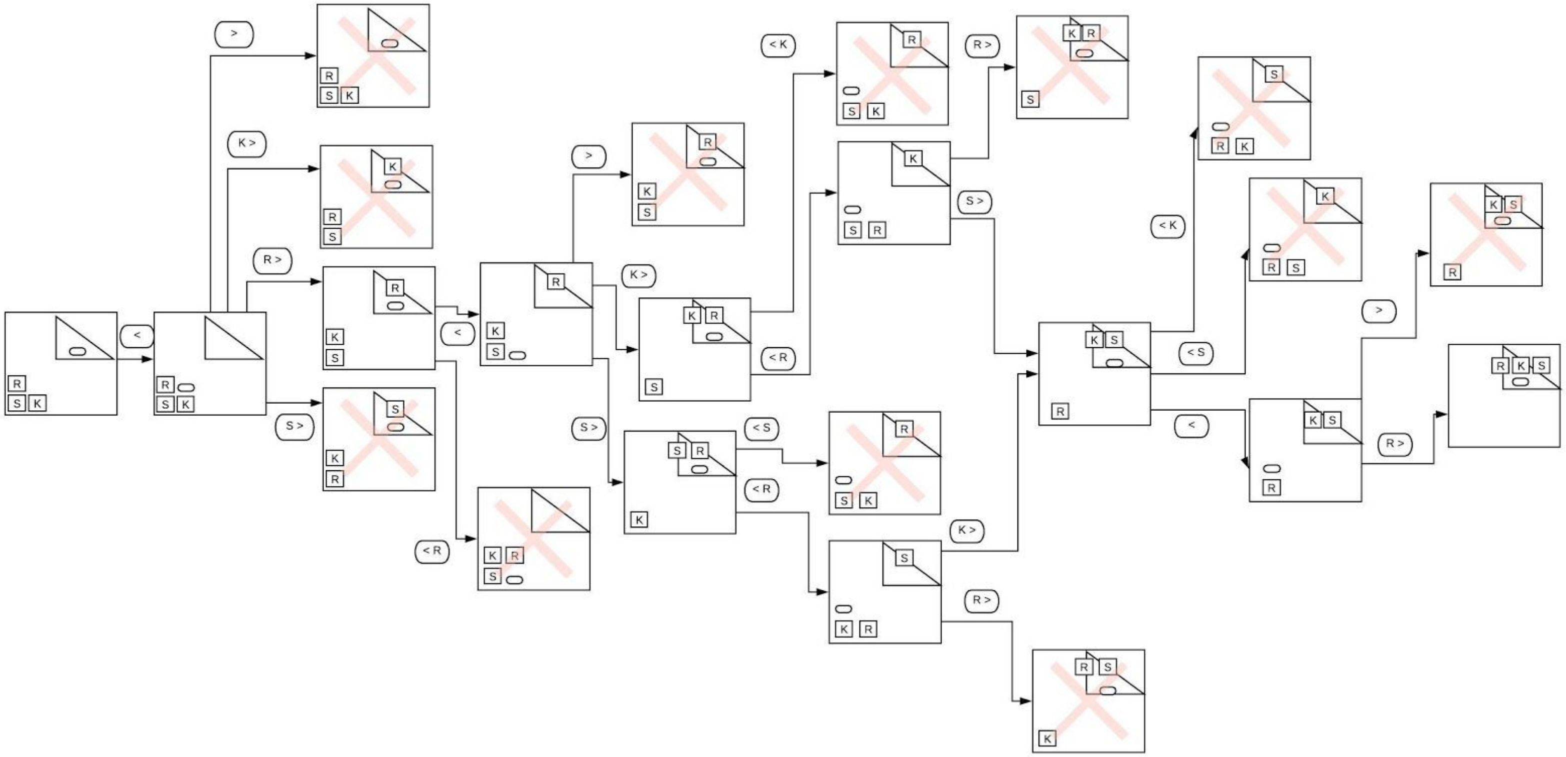
The two depictions are thus essentially addressing the same issue: when AI becomes so advanced as to become indistinguishable from - or even surpass - human intelligence, what will the rights of these AI entities be? Neither depiction portrays AI as inherently good or evil. Indeed, one of the pleasures of *Battlestar Galactica*'s storytelling is the way we regularly confront moral ambiguity on both sides - human and Cylon (there are many sympathetic Cylons, as well as instances of humans committing terrorism or knowingly killing innocents). Instead, the two depictions diverge in the way they address the problem of personhood - either with sober discussion, or war - and this is largely a function of the universes in which they take place. The universe of *Star Trek* - especially the Federation and Starfleet - are guided by Gene Roddenberry's utopian principles of rationality and tolerance. (Roddenberry, for example, instituted a "no conflict among the crew" rule among the writers. See Newitz 2017.) The universe of *Battlestar Galactica*, on the other hand, is less enlightened: it is tribal and deeply steeped in mysticism (for example, the humans worship at a polytheistic pantheon of Olympian gods; the Cylons are monotheistic, praying to an Abrahamic god). As such, the two depictions are compatible - even similar (Data is, after all, basically a Cylon) - but their contextual depictions, the universes in which they live, are not.

References

1. Angwin, J., Mattu, S. & Parris Jr, T. (2016). Facebook Doesn't Tell Users Everything It Really Knows About Them. *ProPublica.* Retrieved from https://www.propublica.org/
2. Burgess, M. (2018, June). What is GDPR? The summary guide to GDPR compliance in the UK. *Wired UK.* Retrieved from https://www.wired.co.uk/
3. Doctorow, C. (2011, August). Google Plus forces us to discuss identity. *The Guardian.* Retrieved from https://www.theguardian.com/
4. Granville, K. (2018). Facebook and Cambridge Analytica: What You Need to Know as Fallout Widens. *The New York Times.*
5. Head, B. (2014). MasterCard to access Facebook user data. *The Age.*
6. Intelligence Squared U.S. Debates (2016). *Should We Trust The Promise of Artificial Intelligence?* Retrieved from https://www.intelligencesquaredus.org/debates/artificial-intelligence-risks-could-outweigh-rewards
7. Karpathy, A. (2017a). Where will AGI come from? *Y Conf 2017.* Retrieved from https://ivenzor.com/wp-content/uploads/2018/07/yconftalk-170902200916.pdf

8. Karpathy, A. (2017b). Software 2.0. *Medium* Retrieved from https://medium.com/@karpathy/software-2-0-a64152b37c35
9. Khatchadourian, R. (2015, November). The Doomsday Invention: Will artificial intelligence bring us utopia or destruction? *The New Yorker.* Retrieved from http://newyorker.com
10. Kurzweil, R. (1999). *The Age of Spiritual Machines: When Computers Exceed Human Intelligence*. Penguin Books.
11. Kurzweil, R. (2005). *The Singularity Is Near: When Humans Transcend Biology.* Penguin Books.
12. Kurzweil, R. (2014a, March). Get ready for hybrid thinking. TED. Retrieved from https://www.ted.com/
13. Kurzweil, R. (2014b, December). Don't Fear Artificial Intelligence. *Time.* Retrieved from http://time.com/
14. LaForgia, M. & Dance, G.J.X. (2018). Facebook Gave Data Access to Chinese Firm Flagged by U.S. Intelligence. *The New York Times.*
15. Lanier, J. (2010). *You Are Not a Gadget.* Vintage.
16. Lanier, J. (2012). *Who Owns the Future?* Simon & Schuster.
17. Lanier, J. (2014) "The Myth of AI." *Edge.* Retrieved from https://www.edge.org/
18. Lanier, J. (2017). *Dawn of the New Everything.* Henry Holt and Co.
19. Lanier, J. (2018). *Ten Arguments for Deleting Your Social Media Accounts Right Now.* Henry Holt and Co.
20. Lev-Ram, M. (2017). Why Futurist Ray Kurzweil Isn't Worried About Technology Stealing Your Job. *Fortune.* Retrieved from http://fortune.com/
21. Newitz, A. (2017). New Star Trek series will abandon Gene Roddenberry's cardinal rule. *Ars Technica.* Retrieved from https://arstechnica.com/
22. Smith, A. (2016). Many Facebook users don't understand how the site's news feed works. *Pew Research Center.* Retrieved from http://www.pewresearch.org/

Appendix: Question 1 - Semantic network, fully solved



<
>
K >
R >
S >
<
< R
>
K >
S >
< K
< R
R >
S >
< S
< R
K >
R >
< K
< S
<
>
R >