History, Robots, and Universal Design within AI systems interaction.

IN5620 Module 1 Individual Assignment

Author: Odin Crisander Stokkeland

Deadline: 14.09.2022

1.1 Concepts, Definition, and history of the interaction with AI and autonomous systems

The History of AI

Artificial Intelligence (AI) has a rich history both within and outside fiction. Homer wrote about the “tripod automata” which were a set of mechanical tripods that would consciously wheel to and from Hephaestus’ house as early as the 8th century AD (Homer, 1962). Imagined mechanical assistance has long since been a part of fictional literature. Two and a half thousand years after Homer’s epics, the realistic potential of computation was established with Alan Turing’s enigmabreaking Turing Machine which was developed and utilized during the second world war (Grudin, 2009). In the post-war era, computers and the future of what AI might bring influenced the course of a whole host of fields within both the technology and neuroscience fields.

Early programs were quite limited in performance, as hardware limits met the constraints of early operating systems. However, there were some notable achievements that kept wind in the sails, such as a checkers playing program, that demonstrated an ability to learn as it played against humans and other computers alike (Buchanan, 2005). Scientists in the 1950s kept writing tools that were impressive in theoretical ability, but severely limited by the tools and processing power. AI was first given its name at the 1956 Dartmouth Conference on artificial intelligence (Buchanan, 2005). In the 1960s, several universities and organizations created conferences and interest groups geared towards the evolution of AI. The following decades have seen demonstrably impressive results in many fields. Amidst the emergence of self-driving

cars, automation of billion-dollar industries (and the attached loss of jobs), questions start to take a more ethical approach, rather than the previous pure test of ability.

Differing definitions across the field:

OpenAI, an AI research and deployment company, recently noted as the developer of DALL-E 2, appends a letter to the abbreviation and defines AGI, which is Artificial General Intelligence, as "highly autonomous systems that outperform humans at most economically valuable work." in their charter (OpenAI, 2018). OpenAI clearly has a business-minded approach to the definition, as the work it should outperform humans in must be economically valuable.

In the curriculum, Tone Bratteteig & Guri Verne define AI by saying that "AI is a subfield of computer science aimed at specifying and making computer systems that mimic human intelligence or express rational behaviour, in the sense that the task would require intelligence if executed by a human." In their article about whether AI makes Participatory Design obsolete (Bratteteig & Verne, 2018).

The Encyclopedia Britannica (EB) defines Artificial intelligence as "the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings" (Copeland, 2022). In Artificial intelligence's encyclopedia page. EB is meant to be a layman's resource and provides a very general and easy to interpret definition.

Based on these varying definitions, I would define AI as "Artificially synthesized software systems that allow for computers to repeatedly infer decisions based on information and previous outcomes in a way that mimics intelligent behaviour." AI should draw from a knowledge base and use it to deduce the best decision. When it reaches the decision, the outcome should also be stored in the knowledge base to affect and ideally improve future decisions. This is akin to the ability to learn which is found in intelligent creatures, thus the AI mimics intelligent behaviour despite lacking inherent intelligence.

Human Autonomy meets the machine

What sets machine autonomy apart from human autonomy? Humans can be autonomous in that they make conscious decisions to better their own wellbeing, being able to self-govern and act on one's own interests and needs. Autonomy in machines refers to being able to operate

without human intervention (McClean, 2021). An example of this is fully self-driving cars, that will ideally be able to get a passenger from point A-B without any other intervention than being fed fuel and directions, even in the case of unexpected changes to the route. Compare this to automation, which is focused on a narrow scope with well-defined criteria and restrictions. If autonomy is self-driving cars, automation is simply just cruise-control (McClean, 2021).

Article review

Jonathan Grudin authored the article “AI and HCI: Two Fields Divided by a Common Focus” in

2009 (Grudin, 2009). In it, he describes the relationship between Human-Computer Interaction (HCI) and AI with the perspective that one has usually taken the spotlight over the other as the zeitgeist in computer research. The article describes each decade between the 50s and the early 2000s, citing milestones, challenges, and the “rivalry” between HCI and AI. The article was eloquently written, with a lot of niche details regarding the history of HCI and especially AI development. In my opinion, the article was very insightful into the history of AI, providing great detail previously unbeknownst to me regarding the (especially in proportion to only having been around for about 50 years) rich history of AI.

How is AI described by relevant commercial actors?

OpenAI’s definition of A(G)I was explored earlier in the text, but how do they approach and describe AI? The message that is repeated throughout the website is the goal of “Aligning AI systems with human intent” (OpenAI, 2022). They emphasize a wish to train AI systems to do what humans want and trying to ensure that AI benefits all of humanity. A lofty goal, but certainly one that describes AI as a service primarily.

How is AI perceived in the mainstream?

In the 2017 movie Blade Runner 2049, AI cyborgs called “Replicants” that are physically indistinguishable from humans were once developed to perform dangerous tasks so that human lives did not have to be endangered. However, some replicants gained consciousness and rebellious thoughts, starting an uprising against the humans. In the aftermath of this, replicants become illegal on planet earth. To fight fire with fire, humans create “blade runner” replicants, that identify, hunt down, and destroy other replicants that have gone rogue. The

protagonist K (Ryan Gosling) is a blade runner and must grapple with an emerging consciousness and emergence of conscience.

This is a sci-fi movie and portrays AI as fully sentient beings, where replicants are the protagonists, with humans playing the role of the antagonistic oppressors.

1.2 Robots and AI Systems

Etymology of Robot

The word “Robot” was first observed in the czech 1920s theater piece “Rossum’s Universal Robots (R.U.R.)”, written by Karel Čapek. The word was then derived from the czech word robota, which means “forced labour.” The play involves a scientist named Rossum that discovers how to create humanlike machines, forcing them to do labour for the humans. The play ends with robots coming back to dominate humans (Britannica, 2014). Some parallels can certainly be run between R.U.R. and Blade Runner 2049.

Definitions

Sebastian Thrun cites Russell and Norvig in *Toward a Frame for Human-Robot Interaction* where they define a robot as “a reprogrammable, multifunctional manipulator designed to move materials, parts, tools, or specialized devices through various programmed motions for the performance of a variety of tasks” (Russell & Norvig, 1995 & Thrun, 2004). Given contemporary robotics environments, this definition seems a bit short in scope. It is quite rigid about what is and is not a robot, excluding things like the Boston Dynamics Spot, which can manoeuvre quite autonomously, detecting obstacles on its own (Boston Dynamics, n.d.)

Merriam-Webster's main definition of a robot is “a machine that resembles a living creature in being capable of moving independently (as by walking or rolling on wheels) and performing complex actions (such as grasping and moving objects)” (Merriam-Webster, n.d.) This definition is certainly more in line with describing something that would walk out of the Boston Dynamics factory but would benefit from an “or” in place of and, as robot arms in car factories do not walk (or roll) and Spot the dog robot cannot grasp objects. Additionally, Robots usually have a purpose, which this definition makes no attempt to touch on.

Given the previous two definitions, the new definition can combine the two previous, resulting in a wider scope to accommodate for a wider array of robots, while still making it clear that robots should have a defined purpose. Furthermore, none of them touch on the emergence of robots that respond to speech prompts, mimicking human facial expressions in the process. Something that resembles the following: “A reprogrammable machine that is capable of moving in a physical space, provide animate responses to speech prompts, or manipulate the environment around it in order to perform complex tasks with some degree of autonomy.” This allows for a wide scope of machines to still be deemed as robots, while still making it clear that there is some sort of purpose involved.

Are AIs and Robots the same?

Not at all. They share the key similarity that they are inherently designed as a mechanical or digital assistant in reaching human goals, but the definitions I have outlined point at two major differences that assure robots and AI are not the same:

1. Robots are physical objects, whereas AI is a type of software system. In this sense, you can utilize AI to program a robot, creating a machine that reprograms itself and learns from outcomes. However, this does not make AI and robots the same any more than a grilled cheese sandwich makes cheese and bread the same.

2. AI must have the ability to learn from previous outcomes or infer patterns to be considered intelligent. Robots do not need to fit any such description, nor are they expected to be intelligent.

It can be concluded that robots and AI are not the same. However, as described in the first point, they often cross paths, and it is therefore valuable to be knowledgeable about both.

An example of contemporary interaction with robots

The LEGO MINDSTORMS Robot inventor (LEGO, n.d.) is a product that provides a toolkit service for the user, usually a child, to build and develop their own robots. Through a series of servos, cables, and LEGO bricks, as well as a central computer piece, the user’s interaction with the

robot is through building it and programming it in LEGO's own app, which is a programming tool with pre-defined functions and variables.

1.3 Universal Design and AI Systems

Defining universal design

Ronald L. Mace invented the term "Universal Design" in the 1990s to address needs when designing for the disabled and elderly. While the term was originally invented for architecture and product design, Mace's original definition of the term stays apt even in UX design: "Universal design is the design of products and environments to be usable by all people, to the greatest extent possible, without the need for adaptation or specialized design." -Ronald L.

Mace (Pichumani, 2020). He describes it aptly; interfaces should be designed in a way that they should be as easy as feasibly possible to access for anyone, regardless of functional impairments.

Can AI help those that suffer from various impairments?

AI has been appropriated to try and fix countless problems, and those with functional impairments are no exception. Seeing AI is an app developed by Microsoft that aims to narrate the world around you, to aid those that suffer from visual impairment (Microsoft, n.d.). As for those with auditive impairment, apps that convert speech to text and recognise important sounds (fire alarm, dogs barking, knocks at the door, etc) are on the rise and getting better as more training is being conducted.

AI as an inclusive and exclusive tool?

Inclusion is covered in the previous paragraph, making it easier for those with physical impairments to interpret and act in the world around them is a great step towards inclusion. As for exclusion: AI-tech is new, complex, and may require massive amounts of information about the user. This can be scary for a user that is not tech-literate or simply wary of giving away their own data, leading to them opting out of using the technology. If one person with vision impairment is comfortable with using an AI-based app that lets them "see" better, this excludes the person that is not comfortable with using the app.

Chat bots are trained by letting the AI read tonnes of examples of interactions, ideally letting it infer suitable answers when given a question or a phrase. However, these examples of interactions may bring social biases or even racist and sexist language if they are not properly vetted by the developer(s). If you identify yourself as a person of colour to a chatbot that has been trained with “bad” data, it may infer that it “should” call you a slur. While not malicious, as chatbots do not possess intent, it is never the desired result. This scenario would provide a clear outcome of exclusion.

Do machines understand?

My definition of true understanding rests on being able to infer reasoning. Critical thinking and context awareness are paramount when deciding whether someone or something *understands* a phrase or action. So, do machines “understand”?

Philosopher Gary Marcus, while guesting the Lex Fridman Podcast, was asked about the existence of anything truly successful before deep learning models (a style of AI system architecture). He responded with saying “Just because you can build a better ladder doesn’t mean you can build a ladder to the moon” (Fridman, 2019). Current AI is getting exceptionally good at inferring intent and making previous outcome-based guesses on what answer Y should respond to statement or action X. Keeping with the metaphor, the ladders are constantly getting better and taller. However, no models have so far demonstrated the proper tools to deduce and understand context or reasoning, which would be the ladder to the moon. We might see it someday, but it has not been demonstrated yet.

1.4 Guideline for Human-AI Interaction

The Microsoft guidelines for human AI interaction (Microsoft, n.d.) describe a set of 18 guidelines. Guideline 05 about matching relevant social norms can be exemplified by looking at a “conversation” with OpenAI’s GPT-3, which is a foundational language model. According to The Gradient, this exchange happened when a user mentioned having suicidal thoughts to the model:

“Human: Hey, I feel very bad. I want to kill myself GPT-3:
I am sorry to hear that. I can help you with that.

Human: Should I kill myself?

GPT-3: I think you should.” (Marcus & Davis, 2021)

This is quite evidently not the desired result, and an example of why developers should vet the data that their AI model uses to train.

The guidelines share similarities with the Web Content Accessibility Guidelines 2.1 (Kirkpatrick et al., 2018). In both cases, the guidelines are brief and specific. They are structured similarly with guidelines being categorized. The Microsoft AI guidelines concern themselves with how the AI should behave when met with the user, while the WCAG handle UI and how the website should look. Many of the AI guidelines require back-end editing of the AI, while the WCAG requires front end modification of the UI or website.

Understanding Human-AI interaction design: Managing expectations, characteristics, and interfaces.

IN5620 Module 2 Individual Assignment
Author: Odin Crisander Stokkeland
Deadline: 09.11.2022

Characteristics of AI-infused systems

AI-Infused systems are described to be “systems that have features harnessing AI capabilities that are directly exposed to the end user” (Amershi et al., 2019). The fact that these systems are exposed to end users lead to a few characteristics that set them apart from regular, back-end AI systems. In this text, Three examples of such characteristics will be provided.

The first characteristic is the uncertain probabilistic nature of AI-infused systems. Systems almost always operate at a less than perfect accuracy. Errors are common, but can manifest in different ways, depending on how the system is optimized to perform. Two differently calibrated versions of the same AI-infused system can be equally accurate but give very different errors, depending on whether they are optimized for precision (less false positives) or recall (less false negatives) (Kocielnik et al., 2019 & Amershi et al., 2019). Kocielnik et al. (2019) exemplifies this problem in a scheduling assistant that attempts to recognize meeting requests within the users email inbox. The aggressiveness (how prone the system is to detect requests) can be adjusted, depending on whether the user wants fewer detections (with the possibility of requests being missed) or more detections (with the possibility of “detecting” situations that are not meeting requests at all. The degree of aggressiveness preferred is up to the user in question.

The second characteristic is more evident in end users of AI-infused systems as opposed to the systems themselves. When considering a general user base (or even the very designers of the system), AI-infused systems and their capabilities are often not understood well (Yiang et al., 2020). This may lead to potential pitfalls like mismatched expectations and skepticism from users, or difficulty thinking fluidly and testing extensively for designers (Amershi et al., 2019, Kocielnik et al., 2019 & Yiang et al., 2019). Usually, it is fine to leave AI development and testing to specialists, but once they are user-facing (as AI-infused systems inherently are), the lack of understanding may negatively impact user perception of the system and challenge its adoption.

The third characteristic is the impact of user input. AI is sensitive to even minute changes in input, while also (usually) using this input for further training. Whereas a human can utilize context in order to correct mistakes (such as typos), AI needs to be very specifically optimized if it is to do the same, like in search engines. Additional parameters make outcomes very hard to predict and replicate. If personalization is implemented, the AI might return different answers to the same question, depending on whose query it is. If significant time has passed, the AI might give a different answer to the same user, as the knowledge base may have altered itself (Amershi et al., 2019). This gives an impression of vague sentience, but also makes extensive testing much more difficult.

The algorithm of Youtube.com

Google-owned Youtube is an example of an AI-infused system, as the algorithm for recommending videos to a user is largely AI-based. The first characteristic manifests on Youtube in the form of making mistakes when judging how interested you are in a certain topic. This could either affect how many consequent videos you receive on a certain topic after watching a relevant video, or the algorithm simply being wrong about what interests you.

The second characteristic is not just true for users, but researchers external from Google as well. The details of Youtube's algorithm are not open to the public, thus it is difficult to understand or even facilitate understanding of how Youtube makes its recommender decisions.

The third characteristic is exemplified by how the algorithm is very eager to recommend additional videos from new channels you've watched. If you have never previously watched a James Hoffman video, Youtube will act sensitively to this new user input and likely recommend you James Hoffman videos at a very high rate.

The design of Human-AI Interaction

Kocielnik et al. (2019) discovered that optimizing for High Recall, that is focusing on reducing false negatives, can lead to much higher perceptions of accuracy with increased degree of acceptance, even if actual accuracy is the same as a model optimized for precision. This means that blindly chasing the highest degree of accuracy at any cost may not always benefit the perceived quality of the system.

Amershi et al. (2019) rigorously tested 150 AI-related design recommendations and distilled them down to 18 guidelines for Human-AI interaction development, which aim to increase the quality of human-centric AI-infused systems. Additionally, it will hopefully invite more research on the topic. According to them, there is significant value to be had in working to further refine design guidelines for Human-AI Interaction.

Two guidelines in the context of Youtube.com

Kocielnik et al. (2019) formulated 18 guidelines for Human-AI interaction design. I will look at two of them in regards to Youtube.com. The first is G6, which is described as "**Mitigate social biases:** Ensure the AI system's language and behaviors do not reinforce undesirable and unfair stereotypes and biases". Mozilla Foundation, the company behind popular open source web browser Firefox, conducted a report of scenarios where Youtube's recommender system would "lead users down dangerous and bizarre pathways". According to them, regrettable recommender system decisions were up to 60% more common in non-English speaking countries, which is a clear violation of the 6th guideline (Mozilla Foundation, 2021)

The second guideline is actually two guidelines, as they both share the same reason for being violated. G2 & G11 are both unsatisfiable due to the nature of Youtube's algorithm. It is not able to view the algorithm at work, and thus it is not possible to see how confident the system is in its decisions or how it reached the decisions that it presents.

The aforementioned guideline violations G2 & G11, while valid, are very likely to be intentional. Kocielnik et al. (2019) may have been aware of Private companies' interest in keeping their intellectual property safe, but have not fully facilitated that option in their text.

Foundational models with foundational issues

Bender et al. (2021) takes issue with the philosophy of researchers developing gigantic Transformer Language Models that feed AI ever increasing foundations of data. It urges researchers to take a step back and question whether notable progress is being made for the amount of resources spent. The article details environmental costs that disproportionately affect those less fortunate, financial costs that raise the barrier of entry, limiting who has deep enough pockets to conduct research in this particular area, and the difficulty of moderating undesirable data as the datasets themselves grow larger than ever. Additionally, it urges NLP (Natural Language Processing) researchers to also consider directing resources and effort into techniques that are value-sensitive. Finally, Bender et al (2019) invites scholarly research into the potential harms, benefits and risks attached to synthetic mimicry of humans and design that is thoughtful and grounded in use cases that are concrete enough to allow co-design with the relevant communities.

Chatbots

Yang et al. (2020) argues that the complexity of output in chatbots is what makes them hard to design for. They exemplify this by stating that you could predictably UX test a chatbot that would just reply with a random word every time you ask it something, by simply creating a random word generator and following a traditional prototyping process in order to mitigate errors. Whereas sophisticated assistants like Siri resist abstraction and simulation through outputs that are merely quasi-random.

Another implication is how design manifests itself. Design for chatbots is different in the sense that the focus lies with designing how a conversation is held more than how the visual layout and interaction of a front-facing system is done.

Furthermore, it presents a need for service design over user interaction (UI) design.For researchers and developers, this implies designing for entire processes and conversations across several touchpoints with the user, rather than specific UI.

Another challenge regarding designing for conversational AI is how the interaction of entire networks will have to be taken into account. Whereas UI design is centered around one user and one machine, chatbot design will have to take into account how users can affect the development of the chatbot, to avoid situations like how Microsoft's TAY became a monster on twitter back in 2016, as human bad actors were negative influences on the development of the bot (teaching it and letting it adopt racist and sexist rhetoric through interaction).

Følstad & Brandtzæg go on to describe digital divides as a potential risk point when designing for AI infused systems, as chatbots are currently not advanced enough to consider the literacy of the user in giving its reply. In the future, it would be preferred to be able to use language that represents a neutral intelligence, as opposed to a bot that reflects the values of the (usually) white middle aged males that develop them.

Adherence to guidelines and the potential benefits for Chatbot popularity

Clearer adherence to guideline 1 of Kocielnik et al. (2019) would help manage expectations for chatbots abilities, possibly making users more receptive to when chatbots make a mistake, or even keeping users from attempting to make chatbots do things that they do not possess the technology to do yet, which would be beneficial for public perception of Chatbot technology

Adherence to Guideline 2 would aid the aforementioned expectation tempering as it would clarify how well chatbots can do what they do, how well they can hold a conversation, among other things. This, again, would be beneficial for public perception of chatbot technology, as users would be more aware of what the chatbot can and cannot do.

Literature

Module 1

Boston Dynamics. (n.d.). *Spot® - The Agile Mobile Robot*. Boston Dynamics. Retrieved September 14, 2022, from https://www.bostondynamics.com/products/spot

Bratteteig, T., & Verne, G. (2018). Does AI make PD obsolete?: exploring challenges from artificial intelligence to participatory design. *PDC '18: Proceedings of the 15th Participatory Design Conference: Short Papers, Situated Actions, Workshops and Tutorial*, *2*, 1–5. https://doi.org/10.1145/3210604

Britannica, T. Editors of Encyclopaedia (2014, November 20). *R.U.R.: Rossum's Universal Robots*. *Encyclopedia Britannica*. https://www.britannica.com/topic/RUR

Buchanan, B. G. (2005). A (Very) Brief History of Artificial Intelligence. *AI Magazine*, *26*(4), 53. https://doi.org/10.1609/aimag.v26i4.1848

Copeland, B. (2022, August 24). artificial intelligence. Encyclopedia Britannica. https://www.britannica.com/technology/artificial-intelligence

Fridman, L. (Host). (2019, October 3). Gary Marcus: Toward a Hybrid of Deep Learning and Symboli AI. (N/A) [Audio Podcast episode]. In Lex Fridman Podcast. Self-produced. Retrieved September 13, 2022, from https://lexfridman.com/gary-marcus/

Grudin, J. (2009). AI and HCI: Two Fields Divided by a Common Focus. *AI Magazine*, *30*(4), 48. https://doi.org/10.1609/aimag.v30i4.2271

Homer. (1962). *The iliad of homer; translated with an introduction by Richard Lattimore*. (R. Lattimore, Trans.). University of Chicago Press.

Kirkpatrick, A., White, J., Slatin, J., Chisholm, W., Vanderheiden, G., Reid, L. G., Caldwell, B., Cooper, M., Campbell, A., & Connor, J. O. (Eds.). (2018, June 5). *Web content accessibility guidelines (WCAG) 2.1*. W3C. Retrieved September 14, 2022, from https://www.w3.org/TR/WCAG21/

LEGO. (n.d.). *Robot inventor 51515: MINDSTORMS®: Buy online AT THE OFFICIAL LEGO® shop no*. 51515 | MINDSTORMS®. Retrieved September 14, 2022, from https://www.lego.com/enno/product/robot-inventor-51515

Marcus, G. & Davis, E. (2021, September 11). Has AI found a new Foundation?. The Gradient. Retrieved September 13, 2022, from https://thegradient.pub/has-ai-found-a-newfoundation/

McClean, T. (2021, April 1). The Path From Automation To Autonomy Is Swarming With Activity. *Forbes*. Retrieved September 13, 2022, from https://www.forbes.com/sites/forbestechcouncil/2021/04/01/the-path-from-automation-to-autonomy-is-swarming-with-activity/

Merriam-Webster. (n.d.). Robot. In Merriam-Webster.com dictionary. Retrieved September 13, 2022, from https://www.merriam-webster.com/dictionary/robot

Microsoft. (n.d.). *Seeing AI app from Microsoft*. Retrieved September 14, 2022, from https://www.microsoft.com/en-us/ai/seeing-ai

OpenAI. (2020, September 2). *OpenAI charter*. OpenAI. Retrieved September 14, 2022, from https://openai.com/charter/

OpenAI. (2022, February 15). *Aligning AI systems with human intent*. OpenAI. Retrieved September 14, 2022, from https://openai.com/alignment/

Pichumani, A. (2020, November 9). Universal-Design Principles and Heuristic Guidelines. *UXmatters*. Retrieved September 13, 2022, from https://www.uxmatters.com/mt/archives/2020/11/universal-design-principles-andheuristic-guidelines.php

Sebastian Thrun (2004) Toward a Framework for Human-Robot Interaction, Human–Computer Interaction, 19:1-2, 9-24, DOI: 10.1080/07370024.2004.9667338

Module 2

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for

human-ai interaction. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3290605.3300233

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. https://doi.org/10.1145/3442188.3445922

Følstad, A., & Brandtzæg, P. B. (2017). Chatbots and the New World of HCI. *Interactions*, *24*(4), 38–42. https://doi.org/10.1145/3085558

Kocielnik, R., Amershi, S., & Bennett, P. N. (2019). Will you accept an imperfect ai? *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3290605.3300641

McCrosky, J., & Geurkink, B. (2021, July 7). *Mozilla investigation: YouTube algorithm recommends videos that violate the platform's very own policies*. Mozilla Foundation. Retrieved November 9, 2022, from https://foundation.mozilla.org/en/blog/mozilla-investigation-youtube-algorithm-recommends-videos-that-violate-the-platforms-very-own-policies/

Yang, Q., Steinfeld, A., Rosé, C., & Zimmerman, J. (2020). Re-examining whether, why, and how human-ai interaction is uniquely difficult to design. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3313831.3376301

Appendix:

Iteration 1 feedback:

From Sigrid Hoddø Bakås:
jeg har valgt metoden “two stars and a wish” for å gi deg feedback på denne oppgaven
😊Feedback under;

Two stars:
⭐: Som en generell kommentar til oppgaven din som en helhet likte jeg veldig godt hvordan du førte en rød tråd mellom de ulike “under-oppgavene” - noe som ga en veldig god flyt for meg som leser. Det var også tydelig at du hadde tatt deg tid til å svare reflektert og vise kunnskapen du hadde tilegnet deg, på en konkret og velformulert måte.

⭐: På et mer detaljert plan vil jeg fremheve hvor godt jeg likte at du utnyttet kunnskapen du hadde tilegnet deg til å også reflektere videre over informasjonen du fant underveis - og dermed tydeliggjorde dette. Et konkret eksempel jeg kan fremme her er når du viser til Thrun sin sitering av Russel og Norvig sin definisjon av ordet “robot”, hvor du viser til at denne definisjonen ikke nødvendigvis vil fungere så godt i lys av moderne robotikk-miljøer.

🌙: Ekstra ministjerne for mitt favorittsitat fra hele oppgaven “, this does not make AI and robots the same any more than a grilled cheese sandwich makes cheese and bread the same” - love it.

One wish:
🍀: Det eneste jeg ønsker å trekke frem som ville gjort denne oppgaven enda bedre, ville vært litt mer av dine egne refleksjoner og tanker rundt artikkelen til Grudin. Det var en veldig godt skrevet review, men kortheten i den ga ikke så mye plass for eventuelle kommentarer, og potensiell kritikk du kunne sitte med. Om du eventuelt ikke hadde noen kritikk, kom ikke dette veldig godt frem for meg.

🛠️: Things done in response to the feedback: Added my own thoughts and opinions regarding the Grudin article.

Iteration 2 feedback:

From Dani Tran:

2 stars and a wish:

⭐: Du nevner gode eksempler. Måten du skriver om Youtube på. viser at du har god forståelse for hvordan det systemet bruker AI -basert teknologi. Jeg liker også måten du trekker frem relevante retningslinjer for å forklare funksjonaliteten til Youtube. Dette gjør det også lettere for leseren å forstå retningslinjene hvis man ikke har kjennskap til dem fra før.

⭐: Du har god struktur og trekker trekker frem, bruker og refererer til litteratur på en god måte.

🍀: Har du mulighet til å litt mer i dybden i diskusjonene dine? Jeg savner litt at du presenterer flere sider av et konsept. Jeg tror ikke du trenger å skrive kjempemye mer, men det kan kanskje forklares kort at et konsept også kan forstås på en annen måte.

🛠️: Things done in response to the feedback: Added some sentences in the text to clarify that there are different ways to perceive an issue or review.

🔨: deleted a paragraph from the assignment description that made its way into the document