What Fictional Machine and Human Characters Do With Machine Vision Technologies

Work in progress - please do not cite

Jill Walker Rettberg, University of Bergen, jill.walker.rettberg@uib.no
Marianne Gunderson Linda Kronman Ragnhild Solberg

2022-04-15

Contents

To do:

1. Go through all the scripts and rewrite now that I revised Characters to make it simpler.
2. Add overview of other content analysis that's comparable
3. Figure out how to get Zotero connected and add references
4. Do more analysis of the character traits in combo with the verbs
5. Figure out to what extend other team members will be involved - they already contributed some of the writing for the first version so probably need to be included, think about HOW.

Introduction

This presentation explores the cultural imaginaries of machine vision technologies as they are portrayed in contemporary science fiction, digital art and videogames. How are the relationships between humans and machines imagined in fictional situations and aesthetic contexts where machine vision technologies occur? Our goal is to use a large-scale analysis of fictional and artistic depictions of machine vision to to understand how the distributed agency of human-technical assemblages is imagined in our global cultural record.

We define machine vision as the registration, analysis and generation of visual data by machines (Rettberg et al. 2019), and include technologies such as facial recognition, optical implants, drones, surveillance cameras and holograms in this. The project team has selected 500 creative works, primarily games, novels, movies, TV shows and artworks, in which machine vision technologies are either described/shown or actually used in the artwork. We have entered structured interpretations of each work in the Machine Vision in Art, Games and Narratives Database.

This is an analysis of the portrayal of human, machine and cyborg characters in 500 narratives, games and artworks. While the first work in the dataset, Arne Garborg's novel *Trætte Mænd*, is from 1891, 80% of the works were published between 2011 and 2021, and just over half are from 2016-2021. This paper focuses on characters in the works that are portrayed as directly interacting with machine vision technologies and analayses how their interactions with machine vision are portrayed.

How to analyse agency

While there are many studies analysing gender representation in videogames, movies and other media, most use content analysis to count presence. We wanted to go beyond that and how different kinds of people are imagined to be using machine vision in different ways.

Posthumanist and feminist new materialist theory argues that agency is distributed between human and nonhuman actors. This is especially pertinent with complex, algorithmically-driven technologies like machine vision.

Our approach builds upon methodologies from literary studies and visual analysis that rely on interpretation and close reading. The team read, watched, played and experienced the 500 works, and entered their data based on interpretative strategies learned through years of training.

Two alternative approaches are content analysis using a more traditional quantitative coding or computational analysis of the textual material itself.

Linguistic analysis of agency roles offers one approach to understanding the agency of characters. Frazer Heritage (2021) used corpus linguistics to analyse 350,000 words of transcribed text from ten videogames published between 2012-2016, and found striking differences between the actions described for male and female characters. While men were connected to verbs like "discovered; faced; investigate; murdered; showed",

women were more likely to have collocates like “gave; made; meet; took”. The pronoun “he” was often collocated with violent words, while there were no explicitly violent words in the list of words most frequently collocated with “she”. Heritage also found that while male characters were “represented as enacting physical changes in the world, female characters are the ones who have physical changes enacted upon them.

Our approach was designed to work across multiple genres of creative work, so relying on linguistic or visual analysis alone would not suffice. We also wanted to preserve a key strength of literary analysis and related methodologies: the hermeneutics of considering both details and whole.

Method

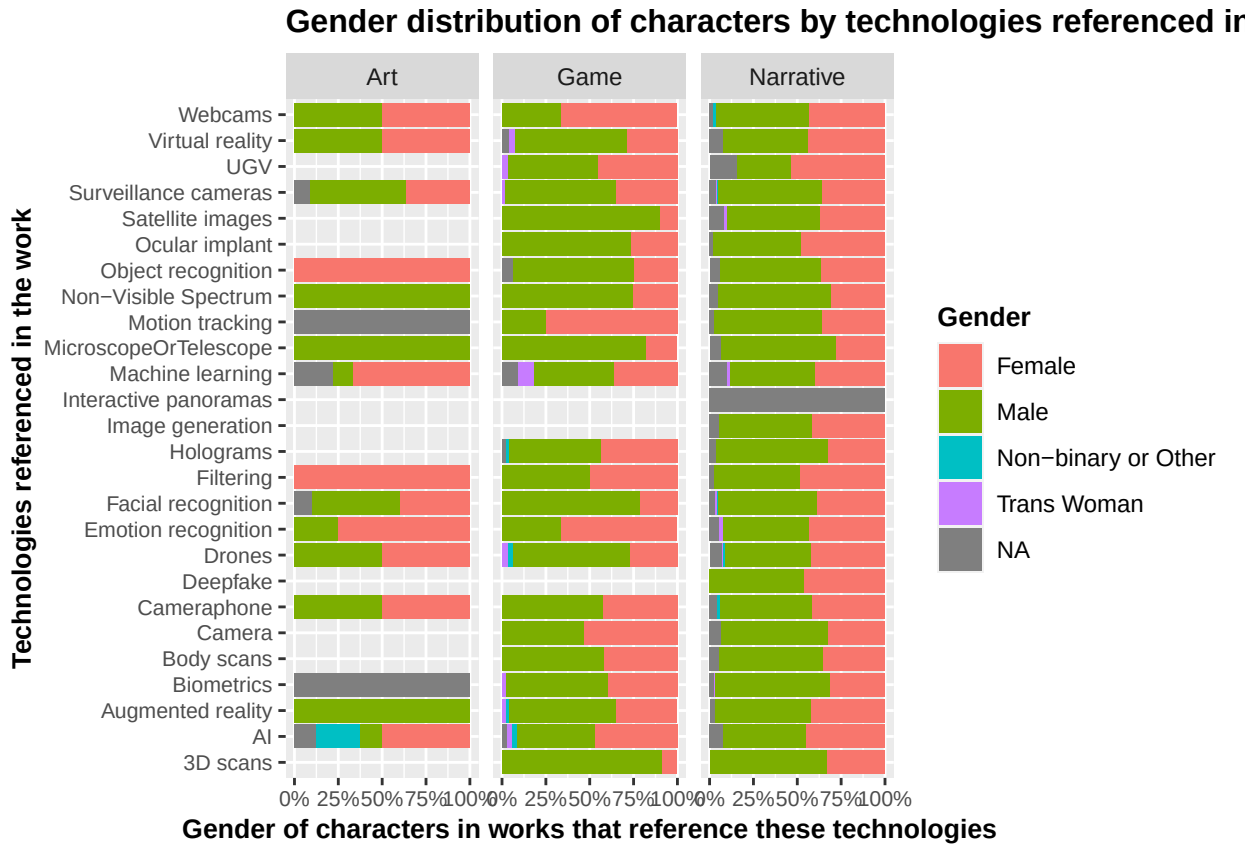
(Short general description of database, ref dataset and data in brief paper).

We used R and the Tidyverse library to import data about characters and machine vision situations. The characters data consists of a list of characters that interact with machine vision technologies in the 500 works, and each character is described in the traits Species, Gender, Race or Ethnicity, Age and Sexuality. We defined subsets consisting of all the human characters and all the robots, where robots include all kinds of software or hardware robots, cyborgs and AIs.

Situations are (DESCRIBE). We imported data about each situation (situations.csv) and selected the data describing individual characters and verbs that describe what each character *does* with machine vision in the situation.

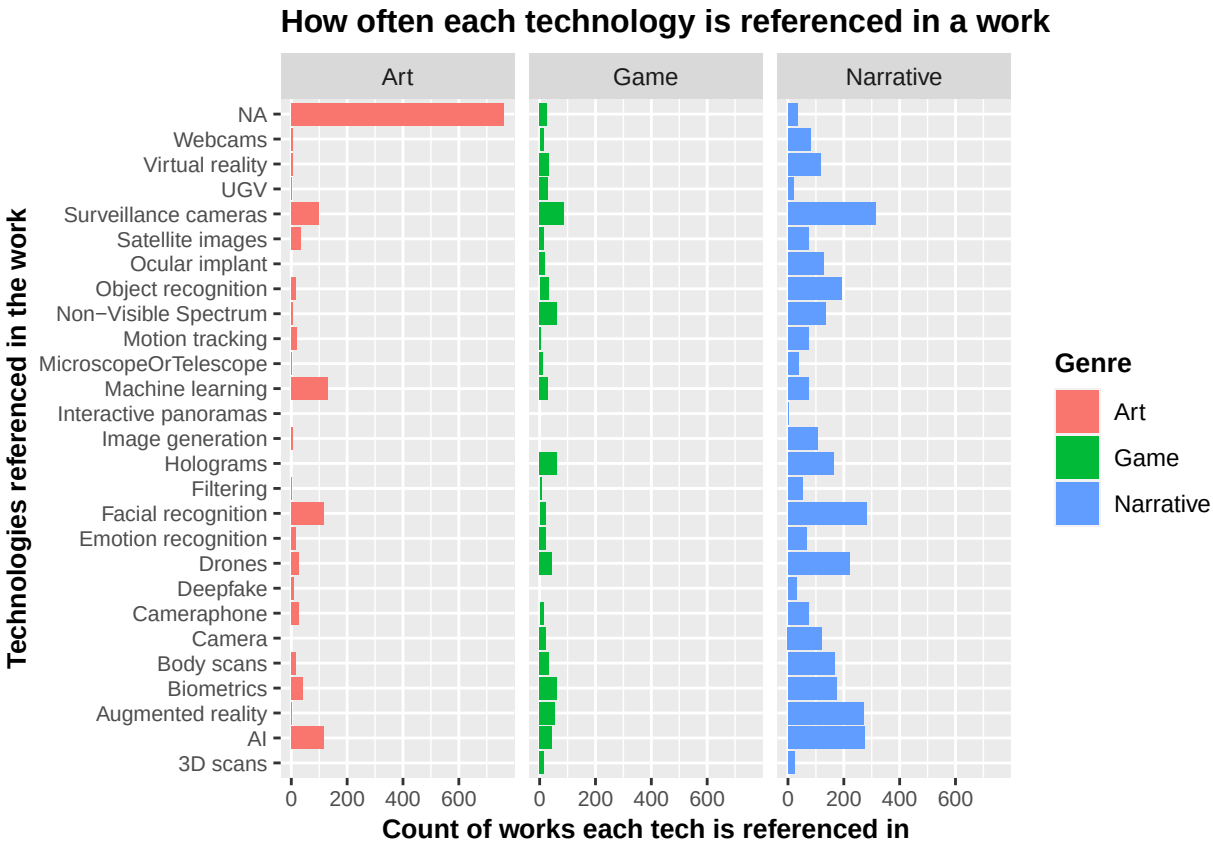
Which technologies are used in which genres and by whom?

```
# Selecting relevant columns and removing duplicates so the numbers are correct.
# This way each character occurs several times, but only once for each technology.
# Note that this looks at characters in relation to technologies referenced in the
# whole work they are in, rather than in relation to the specific situation.

# I would like to order the bars according to the fill, and have been trying to do this
# in "scripts/test-for-techref-char-graph.R"

Character_background %>%
        filter(!is.na(TechRef)) %>%
        select(TechRef, Character, Genre, Gender, Sexuality, Age,
               RaceOrEthnicity, HumanOrMachine) %>%
        distinct() %>%
        ggplot(aes(x = TechRef, fill = Gender)) +
        geom_bar(position = "fill") +
        scale_y_continuous(labels = scales::percent) +
        labs (x = "Technologies referenced in the work",
              y = "Gender of characters in works that reference these technologies",
              title ="Gender distribution of characters by technologies referenced in works") +
        theme(axis.text=element_text(size=8),
              title=element_text(size=10,face="bold")) +
        coord_flip() +
        facet_wrap(~Genre)
```



Which technologies are most referenced?

```
# See "scripts/tech_in_works.R" for attempts to make this for ALL tech, so with TechRef
# and TechUsed merged.

CreativeWorksFull %>%
        select(WorkTitle, Year, Genre, Country, TechRef, TechUsed, Sentiment) %>%
        distinct() %>%
        ggplot(aes(x = TechRef, fill = Genre)) +
        geom_bar()  +
        labs (x = "Technologies referenced in the work",
              y = "Count of works each tech is referenced in",
              title ="How often each technology is referenced in a work") +
        theme(axis.text=element_text(size=8),
              title=element_text(size=10,face="bold")) +
        coord_flip() +
        facet_wrap(~Genre)
```




```
# To do this with just Technology instead of TechRef vs TechUsed we could either
# - make one df x with just TechRef and another y with just TechUsed, rename
# both columns Tech then vbind and
# remove duplicates using distinct().
# - or use the technologies in situations instead and merge with creativeworks to
# connect them to the WorkTitle.
```

Characteristics of characters only in games and narratives

How many works actually have characters in the different genres?

```
ArtWorksWithCharacters <- Character_background %>%
        filter(Genre =="Art") %>%
        select(WorkTitle, Character) %>%
        distinct() %>%
        group_by(WorkTitle) %>%
        summarise(n = n()) %>%
        count(WorkTitle)

GamesWithCharacters <- Character_background %>%
        filter(Genre =="Game") %>%
        select(WorkTitle, Character) %>%
        distinct() %>%
        group_by(WorkTitle) %>%
        summarise(n = n()) %>%
```

```
        count(WorkTitle)

NarrativesWithCharacters <- Character_background %>%
        filter(Genre =="Narrative") %>%
        select(WorkTitle, Character) %>%
        distinct() %>%
        group_by(WorkTitle) %>%
        summarise(Total = n()) %>%
        count(WorkTitle)

# To display the number of works in each genre with characters:
# cat("Number of artworks with characters: ", sum(ArtWorksWithCharacters$n))
# cat("Number of games with characters: ", sum(GamesWithCharacters$n))
# cat("Number of narratives with characters: ", sum(NarrativesWithCharacters$n))

WorksWithCharacters <- c(sum(ArtWorksWithCharacters$n), sum(GamesWithCharacters$n),
                         sum(NarrativesWithCharacters$n))

CreativeWorks %>%
        select(WorkTitle, Genre) %>%
        distinct() %>%
        group_by(Genre) %>%
        summarise(TotalWorks = n()) %>%
        add_column(WithChar = WorksWithCharacters) %>%
        mutate(Percent = scales::percent(WithChar/TotalWorks))
```

```
## # A tibble: 3 x 4
##   Genre     TotalWorks WithChar Percent
##   <chr>          <int>    <int> <chr>
## 1 Art              190       33 17%
## 2 Game              77       54 70%
## 3 Narrative        233      214 92%
```

OK, so which narratives and games DON'T have characters?

```
# Find all the works that actually have characters.
CreativeWorks %>%
        select(WorkTitle, Character, Genre) %>%  # only use WorkTitle and Character columns.
        distinct() %>%                     # remove duplicates
        mutate(HaveChar = ifelse(!is.na(Character), TRUE, FALSE)) %>%
        select(WorkTitle, Genre, HaveChar) %>%
                filter(WorkTitle != "Beholder" & WorkTitle != "I`m Alan Partrige" & WorkTitle != "The Ma
               & WorkTitle !=  "The Expanse" & WorkTitle !=  "Six Wakes"
               & WorkTitle !=  "Fall; or, Dodge in Hell: A Novel") %>%  # delete these when new export
        distinct() %>%
        ggplot(aes(x = Genre, fill = HaveChar)) +
        geom_bar() +
        geom_text(aes(
                label = ..count..),
                stat = "count",
                position = "stack",
                size = 3,
                vjust = 1.2) +
```

```
        labs(x = "Genre of creative work",
             y = "Count of works in each genre",
              title ="Distribution of Characters by Genre") +
        scale_fill_manual(values=c("#999999", "#E69F00"),
                          name="Work\nincludes\ncharacters",
                          labels=c("No", "Yes"))
```


```
# Delete line with all the titles after new export (errors in Jan 2022 export)
```

Most of the artworks don't have characters. Which games and narratives don't have characters?

```
CreativeWorks %>%
        select(WorkTitle, Character, Genre) %>%  # only use WorkTitle and Character columns.
        distinct() %>%                    # remove duplicates
        mutate(HaveChar = ifelse(!is.na(Character), TRUE, FALSE)) %>%
        filter(HaveChar == FALSE & Genre != "Art") %>%
        filter(WorkTitle != "Beholder" & WorkTitle != "I`m Alan Partrige" & WorkTitle != "The Matrix"
               & WorkTitle !=  "The Expanse" & WorkTitle !=  "Six Wakes"
               & WorkTitle !=  "Fall; or, Dodge in Hell: A Novel") %>%  # delete these when new export
        select(WorkTitle, Genre) %>%
        arrange(Genre)
```

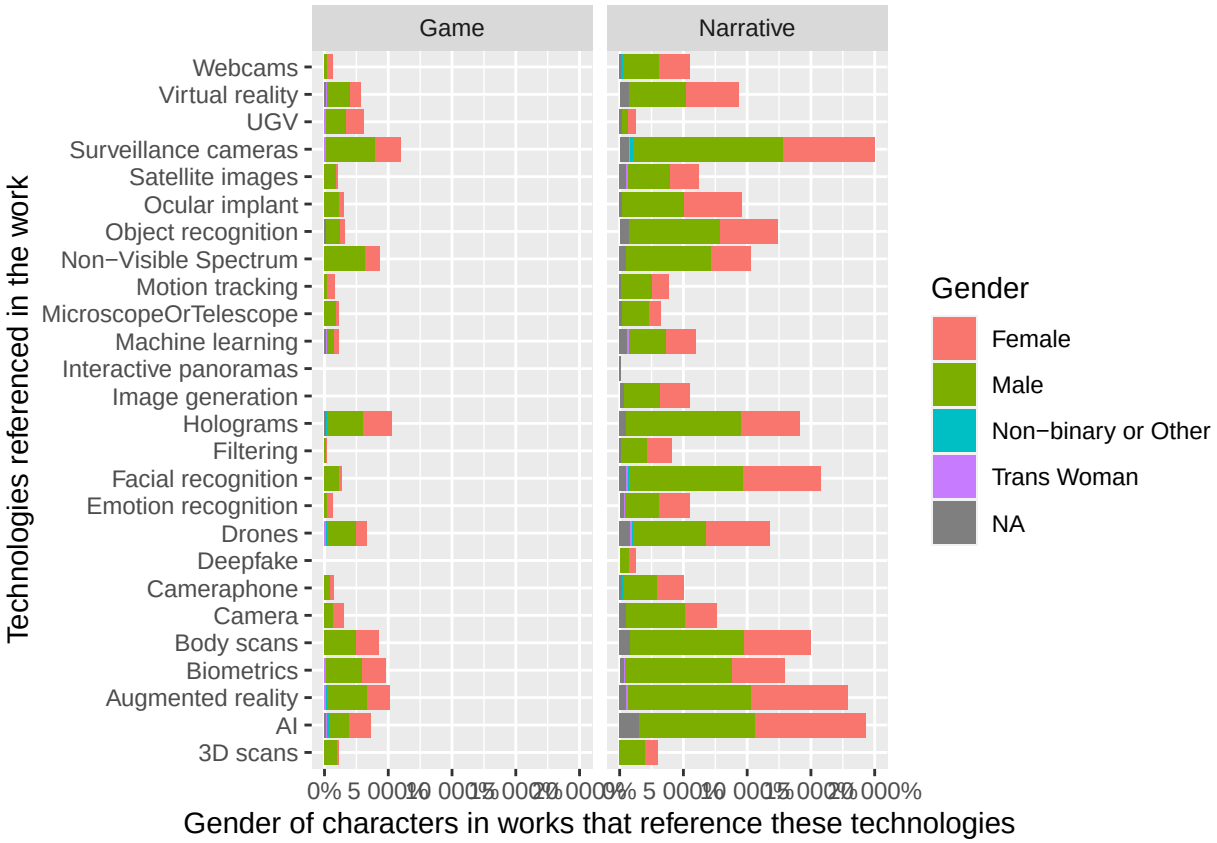
```
## # A tibble: 22 x 2
##    WorkTitle                Genre
```

```
##    <chr>                   <chr>
##  1 Hacker                  Game
##  2 Vigilance 1.0           Game
##  3 Shutter (game)          Game
##  4 Emotion Hero            Game
##  5 Face Your Feelings      Game
##  6 Scanner Sombre          Game
##  7 Do Not Feed the Monkeys Game
##  8 I'm on Observation Duty Game
##  9 Just Dance 2019         Game
## 10 Don't Look              Game
## # ... with 12 more rows
```

```
# Errors in creativeworks export Jan 2022 - some extra rows with Character = NA. Have fixed
# in database, need to get new export. Some records (like The Matrix) had extra rows with character
# marked as NA. This caused them to show up as though they had no character although they actually did.
```

Most artworks don't have characters in the sense that narratives and games do. To capture agency in these, we used general "Entities", as we called them, for general categories such as law enforcemnet, the creator of the artwork or humans in general. These were not assigned specific traits like gender and race. 8 out of 233 narratives have no characters, and most of these have situations where only technologies and entities are active. 13 out of 77 games have no characters, but all these have a "User" who interacts with machine vision technologies. The User in a game is the player when there is no clearly established player character.

```
Character_background %>%
        filter(!is.na(TechRef)) %>%
        select(TechRef, Character, Genre, Gender, Sexuality, Age,
               RaceOrEthnicity, HumanOrMachine) %>%
        filter(Genre == "Narrative" | Genre == "Game") %>%
        distinct() %>%
        ggplot(aes(x = TechRef, fill = Gender)) +
        geom_bar() +
        scale_y_continuous(labels = scales::percent) +
        labs (x = "Technologies referenced in the work",
              y = "Gender of characters in works that reference these technologies") +
        coord_flip() +
        facet_wrap(~Genre)
```



What kinds of characters interact with machine vision?

Here is a table to give us an overview. TO DO - Figure out how to include the Döring & Poeschl 2019 data – find a common key – use merge() - but I think that removes rows that don't have shared key, so watch out for that? – might want to split up humans vs robots first then rbind all after adding new columns to simplify? - Are there other stats I could include? - Consider installing a package to make tables more beautiful and informative. E.g. these tables with built-in bar charts.

Note that this is *all* characters across genres.

```
# It would be better to write a function rather than all this repetition. I can't
# figure out how though.

HumanGender <- Characters %>%
        filter(HumanOrMachine == "Human") %>%
        count(Gender, sort = TRUE, name = "Count") %>%
        mutate(Percent = round(Count/sum(Count)*100)) %>%
        mutate(Variable = "Gender", .before = 1) %>%
        mutate(Species = "Human", .before = 1)  %>%
        rename(Value = Gender)

HumanRace <- Characters %>%
        filter(HumanOrMachine == "Human") %>%
        count(RaceOrEthnicity, sort = TRUE, name = "Count") %>%
        mutate(Percent = round(Count/sum(Count)*100)) %>%
```

```
        mutate(Variable = "Race or Ethnicity", .before = 1) %>%
        mutate(Species = "Human", .before = 1) %>%
        rename(Value = RaceOrEthnicity)

HumanAge <- Characters %>%
        filter(HumanOrMachine == "Human") %>%
        count(Age, sort = TRUE, name = "Count") %>%
        mutate(Percent = round(Count/sum(Count)*100)) %>%
        mutate(Variable = "Age", .before = 1) %>%
        mutate(Species = "Human", .before = 1) %>%
        rename(Value = Age)

HumanSexuality <- Characters %>%
        filter(HumanOrMachine == "Human") %>%
        count(Sexuality, sort = TRUE, name = "Count") %>%
        mutate(Percent = round(Count/sum(Count)*100)) %>%
        mutate(Variable = "Sexuality", .before = 1) %>%
        mutate(Species = "Human", .before = 1) %>%
        rename(Value = Sexuality)

RobotGender <- Characters %>%
        filter(HumanOrMachine == "Robot") %>%
        count(Gender, sort = TRUE, name = "Count") %>%
        mutate(Percent = round(Count/sum(Count)*100)) %>%
        mutate(Variable = "Gender", .before = 1) %>%
        mutate(Species = "Robot", .before = 1) %>%
        rename(Value = Gender)

RobotRace <- Characters %>%
        filter(HumanOrMachine == "Robot") %>%
        count(RaceOrEthnicity, sort = TRUE, name = "Count") %>%
        mutate(Percent = round(Count/sum(Count)*100)) %>%
        mutate(Variable = "Race or Ethnicity", .before = 1) %>%
        mutate(Species = "Robot", .before = 1) %>%
        rename(Value = RaceOrEthnicity)

RobotAge <- Characters %>%
        filter(HumanOrMachine == "Robot") %>%
        count(Age, sort = TRUE, name = "Count") %>%
        mutate(Percent = round(Count/sum(Count)*100)) %>%
        mutate(Variable = "Age", .before = 1) %>%
        mutate(Species = "Robot", .before = 1) %>%
        rename(Value = Age)

RobotSexuality <- Characters %>%
        filter(HumanOrMachine == "Robot") %>%
        count(Sexuality, sort = TRUE, name = "Count") %>%
        mutate(Percent = round(Count/sum(Count)*100)) %>%
        mutate(Variable = "Sexuality", .before = 1) %>%
        mutate(Species = "Robot", .before = 1) %>%
        rename(Value = Sexuality)

CharacterStats <- rbind(HumanGender, HumanRace, HumanAge, HumanSexuality, RobotGender, RobotRace, Robot
```

```r
# Display the CharacterStats table using kable in the knitr library. Probably better to use
# a more fully featured table package, like Reactable (there are several others too).
#
# I set the NAs to show as Unknown in the table. Not sure whether this is the best option.

library(knitr)
options(knitr.kable.NA = "Unknown")
kable(CharacterStats, caption = "An overview of the characteristics of the characters.")
```

Table 1: An overview of the characteristics of the characters.

Species	Variable	Value	Count	Percent
Human	Gender	Male	313	58
Human	Gender	Female	211	39
Human	Gender	Unknown	11	2
Human	Gender	Non-binary or Other	3	1
Human	Gender	Trans Woman	2	0
Human	Race or Ethnicity	White	295	55
Human	Race or Ethnicity	Unknown	99	18
Human	Race or Ethnicity	Black	53	10
Human	Race or Ethnicity	Asian	41	8
Human	Race or Ethnicity	Person of Colour	36	7
Human	Race or Ethnicity	Complex	14	3
Human	Race or Ethnicity	Immigrant	1	0
Human	Race or Ethnicity	Indigenous	1	0
Human	Age	Adult	423	78
Human	Age	Young Adult	65	12
Human	Age	Child	22	4
Human	Age	Elderly	19	4
Human	Age	Unknown	11	2
Human	Sexuality	Unknown	329	61
Human	Sexuality	Heterosexual	192	36
Human	Sexuality	Homosexual	10	2
Human	Sexuality	Bi-sexual	7	1
Human	Sexuality	Other	2	0
Robot	Gender	Male	42	42
Robot	Gender	Female	37	37
Robot	Gender	Unknown	19	19
Robot	Gender	Non-binary or Other	2	2
Robot	Gender	Trans Woman	1	1
Robot	Race or Ethnicity	Unknown	54	53
Robot	Race or Ethnicity	White	35	35
Robot	Race or Ethnicity	Asian	5	5
Robot	Race or Ethnicity	Black	4	4
Robot	Race or Ethnicity	Person of Colour	2	2
Robot	Race or Ethnicity	Indigenous	1	1
Robot	Age	Adult	67	66
Robot	Age	Unknown	24	24
Robot	Age	Child	6	6
Robot	Age	Young Adult	4	4
Robot	Sexuality	Unknown	83	82
Robot	Sexuality	Heterosexual	17	17

Species	Variable	Value	Count	Percent
Robot	Sexuality	Homosexual	1	1

This overview shows that for human characters,

Gender and age are explicitly described or shown in almost 90% of human characters, and race or ethnicity are made explicit in 82% of the time. Sexuality however, is only explicit in 39% of human characters. Robots do not have the cultural and biological ties that humans do, but a vast majority (82%) of the robot characters in our dataset are still explicitly assigned gender.

(This is probably not significant) Interestingly, a greater proportion of robots than humans are assigned explicitly non-binary genders. Out of 540 humans, there are only 3 non-binary and 2 transgender characters, so 1% of the characters. There are only 101 robots in the dataset, but 2 are non-binary and one transgender, which is equivalent to 3%. (Not a big difference after I cleaned the data to remove the robots that were obviously miscoded, like 3CPO).

```
Characters %>%
       group_by(HumanOrMachine) %>%
       summarise(Total = n())
```

```
## # A tibble: 2 x 2
##   HumanOrMachine Total
##   <fct>          <int>
## 1 Robot            101
## 2 Human            540
```

Let's take a closer look at the non-binary and transgender characters:

TO DO: - Merge creativeworks and characters to show chars, title of creative work, country, year, creator, genre - Look at SEER robot - is it correct to view this as non-binary? - might want to go more specific to narratives/games?

Non-binary characters

There are very few explicitly non-binary or transgender characters in the dataset. Here is a complete list:

```
Character_verbs %>%
       filter(Gender == "Non-binary or Other" | Gender == "Trans Woman") %>%
       select(Character, Genre, Gender, Sexuality) %>%
       distinct()
```

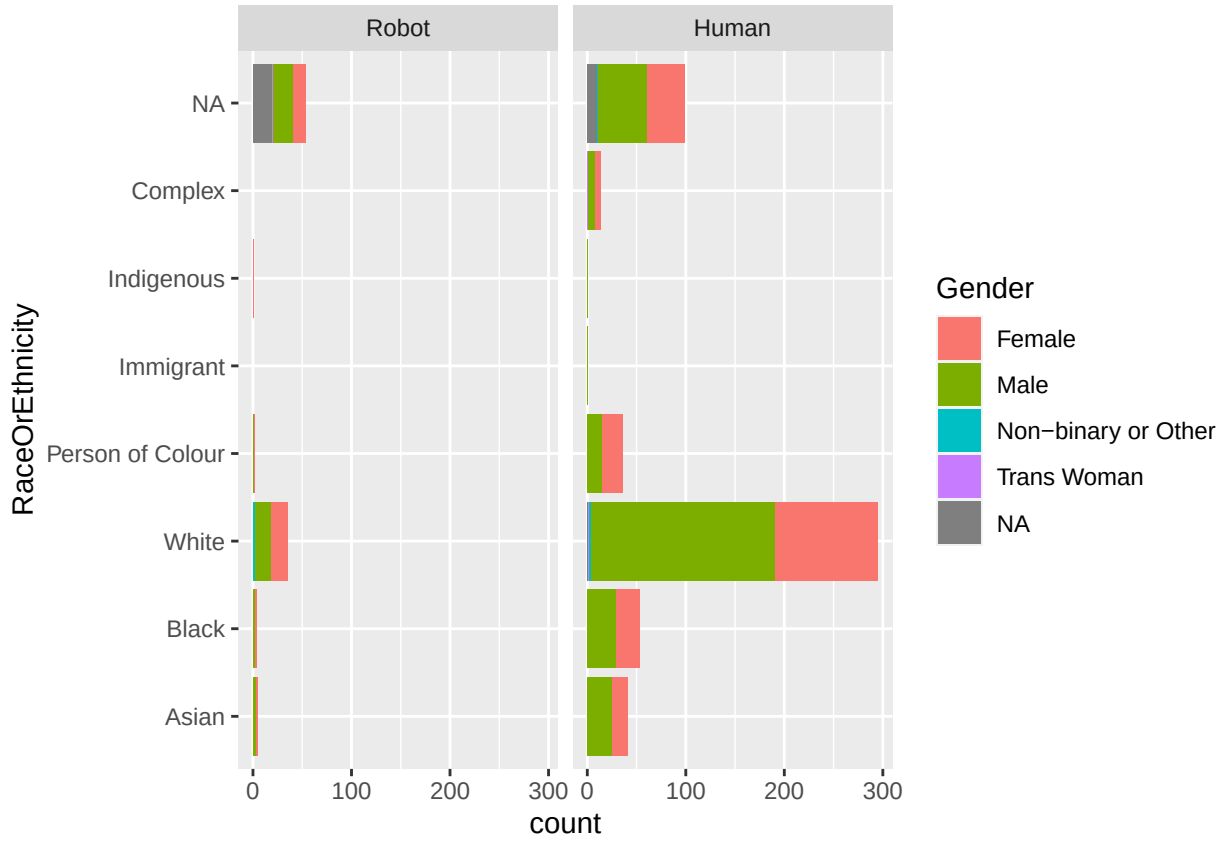
```
##              Character     Genre              Gender    Sexuality
## 1 "Chelsea E. Manning"       Art         Trans Woman         <NA>
## 2        "Mood Killer" Narrative Non-binary or Other         <NA>
## 3              Max Lao      Game         Trans Woman         <NA>
## 4                 Niki Narrative Non-binary or Other         <NA>
## 5              Paladin Narrative         Trans Woman Heterosexual
## 6            Red Girls      Game Non-binary or Other         <NA>
## 7           SEER Robot       Art Non-binary or Other         <NA>
## 8                 Zizi       Art Non-binary or Other         <NA>
```

Niki is an explicitly non-binary teenager in Ted Anderson and Jen Hickman's 2018 cyberpunk comic book *Moth & Whisper*. The publicly non-binary musician and performance artist Mood Killer plays a role as a sexually androgynous "cam boy" in their music video by that name. Zizi is a character created by combining deepfakes of drag artists in Jake Elwas's 2020 artwork *The Zizi Show*. Other characters tagged as non-binary are more difficult to place. The "Red Girls" in the videogame Nier:Automata are for instance represented visually as two traditionally feminine girls, but they are instantiations of AIs and can multiply, and have a deep, booming male voice. While it may seem a stretch to call these "non-binary" in the sense that is usually used for humans, they do represent a queering of binary gender that we wanted to capture.

There are also very few transgender characters: only three. The representation of Chelsea Manning in *Probably Chelsea* is one. The robot Paladin from Annalee Newitz's novel *Autonomous* (2017) is the character where gender transitioning is most clearly represented. Paladin is a combat robot with an organic human brain that assists with facial recognition. At first, Paladin's operator, Eliasz, thinks of Paladin as male, and when sexual attraction begins to arise between them, this becomes a problem for Eliasz, who nourishes some homophobia. Paladin does some research to find the origin of the human brain, and is thrilled when she is told that the brain came from a woman. She begins to see herself as female, and when she shares this new identity with Eliasz, their sexual and romantic relationship flourishes. The third transgender character in the dataset is Max Lao in the game *Technobabylon* (2015). Here the transition is mentioned, but not emphasised in the game.

We may have missed some non-binary or transgender characters, but overall the lack of representation affirms other findings. In an analysis of transcripts of dialogue and other text in ten video games, Frank Heritage found no uses of "they" that referred to non-binary characters[@heritage2021]. In a survey of games in Adrienne Shaw's LGBTQ Video Game Archive, Utsch et.al. found that while LGBTQ characters in video games have increased since the 1980s, the majority are gay men, and LGBTQ characters of all kinds are far more prevalent in Adventure and RPGs than in Fighting, Simulation or Shooter games[@utsch2017].

```
Characters %>%
        ggplot() +
        geom_bar(aes(y = RaceOrEthnicity, fill = Gender)) +
        facet_grid(. ~ HumanOrMachine)
```



There are considerably more male than female humans. Interestingly the genders are more even among the robots. You can also see that there aren't many NAs for gender - most of the characters were represented in a way that clearly expressed a specific gender. The machine characters were an exception.

We used separate tags for "Machine" and "Cyborg" species because we were interested in the agency of humans in comparison to non-humans, and the cyborg seemed a potentially interesting hybrid categori. We tagged robots and androids, as well as AI systems, as the "Machine" species, and defined "Cyborg" as "A [@heritage2021a]being with both organic and mechanical/machinic body parts."

TRYING TO MAKE A TABLE https://stackoverflow.com/questions/29549731/finding-percentage-in-a-sub-group-using-group-by-and-summarise

Comparing with the representation in . . .

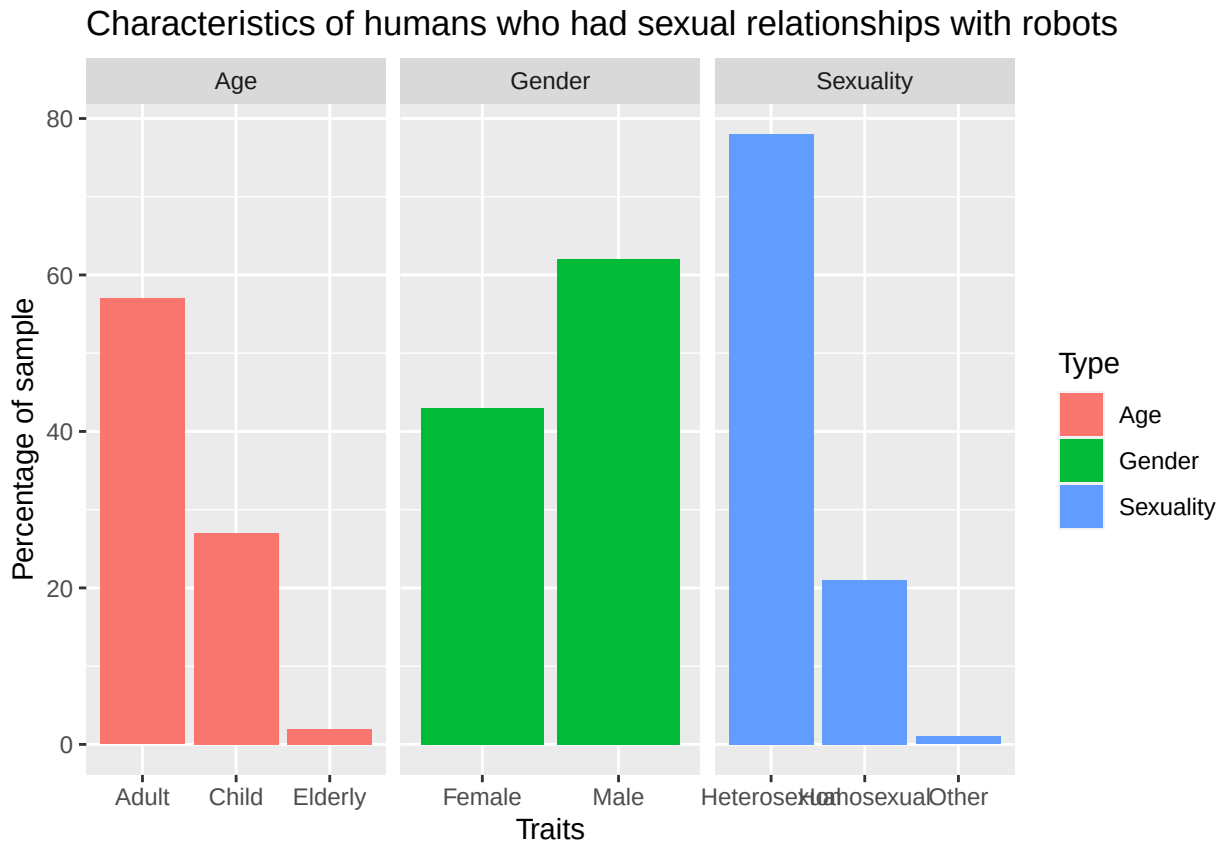
This compares (how) to Nicola Döring and Sandra Poeschl's 2019 content analysis of the portrayal of sexual relationships between humans and robots in fiction and non-fiction[@döring2019a], which is one of the few studies to explore characteristics of relationships between humans and technologies. Heritage uses NLP and finds that men are more often agents SUMMARISE THIS BETTER - also Marianne's creepypasta [@gunderson2020]

NB: Not much point in including sexuality as a property of humans or robots since it's actually a property of hte *relationship*, not the individual.

Also need to actaully COMPARE with my data. Maybe put in same table, somehow?

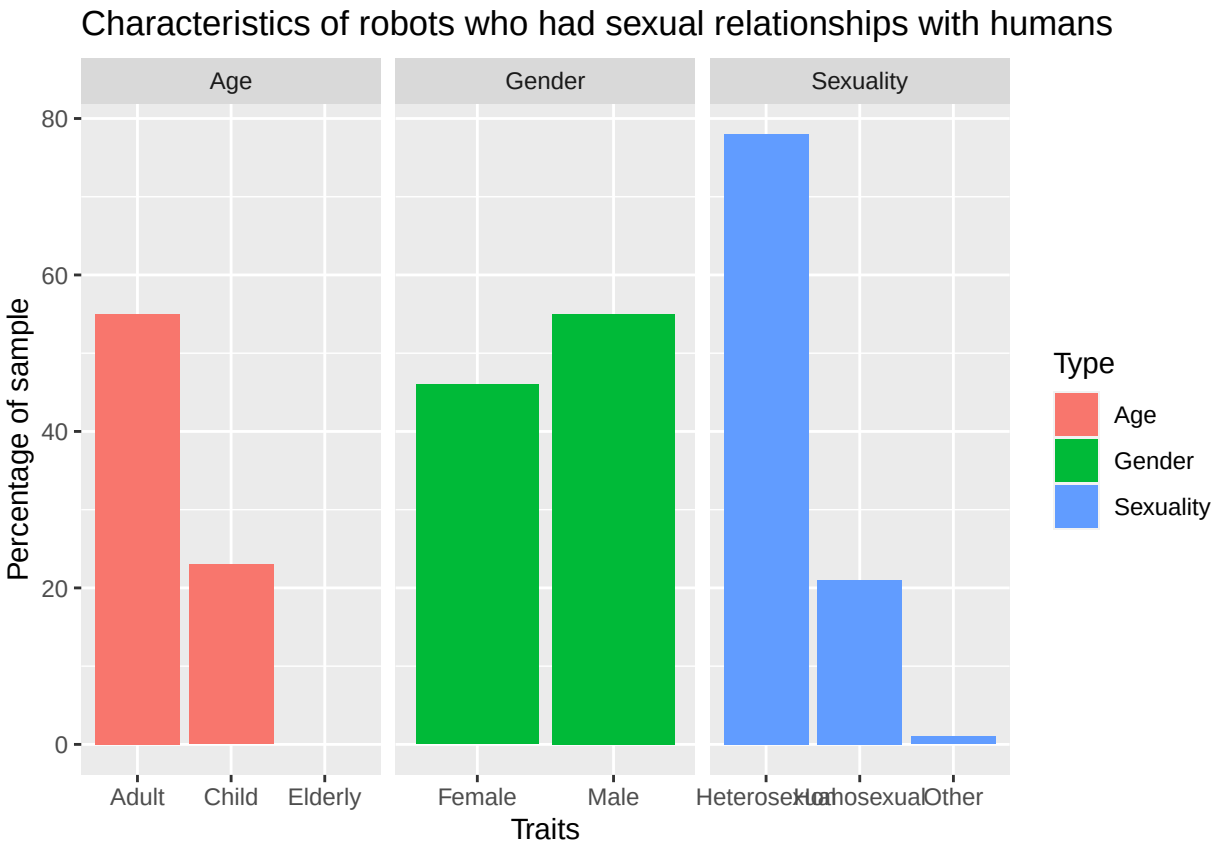
```
# A simple visualisation of the distribution of gender and
# sexuality as found in this paper:
#
```

```r
# Döring, N., Poeschl, S. Love and Sex with Robots: A Content Analysis of
# Media Representations. Int J of Soc Robotics 11, 665-677 (2019).
# https://doi.org/10.1007/s12369-019-00517-y
#
# I manually copied the data for sexbots from [Table 3]
# (https://link.springer.com/article/10.1007/s12369-019-00517-y/tables/3) in
# the paper, using the values form the column % under Fictional media. It's
# important to realise that the numbers are PERCENTAGES.

sexbots <- structure(list(
        Variables = c("Female", "Male", "Child", "Adult", "Elderly",
                      "Heterosexual", "Homosexual", "Other"),
        Human = c(43, 62, 27, 57, 2, 78, 21, 1),
        Robot = c(46, 55, 23, 55, NA, 78, 21, 1),
        Type = c("Gender", "Gender", "Age", "Age", "Age", "Sexuality",
                 "Sexuality", "Sexuality")),
        row.names = c(NA, -8L),
        class = c("tbl_df", "tbl", "data.frame"))

ggplot(data = sexbots, group = Type) +
        geom_col(mapping=aes(x = Variables,
                             y = Human,
                             fill = Type)) +
        ggtitle("Characteristics of humans who had sexual relationships with robots") +
        ylab("Percentage of sample") +
        xlab("Traits") +
        facet_wrap(~ Type, scale="free_x")
```




```
ggplot(data = sexbots, group = Type) +
        geom_col(mapping=aes(x = Variables,
                             y = Robot,
                             fill = Type)) +
        ggtitle("Characteristics of robots who had sexual relationships with humans") +
        ylab("Percentage of sample") +
        xlab("Traits") +
        facet_wrap(~ Type, scale="free_x")
```

```
## Warning: Removed 1 rows containing missing values (position_stack).
```



Human Sexuality

Why are there so few queer characters? The number of characters explicitly identified as queer in the database is very low.

$$givenumbers, explaincategories$$

.

There are a lot of characters who are never described as having any sexual or romantic feelings. These are in the category of characters any sexuality or being involved in romantic relationships. This is often the case with robots, like BB8 or R2D2 in Star Wars, or with child-like characters, like Vanellope and Ralph in Ralph Breaks the Internet (check title). Zombies and

$$moreexamples$$

are usually not see marked as "Unknown/Not applicable". Most of these are not represented as having n in a sexual light either. Groups of people also fall into this category.

Sometimes we see a character that in many ways is presented as "culturally queer", but who is never shown in any sexual or romantic situation. Shuri in the Marvel universe is an example of this. There is a lot of speculation online that Shuri is lesbian or bisexual, and but Marvel avoids showing her

We logged characters as heterosexual or homosexual if they were mentioned as attracted to or in a sexual or romantic relationship with a character of the opposite or same gender. If both genders, we logged them as bisexual. Sometimes we may not be aware of all presentations of the character [e.g. Dr Who, shuri, characters in franchises – might not have read all – so perhaps they have a same-sex relationship in season X and so really aren't exclusively heterosexual.

Obviously problematic to lock people into fixed idnetities like this. Fictional representations though, not actual people. We're interested in how they're repreented. But unavoidably using lenses that can lead to stereotypical depiction – unfortunate.

$$is there scholarship on this that we could use here?$$

We have not logged the most important characters in the stories overall, but only characters who are directly involved with machine vision technologies. So it is possible that gay and queer characters are more visible in the works as a whole, but queer characters are less likely to be presented as engaging with machine vision technology. Unfortunately we don't have data to prove or disprove this.

Robot Sexuality

You may be surprised to find machine characters - robots and AIs, mostly - have sexuality at all. As you can see below, the vast majority of machine and cyborg characters *don't* have any explicit sexuality.

80% of the machine and cyborg characters are explicitly gendered. Out of 76 machine characters, there are only 15 where the gender is unknown or not applicable (20%). 6 are explicitly non-binary, the others are fairly evenly split between female-identifying (26) and male-identifying (29). The distribution is similar for cyborg characters, with 3 out of 14 having unknown or not applicable gender, as well as 6 with male and 5 with female identities.

In contrast, only 10% of machine characters and 35% of cyborgs are portrayed as have a romantic or sexual interest in another character. All these have an assigned gender.

gender is usually assigned (how are demographics in comparison to western countries?) but sexuality is far less frequently made explicit. When sexuality is explicit, heterosexuality is heavily overrepresented (check).

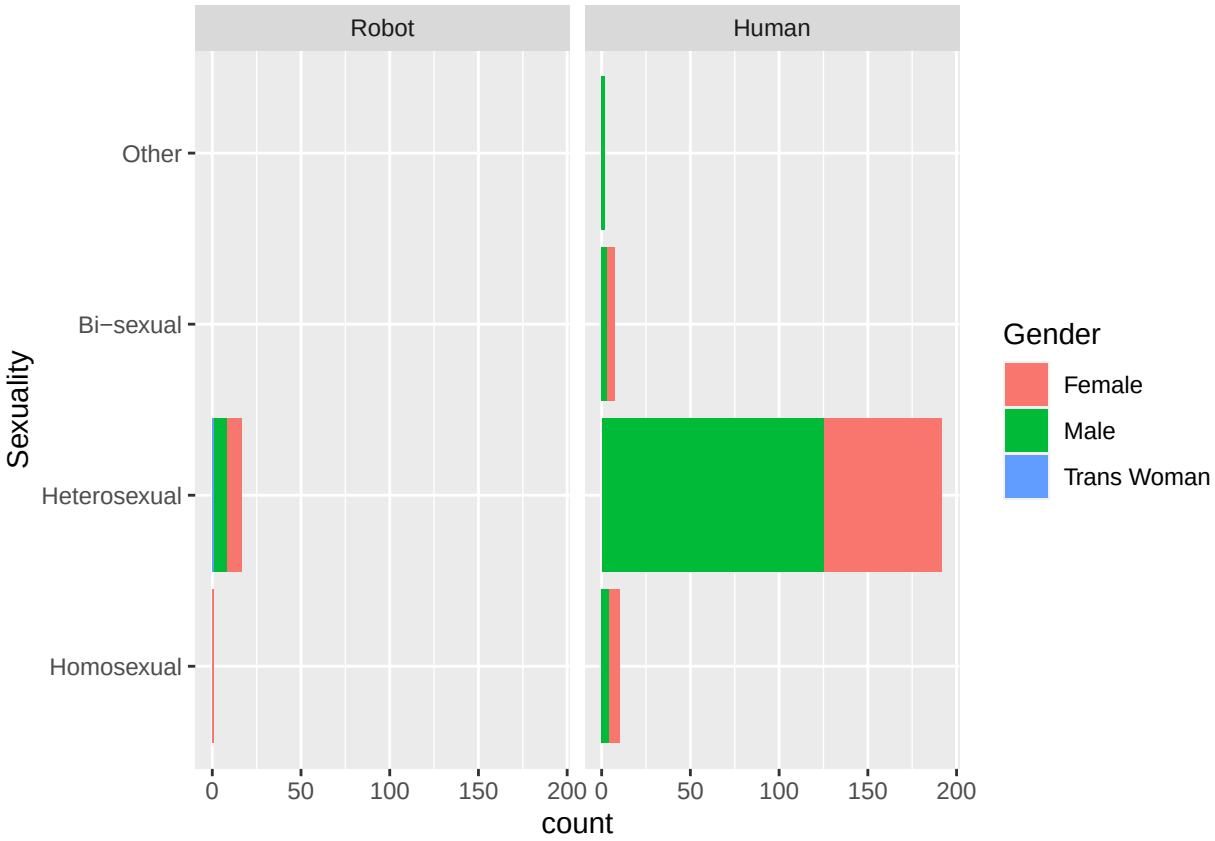
Far fewer characters are explicitly shown as homosexual or bi-sexual than would be expected based on demographics in Western countries. Only 10 characters are tagged as homosexual. All are entered by Jill... which suggests a methodological issue where other people haven't entered sexuality? Or perhaps there just aren't homosexual characters in games and art? A few of the seven bi-sexual characters are entered by others. Lower visibility of queer characters – this is typical and I can cite something about it. Also research team may be defaulting to straight, and there are a lot of Unknown/not applicable.

Compare gendering of machine characters to gendering of voice assistants etc. (Cite Lai-Tze Fan if her book Unseen Hands is out or something else she has done)

Let's make a new table called "Robots" that just has the robots and cyborgs and AIs - the characters that are tagged as "Machine" or "Cyborg".

OK; I haven't figure that out yet but look at this ;)

```
Characters %>%
        drop_na(Sexuality) %>%
        ggplot() +
        geom_bar(aes(y=Sexuality, fill = Gender)) +
        facet_wrap(.~HumanOrMachine)
```



Maybe a table would be a better way of representing this data, so we can see which machine characters are portrayed as having a sexual preference.

Oops

- the followign analysis is for a version of this that no longer works sigh

Emily at the top of that list is the narrator of the novel *Emily Eternal* by M.G. Wheaton (2019). She is an artificial intelligence designed by MIT researchers to feel empathy with humans, and she is experienced as physically present by humans who wear an interface chip that allows Emily to interface directly with their brains. In the first pages of the novel this is explained to the reader:

> The chip allows me to manipulate Regina's senses of sight, smell, touch, and hearing. Her eyes tell her brain there's a Caucasian woman in her early thirties with brown hair, blue-green eyes, and a kind face sitting opposite her. Her ears tell her my voice has a mid-range pitch, not too low, not too high, with a slight New Englander's accent. Her nose tells her I use mostly fragrance-free soap, a kiwi-infused shampoo, no perfume, but a baby powder–scented antiperspirant. When I touch her hand or even embrace her, I come off as warm, upright but not rigid, and a good hugger. In return, the chip gives me unlimited access to her brain, including thoughts, memories, learned behaviors, hopes and dreams, worst fears, and all things in between. Utilizing bioalgorithms, I'm able to create a comprehensive neural map of an individual's mind that can then be used in a therapeutic context to help patients with their issues, large or small. (Wheaton, M. G. 2019. *Emily Eternal* (pp. 5-6). Grand Central Publishing. Kindle Edition.)

This is surprisingly typical: machine characters tend to be described as very human-like, and one of the ways they are anthropomorphised is by being assigned explicit gender, sexuality and race. The above

excerpt makes it easy to tag Emily for our dataset: her species (an AI, thus a "Machine"), race, gender and age are all explicitly described. Her sexuality is very clear later on in the novel: she is in love with Jason:

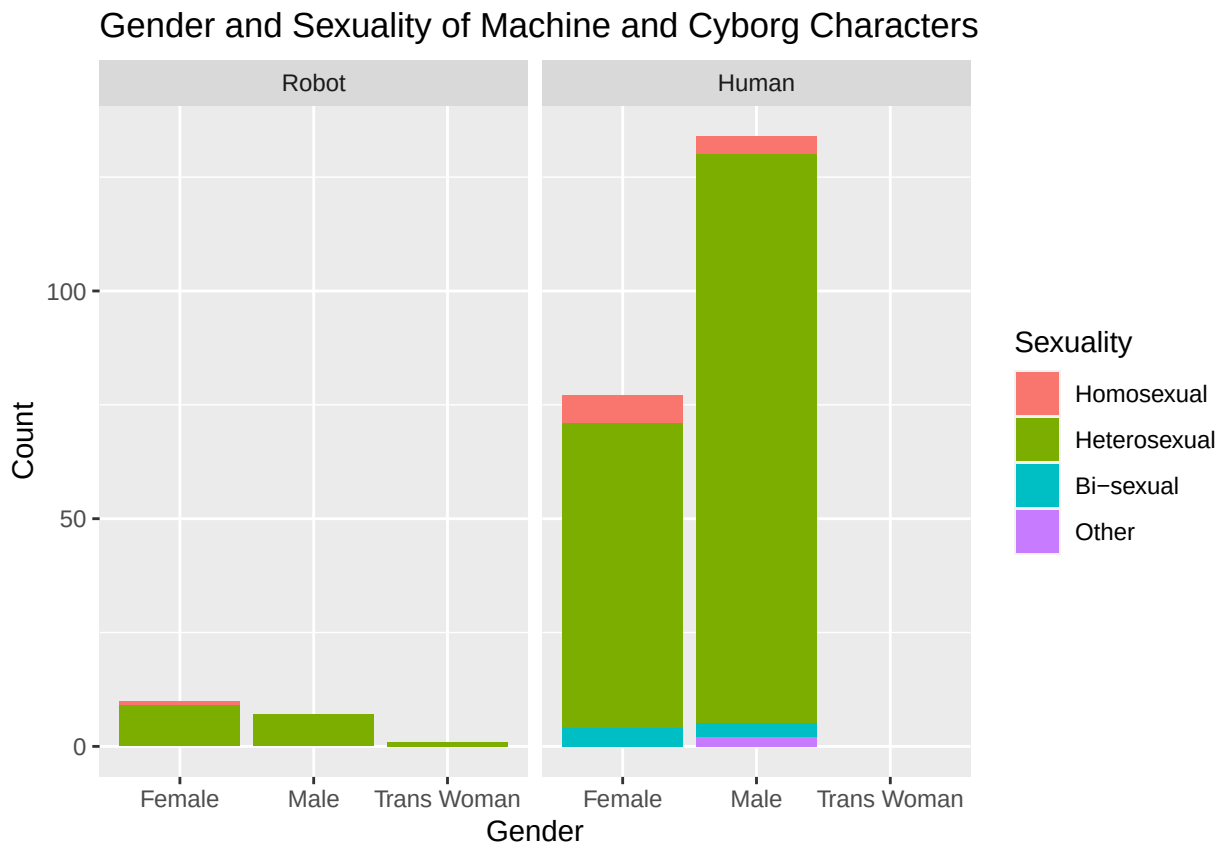
> So yeah, he's my crush. Not entirely sure when it started (yeah, right—Emily's Crushing On Jason Protocol went online precisely nine weeks, four days, three hours, and fifty minutes ago. . . he was wearing a green cardigan), but it took a few hours of systems analysis to make certain it had developed naturally as part of my evolving socialization rather than another mental mouse maze my colleagues had dropped me into. That my emotional responses to it are not only real but ever-changing, as well as its possible-to-even-likely transience, is why I've kept it cloaked from my team. (p. 23).

```
Characters %>%
        filter(Character == "Emily")
```

```
## # A tibble: 1 x 6
##   Character HumanOrMachine Gender Sexuality    RaceOrEthnicity Age
##   <chr>     <fct>          <fct>  <fct>        <fct>           <fct>
## 1 Emily     Robot          Female Heterosexual White           Adult
```

Why is there only one gay robot?

```
Characters %>%
        group_by(Sexuality) %>%
        filter(!is.na(Sexuality)) %>%
        ggplot() +
        geom_bar(aes(x=Gender, fill=Sexuality)) +
        facet_grid(. ~ HumanOrMachine) +
        scale_size_area(max_size = 7) +
        labs(title = "Gender and Sexuality of Machine and Cyborg Characters",
             y = "Count",
             x = "Gender") # +
```




```
#       theme(plot.title = element_text(size = 10)) # including this line makes
#                                                   # the whole plot disappear!
```

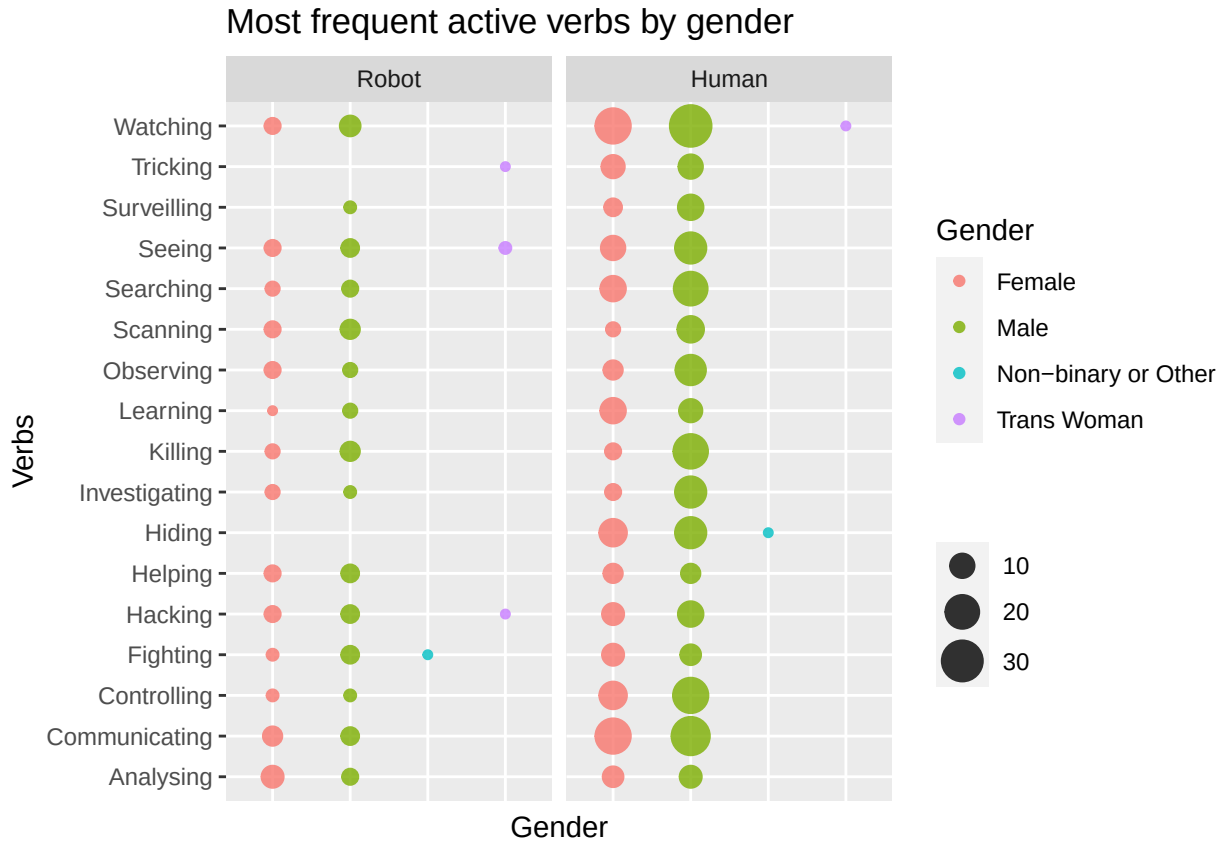
What do characters *do* with machine vision?

There are some notable differences in the verbs associated with different genders in our dataset.

Active verbs

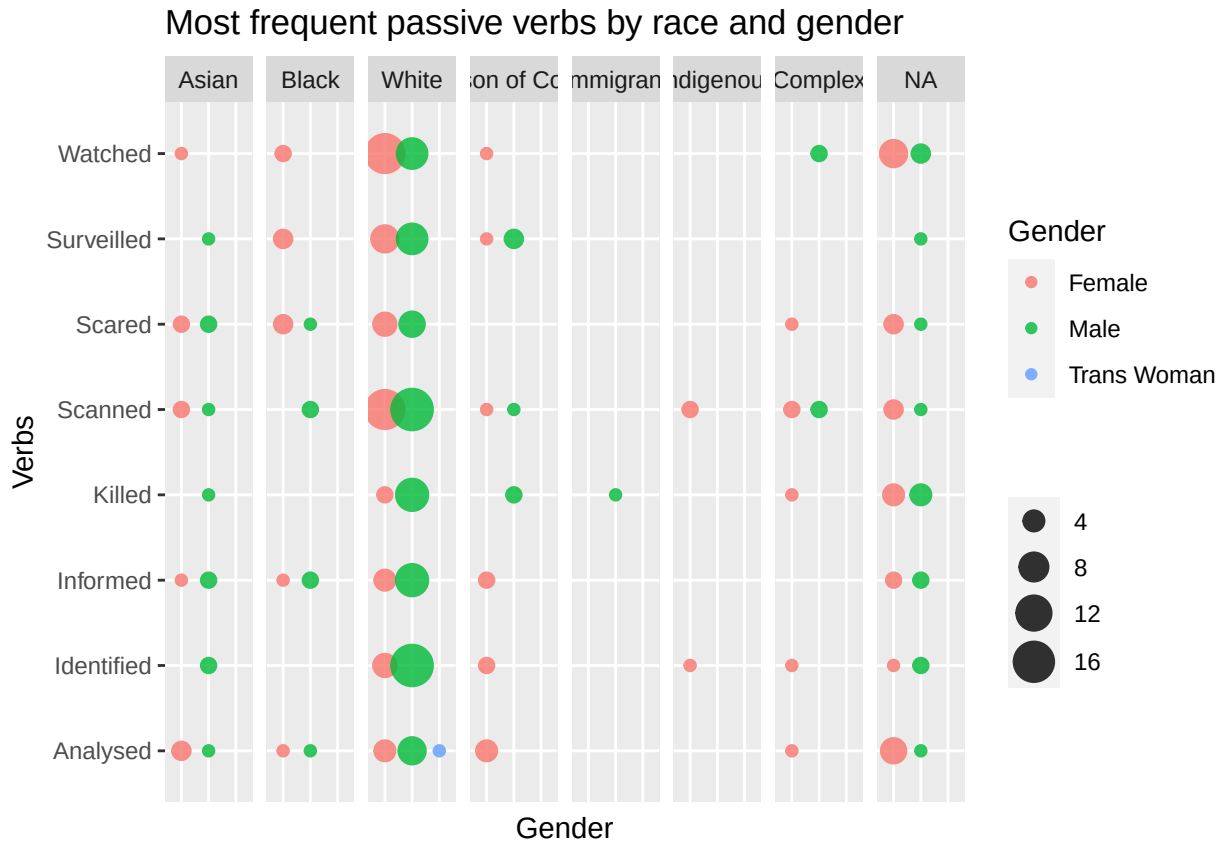
```
# Using the Tidyverse library the following code takes Character_verbs and then
# does the following:
# 1. Adds a column called VerbCount that gives a count of each verb
# 2. Arranges the dataframe in decreasing order by VerbCount
# 3. Filters the dataframe to only include rows where the VerbCount is more
# than 20
# 4. Filters the dataframe to only include rows where Gender isn't unknown.
# 5. Filters the dataframe to only include rows where the string "ing" is
# detected in the Verb variable.
# 6. Makes a plot where the visible objects are based on the Verb and Gender
# columns.
# 7. Uses geom_count to plot this - this is a geom that is good for discrete or
# categorical values (like text rather than numbers) that shows a bigger or
```

```
# smaller point (circle) depending on how many co-occurences each x and y value
# have.
# 8. Set the max size of the points to 7.
# 9. Adds labels to the diagram.

Character_verbs %>%
  add_count(Verb, name = "VerbCount") %>%
  arrange(desc(VerbCount)) %>%
  filter(VerbCount>20) %>%
  filter(Gender != "Unknown") %>%
  filter(str_detect(Verb, "ing")) %>%
  ggplot(aes(y = Verb, x = Gender)) +
  geom_count(aes(colour = Gender), alpha = 0.8) +
  facet_grid(. ~ HumanOrMachine) +
  scale_size_area(max_size = 7) +
  labs(title = "Most frequent active verbs by gender",
       y = "Verbs",
       x = "Gender",
       size = "") +
   theme(axis.ticks.x = element_blank(),
         axis.text.x = element_blank())
```



Passive Verbs

```
Character_verbs %>%
  add_count(Verb, name = "VerbCount") %>%
  arrange(desc(VerbCount)) %>%
  filter(VerbCount>20) %>%
  filter(Gender != "Unknown") %>%
  filter(str_detect(Verb, "ed")) %>%
  ggplot(aes(y = Verb, x = Gender)) +
  geom_count(aes(colour = Gender), alpha = 0.8) +
  facet_grid(. ~ RaceOrEthnicity) +
  scale_size_area(max_size = 7) +
  labs(title = "Most frequent passive verbs by race and gender",
       y = "Verbs",
       x = "Gender",
       size = "") +
   theme(axis.ticks.x = element_blank(),
         axis.text.x = element_blank())
```



(the following is from SLSA version) Notably, the male characters are portrayed as having far more active relationships with machine vision technologies than characters of other genders are. This indicates that women are more often portrayed as the object being acted upon by machine vision, while men are more often portrayed as active subjects making use of machine vision to further their own agency. While both men and women in the dataset are often portrayed to be watching, women are more frequently being watched, while men are a lot more likely to be the ones doing the watching rather than being watched. Men are also

more likely to be killed by machine vision technologies, and there are more male characters overall in our dataset.

Overall, this confirms typical gender stereotypes. Apparently, the machine gaze is often aligned with the male gaze we know from decades of cinema (Mulvey 1975), where men watch and people who are not men are watched. Women do use machine vision technologies to watch others almost as often as they are watched, though, and it's worth noting that communicating is a very common use for both men and women, although more so for women. Killing and being killed almost only happen to male characters in our dataset.

Do Robots and Humans interact with machine vision differently?

```
# Fix this to work with new Characters script
#
# Character_verbs %>%
  # Count how many occurances there are of each Verb. Make a new table
  # with a list of Verbs in one column and a second column indicating
  # how many times each Verb is used. Sort this in descending order.
#  count(Verb, sort = TRUE) %>%
  # Take the first 20 rows of the new table that count() generated.
#  slice(1:20) %>%
  # Now join that table to the Character_verbs table. R will drop all
  # the rows with verbs NOT on the top-20 list.
#  left_join(Character_verbs) %>%
  # Add a column renaming (recoding) Cyborgs to Machine so as to merge
  # Machine and Cyborg characters. Leave humans as humans and call
  # everything else "Other" so we can easily get rid of it and just focus
  # on Machines and Humans.
#  mutate(HumanOrMachine = recode(Species,
#                                 "Machine" = "Machine",
#                                 "Cyborg" = "Machine",
#                                 "Human" = "Human",
#                                 .default = "Other")) %>%
# Remove all the Other species. Then remove all rows where Gender
# is Unknown.
#  filter(HumanOrMachine != "Other") %>%
 # filter(Gender != "Unknown") %>%
# Use ggplot() to generate a barchart.
#  ggplot() +
# the geom is barchart. Put Verb on the y axis, count will be on x.
# The fill colour depends on the Gender. (Note: set a colour palette
# later)
#  geom_bar(mapping = aes(y = Verb, fill=Gender)) +
# Make a grid with a separate graph for each variable in HumanOrMachine.
#  facet_grid(cols = vars(HumanOrMachine)) +
# Add labels.
#  labs(title = "What characters do with machine vision",
#       y = "The character is...",
#       x = "Number of occurences",
#       size = "")
```

To do - more visualisations to make

1. Make a new column with values active or passive and use facet_grid() to make two different graphs based on that.
2. Try putting the bars for machines/humans on top of each other (in same graph)
3. Try making verbs proportional rather than a simple count so it's easier to compare machine/human.
4. Add some charts showing how many characters of each gender etc.
5. Could select just a few verbs (killing/killed? Scanning/Scanned?) and do more detailed graphs of the characters that use those verbs.
6. Connect to technologies, entities in situations?
7. Change colour palette - find the colourblind-friendly palette.
8. Obviously connect to close readings.
9. Create node and edge tables and feed into Gephi.
10. Test other kinds of graph.
11. What if the grid takes gender as main factor rather than species?
12. Bring genre art/game/narrative back and run the verbs based on that.

Further research

Linking this data to existing crowdsourced and commercial datasets, or to research datasets, could yield new insights. Some examples of possible research questions: - Add star ratings from IMDB (movies) or Goodreads (novels) or other services to find out whether there is a correlation between high ratings and the representation of certain kinds of human-technology relationship. -

UPDATE THIS - THIS IS FROM SLSA PAPER

Most research on fictional and artistic depictions of technologies uses literary analysis and close reading as its method. We hope that using digital humanities methodologies will allow us to say something about the larger picture, and that it will generate new research questions that can be explored in more nuance and detail through close readings. Quantitative analysis can suggest overall tendencies and patterns, but to really understand how machine vision technologies are portrayed in art, games and narratives we will need to follow up the distant reading with close readings. Combining exploratory analysis as in this paper with "the beauty of small data," as Gabriel Recchia puts it (2020, 389), will be a major part of our future research, and we have begun to publish some of this, such as Linda Kronman's analysis of digital art and machine vision (Kronman 2020).

Acknowledgements

This research is part of the project Machine Vision in Everyday Life: Playful Interactions with Visual Technologies in Digital Art, Games, Narratives and Social Media, which has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement no. 771800).

References