The Invisible Machine: Big Data and You

This lesson is part of USE, UNDERSTAND & ENGAGE: A Digital Media Literacy Framework for Canadian Schools: http://mediasmarts.ca/teacher-resources/digital-literacy-framework.

Overview

In this lesson, students examine a fictional social network profile to learn how online platforms collect data about their users. They then read an article that explains how platforms use this data and explores some of the issues this raises. Finally, they create a mind map of their own online data profile and reflect on how the data they post may be collected and used by others.

Learning Outcomes

Students will:

- Learn how and why internet platforms collect and use personal data
- Understand the commercial value of their data
- Learn strategies to help them control the collection of their data and the impact of algorithms on their online experience
- Reflect on the impact of data collection and algorithms on their lives and on society and what governments and/or corporations should do to change things

Preparation and Materials

Prepare to project the slideshows *Sada's Profile (Public View)* and *Sada's Profile (Data View)*

Photocopy the article *The Invisible Machine: Big Data and You* and the assignment sheet *Mapping Your Data Profile*

Optional: Prepare to project or distribute the handout *Defining the Invisible Machine*

Procedure

Sada's Profile

Start by asking students how many of them have an account on Instagram or another social network. Then ask them if they know how social networks like Instagram, Snapchat, etc., make money. (Short discussion - no definitive answers, but make sure advertising comes up.)

Now ask students how much they think people know about them, based on their social network accounts. Have students give some examples of things people might know (or think they know) about them from their online profile. How many have ever posted something that they *intended* to give people a particular impression of themselves?

Tell students that you are going to show them a fictional Instagram profile, then project *Sada's Profile (Public View)* and ask students what conclusions they can draw about Sada based on her profile.

After students have shared their responses, ask if they know:

- Whether or not Sada Moore is the person's real name?
- Where Sada lives?
- Where Sada works?
- Where Sada saw the movie?
- As what race or ethnic origin, if any, does Sada identify?
- As what gender, if any, does Sada identify?

In some cases students will be able to glean some of this information (the city Sada lives in, that Sada likes black and white photography) but point out that in general Sada's profile doesn't reveal much about Sada; ask students if they take steps not to give out directly identifying information.

Now tell them that you are going to show them what Sada's profile looks like *to the company that owns the social media platform*. Project *Sada's Profile (Data View)* and ask them if they can now answer the same questions:

- Is Sada Moore the person's real name? *No, it's Reilly Smith*
- Where does Sada live? *24 2nd Street W., Laughia-on-the Lake*
- Where does Sada work? *BoxStore*
- Where did Sada see the movie? *Movie Mania Cinema*
- As what race or ethnic origin, if any, does Sada identify? *Discuss the social media company's conclusion that Sada is "Black"*
- As what gender, if any, does Sada identify? *Discuss the social media company's conclusion that Sada is "female"*

Now ask how the social media company knew different things about Sada:

- Where Sada was in the bus photo *GPS (Global Positioning System) information from her phone.*
- Where Sada was in the photo with the broom *Which WiFi network Sada was using. Point out that since the bus Sada was getting onto in the bus photo had WiFi, the social media company would also know where Sada went on her trip.*
- How it was Dylan in the picture at the art gallery? *A facial recognition algorithm, which automatically tagged Dylan in the photo.*

Point out that all these bits of data can also add up to more information about Sada: for example, the social media company concluded that Sada worked at BoxStore because of how often Sada was there, and things like Sada's age and where Sada goes can be used to identify what Sada is interested in. Also discuss the chance that the way the social media company assesses the bits of data to draw conclusions about Sada may not be the same conclusions Sada would draw (e.g. Sada might identify as gender queer, not female, and as Latino rather than Black).

The Invisible Machine

Now distribute the article *The Invisible Machine: Big Data and You* and have students read it and answer the questions. (This can be done for homework if you wish.) Take up the questions with the class by drawing a mind map with the word "algorithms" at the centre and the following branches:

- Who uses algorithms?

- What are algorithms used for?

- Where do they get data?

- How can algorithms be unfair?

- How can we limit data collection?

If your students are not familiar with mind maps, explain that they are a way of thinking about a topic by making "branches" out of different topics and further exploring each one. If you like you can use one of the min mapping tools listed below. You may also choose to simply project or distribute the handout *Describing the Invisible Machine* and walk students through it.

Assessment/Evaluation Task: Mapping Your Data Profile

Tell students that they're going to try to look at their own online presence the way they looked at Sada's. Distribute the assignment sheet *Mapping Your Data Profile* and have students create a mind map with one of their social network accounts in the middle. (If they don't have a social network, they can use a search engine like Google or a video site such as YouTube or Netflix, all of which track and use data in the way described in the article) and, based on what they've learned from the article and Sada's profile, add at least five "branches" representing different kinds of data the platform knows about them. Next, have them draw a third layer that represents broader conclusions about them (ethnicity, interests, etc.) connected to one or more of the kinds of data.

- If you feel your students need more guidance in creating their mind maps, you can direct them to resources such as Mind Mapping for Kids (http://www.mindmapsforkids.com/mindmappingresources.html) or Concept Mapping in the Classroom (http://www.schrockguide.net/concept-mapping.html)

- You may want students to use online mind-mapping tools such as MindMup (https://www.mindmup.com), Coggle (https://coggle.it/) or MindMeister (https://www.mindmeister.com/) instead of doing their mind maps on paper. If so, make sure they save or print a copy of the final product for you to evaluate.

Once students have completed their mind maps, have them partner with another student and try to identify what the platform might know about them based on linking their two profiles together.

Finally, have students go back to the list of steps you can take to limit online data collection and pick **three of them.** Tell students to match each of them two one of the "branches" in the middle layer of their mind maps or to the outer layer. On a separate page, have students explain why they matched each tool to that particular branch or layer and how it helps to limit that kind of data collection.

Sada's Profile (Public View)

Sada's Profile (Public View)

Sada_Gem17

36 likes

Sada_Gem17 I definitely spend most of my day commuting #bored #glam
Charlie I totally just saw you on the subway!
Sada_Gem17 ahhh no way, @Charlie can't believe I missed you
Frankie @Sada_Gem17 do you have class today, let's grab coffee!
Jordan I need to buy a car so badly someone behind me on the bus won't stop singing #awks
Sada_Gem17 just gotta sing right back to them! Lol

Sada's Profile (Public View)

Sada_Gem17

77 likes

Jordan you working this weekend?
Sada_Gem17 yup @Jordan I live here now lol
Robin LOL same
Avery let's just retire now

Sada's Profile (Public View)

Sada_Gem17

♡ ○ ▷   ⬆ 🔖

29 likes

Sada_Gem17 If I'm being honest I only come here for the food #allthesingleladies #forgetthe heartbreak

Riley I'm still mad we didn't get the nachos #ragepost

Sada_Gem17 popcorn over everything

Jordan seeing this movie tomorrow!

Robin SAME @Jordan we should meet up!

Pat Hey @Sada_Gem17, I need to talk to you about things

Sada_Gem17 @Pat Sorry, I really don't want to talk right now

Pat @Sada_Gem17, Come on, I just want to see you

Sada_Gem17 @Pat Please, not tonight I need some space for a while

Pat @Sada_Gem17, C'mon – I just want to talk. Where are you – the Movie Mania?

Sada's Profile (Public View)

Sada_Gem17

♡ ◯ ◁   ⬆ ⊓

21 likes

Sada_Gem17 I am so adventurous. Photo taken by my bestie **@Robinblue**
Dylan I love your hair
Sada_Gem17 I love you **@Fin**
Robin don't you mean we **@Spencer** #rude
Spencer unfollowing
Robin eaving your trouble behind, hope those views are awesome! #bye #seeyouintwoweeks #worthit
Morgan such a beautiful picture!
Robin thank you!
Spencer we are so adventurous
Robin hahahaha
Sada_Gem17 you two are the worst lol
Quinn wow I haven't been out on a trip in forever, looks like fun!
Spencer you should come next time @Quinn

Sada's Profile (Public View)

Sada's Profile (Data View)

Sada_Gem17

♡ ◯ ◁   ↥ 🔖

41 likes

Sada_Gem17 Enjoying the afternoon downtown **@KraveCoffee** -need some caffeine to stay fueled #caffeine4life #esppressoornothing
01010011 01110000 01111001 @Sada_Gem17 profile created as IP3243.5465.433 Real Name: Reilly Smith … <calculating> … geo-location in process … geo-location completed … 24 2nd Steet W., Laighton-on-the-Lake, Canada … searching nearby businesses … socio-economic calculation running … <calculating> … category determined: Urban Asian
Alex That place is the best! #thatmacchiatolife
Jesse I love that place! I'm there every morning on my way to work! #wakeupforwork #needajolt
01010011 01110000 01111001 @Alex profile created … <calculating> … linking to **@Sada_Gem17, @Jesse** … updating location data … location added
01010011 01110000 01111001 @Jesse profile updating … <calculating> … linking to **@Sada_Gem17, @Alex** … updating location data … location added … <calculating> … updating frequency … frequency updated

Sada's Profile (Data View)

Sada_Gem17

♡ ○ ◁   ⬆ ⊓

36 likes

Sada_Gem17 I definitely spend most of my day commuting #bored #glam
11110101 10111101 01010000 Acquiring data … <calculating> … accessing speedometer … OUTPUT 45 km/h … <calculating> GPS data acquired … <calculating> … geo-location in process … geo-location completed … 115 17th Street, Laighton-on-the-Lake, Canada … <calculating> … frequent location identified … storing location data … <calculating> … searching nearby businesses … socio-economic calculation running … <calculating> … category determined: Urban Liberal
Charlie I totally just saw you on the subway!
Sada_Gem17 ahhh no way, @Charlie can't believe I missed you
Frankie @Sada_Gem17 do you have class today, let's grab coffee!
Jordan I need to buy a car so badly someone behind me on the bus won't stop singing #awks
Sada_Gem17 just gotta sing right back to them! Lol
11110101 10111101 01010000: @Charlie profile created … <calculating> … linking **@Charlie** to **@Sada_Gem17**
11110101 10111101 01010000 @Frankie profile created … <calculating> … linking **@Frankie** to **@Sada_Gem17**
11110101 10111101 01010000 @Jordan profile created … <calculating> … linking **@Jordan to @Sada_Gem17**
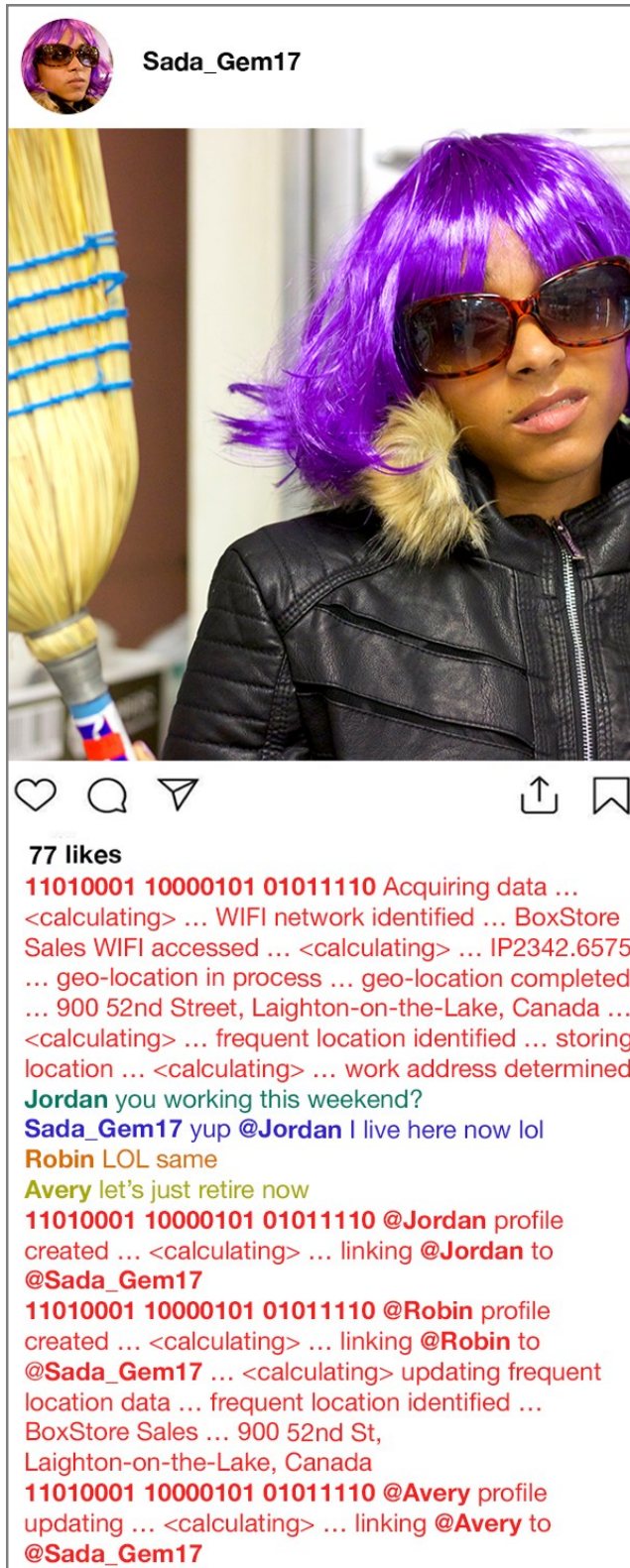
Sada's Profile (Data View)

Sada_Gem17

♡ ○ ◁   ⬆ 🔖

77 likes

11010001 10000101 01011110 Acquiring data … <calculating> … WIFI network identified … BoxStore Sales WIFI accessed … <calculating> … IP2342.6575 … geo-location in process … geo-location completed … 900 52nd Street, Laighton-on-the-Lake, Canada … <calculating> … frequent location identified … storing location … <calculating> … work address determined
Jordan you working this weekend?
Sada_Gem17 yup **@Jordan** I live here now lol
Robin LOL same
Avery let's just retire now
11010001 10000101 01011110 @Jordan profile created … <calculating> … linking **@Jordan** to **@Sada_Gem17**
11010001 10000101 01011110 @Robin profile created … <calculating> … linking **@Robin** to **@Sada_Gem17** … <calculating> updating frequent location data … frequent location identified … BoxStore Sales … 900 52nd St, Laighton-on-the-Lake, Canada
11010001 10000101 01011110 @Avery profile updating … <calculating> … linking **@Avery** to **@Sada_Gem17**

Sada's Profile (Data View)

Sada_Gem17

♡ ○ ◁   ⬆ 🔖

29 likes

Sada_Gem17 If I'm being honest I only come here for the food #allthesingleladies #forgetthe heartbreak
01010111 00110100 01011011 Acquiring data … <calculating> … GPS data acquired … location determined … Movie Mania Cinema, Canada … <calculating> … WIFI network identified … WIFI accessed … <calculating> … geo-location in process … geo-location completed … 45 West Avenue, Laighton-on-the-Lake, Canada <calculating> … linking to **@Sada_Gem17**, **@Riley**…OUTPUT image posted
Riley I'm still mad we didn't get the nachos #ragepost
Sada_Gem17 popcorn over everything
Jordan seeing this movie tomorrow!
Robin SAME @Jordan we should meet up!
01010111 00110100 01011011 @Riley profile created … <calculating> … linking **@Riley** to **@Sada_Gem17** …
01010111 00110100 01011011 @Jordan profile updating … <calculating> … linking **@Jordan** to **@Sada_Gem17, @Riley**
01010111 00110100 01011011 @Robin profile created … <calculating> … linking **@Robin** to **@Sada_Gem17, @Riley**
Pat Hey **@Sada_Gem17**, I need to talk to you about things
Sada_Gem17 @Pat Sorry, I really don't want to talk right now
Pat @Sada_Gem17, Come on, I just want to see you
Sada_Gem17 @Pat Please, not tonight I need some space for a while
Pat @Sada_Gem17, C'mon – I just want to talk. Where are you – the Movie Mania?

Sada's Profile (Data View)

Sada_Gem17

21 likes

Sada_Gem17 I am so adventurous. Photo taken by my bestie **@Robinblue**
11010001 10010101 01010000: Acquiring data … <calculating> … geo-location in process … geo-location completed … 25 River Road, Laighton-on-the-Lake, Canada … linking to @Robin … linking to **@Spencer** … searching nearby businesses … socio-economic calculation running … <calculating> … category determined: Urban Asian … <calculating> … accessing pedometer … OUTPUT 7303 steps taken … 5 kilometers travelled
Dylan I love your hair
Sada_Gem17 I love you **@Fin**
Robin don't you mean we **@Spencer** #rude
Spencer unfollowing
11010001 10010101 01010000 @Dylan profile created … <calculating> … linking @Dylan to @Sada_Gem17
11010001 10010101 01010000 @Robin profile created … <calculating> … linking to **@Sada_Gem17** … updating location data … location added
11010001 10010101 01010000 @Spencer profile created … <calculating> … linking to **@Sada_Gem17** … updating location data … location added

Robin eaving your trouble behind, hope those views are awesome! #bye #seeyouintwoweeks #worthit
01011001 11110101 01110001 Acquiring data … <calculating> … geo-location in process … geo-location completed … 25 River Road, Laighton-on-the-Lake, Canada … linking to **@Sada_Gem17** … linking to **@Spencer**
Morgan such a beautiful picture!
Robin thank you!
01011001 11110101 01110001 @Morgan profile created … <calculating> … linking **@Morgan** to **@Robin**

Spencer we are so adventurous
01011001 11110101 01110001 Acquiring data … <calculating> … geo-location in process … geo-location completed … 25 River Road, Laighton-on-the-Lake, Canada … linking to **@Sada_Gem17** … linking to **@Robin**
Robin hahahaha
Sada_Gem17 you two are the worst lol
Quinn wow I haven't been out on a trip in forever, looks like fun!
Spencer you should come next time @Quinn
@01011001 11110101 01110001: @Robin profile created … <calculating> … linking to @Spencer … updating location data … location added
@01011001 11110101 01110001: @Sada@Sada_Gem17 profile created … <calculating> … linking to @Spencer … updating location data … location added
@01011001 11110101 01110001: @Quinn profile created … <calculating> … linking to @Spencer

Sada's Profile (Data View)

Sada_Gem17

♡ ○ ◁   ⬆ ⬚

19 likes

Sada_Gem17 Loving this art exhibition #retro #black&whitephotography #filmisnotdeadstudy date with a bunch of my favourites @MetroMuseum
00010001 10011101 01010011 Acquiring data … <calculating> … WIFI network identified … Metro Museum WIFI accessed … <calculating> … accessed … IP2342.6575 … geo-location in process … geo-location completed … 222 Arts Boulevard59 3rd Stree E., Laighton-on-the-Lake, Canada … <calculating> … frequent location identified … storing location … <calculating> … socio-economic calculation running … <calculating> … category determined: Upper Middle Class … education level calculation running … <calculating> … category determined: Undergraduate
Frankie how long will you be there for? I am on my way now #foreverlate
Sada_Gem17 until 4 see you soon!
Avery can you even say you went to the metro if you didn't take a photo lol
Bailey I didn't know you could study there, so cool!
Sada_Gem17 I know right **@Bailey** plus its free!
00010001 10011101 01010011 @Frankie profile created … <calculating> … linking **@Frankie** to **@Sada_Gem17**
00010001 10011101 01010011 @Avery profile created … <calculating> … linking **@Avery** to **@Sada_Gem17**
00010001 10011101 01010011 @Bailey profile created … <calculating> … linking **@Bailey** to **@Sada_Gem17**

01010011 01110000 01111001 Faces identified … running facial recognition … <calculating> … OUTPUT **@Sada, @Dylan** … ethnicity detection … <calculating> … category determined … OUTPUT Black, White … running gender classification … <calculating>…OUTPUT **@Sada_Gem17,** Female … **@Dylan**, Male … searching mutual contacts … mutual contacts determined … linking **@Sada, @Dylan** to **@Robin**
Robin WTH??!! Are you moving in on my ex-bae?
Sada What?! Of course not!
Robin That's certainly what it looks like!
Sada That is not what's happening!
01010011 01110000 01111001 Acquiring data … <calculating> … linking **@Robin** to **@Sada, @Dylan**

The Invisible Machine: Big Data and You

There is a machine that affects your whole life, but which you probably don't even know exists. This machine is called an *algorithm*, and it affects everything you do online – from what videos you see, to what search results you get, to what you see on your social networks. More and more, algorithms have an impact on our offline lives, too.

An algorithm is basically a series of steps or instructions for doing something. When we talk about online algorithms we mostly mean one specific type, *sorting algorithms*, which sort things (or people) into different categories. If you have a social network account, for example, the company has *sorted* you based on what they know, or think they know, about your age, your ethnic background, your gender, where you live, your interests, and dozens or even hundreds of other bits of data.

Where does this data come from? Apps or websites constantly collect data on what you do there: what posts you like, what you buy on a shopping site, what you search for, what videos you watch and for how long. They also get information on things like where you live (which might come from GPS, WiFi, your Internet Protocol address or what cell phone towers you're nearest to) or what browsers and devices you use to go online. Your friends are another source: their data can be used to double-check what the algorithm thinks it knows about you. All of this information gets used to create a *profile* of you that will be used for matching. It also often gets sold to *data brokers*, who collect data from lots of different sources to build a more complete profile of you.

What is it all for? Mostly, to give you stuff you want to see online. Your social networks use algorithms and your data to decide which of your friends' posts to show you, video sites use them to decide what videos to suggest you watch next, search engines use them to decide what search results are most relevant to you. More importantly, from their point of view, they almost all use algorithms to show you ads that you're more likely to respond to. That's why these platforms are mostly free: they make money by showing you ads that have been matched with your profile.

That may not seem so bad. After all, if you have to see ads, it's better to see ones for things you're actually interested in. But because algorithms show us what they think we want to see, they can keep us from seeing the whole picture. You may miss an important post from one of your friends because the algorithm doesn't think you'll like it. You may not get the best or most reliable results from a search engine because its algorithm thinks you'll be more interested in different sources. Algorithms on social networks and video sites also usually prefer whatever people have interacted with the most, which means that hoaxes, conspiracy theories and misinformation often spread more easily than reliable information and the loudest voices can seem like the majority.

Algorithms also use your data in other ways that might affect you without you even knowing it. Shopping and travel sites often offer different prices depending on things like what device or browser you're using, and sites that show job ads use your profile to decide which ads to show you. Your online data can affect things offline, too. A lot of employers use it to sort job applications, so even if you apply for a job your online profile might keep you from getting an interview. Banks and credit card companies use it to decide whether to give you a loan, and how much interest to charge you. More and more, algorithms are even being used to decide how long someone convicted of a crime should be sentenced for, or whether they should get parole, based on whether they've been sorted into a category of people who are more likely to re-offend.

That's obviously unfair if the algorithm didn't sort you correctly, and that happens pretty often: a study found that a quarter of people felt that Facebook, which has access to more data than just about anyone in the world, had classified

them in a way that wasn't accurate. But even if you are sorted correctly, the results are often unfair. That's because most algorithms aren't programmed but *trained*: they're given data that already exists and use how that data was sorted to decide how they should sort the data that they're given. Because of that, algorithms can copy *unfair* patterns caused by racism, sexism and other kinds of discrimination. For example, an algorithm that was trained on a thousand resumes, some of which led to applicants getting job interviews and some of which didn't, might end up ranking men's resumes higher than women's, or ranking names that sound Black lower than those that sound White, because that's what the humans who had sorted those original resumes had done. (Both of those examples actually happened.) That also means that a lot of the time even the people who created the algorithm don't know exactly how it makes decisions.

So how can we take charge of the invisible machine? You've already taken the first step: you know that your search engine results, social network posts and video recommendations have all passed through it on the way to you.

You can also take steps to limit how much data you give out to platforms and data brokers:

- Use non-tracking search engines like Startpage or DuckDuckGo
- Use browsers like Firefox (Firefox Focus on mobile) with strong privacy protection, and select Do Not Track in your settings
- Use browser plugins like Privacy Badger or Ghostery to limit how much websites can track about you
- Turn off personalization in your browser and social networks
- Turn off autoplay and search for what you want instead of letting apps recommend things
- Turning off geolocation on mobile devices
- Only give apps permission to access things on your device that they actually need

There's one more thing you can do that might have an even bigger impact, though not right away: you can tell companies (and governments) that you want them to use algorithms fairly, and tell them if there's a particular thing you think is unfair. (You can do the same thing with big companies that advertise on places like social networks or video sites, just make sure you tell them which platform you're talking about.) It may seem like you don't have any power compared to them, but companies as big as Facebook and Google *have* changed how they do business because of complaints from users and advertisers.

Questions

List three kinds of data that online platforms collect.

1. _____

2. _____

3. _____

List three different ways that algorithms can affect your life, online or offline.

1. _____

2. _____

3. _____

List two possible ways that algorithms might lead to inaccurate or unfair results.

1. _____

   _____

2. _____

   _____

Look at the list of things you can do to limit online data collection. List the two that you think would be most important for you to do and briefly explain why.

1. _____

   _____

2. _____

   _____

Mapping Your Data Profile

For this assignment, you will create a mind map with one of your social network accounts in the middle. (If you don't have a social network account, you can use a search engine like Google or a video site such as YouTube or Netflix, all of which track and use your data.)

Based on what you've learned from the article and Sada's profile, add at least five "branches" representing different kinds of data the platform knows (or thinks it knows) about you. For example, one branch might be the things you intentionally post on the platform, like photos or videos. Next, draw a third layer that represents broader conclusions about you (ethnicity, interests, etc.) connected to one or more of the kinds of data.

Your final mind map will have three levels or layers:

Inner: Your account

Middle: What that company knows (or thinks it knows) about you

Outer: Guesses or connections the company could make about you based on the other two layers

Once you have completed your mind map, go back to the list of steps you can take to limit online data collection. Pick **three** and match each of them two one of the "branches" in the middle layer of your mind map or to the outer layer.

On a separate page, explain why you matched each tool to that particular branch or layer and how it helps to limit that kind of data collection.

ALGORITHMS

What are algorithms used for?
- Decide what videos to recommend or show you
- Decide what ads to show you
- Search engine results
- Decide how much to charge you
- Sorting job applications
- Loan applications and interest rates

How can we limit data collection?
- Use non-tracking search engines
- Use browsers with privacy protection
- Using tracking-blocker plugins
- Turn off personalization
- Turn off autoplay
- Turn off geolocation
- Only give apps absolutely necessary permissions
- Tell companies you want fair algorithms

Who uses algorithms?
- Social networks
- Video sites
- E-commerce sites
- Social networks
- Banks and credit card companies

How can algorithms be unfair?
- Trained on unfair data
- Reproduce discrimination
- Nobody knows how a decision was made

Where do they get data?
- What you post on apps and websites
- What you do on apps and websites
- GPS, WiFi, IP address
- What your friends do
- Data brokers