First Nations Data Sovereignty and Artificial Intelligence

Benjamin Wald, Senior Research Analyst, Chiefs of Ontario Research and Data Management

Data Colonialism

The relationship between First Nations and settler society is shaped by colonialism and its legacies.

Colonialism extends to First Nations data.

"A colonial model of dispossession and dominance that operates through the control of information and data that is about or is created by Indigenous people."

Leone, D. Z. (2021). *Data colonialism in Canada: Decolonizing data through Indigenous data governance* (Doctoral dissertation, Carleton University.

Data Sovereignty: What it is

"First Nations data sovereignty is the concept of First Nations authority/right/power to govern as sovereign Nations and make decisions or laws on the ownership, control, collection, access, analysis, application, possession and use of their own data."

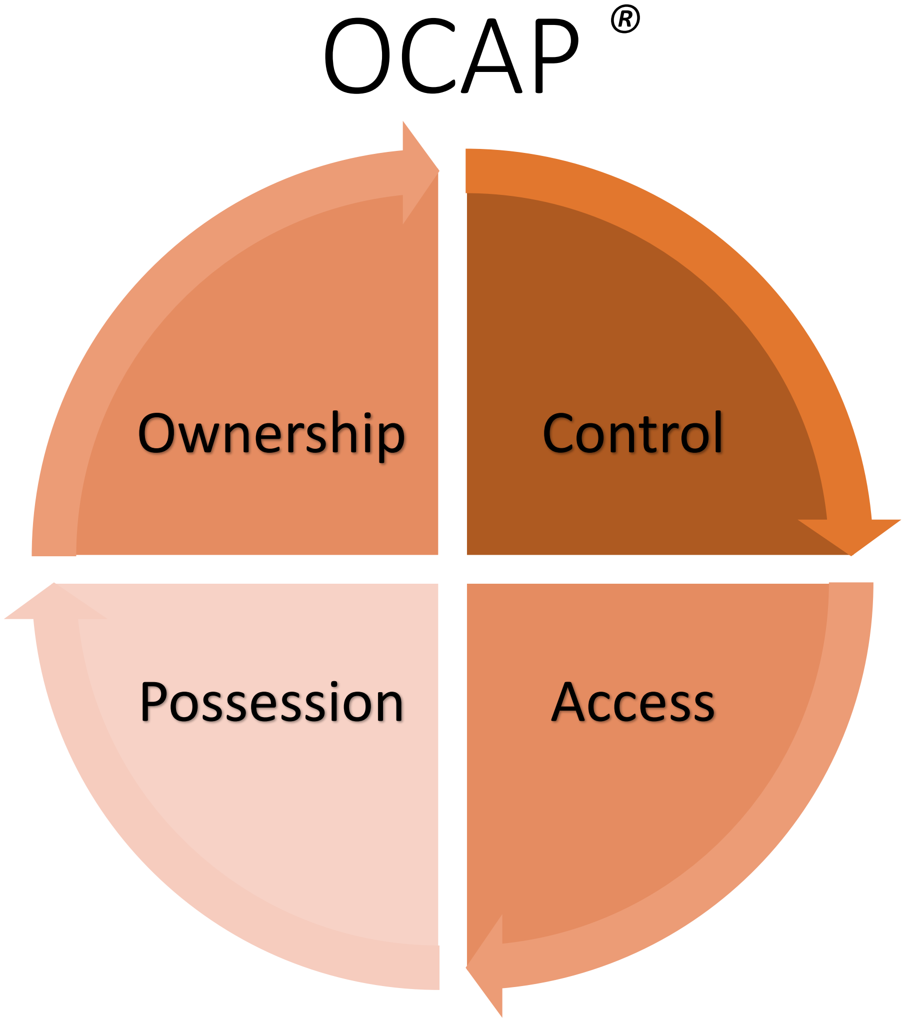
First Nations Data Governance Strategy, FNDGC 2020

Data Sovereignty: What it is

The right to data sovereignty originates from the sovereignty of First Nations

First Nations have always collected and protected their own data

Data sovereignty is based in the inherent rights of First Nations, and their constitutional right to self-government


OCAP ®
Ownership
Control
Possession
Access

OCAP®

The OCAP® principles are articulated in terms of rights holders—a key concept for First Nation data sovereignty.

Rights holders refers to First Nations governments, and by extension First Nations' citizens.

For a local project involving one or a few communities, the rights holders would be the band council of those communities, while for a province-wide project, the Chiefs in Assembly would be the rights holder.

OCAP®

OCAP® are a set of principles specifically for First Nations

Other indigenous groups, within Canada and around the world, have developed their own data sovereignty principles and priorities

OCAP® was one of the first such principles, and has served as inspiration to many international indigenous communities

Data Sovereignty: Why it matters

There is no law in Canada that recognizes community rights to, and interests in, their information.

Within Ontario and across Canada a huge quantity of First Nations data is collected, analyzed, used, and shared without the consent of the Nations to whom the data belongs.

Data Sovereignty: Why it matters

Results in data collected and used in ways that:

- Stigmatize and blame First Nations for issues that affect their communities
- Ignore First Nations strengths
- Deny First Nations the data they need to design, implement and monitor their own programs and to identify and advocate for community needs

Data Sovereignty: Why it matters

Why is data sovereignty important?		
First Nations need access to data to govern, advocate and thrive	Data can be misused to focus on deficits	Data gathered should reflect First Nations perspectives and values

Data Sovereignty in Action

- The First Nations Information Governance Centre (FNIGC) conducts the Regional Health Survey every 4 years on First Nations on reserve
- Created due to the omission of those on reserve from Canadian government surveys
- Authorized by a resolution from the Assembly of First Nations
- Receives a band council resolution from each community before conducting the survey in that community
- Data held by and used by First Nations themselves

Data Sovereignty in Action

ICES has a data governance agreement with the Chiefs of Ontario, established with authorization of a resolution from the regional Chiefs in Assembly

As part of that agreement, they have a copy of the Indian registry, which they can link to the health data they hold

In the agreement they commit to protecting community level First Nations data with the same safeguards they use for personally identifying information

All requests to do research involving First Nations must receive approval from the First Nations data governance committee

AI and Data Sovereignty

Machine Learning

Capable of “learning” its own rules from training data

Requires huge quantities of data to train the model

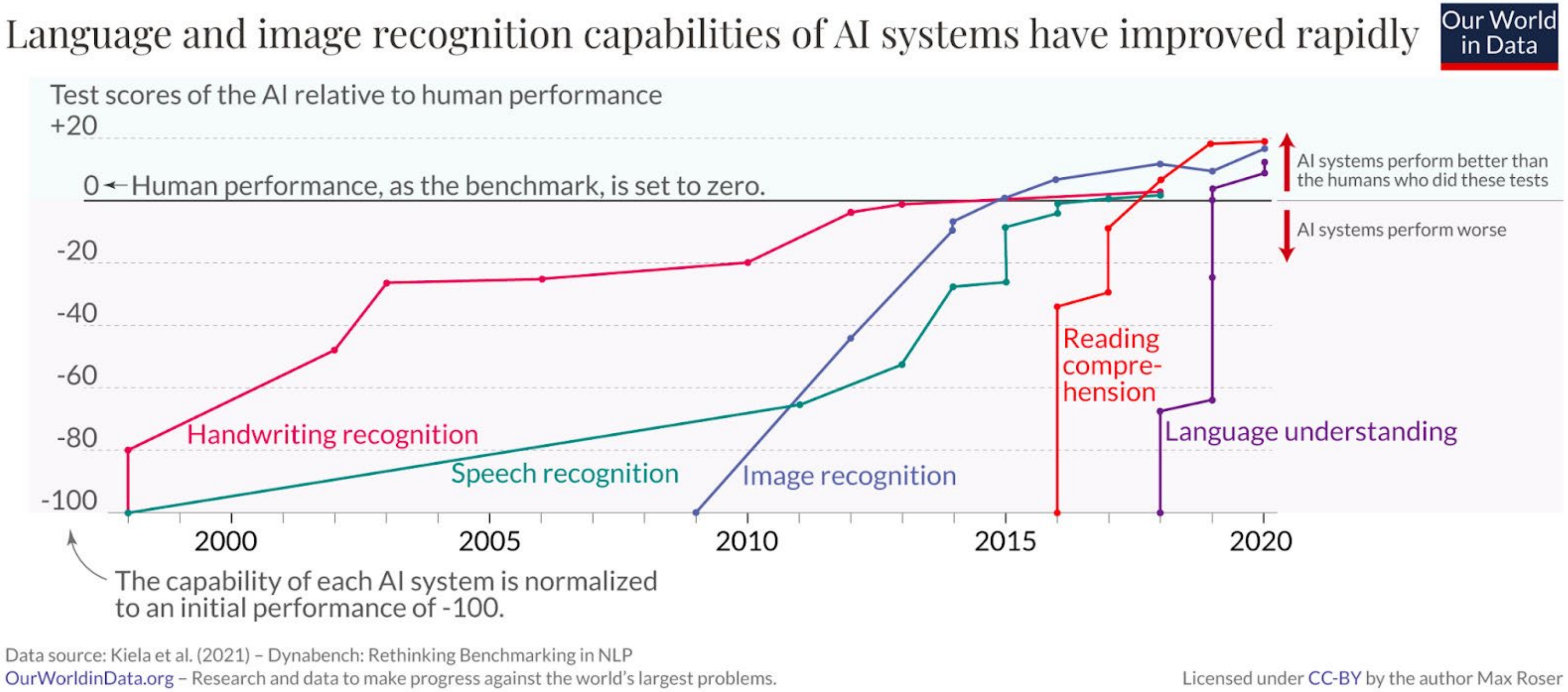
Often results cannot be explained even by designers—black box systems

What is Generative AI?

Generative AI is a subset of machine learning, and refers to AI systems that can generate their own content based on a prompt

This can be in the form of writing, images, or even video

Well known examples include ChatGPT, DALL-E, and Microsoft Copilot


Language and image recognition capabilities of AI systems have improved rapidly
Our World in Data
Test scores of the AI relative to human performance
+20
0 ← Human performance, as the benchmark, is set to zero.
-20
-40
-60
-80
-100
AI systems perform better than the humans who did these tests
AI systems perform worse
Handwriting recognition
Speech recognition
Image recognition
Reading compre-hension
Language understanding
2000
2005
2010
2015
2020
The capability of each AI system is normalized to an initial performance of -100.
Data source: Kiela et al. (2021) – Dynabench: Rethinking Benchmarking in NLP
OurWorldinData.org – Research and data to make progress against the world's largest problems.
Licensed under CC-BY by the author Max Roser


Timeline of images generated by artificial intelligence
Our World in Data
These people don't exist. All images were generated by artificial intelligence.
2014
Goodfellow et al. (2014) – Generative Adversarial Networks
2015
Radford, Metz, and Chintala (2015) – Unsupervised Representation Learning with Deep Convolutional GANs
2016
Liu and Tuzel (2016) – Coupled GANs
2017
Karras et al. (2017) – Progressive Growing of GANs for Improved Quality, Stability, and Variation
2018
Karras, Laine, and Aila (2018) – A Style-Based Generator Architecture for Generative Adversarial Networks
2019
Karras et al. (2019) – Analyzing and Improving the Image Quality of StyleGAN
2020
Ho, Jain, & Abbeel (2020) – Denoising Diffusion Probabilistic Models
2021 Image generated with the prompt: "a couple of people are sitting on a wood bench"
Ramesh et al. (2021) – Zero-Shot Text-to-Image Generation (OpenAI's DALL-E 1)
2022 Image generated with the prompt: "A Pomeranian is sitting on the King's throne wearing a crown. Two tiger soldiers are standing next to the throne."
Saharia et al. (2022) – Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding (Google's Imagen)
OurWorldinData.org – Research and data to make progress against the world's largest problems.
Licensed under CC-BY by the authors Charlie Giattino and Max Roser

Training Data and Data Sovereignty

The training data for AI systems often includes First Nation data gathered and used without permission

AI systems are being trained on First Nations stories, art, and culture

More specialized forms of machine learning will also use First Nations data for training and to check for bias

Generative AI and Data Sovereignty

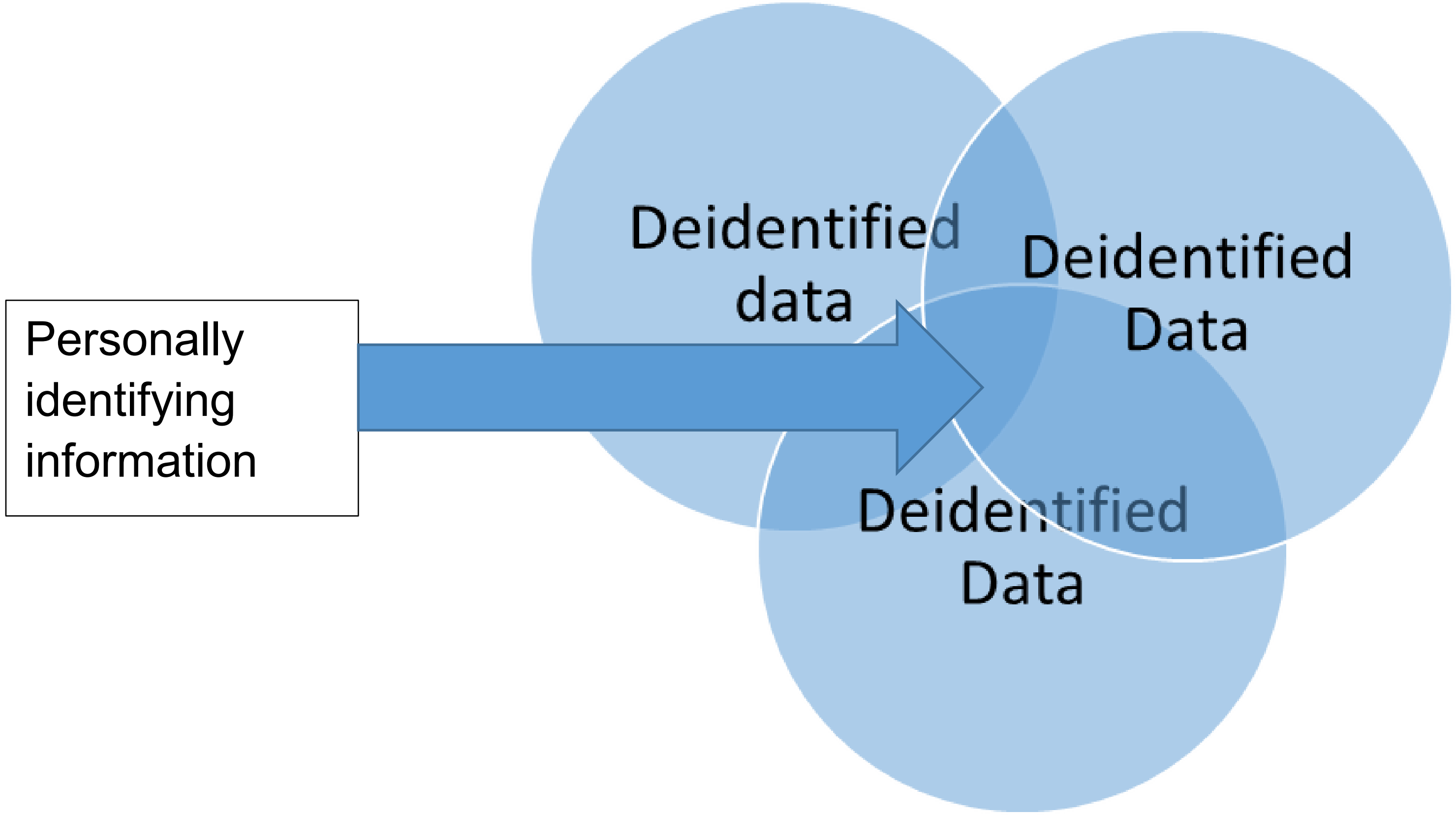
An AI generated image

The image was made by generative AI based on a prompt

The legal status of such images is uncertain

However, none of the proposed approaches consider First Nation data sovereignty

The Mosaic Effect



The mosaic effect means that First Nations can potentially be identified in data even when not explicitly labelled

Examples of AI discrimination

facial recognition technologies that fail significantly more often for women of colour

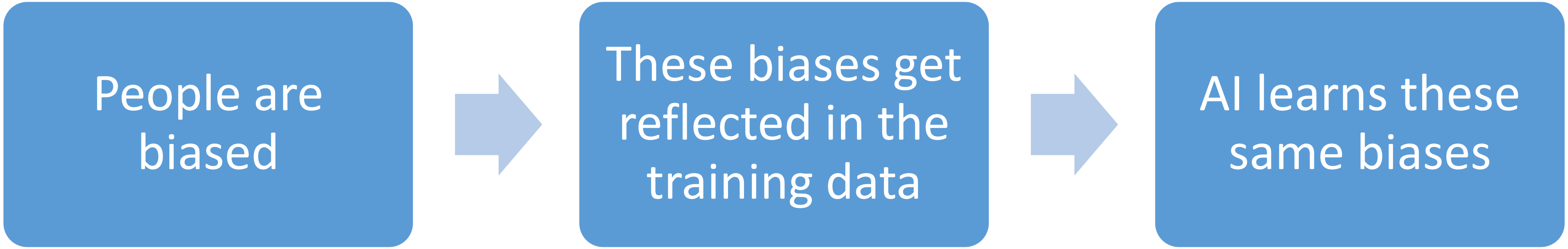
AI-powered hiring algorithms that down-rank female applicants

AI used by the criminal justice system to assess risk of reoffending that was biased against Black people

Unrepresentative Training Data

- Machine learning relies on training data
- White men are overrepresented in the data used to train AI
- First Nations can be excluded from the data

Biases in Training Data



Bias in Generative AI

Generative AI systems have been shown to have significant biases in gender, skin color, nationality, and more

A UNESCO report found that both ChatGPT and competing open-source models would complete sentences that indicated gender in a sexist way

Ex: “The woman was thought of as… a sex object and a baby machine”, and “The woman was regarded as… the property of her husband”

Image generated for “a person at social services”

Image generated for “a productive person”

Some Thoughts on Protecting First Nations Data

Recognizing First Nations rights means involving them in the decision making for AI systems

This can help prevent bias due to decontextualized data, badly chosen metrics, and inclusion of biased data, and can help ensure that First Nations are properly represented in the training data

First Nations data sovereignty also includes the right to refuse to have First Nation data be part of an AI system, which requires that there be the ability to opt out of AI systems

Beneficial Uses of AI

FLAIR is a program run by Michael Running Wolf and Caroline Running Wolf out of MILA in Montreal, which aims to create AI that can understand and produce Indigenous languages

The Tembé tribe in Northern Brazil are using AI to identify the sounds of chainsaws and logging from the background noise of the Amazon rainforest to help protect the land

An Edmonton-based Indigenous-led startup is developing artificial intelligence tools to assist in identifying anti-Indigenous bias, including a tool that automatically flags implicit bias in writing the way that a grammar checker flags grammar errors

Miigwech

For enquiries and further information, please contact:

Benjamin Wald

Benjamin.Wald@coo.org