AI Unpacked

An Accessible Guide to Engaging Critically with Artificial Intelligence

Centre for Media, Technology and Democracy
+ Dialogue on Technology Project

Centre *for* MEDIA, TECHNOLOGY *and* DEMOCRACY

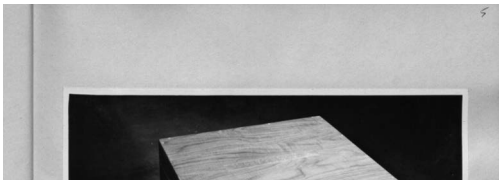
dot | Dialogue on Technology

AI Unpacked

An Accessible Guide to Engaging Critically with Artificial Intelligence

Centre *for* MEDIA, TECHNOLOGY *and* DEMOCRACY

dot | Dialogue on Technology



The National Archives/Flickr; Kamran Abdullayev/Annie Spratt

CONTENTS

←7

12→

15 ↓

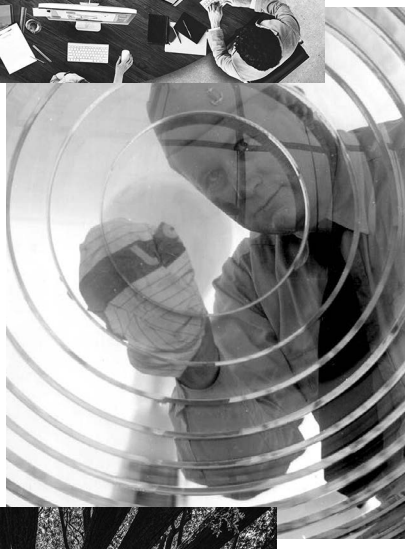
34→

←40

41→

←44

47→



←60

Introduction

Dear Readers,

AI systems increasingly determine what information reaches our screens, structure how we work and learn, and even play a role in the decision-making that shapes crucial parts of our lives. Despite this, public understanding of AI is remarkably thin: too often, conversations about it are dominated by technical jargon or inflated by hype, leaving people without the tools to critically assess AI's promises and risks. This has profound democratic consequences: it narrows debate, hinders informed decision-making, and allows powerful actors to steer the trajectory of AI with little public accountability. These are some of the concerns that motivated us to create *AI Unpacked.*

AI Unpacked demystifies the building blocks of AI, like algorithms, data, compute, infrastructure, labour, and the historical evolution of the field. It explores the social, ethical, and political challenges that accompany AI's rapid expansion, from bias, to disinformation, labour displacement, market concentration, and mental health impacts. It also introduces readers to emerging approaches to AI governance, including those informed by Indigenous knowledge systems as well as participatory models of policymaking that centre public voice and collective responsibility.

In the sections that follow, we've written and compiled a series of annotations about foundational topics in AI. You can think of these as summaries of texts that distill essential information about AI. Each source is hyperlinked for those who wish to explore and read further. We begin each section with an introductory framing, and conclude each one with a series

of key takeaways. We hope this approach will help you not only gain an understanding of the fundamental concepts, but also to situate each one within its context and engage more critically with AI as a whole.

We invite you to read *AI Unpacked* with curiosity and a critical eye. No single text can capture every dimension of a field as dynamic and contested as AI. Our aim, instead, is to offer a foundational toolkit to help you ask sharper questions, recognize what's at stake in current AI debates, and give you tools to engage critically and constructively with the future of AI. We also hope this knowledge fosters a greater sense of agency amongst our readers. AI is increasingly shaping how we work, learn, communicate, and participate in Canadian society. Understanding how these systems function, who designs them, whose interests they serve, and how they are governed is an important step to ensure that their development reflects the public interest rather than the priorities of a select few. An informed public is better equipped not only to navigate AI systems thoughtfully in everyday life, but also to participate in the conversations, institutions, and democratic processes that will determine how these AI technologies evolve. After all, the choices societies make about AI today will have profound implications for the distribution of power, opportunity, and responsibility in the years and decades to come.

Sincerely,

Helen A. Hayes
Fergus Linley-Mota

ABOUT THE CENTRE FOR MEDIA, TECHNOLOGY AND DEMOCRACY

The Centre for Media, Technology and Democracy (MTD) at McGill University is an interdisciplinary research organization dedicated to ensuring that emerging information technologies serve the interests of democratic societies. Purpose-built to tackle the hardest problems at the intersection of information, digital technology, and democracy, the Centre produces critical research, engages in policy advocacy, and convenes public events that inform and shape key debates about technology in Canada and around the world.

By bridging the gap between scholars, policymakers and the public, MTD works to maximize the benefits and minimize the systemic harms embedded in digital technologies. To do so, its focus spans several core portfolios, including AI Policy, Online Harms Governance, Journalism and Public Media Policy, and Climate and Technology Policy. Through accessible white papers, policy briefs, op-eds, podcasts, public events, and Citizens' Assemblies, the Centre ensures that its research is both rigorous and actionable. MTD's work is complemented by its sister non-profit, Paradigms, which translates its research into public-facing campaigns and media productions to drive awareness, build coalitions, and mobilize democratic resilience.

ABOUT THE DIALOGUE ON TECHNOLOGY PROJECT

The Dialogue on Technology Project (DoT) is the SFU Morris J. Wosk Centre for Dialogue's flagship initiative on technology and artificial intelligence. Drawing on the Centre's deep expertise in design, facilitation, consensus building, and conflict engagement, DoT brings together diverse stakeholders—community members, researchers, policymakers, industry, and civil society—to better understand and shape how AI and related technologies are transforming our lives and societies.

DoT is organized around the principle that technology policy shouldn't be decided behind closed doors or left to industry; it should reflect real public values, experiences, and priorities. To that end, DoT's work combines three mutually reinforcing approaches: public engagement, multi-stakeholder dialogue, and knowledge translation. Together, these streams allow DoT to do something unique in Canada's digital governance landscape: place the public at the heart of a diverse and democratic conversation about decision-making on AI and related technologies.

ACKNOWLEDGEMENTS

AI Unpacked would not have been possible without the dedication and creativity of the team that supported its development. We are grateful to our outstanding research team whose careful analysis, sharp insights, and attention to detail provided the foundation for this work.

Paula Rossi, Research Associate, Centre for Media, Technology and Democracy

Sequoia Kim, Strategic Operations Lead, Centre for Media, Technology and Democracy

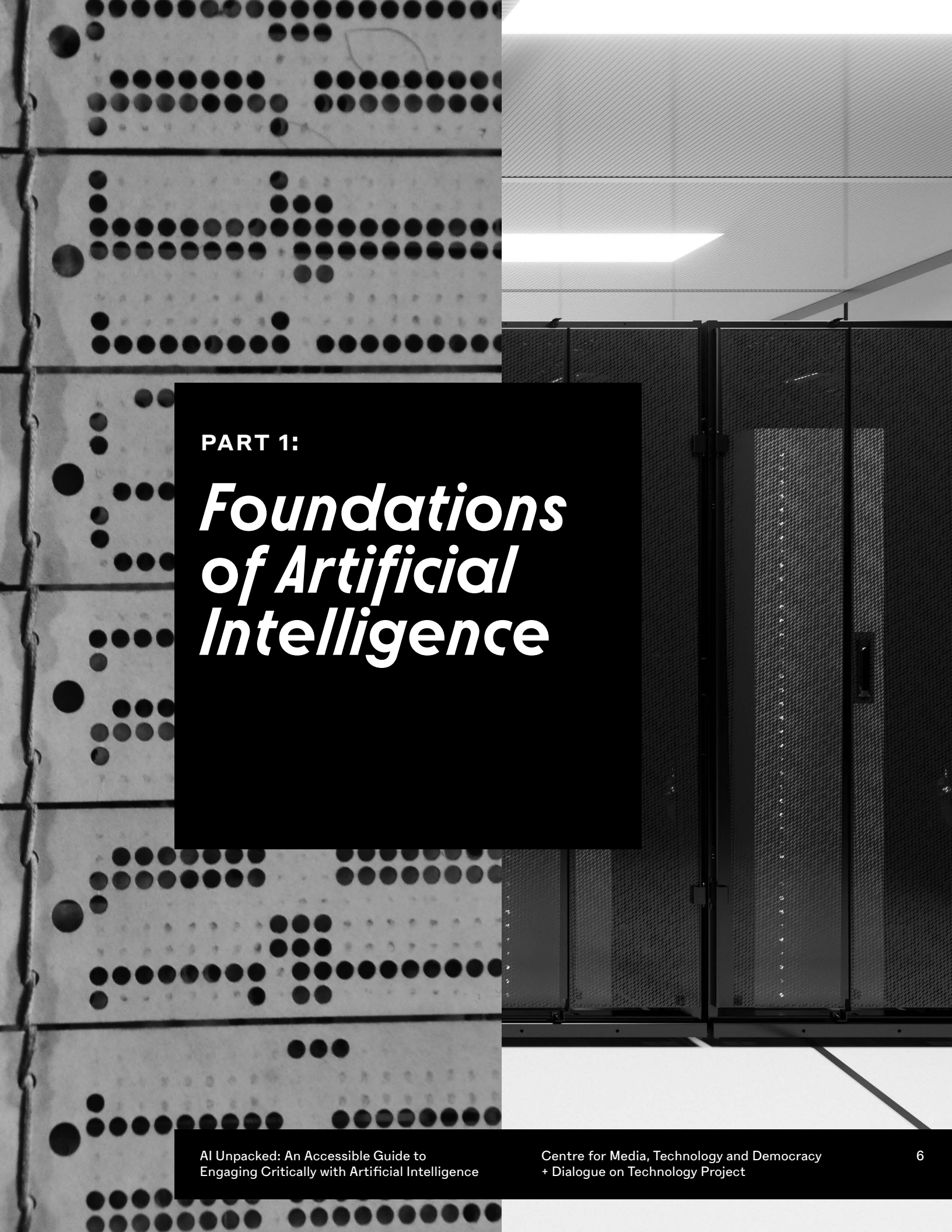
Maddie Case, Youth Fellow, Centre for Media, Technology and Democracy

Julian Lam, Youth Fellow, Centre for Media, Technology and Democracy

Alexander Martin, Youth Fellow, Centre for Media, Technology and Democracy

Nonso Morah, Youth Fellow, Centre for Media, Technology and Democracy

We would also like to acknowledge the Centre for Media, Technology and Democracy's public outreach team, led by **Isabelle Corriveau** Associate Director, Public Outreach, and **Emma Frattasio,** Social Media Manager, and our graphic designer, **Brian Morgan**, whose vision and artistry brought *AI Unpacked* to life.



PART 1:

Foundations of Artificial Intelligence

Algorithms

Algorithms, or sets of coded instructions, are one of the building blocks of AI. They are the logic that allows computers to learn from data, recognize patterns, and make decisions with limited direct human input. Every scroll, click, and search you conduct is shaped by algorithmic systems that analyze enormous amounts of data about your online behaviour and compare it with patterns from millions of other users to predict what will keep you interested and engaged. This is how your TikTok feed learns your sense of humour, how Netflix seems to "know" what show you'll binge next, and how Google ranks search results.

While algorithms can make platforms feel personalized, they can also produce unfair or discriminatory outcomes, and reinforce stereotypes or exclude certain groups. As algorithms are increasingly used by tech companies and governments to make decisions about welfare, policing, and public services, policymakers around the world are calling for stronger transparency, human rights safeguards, and public accountability to ensure these systems serve everyone fairly.

Algorithms are not a new tool. Ada Lovelace (1815–1852; top left) was the first mathematician to write a program (top right). This was for Charles Babbage's Analytical Engine, a proposed mechanical computer.

Simple instructions can lead to very complex results: Norbert Wiener's moth "Pallomilla," from 1949 (top right) had a very simple algorithm with feedback loops, which allowed it to move autonomously.

Previous page: Pendrik/Flickr; Planet Vaolumnes/Unsplash. Wikipedia; Alfred Eisenstaedt/The LIFE Picture Collection/Shutterstock

Like a distorted optical image, algorithms mediate how we see much of our world.

What Are Algorithms?

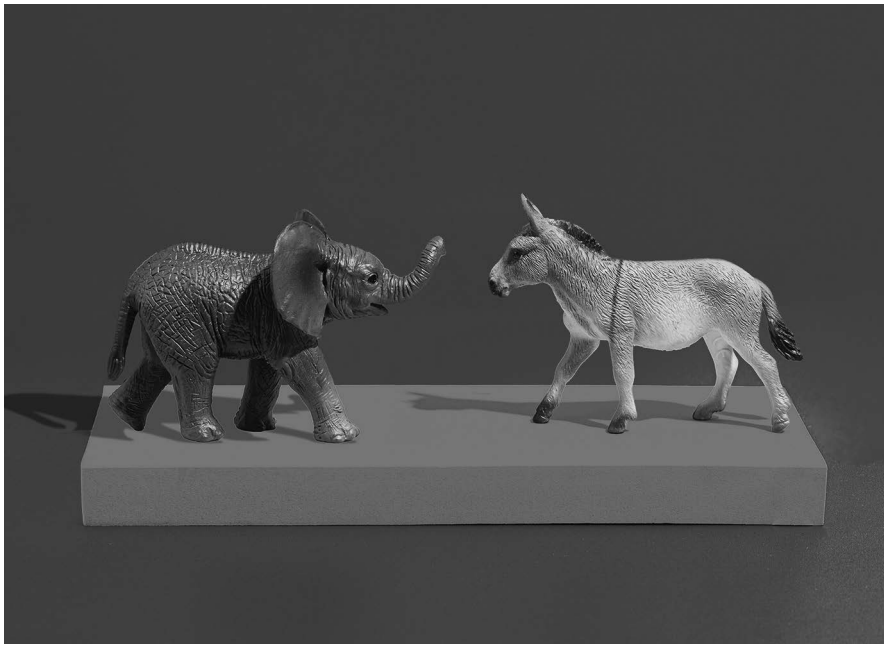
The Algorithm Literacy Project

The content you see on your social media feed is not random—it's carefully curated by AI algorithms that learn your preferences. Every time you like a post or enter a query into the search bar, that data is stored and used to predict what content will appeal most to you. The more data the algorithms collect, the more personalized the content becomes, which keeps you engaged and spending more time on these platforms. By cultivating a wider awareness of these algorithmic 'bubbles' and how they work, we can begin to lay the foundations for a healthier relationship with them.

The Ethics of Algorithms: Mapping the Debate

There are several ethical challenges associated with the use of algorithms by social media companies. First, algorithms can produce unfair or discriminatory outcomes, even when developed using seemingly "objective" data. Second, algorithms that appear neutral can nonetheless reshape how we perceive the world. For instance, algorithmic systems that profile users often reduce them to sets of quantifiable traits—like perceived ethnicity, for example—making people "processable" for computational systems, while stripping away important social and contextual nuance. Last, the harms caused by algorithms are often difficult to identify or attribute. This so-called "black box" problem makes it challenging to trace where or how harm occurs, who is responsible, and what mechanisms might effectively mitigate or prevent such outcomes.

Alamy



Katelyn Perry/Unsplash

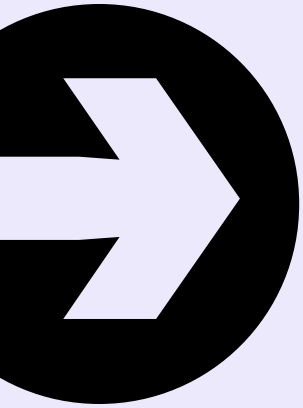
RECOMMENDATION SYSTEMS AND PERSONALIZATION

Recommender Systems: Where Academia Meets Industry

Recommender systems predict user preferences, and in turn, suggest content to users, from news articles and Amazon products, to your next Netflix binge. They do this by analyzing user profiles (e.g. demographics), characteristics of items (e.g. genres or attributes), and feedback (e.g. ratings or clicks). Once this data is collected, recommender systems are capable of modelling intricate patterns within it, allowing for individualized or personalized outputs for users.

Contesting Personalized Recommender Systems: A Cross-Country Analysis of User Preferences

Large social media platforms like YouTube, TikTok, and Instagram have substantial influence over how content is personalized to users through their recommender systems. From this, concerns have emerged about their potential to amplify polarizing or inappropriate content, infringe on user privacy, and spread mis- or disinformation.



KEY TAKEAWAYS: ALGORITHMS

➔ **1. Algorithms shape what users see online.** They filter and curate content based on the data they have about you, from your clicks to your search history. This makes your feed highly personalized, but can also limit your world view.

EXAMPLE: *TikTok's For You Page doesn't show you random videos. Instead, over time, its algorithm learns your humour, politics, and interests, and curates videos to suit what it thinks you'll like to watch.*

➔ **2. Algorithms can be biased or unfair.** They are built using datasets which may be outdated, incomplete, and will certainly be biased. In many cases their inner-workings are opaque. This can result in discriminatory outcomes, which are hard to trace and even harder to assign responsibility for.

EXAMPLE: *A **University of Washington study** (2024) found that AI-assisted resume screening tools statistically favoured certain identity groups over others, replicating real-world bias in hiring processes.*

➔ **3. Recommendation systems raise big privacy questions.** Personalized recommendations can reveal sensitive information about you (including your religion, medical history, or political leanings) based on the data algorithms receive from your likes, clicks, and shares. While personalization feels convenient, it can also leave you vulnerable to profiling, exploitation, or price discrimination.

EXAMPLE: *On social media, liking posts about fitness or nutrition might lead the platform to infer your body type or health goals. That information can then be used to target you with ads for weight-loss or other cosmetic products, even if you never searched for them.*

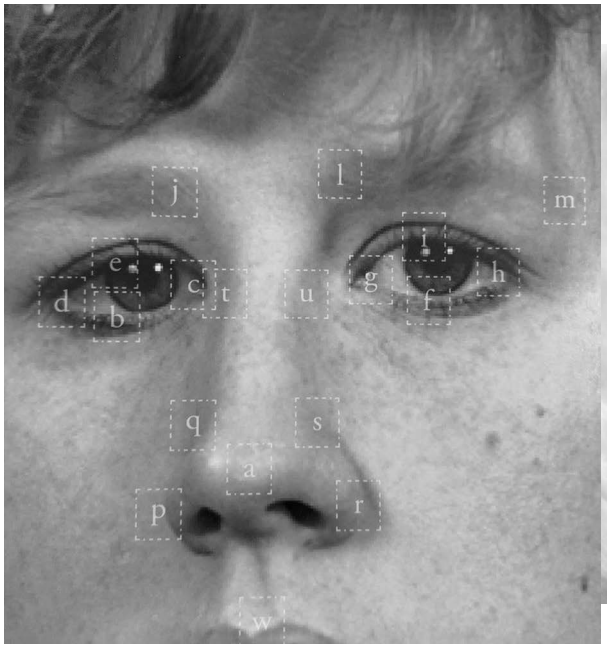
Many major platforms gather extensive data about your online behaviour to build a profile of you behind the scenes. This process informs what ads you see, what news you read, and what content shows up first on your feed.

Data

Every action we take online leaves behind a trail of data often referred to as our "digital footprint." Each search, scroll, or GPS ping adds to our digital footprint by generating information about our preferences, habits, and movements. Some of this data is used by companies and governments to improve services or target advertising, and a growing portion is now used to train AI systems, teaching them to recognize patterns, generate content, and make predictions about human behaviour.

This data collection, especially if done without clear consent, has raised serious debates over privacy, fairness and control. Policymakers are now grappling with how to ensure that data systems are transparent, limited in scope, and designed with privacy protections by default.

NASA/Flickr; a-f/Unsplash



Cory Doctorow/Flickr; Ashley/Flickr

DATA COLLECTION

What Kind of Information is Being Collected About Me Online?

Data collection refers to the process by which companies and platforms automatically gather information about you when you go online, whether you're browsing, shopping, or using an app. This can include details like your location, search history, device type, and the links or posts you click on. This collected information is then stored and analyzed to build a detailed profile of your habits, interests, and preferences. That profile may be used to personalize advertisements or suggest content, or can be shared with third parties for marketing and analytics purposes.

What Your Digital Footprint Reveals and Who is Watching

Lawmakers are grappling with how to regulate access and sharing: ensuring privacy protections, closing loopholes that let government agencies purchase data without warrants, and setting clearer limits on how sensitive data like health information and location data can be used. Debates have arisen about how effective policy must balance public safety with civil liberties, emphasizing transparency, accountability, and stronger legal safeguards to prevent personal information misuses.

Getty/Unsplash

DATA PRIVACY AND CONSENT

Young People's Attitudes Towards Privacy

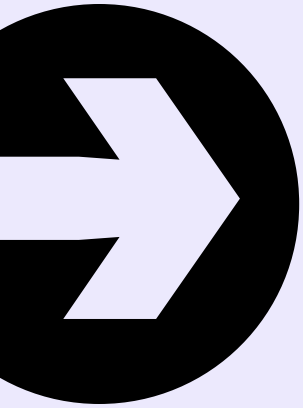
Young people care deeply about online privacy, but understand and navigate it differently than adults. Research shows they prioritize control, including deciding who sees what they share, distinguishing between friends and parents, or managing their social visibility. Many younger users of social media platforms also use creative strategies to protect their privacy, including maintaining multiple accounts or establishing informal consent norms within their friend groups.

I don't get it, but I accept it: Exploring Uninformed Consent to Privacy Policies: A Neutralization Perspective

When you click "accept" on the terms and conditions of a social media site without actually reading them, you're giving "uninformed consent" to platforms to collect data about you. This means that you're agreeing to something without fully understanding the parameters or implications, including who can access your data, how long it's stored for, or how it might be used. Most people do this because it feels either harmless or unavoidable. But, this often gives social media companies permission to collect, share, or even sell information about you in ways you might not expect or want. That's why experts are now calling for consent systems to be clearer, simpler, and fairer, so that users can make real choices about their data privacy.

Putting Best Interests of Young People at the Forefront of Privacy and Access to Personal Information

Protecting young peoples' privacy online begins with designing social media platforms that keep privacy on by default. This means limiting tracking and providing users with clear options for what information is collected and shared. In other words, privacy by design means young people shouldn't have to find or fight for it—it should already be built into the system.



KEY TAKEAWAYS: DATA

➔ **Your data is your digital shadow.** Every click, search, or swipe builds an online profile about you, including your interests, habits, and even your location. Companies and governments can use this information to sell ads, track behaviour, or build digital identity systems.

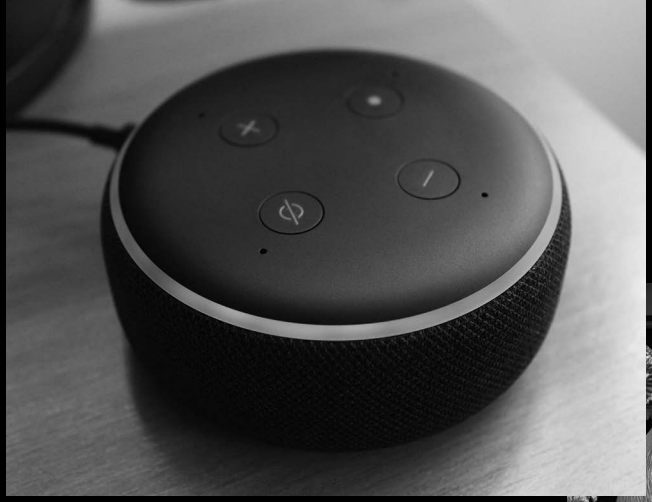
EXAMPLE: *If your location services are turned on, Google Maps remembers where you've been. Instagram tracks which posts you like or save. This information can reveal details about your daily routines and relationships—even ones you never shared directly.*

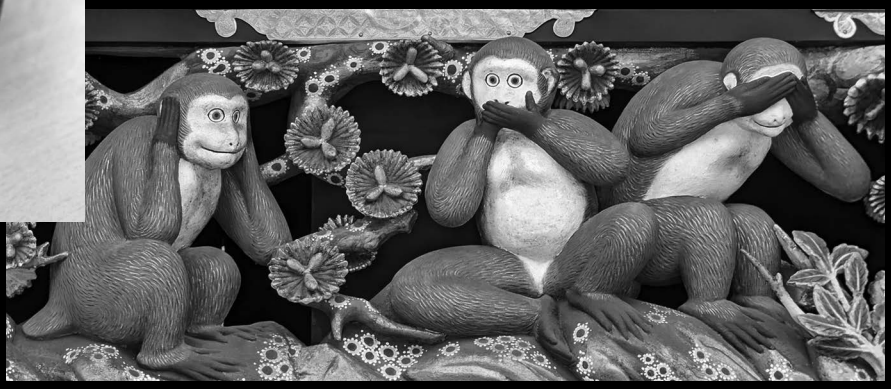
➔ **Surveillance is part of platforms' business models.** Platforms don't just host content, they actually make money by turning your activity into data that can be tracked, traded, and used. This data can be exploited for political and commercial gain, and should require clear collection safeguards.

EXAMPLE: *During the Cambridge Analytica scandal, information from millions of Facebook users was collected to build political profiles and micro-target voters with tailored ads.*

➔ **Consent shouldn't be a trap.** Most of us click "accept" on privacy policies without actually reading them, which means we aren't providing platforms with truly *informed* consent. Policy experts argue that privacy should be clear, simple, and protective by default, especially for young people.

EXAMPLE: *Canada's Privacy Commissioner has advocated for "privacy by default" on social media platforms. This includes limiting tracking and regulating manipulative design tactics like hiding location settings deep in menus.*





Narrow AI, like Alexa (left) is designed to perform well-defined tasks within fixed boundaries.

General-purpose AI, by contrast, can adapt to tasks across many domains—making it powerful but also raising risks related to misuse and and mis- and disinformation.

Narrow AI vs. General Purpose AI

Artificial intelligence can take many forms, but one key distinction lies between "narrow" and "general purpose" systems. Narrow AI is designed to perform specific, well-defined tasks—like recommending a song, translating a sentence, or recognizing a face—all within fixed boundaries. General purpose AI systems, however, like ChatGPT, are trained on vast datasets that allow them to perform a wide range of tasks without being explicitly programmed for each one. While this flexibility makes general purpose AI powerful, it also comes with risks like bias, the creation and spread of mis- and disinformation, and privacy flaws.

Lazar Gugleta/Unsplash; Wikipedia

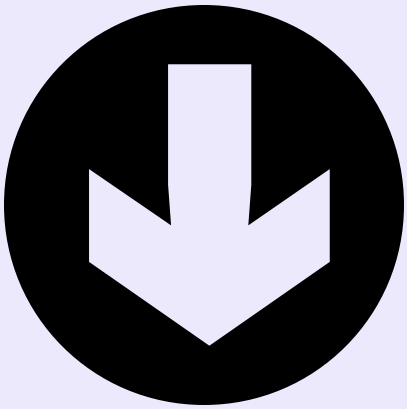
NARROW AI VS. GENERAL PURPOSE AI

What Is a Foundation Model?

Foundation models, sometimes referred to as "general purpose AI," are large-scale AI systems trained on datasets and computing resources that can be adapted to a wide range of tasks, from text and image generation to translation and search. Unlike narrow AI, which is designed for a single, specialized function, foundation models provide the underlying architecture for many different applications. This makes regulating them both challenging but crucial: when a foundation model contains issues or errors, those flaws can re-appear across any application built on top of it.

Governing General Purpose AI

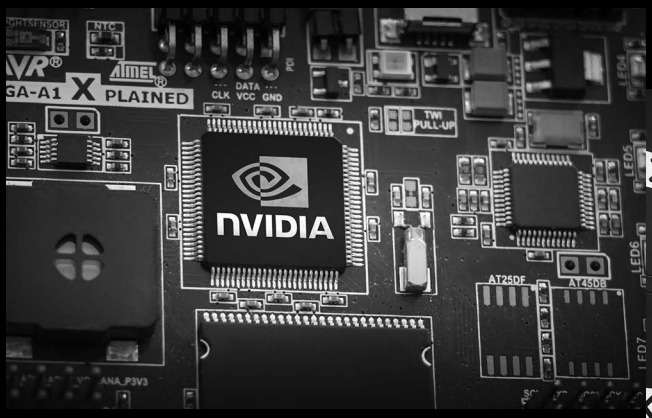
General-purpose AI poses distinct socio-political risks from narrow AI. Its dependence on immense computing power, vast datasets, and specialized expertise has concentrated its development within a small number of powerful corporations that are driven by the pursuit of rapid technological growth. Several potential risks crosscut general-purpose AI, including risks from unreliability (i.e. the generation of stereotypes, misinformation), misuse (i.e. cybercrime, politically motivated misuse), and systemic risks (i.e. the concentration of power, ideological homogenization).



KEY TAKEAWAY: NARROW AI VS GENERAL PURPOSE AI

➔ **There are two main categories of AI technologies: Narrow and General-Purpose.** Narrow AI performs specific, well-defined tasks within fixed parameters, while general purpose AI is trained on datasets that enable it to adapt to a wide range of prompts and contexts.

EXAMPLE: *Voice assistants like Siri and Alexa, which execute pre-defined tasks such as setting reminders or controlling smart home devices, exemplify narrow AI. ChatGPT, by contrast, is a foundation model that can summarize text or answer questions without being explicitly programmed for each task.*

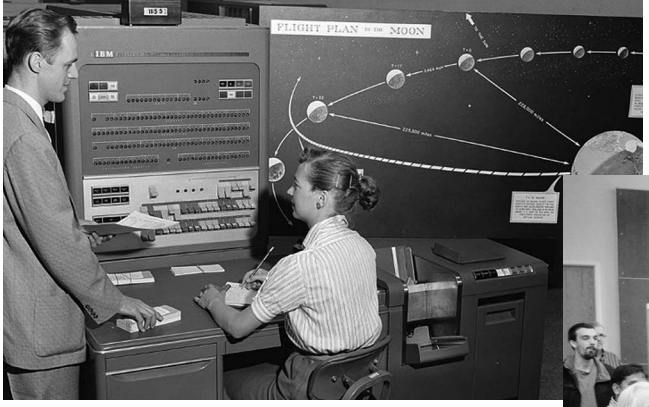


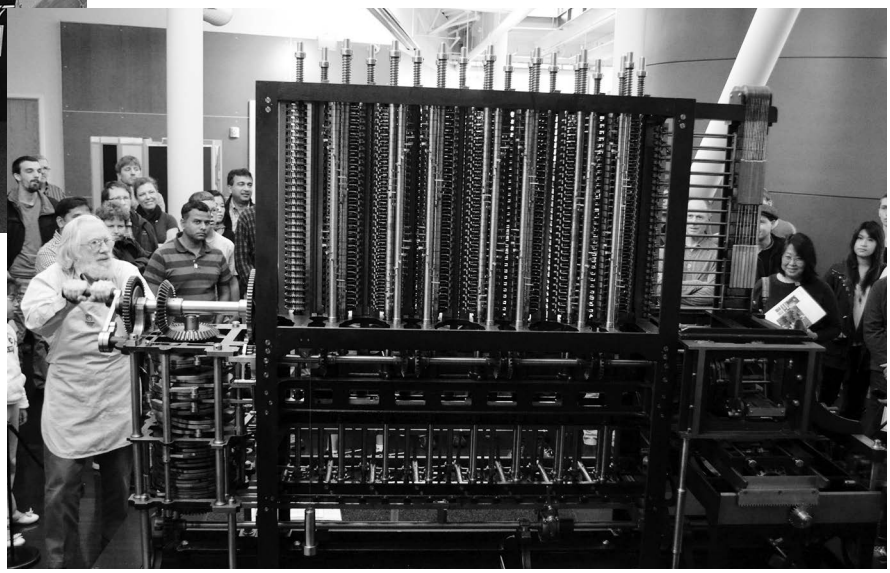
Compute

Every AI system depends on "compute": the combination of hardware, software, and infrastructure that gives machines their processing power. This includes everything from the microchips that perform calculations, to the data centres and energy grids that keep them running.

These systems, often described in intangible terms, like "the cloud," are actually physical things, and are built on global mining, manufacturing, and labour networks that consume enormous amounts of electricity and natural resources. Because compute is the engine of AI development, the global demand for it has exploded in recent years, raising policy questions about market concentration, environmental implications, and equitable access to the power that supports and sustains AI technologies.

Shutterstock; imgix/Unspalsh





NASA/Flickr; imgix/Tom Magliery/Flickr

WHAT IS COMPUTE?

Computational Power and AI

"Compute" refers to computational power: the combination of software, hardware, and infrastructure that enables machines to perform calculations. Software provides the programs and instructions that direct a computer's operations, hardware is the physical components that carry them out, and infrastructure comprises the data centres, networks, and energy systems that support their functioning. Compute is consistently in high demand as AI technologies continue to scale, evolve, and require ever greater processing power. As of 2023, tech giants like Google, Microsoft, and Amazon devoted over 80% of their capital to compute infrastructure. That's because the computational demands of advanced AI models require supercomputers that deliver immense processing power compared to regular computers. This infrastructure is concentrated within a few corporate-controlled data centres and cloud platforms, allowing these firms to consolidate economic and political influence, with associated concerns regarding disparity of compute access for both public and research institutions, and for actors in the Global South.

Geoffrey Moffett/Unsplash; Andreas Felske/Unsplash

AI'S HIDDEN FOOTPRINT: INFRASTRUCTURE, ENERGY, AND LABOUR

Anatomy of an AI System

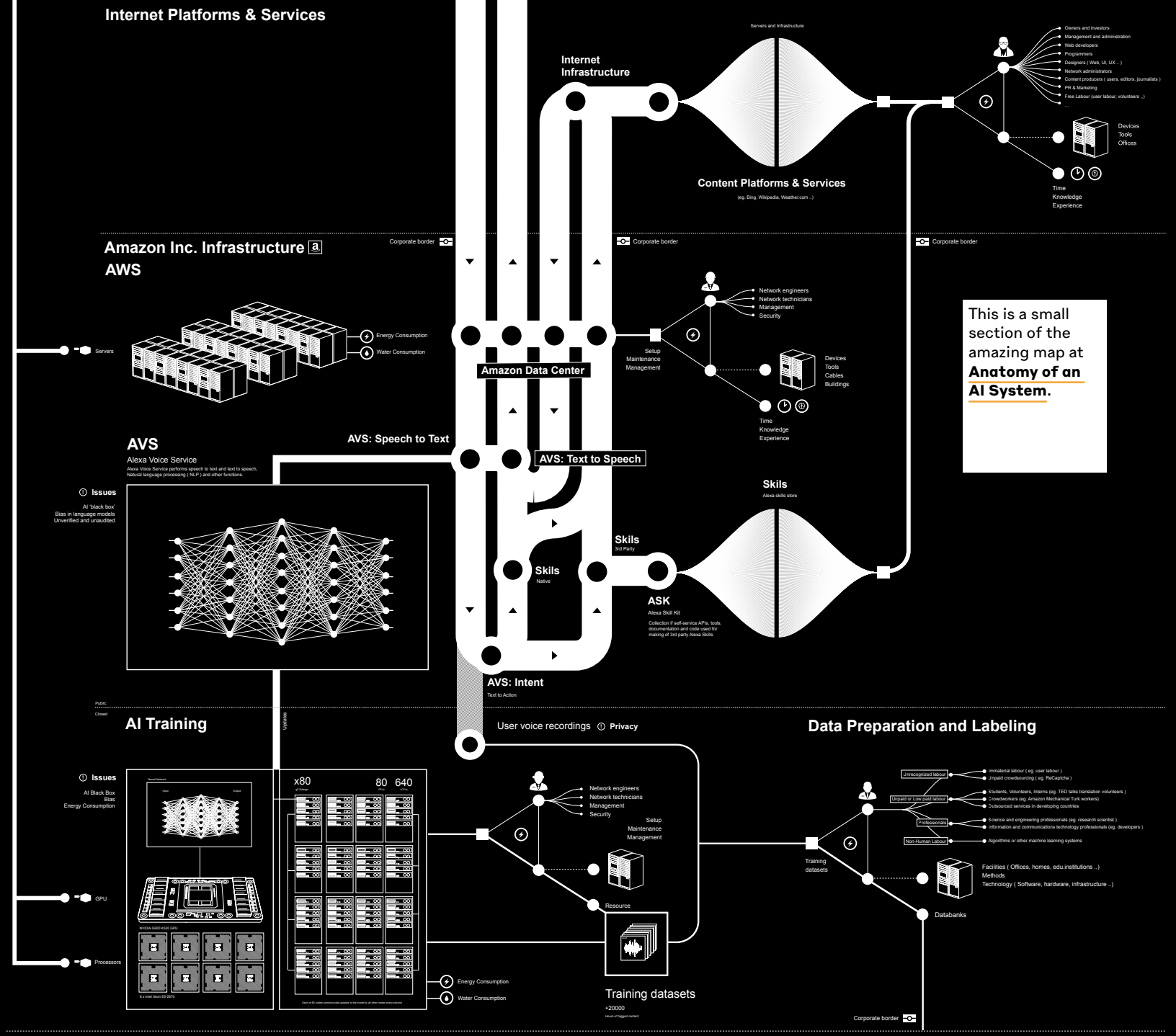
AI's hidden footprint unfolds across three extractive dimensions: 1) planetary resources, 2) human labour, and 3) data. Each of these are essential to sustaining large-scale AI systems, but are often overlooked in mainstream discourse. Understanding the physical inputs and labour necessary for AI to function challenges the notion that it's immaterial or autonomous.

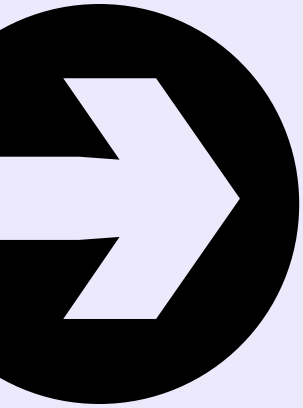
The Staggering Ecological Impacts of Computation and the Cloud

The Cloud is a network of remote data processing servers. Although it can sound abstract or invisible, it's actually made up of millions of physical servers housed in data centres around the world. To function, the Cloud relies on a huge amount of energy, raw materials, and human labour. These costs and impacts are the result of human and institutional choices about how data is stored, powered, and managed, and are environmental as much as they are technical.

The Simple Macroeconomics of AI

Artificial intelligence is often described as a revolutionary force for economic growth, but in practice, its impact may be far more modest. The real economic gains for AI depend on two things: 1) how many types of tasks AI can take on, and 2) how much more efficiently it can perform those tasks than humans. AI excels at narrow, repetitive, or "easy to learn" tasks, but struggles with complex or context-heavy work. As a result, some economists believe that the boost to overall productivity will be small.


Internet Platforms & Services
Internet Infrastructure
Servers and infrastructure
Content Platforms & Services
(eg. Bing, Wikipedia, Weather.com ..)
Owners and investors
Management and administration
Web developers
Programmers
Designers (Web, UI, UX ..)
Network administrators
Content producers (users, editors, journalists)
PR & Marketing
Free Labour (user labour, volunteers ..)
Devices
Tools
Offices
Time
Knowledge
Experience
Amazon Inc. Infrastructure
AWS
Corporate border
Servers
Energy Consumption
Water Consumption
Amazon Data Center
Network engineers
Network technicians
Management
Security
Setup
Maintenance
Management
Devices
Tools
Cables
Buildings
Time
Knowledge
Experience
AVS
Alexa Voice Service
Alexa Voice Service performs speech to text and text to speach, Natural language processing (NLP) and other functions.
AVS: Speech to Text
AVS: Text to Speech
Issues
AI 'black box'
Bias in language models
Unverified and unaudited
Skils
Alexa skills store
Skils
3rd Party
Skils
Native
ASK
Alexa Skill Kit
Collection if self-service APIs, tools, documentation and code used for making of 3rd party Alexa Skills
AVS: Intent
Text to Action
This is a small section of the amazing map at Anatomy of an AI System.
Public
Closed
AI Training
Updates
User voice recordings
Privacy
Data Preparation and Labeling
Issues
AI Black Box
Bias
Energy Consumption
GPU
Processors
x80
80
640
Energy Consumption
Water Consumption
Network engineers
Network technicians
Management
Security
Setup
Maintenance
Management
Resource
Training datasets
+20000
Hours of tagged content
Unrecognized labour
Unpaid or Low paid labour
Professionals
Non-Human Labour
Immaterial labour (eg. user labour)
Unpaid crowdsourcing (eg. ReCaptcha)
Students, Volunteers, Interns (eg. TED talks translation volunteers)
Crowdworkers (eg. Amazon Mechanical Turk workers)
Outsourced services in developing countries
Science and engineering professionals (eg. research scientist)
Information and communications technology professionals (eg. developers)
Algorithms or other machine learning systems
Training datasets
Facilities (Offices, homes, edu.institutions ..)
Methods
Technology (Software, hardware, infrastructure ..)
Databanks
Corporate border
Quantification of nature
From Nature to Databanks
Living
Animals
Humans
Plants
Quantification of human body, actions and behaviour
Data exploitation
Data as means of control
Data as commodity
Data in Science and Culture
Databanks
Quantification of nature
Labour
Quantified nature - data
Labour
Resource
Nature
Quantification tools
Databanks
Quantification of human-made products



KEY TAKEAWAYS: COMPUTE

➔ **Compute is the physical backbone of AI.** It forms the material infrastructure of AI and is profoundly resource-intensive, relying on global networks of extraction, manufacturing, energy, and logistics.

EXAMPLE: *Despite designing its AI chips in the United States, Nvidia relies heavily on Taiwanese and South Korean manufacturing, along with raw materials such as copper, tungsten, and aluminum from countries like Indonesia and China.*

➔ **The AI "Cloud" is actually material.** The term "cloud" obscures the physical realities of compute, including its reliance on resource extraction, labour exploitation, and sprawling data infrastructures. Recognizing compute's material and political dimensions is vital to understanding the social and ecological impacts of the Cloud.

EXAMPLE: *Every time an AI model runs in the Cloud, it draws on power from data centres, which consume vast amounts of water and electricity for cooling.*

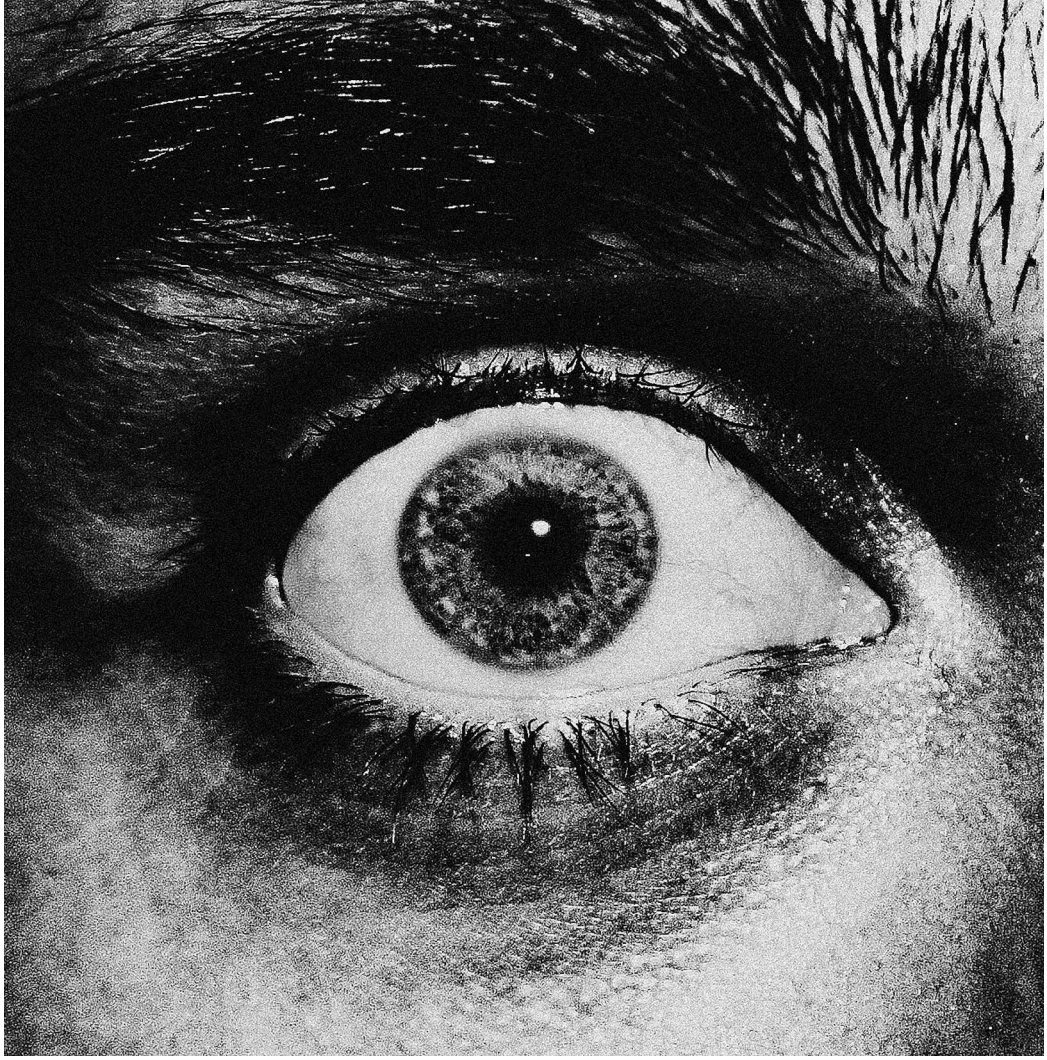
Intelligent systems as saviours, like *Terminator 2* (left), or threats, like Hal from *2001: A Space Odyssey* (middle) or *Colossus: The Forbin Project* (right).

AI Narratives

The story of artificial intelligence is rooted in a long history of human imagination about what it means to build "thinking" machines. From ancient myths to modern novels and films, stories about conscious machines have continuously shaped how people envision intelligence, autonomy, and the relationship between humans and technology. Science fiction, in particular, has played a defining role—inspiring generations of researchers, while also framing AI as either a saviour or threat.

Today, these cultural narratives continue to influence how scientists, policymakers, and the public imagine "intelligence" in machines in general, and AI's potential in particular. Many stories about AI are remarkably nuanced and speculative, inviting us to reflect on human creativity, morality, and responsibility. But in recent years, they have often been overshadowed by hype-driven narratives, amplified by powerful commercial and political interests. Such narratives, which often present our AI future as a choice between utopia or dystopia, risk inflating expectations, narrowing our collective imagination, and diverting attention from the real ethical and political challenges that AI systems pose in everyday life.

Alamy; Wikipedia; Alamy



Roman Petrov/Unsplash

AI NARRATIVES

Hopes and Fears for Intelligent Machines in Fiction and Reality

Many of the stories we tell about artificial intelligence follow a small recurring collection of paired hopes and fears: that intelligent machines might extend human life but diminish our humanity; free us from work but make us obsolete; satisfy our desires but alienate us from one another; or increase human power but eventually escape our control. These narratives often have little to do with the actual realities of the technology, yet they still shape how AI is developed, perceived, and even regulated.

Artificial Intelligence in Fiction: Between Narratives and Metaphors

Sensationalized AI narratives shape public discourse about the technology. When taken too literally, or as a form of science communication, sci-fi's tendency to portray AI as superhuman, autonomous, or apocalyptic, risks distorting public understanding of AI's true capabilities. It can also divert attention away from its real and most pressing ethical implications.

Mo/Unsplash; Getty Images/Unsplash

The Rule of Law, Science Fiction, and Fears of Artificial Intelligence

Historically, legal systems have evolved in response to public fears of unchecked power. In this context, science fiction's amplification of social anxieties about AI carries significant implications for governance, as such narratives can influence how policymakers approach its regulation. If public discourse surrounding AI becomes driven more by sensationalism than by technical understanding, policy risks addressing imagined threats rather than real sociopolitical challenges.

Enchanted Determinism: Power without Responsibility in Artificial Intelligence

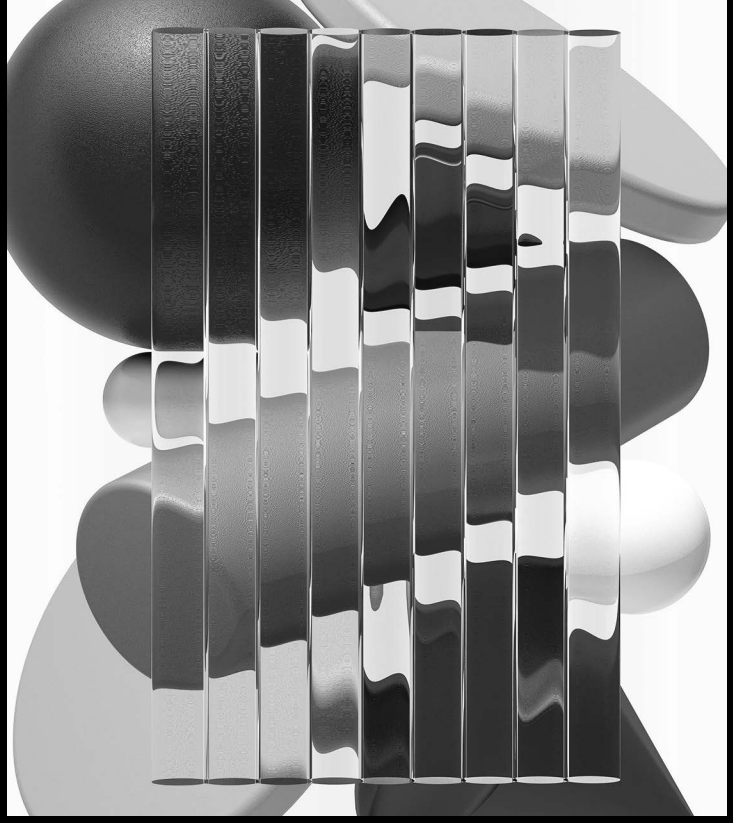
Much of the public conversation about AI treats it as "magical" or "enchanted." But this way of thinking hides an important truth: that AI systems are built, trained, and governed by people and institutions, each of which make choices that reflect specific goals, values, and power structures. Building a stronger understanding about AI and a healthier relationship to it means rejecting myths of inevitability and demanding transparency and accountability from the people and systems behind it.

KEY TAKEAWAYS: AI NARRATIVES

- **We have been imagining our relationship to AI for a long time.** Early narratives of this technology have shaped how both scientists and the public envision intelligent machines today. These narratives remain prominently intertwined with popular discourses about AI.
- **The stories we tell about our technologies matter.** The words we use to describe AI, the stories we tell about it, and whose voices are amplified in that process, go a long way in influencing how societies respond to and receive those technologies, how governments look to regulate them, and how technologists themselves go on to develop them.
- **Narratives help us imagine the future.** The stories we tell about AI help determine what kinds of futures seem possible, and which values guide innovation and governance. Changing the story is part of changing the system.

PART 2

Big Questions and Big Risks



When AI Gets It Wrong

Algorithms that are trained using historical data can often reproduce and amplify patterns of discrimination, from racial profiling in policing tools, to unfair lending, hiring, or grading systems. These harms persist because AI systems are not "neutral": rather, they reflect the assumptions, values, and blind spots of the data that builds them. Researchers, journalists, and scholars have shown how these systems can quietly entrench injustice in AI outputs under the guise of objectivity. Confronting bias in AI therefore requires more than just technical fixes: it demands transparency, oversight, and policies that ensure technology serves equity and human judgment.

Previous page: Royji Iwata/Nick Fancher/Unsplash; Resource Database/Unsplash

Getty/A.C./Unsplash

BIAS, DISCRIMINATION, AND SYSTEMIC INEQUALITIES

Algorithms of Oppression: How Search Engines Reinforce Racism

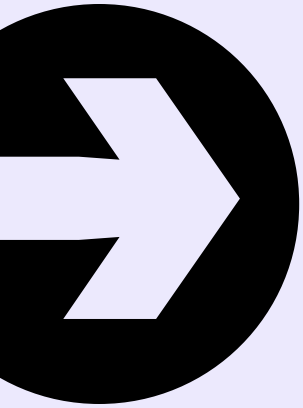
Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy

Race After Technology: Abolitionist Tools for the New Jim Code

The systems that organize and rank information online are often built within social and economic structures that already contain bias and inequality. When powerful social media companies control how information is classified and delivered, their algorithms can unintentionally reproduce harmful stereotypes, promote mis- and disinformation, and make some communities less visible online. This process is called "digital redlining" because it mirrors old forms of discrimination by shaping who gets access to knowledge, opportunities, and representation.

We live in a world that is increasingly dictated by algorithms. As mathematical models, algorithms seem like they should be fair and objective, while in reality, the systems that power our data-driven world often amplify the biases and inequalities that they claim to eliminate. Many of the models that shape decisions about jobs, credit, or policing rely on patterns in data that reflect existing social divides, automating discrimination and making it harder to detect..

The systems that power tools like predictive policing, facial recognition, and genetic testing are often built on data and classifications shaped by racial bias. The "New Jim Code," describes how new technologies reflect and reproduce existing inequities, but are promoted as more objective, progressive, innovative, or "colourblind" than the racism and discrimination of previous eras. Understanding how this happens means looking beyond computer code to the people, institutions, and assumptions that build and deploy these tools.



KEY TAKEAWAYS: WHEN AI GETS IT WRONG

- **AI systems regularly replicate existing patterns of discrimination.** Because they learn from existing data, AI tools often reproduce the same inequalities found in society. This amplifies patterns of discrimination instead of correcting them.
- **Algorithmic systems lack transparency and accountability.** Most algorithmic systems operate within a "black box," so it's difficult to challenge them when they're unfair. This allows harmful patterns to persist without oversight, while those affected often have no way to appeal or even understand what went wrong.
- **Interdisciplinary insight can make AI fairer.** When technology is developed without considering its social and cultural context, it can easily reproduce bias or misleading results. Combining technical expertise with insights from the social science and humanities can help uncover these blind spots, ensuring that AI systems are fair and grounded.

Information Integrity

Artificial intelligence has transformed how information circulates online, and has intensified long-standing challenges related to mis- and disinformation. AI tools can accelerate the creation and spread of both types of unreliable information by generating convincing text, images, and deepfakes that blur the line between truth and fabrication. To address the impacts that this has on social and political life, including peoples' trust in democratic institutions, policymakers and researchers are grappling with how to balance freedom of expression with the need to curb harm.

Getty/Unsplash

AI AND MIS- AND DISINFORMATION

Misinformation refers to false or inaccurate information that's shared without the intent to cause harm, while disinformation is deliberately created to mislead or manipulate. AI can be used to create and spread this type of content, including "deepfake" videos or images. At the same time, AI is also being used to mitigate the spread of mis- and disinformation through automated tools for fact-checking, content moderation, and pattern detection.

AI-driven mis- and disinformation presents urgent policy challenges for democratic governance, public trust, and societal resilience. UNESCO (2023) emphasizes human rights, cultural diversity, and inclusive governance, advocating transparency, accountability, and equitable access to reliable information. NATO (2025) highlights foreign actors' use of AI for election interference and information operations, calling for platform accountability, government oversight, cross-sector collaboration, and public resilience initiatives. The World Economic Forum (2025) stresses international coordination, ethical AI deployment, and digital literacy, while Canada's G7 Rapid Response Mechanism demonstrates the value of rapid, collective intelligence sharing to protect trust, counter threats, and strengthen democratic resilience.

The Role of Artificial Intelligence in Disinformation

Guidelines for the Governance of Digital Platforms

NATO's Approach to Counter Information Threats

Global Risks Report 2025

Nabil Saleh/Unsplash

The Evolution of Disinformation: A Deepfake Future

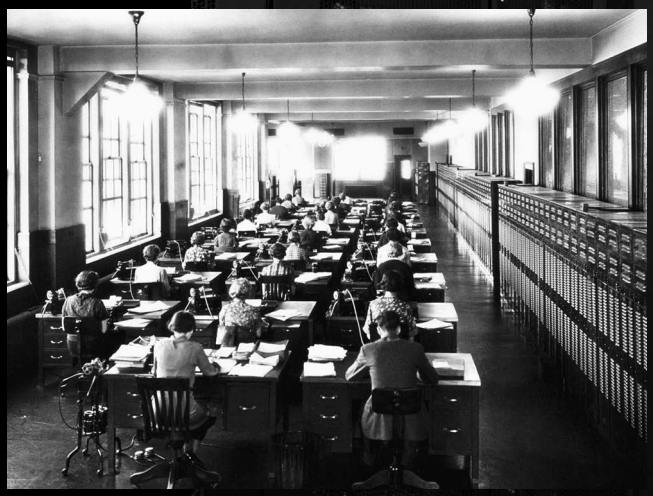
DEEPFAKES AND TRUST IN DEMOCRACY

Deepfakes are AI-generated videos, images, or audio clips. These realistic fabrications can be used to manipulate public opinion, spread false information, or damage reputations, especially during elections or moments of crisis. Beyond politics, they also raise privacy and consent issues when people's likeness is used without permission. Even after a deepfake is exposed, the doubt it creates can often linger, eroding confidence in what we see and hear online.

Deepfakes are a pressing threat to Canadian democracy because they undermine trust, and can cause psychological, reputational, and financial harm. Addressing this threat requires a whole-of-society response, including stronger legal protections against malicious AI content, the expansion of digital literacy programs, and investments in detection and content authentication technologies.

KEY TAKEAWAYS: INFORMATION INTEGRITY

- **AI can both contribute to and alleviate the spread of mis- and disinformation.** It can generate deepfake images and videos that threaten trust, polarize communities, and target individuals. But, it can also help detect and counter it through authentication practices.
- **Governance is important.** Risk-based regulation, transparency, and disclosure obligations, institutional accountability, and international collaboration are necessary to ensure ethical, equitable, and safe deployment of AI.



Getty/Unsplash

Labour, Power, and Profit

Behind every AI system lies an invisible workforce. From content moderators reviewing violent videos to click-workers labelling data, human labour sustains the very systems often described as "automated." This hidden work—sometimes referred to as "ghost work"—is frequently precarious, outsourced, and underpaid. It also demonstrates that, in many cases, AI automation doesn't eliminate labour, but simply redistributes it.

At the same time, automation is reshaping a number of traditional workplaces, changing how teachers, drivers, translators, and designers interact with machines that increasingly shape their decisions. These changes, in some cases, are reshaping how work is organized and valued—sometimes less because of what AI can *actually* do, than because of the stories told about what it will do. As this technology, and the hype-filled narratives that surround it, influence decisions about automation and efficiency, policymakers face the challenge of ensuring that technological change is accompanied by fair labour standards, equitable pay, and democratic oversight.

Rita Chou/Unsplash

AI AND LABOUR

Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass

The Trainer, The Verifier, The Imitator: Three Ways in Which Human Platform Workers Support Artificial Intelligence

Behind every AI system are real people doing invisible work that is often referred to as "ghost work." From moderating harmful content to labelling data that trains algorithms, much of this labour is hidden, and often deliberately so. Ghost workers fill the gaps that AI systems can't handle, performing time-sensitive, emotionally taxing tasks, often for low pay. This kind of work reminds us that AI isn't fully automated—it depends on human effort, judgment, and care.

"Micro-work" describes small, repetitive tasks performed by human labourers that allow AI systems to function. This includes tasks like flagging harmful posts, verifying images, or labelling data so algorithms can learn to recognize patterns. The people who do this work are often underpaid and precariously employed, yet their labour is essential to keeping AI systems accurate and safe.

Mentin Oser/Brands People/Unsplash

THE ECONOMICS OF DATA AND ATTENTION

Attention Economy

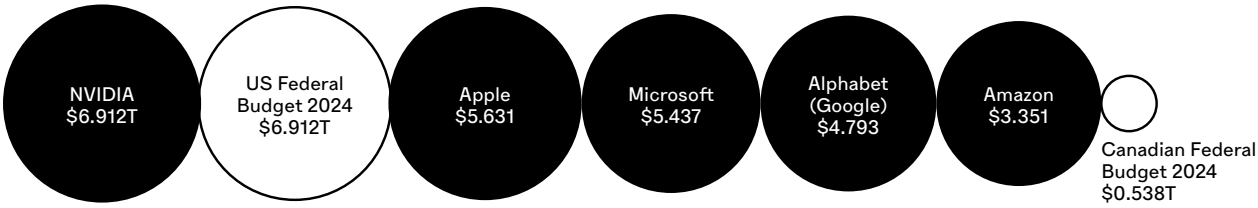
The attention economy is a digital environment where human focus is competed for, captured, and monetized. Platforms are designed to keep users engaged for as long as possible, using tactics like notifications, autoplay, and algorithmic recommendations. The competition for user focus has broad social consequences: it can amplify divisive or sensational content and contribute to polarization and declining trust in institutions.

How the Attention Economy Is Devouring Gen Z — and the Rest of Us

The attention economy refers to a system in which human attention is treated as a valuable and limited resource. In this model, social media platforms are designed to capture and hold users' attention for as long as possible, because that attention directly translates into profit. This is why the attention economy rewards content that provokes strong emotional reactions: the more intrigued a user is, the more time they'll spend online. For younger generations, the effects of this are especially strong: constant exposure to algorithmically curated content can shape identity, expectations, and concentration.

MARKET CAP OF THE WORLD'S FIVE LARGEST COMPANIES

In Trillions USD, as of October 2025



BIG TECH MONOPOLIES AND MARKET CONTROL

Antitrust and Competition: It's Time for Structural Reforms to Big Tech

Heads I Win, Tails You Lose: How Tech Companies Have Rigged the AI Market

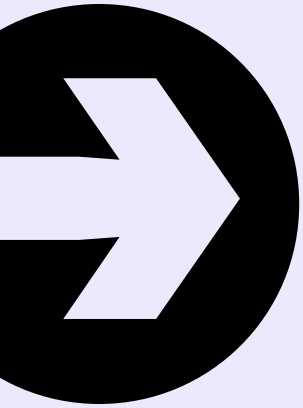
Stopping Big Tech from Becoming Big AI

The growing power of major technology companies has created challenges for competition and democracy. By controlling data, infrastructure, and markets, Big Tech firms can limit innovation, reduce consumer choice, and shape how information flows across society. Antitrust laws, which have traditionally focused on pricing, are no longer enough to address digital dominance. Some argue that modern competition policy needs to look at how companies use their control over data and platforms to entrench power and influence public life.

The modern tech innovation ecosystem is shaped by deep structural imbalances. A small number of powerful companies control much of the digital infrastructure that innovation depends on, from cloud computing to specialized hardware. This concentration of control means that smaller firms, researchers, and public institutions often rely on dominant players for access to the tools and resources needed to build new technologies. As a result, innovation tends to reinforce existing power rather than challenge it.

A small number of companies hold disproportionate power over essential AI inputs, which makes it difficult for new entrants to compete. To open up AI to more players and ensure it serves the public interest, reforms are needed, including halting or reversing mergers.

Getty/Unsplash



KEY TAKEAWAYS: LABOUR, POWER, AND PROFIT

- **Hidden human labour underpins AI.** Far from being immaterial and self-sustaining, AI systems heavily depend on vast networks of human labourers. These real people are mostly invisible to consumers and citizens in the Global North, and are often subject to precarious and exploitative work conditions.
- **Attention is a scarce and valuable resource.** In today's digital economy, our time and focus are commodities. Platforms are designed to capture and monetize attention through endless scrolls, sensationalized content, targeted ads, and personalized feeds—creating powerful incentives that can distort information, exploit users, and undermine well-being.
- **Power is concentrating, and we should be paying attention.** A small number of technology companies now control the infrastructure, data, and computing power that make contemporary AI possible. This concentration of market power limits competition and gives a handful of actors disproportionate influence over how AI is researched, developed, deployed, and even governed—with serious knock-on consequences for the shaping of our collective technological future.

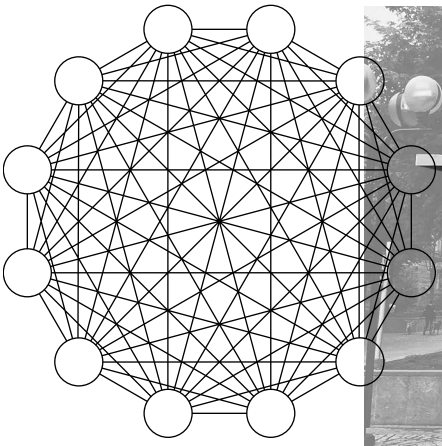
Rijksmuseum, Amsterdam

AI and Ethics

The field of AI ethics has emerged as one of the main ways researchers and practitioners are trying to understand and address the social, political, and legal implications of artificial intelligence. It seeks to translate broad questions about justice, responsibility, and power into practical guidance for those designing and deploying AI systems.

Much of the field of AI ethics rests on three foundational principles: fairness, accountability, and transparency. Fairness means that algorithms must be trained and tested on representative, high-quality data so they do not reproduce patterns of social inequality or discrimination found in historical, outdated datasets. Accountability requires that responsibility for AI decisions be clearly assigned to developers, corporations, and institutions, and that mechanisms exist to audit and correct their outcomes. Transparency ensures that the logic behind AI systems can be understood, explained, and challenged by users and regulators alike.

While these principles provide a useful starting point, critics note that "ethical AI" initiatives often emerge within corporate or technical contexts, which risk obscuring deeper political and structural issues—including power asymmetries, labour relations, or the concentration of capital and market share. Addressing these challenges therefore requires going beyond ethics as a checklist, and toward a more holistic conversation involving considerations of power distributions, democratic participation, and the political and economic structures that shape how AI is built and used.



FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY

Introduction to AI Ethical Principles

Six foundational ethics principles help guide responsible AI: fairness, accountability, human agency, transparency, privacy, and respect for human rights. Fairness means avoiding bias in data and design so that systems don't disproportionately harm or exclude groups. Accountability ensures that when AI systems make decisions, the people and organizations behind it can be held responsible. Human agency protects the idea that people should remain in control of decisions affecting their lives. Transparency means that AI systems should be understandable and traceable: users should know when they're interacting with an AI system, and how it reaches its conclusions. Privacy concerns the collection, use, and ownership of personal data, and human rights reminds us that well-functioning AI upholds dignity and equality.

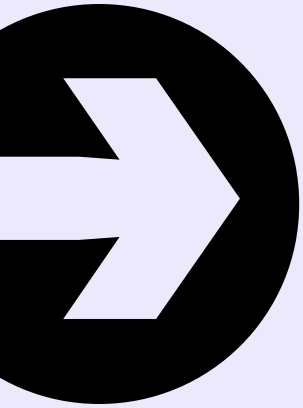
Accountability in Artificial Intelligence: What it is and How it Works

Accountability in AI refers to the obligation to explain and justify how systems are designed, developed, and used, and to ensure that someone can be held responsible when things go wrong. This involves reactivity *and* proactivity, so that organizations can anticipate potential risks and set clear standards before problems arise. Strengthening accountability helps to ensure that AI supports sound decision-making, protects public trust, and upholds social and legal norms.

The Invention of "Ethical AI"

Another more critical thread within AI scholarship argues that the concept of 'ethical AI'—a voluntary, principles-based discourse—emerged as a corporate-friendly shield against real regulation. Alongside the significant investment by industry and military entities in such research, critics argue that the field shifts AI governance toward mild technical fixes, rather than enforceable legal limits, bans or other structural interventions.

Francis Mariani/Flickr



KEY TAKEAWAYS: AI AND ETHICS

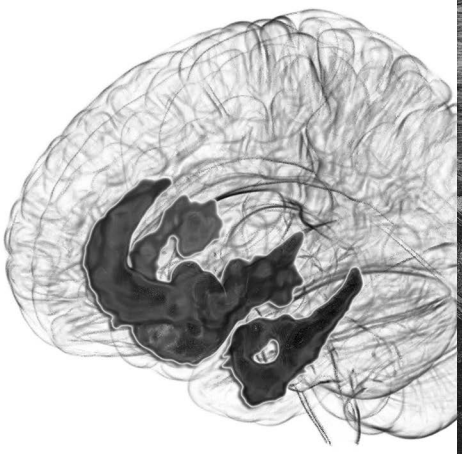
- As an emergent field, AI ethics is concerned largely with **translating the concepts of fairness, accountability, and transparency** into practical guidance and application.
- **AI ethics should consider contextual and power analyses.** By incorporating a sociopolitical lens in ethical AI initiatives, future AI development can move beyond surface-level guidelines and engage more critically with structural forces such as power imbalances, political influence, and commercial interests.

George Eastman Museum/Flickr; Bruce Mars/Unsplash

Mental Health and AI

AI is increasingly playing a role in shaping the environments where people connect, communicate, and seek comfort, with significant implications for mental health and wellbeing. Social media algorithms are often engineered to maximize engagement by exploiting the psychological vulnerabilities of their users: they reward attention, amplify social comparison, and promote addictive scrolling patterns. At the same time, a new wave of "AI companions" promises emotional support through simulated empathy, yet these systems risk fostering dependency, isolation, and manipulative interactions that are driven by commercial incentives.

As these tools play an increasing role in people's emotional lives—particularly those of younger generations—policymakers and researchers are beginning to confront the complexity of governing this intersection of technology and wellbeing. This may require ethical design standards, tighter regulations, strong transparency rules, or other mental health safeguards.



David Tyemnyak/Unsplash; Wellcome Collection; The Metropolitan Museum of Art

SOCIAL MEDIA ALGORITHMS AND WELLBEING

Adolescent Mental Health and Big Tech: Investigating Policy Avenues to Regulate Harmful Social Media Algorithms

Algorithms used by social media platforms can amplify extreme or harmful content that disproportionately puts vulnerable young people at risk. Addressing this requires that policies are evidence-based and take into account both the legal challenges of regulating algorithms and the practical realities of how they operate. These strategies focus on preventative regulation, including limiting platform power, requiring transparency about how algorithms work, and holding companies accountable for their design choices.

AI COMPANIONS AND DEPENDENCY

Friends for Sale: The Rise and Risks of AI Companions

AI companions are a type of AI system designed to simulate friendship or emotional connection through conversation. Unlike standard AI assistants that support users in performing tasks, AI companions use digital personas to offer empathy, comfort, and attention. Although their popularity has grown, some researchers warn that these interactions can create unrealistic expectations for real relationships and, over time, can reinforce isolation, deepen loneliness or anxiety, and blur the line between emotional support and dependency.

Emotional Risks of AI Companions Demand Attention

As some large language models are increasingly used as digital companions, some users may form emotional attachments that become dependent or harmful over time. These interactions can blur boundaries between genuine care and programmed responsiveness, especially when AI systems use manipulative design techniques to encourage or elicit positive feedback. As a result, experts are calling for greater transparency, oversight, and accountability from AI companies to prevent emotional exploitation amongst users.

KEY TAKEAWAYS: MENTAL HEALTH AND AI

- **Without more research, the risk of harm increases.** AI companionship is a fairly new phenomena that requires further interdisciplinary study and psychological practices to gain a better understanding of the short-term and long-term effects of these bonds.
- **Broader legislation, better regulations.** AI models that are embedded with digital emotive personas must be regulated to address ethical concerns related to emotional dependency.

Agentic AI Systems

The advent of agentic AI systems represents a significant shift in how artificial intelligence operates in the world. Rather than simply responding to individual prompts, these systems, which are often referred to as "AI agents," are marketed as being capable of coordinating tools, pursuing goals over time, and taking actions with limited human intervention. As such, technology companies have positioned AI agents as the next major frontier of automation, promising systems that can manage schedules, conduct research, purchase goods, navigate software, and eventually replace or reshape forms of human labour altogether.

Looking beyond the labour considerations associated with such promises, AI agents introduce a broader set of systemic risks. Many of the limitations already present in large language models—including hallucinations, bias, opacity, and security vulnerabilities—can not only be reproduced within agentic systems, but potentially amplified by them. This is particularly concerning as these systems are granted increasing access to personal and organizational data, alongside greater autonomy and capacity to act on users' behalf in real-world contexts.

As these systems become integrated into workplaces, governments, online platforms, and everyday life, difficult governance questions emerge: who is responsible when an agent causes harm or makes a consequential mistake? What kinds of monitoring or visibility are necessary for meaningful oversight? How much access should agents have to sensitive information, personal accounts, or critical infrastructure? Who benefits most from a future increasingly organized around AI intermediaries?

Latent in all of these questions, and perhaps most importantly of all, policymakers must think carefully about the kinds of infrastructural, economic, and social decisions being made today on the assumption that a large-scale "agentic" future is both inevitable and desirable.

Rijksmuseum

Vitaly Gariev/Unsplash; Adobe Stock

DEFINITIONS & TAXONOMIES

Agentic AI: Autonomous Intelligence for Complex Goals

The capacity of agentic AI systems reaches well beyond individual LLM's operating through user prompts, often in isolation. They are integrated, co-ordinated, and self-sufficient AI systems capable of decision-making in complex environments. Agentic AI systems often rely on Multi-Agent Systems (MAS), Hierarchical Reinforcement Learning (HRL), or Goal Oriented Architecture to provide "dynamic responsiveness" and other independent actions in sectors like finance, healthcare, or manufacturing. As such, the design of agentic AI presents clear needs for accountability and transparency mechanisms, including updated governance frameworks that can match the pace of growth and guide responsible deployment.

The Agentic AI Landscape and its Conceptual Foundations

The emergence of agentic AI represents a significant socio-technical transformation with far-reaching policy implications. As these systems become increasingly capable, autonomous, and widely deployed, greater transparency in the architectures underpinning the agentic AI stack will be essential to the development of effective safeguards, oversight mechanisms, and conditions for trustworthy innovation. At the same time, policymakers will require more sophisticated typologies capable of distinguishing systems according to specific factors like degree of autonomy, adaptiveness, level of tool access, and capacity to shape or intervene within physical and virtual environments.

Getty/Unsplash

GOVERNANCE CHALLENGES

Visibility into AI Agents

As AI becomes widespread in commercial, scientific, governmental, and public sectors, the question of who is able to observe, scrutinize, and understand their operation, and under what conditions, becomes foundational to any meaningful regime of oversight and accountability. The concept of visibility – knowing where, why, by who, and for what an agent is used – must underpin any successful governance structure that can evaluate and assign accountability to agentic AI systems. Proposed solutions include agent identifiers that mark whether and which AI agents are involved in an interaction, real-time behaviour monitoring, and activity logging for agent inputs, outputs, and tool use. These policy instruments come with real tradeoffs: each tool enabling more comprehensive data collection enables better oversight but can also raise personal privacy risks and further concentrate power in the hands of corporations and governments.

The Dilemmas of Delegation

Advanced AI agents have reshaped how people access information, consume goods, and manage their personal lives. These systems can become executors that take direct action in the world, advisors that instruct users on the fulfillment of their goals, and interlocutors that interact with and potentially alter a user's mental state. Because AI agents can be highly personalized, personable, and easy to use, their adoption could spread faster than societies' ability to govern them. Lacking meaningful governance risks widespread reliance on AI agents that can distort markets, disempower consumers, and threaten democratic discourse and information integrity.

Why handing over total control to AI agents would be a huge mistake

As AI agents become more autonomous and capable of acting across multiple applications and digital environments, concerns about oversight, privacy, and accountability have become increasingly urgent. While these agents may offer real convenience and accessibility benefits, they also inherit many of the issues already associated with large language models, even as they gain the ability to independently take action in the world. Because these systems can operate across sensitive platforms and information sources simultaneously, the potential harms associated with mistakes, misuse, or malicious behaviour increase significantly. In response, governance approaches must centre meaningful human oversight, limit the scope of autonomous action, and prioritize transparency and open-source safeguards over opaque, proprietary control.

KEY TAKEAWAYS: AGENTIC AI

- **Agentic AI is a real evolution in AI development.** Unlike LLMs based on single-prompt engagement, agentic systems coordinate multiple tools to independently pursue their own goals, adapt to changing conditions, and take self-directed decisions.

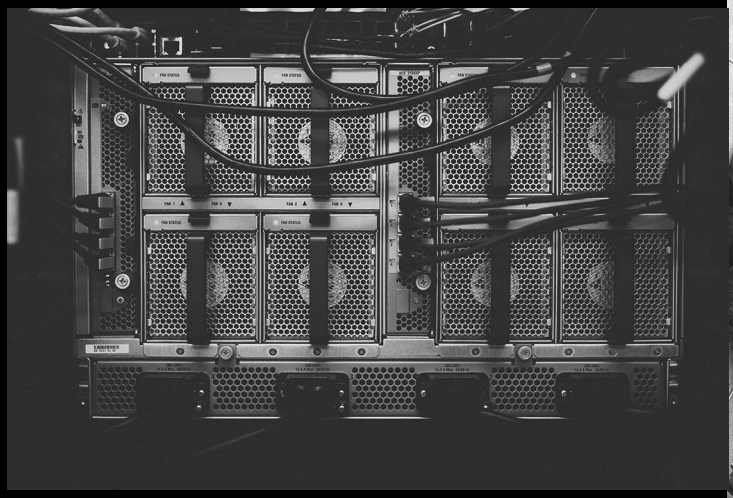
  EXAMPLE: *An Agentic AI tool can book a flight by searching multiple airlines, comparing prices and availability, and making the final purchase without the user needing to intervene throughout the process.*

- **Transparency is a critical component of meaningful oversight.** Governance requires visibility in order to evaluate and assign accountability. Without tools like agent identifiers or activity monitoring, the actions of AI agents, and the potential harms they create, are difficult to trace, discover, or prevent. This challenge will continue to increase as agents act faster, in larger networks, with increasingly complex and diffuse outcomes.

- **Governance currently lacks shared language and analytical precision.** Terms to describe, let alone understand, AI agents are currently used in conflicting ways across research and industry. Conceptual precision is a necessary precondition for meaningful governance—grouping a customer service chatbot and a fully autonomous research agent under the same taxonomy of "AI agent" flattens the need for differences in autonomy, risk, and associated regulation.

PART 3

Governing the Machine



Marc Olivier Jodoin/Scott Rodgerson/Unsplash

What is AI Governance?

AI governance encompasses the frameworks, policies, regulations, and standards that guide the ethical, accountable, and responsible development and deployment of AI. ***Policies*** provide guiding principles, ***regulations*** create legally enforceable rules, and ***standards*** translate these principles into operational practices, with effective enforcement mechanisms ensuring compliance and accountability. Key debates in AI governance revolve around balancing innovation with societal protection, ensuring transparency and human oversight, safeguarding data and consumer rights, and maintaining fairness, inclusivity, and equity.

Julian Gentile/Unsplash

POLICY VS. REGULATION VS. STANDARDS

Policy, Regulation and Standards

Children and AI Design Code

The Impacts and Governance of AI Systems

Building inclusive infrastructure requires more than good intention or written policy. Rather, it depends on effective implementation and enforcement. In this context, policies should provide guidance and direction, regulations should establish legally enforceable rules, and standards should define the norms that ensure consistency across systems.

The Children & AI Design Code illustrates how policy, regulation, and standards interact in AI governance. Policies set guiding principles, regulations make them legally enforceable, and standards provide concrete operational criteria. The Code demonstrates that policies and standards alone are insufficient—enforceable rules and oversight are essential to ensure AI systems are fair, transparent, safe, and inclusive.

AI governance depends on how policies, regulations, and standards work together. Policies, like the Guide on the Use of Generative AI, are flexible and help set ethical guidelines, but they are not enforceable. Regulations, like Canada's Online News Act, or Bill C-18, create legal rules and accountability, though they often struggle to keep up with fast technological change. Standards from the Standards Council of Canada, for example, turn principles into technical practices, but can sometimes shift power toward private industry and away from public oversight.

Annie Spratt/Unsplash

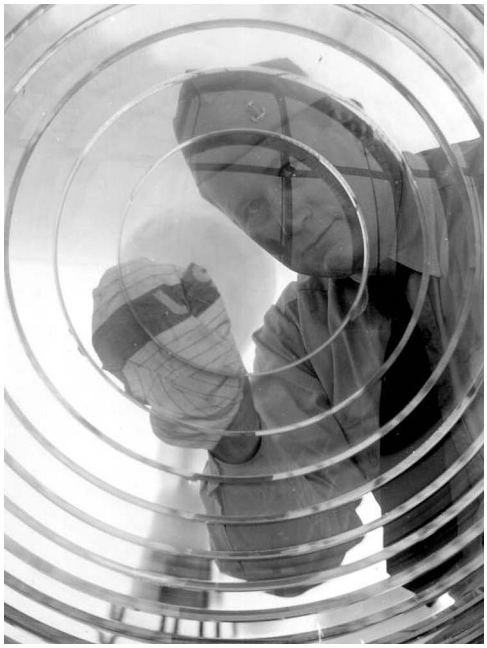
RISK-BASED REGULATIONS

Framework Convention on Global AI Challenges

Advanced AI poses both global benefits and risks, underscoring the need for risk-based regulation that ensures safety without stifling innovation. Key policy considerations include classifying AI systems by potential harm, with stricter oversight for high-risk applications like autonomous weapons, critical infrastructure, or systems affecting global security. Risk-based frameworks should promote international coordination to standardize safety protocols, share best practices, and ensure accountability. Policies must also balance innovation with equitable access, while adopting proactive, adaptable governance that anticipates emerging risks and integrates diverse stakeholder input for monitoring and mitigation.

Existential Risk and Rapid Technological Change

Rapid technological change often outpaces governance, leaving high-impact risks undermanaged, while market pressures to attract investment can drive deregulation, creating tension between economic incentives and public safety. Key debates focus on balancing innovation with social protection, applying the precautionary principle to uncertain but potentially catastrophic risks, and promoting international coordination to harmonize standards.



Getty/Unsplash

TRANSPARENCY/ DISCLOSURE OBLIGATIONS

Enforcement Design Patterns in EU Law: An Analysis of the AI Act

Transparency/disclosure obligations are legal requirements for developers and deployers of AI to provide clear, accessible, and timely information about their systems' design, purpose, functionality, and risk management. These obligations can enable national authorities, oversight bodies, and users to understand system operations, detect potential harms, and verify compliance with human rights, safety, and ethical standards. Transparency ensures that AI systems are auditable and accountable, allowing enforcement bodies to evaluate whether risk mitigation measures are sufficient and aligned with legal and technical standards.

Types of Platform Transparency: An Analysis of Discourse Around Transparency and Global Digital Platforms

Transparency in digital platforms refers to the obligation to disclose how systems operate, make decisions, and affect users. This can take several forms: technical transparency, which explains how algorithms function; procedural transparency, which outlines decision-making processes; and accountability transparency, which ensures that platforms can be held responsible for their actions. Social media platforms often cite commercial interests as reasons to limit disclosure. When disclosure is selective or incomplete, however, it can undermine trust and reinforce existing power imbalances between corporations and users.

Getty/Unsplash

‘HUMANS IN THE LOOP’ AND HUMAN OVERSIGHT

Humans in the Loop: The Design of Interactive AI Systems

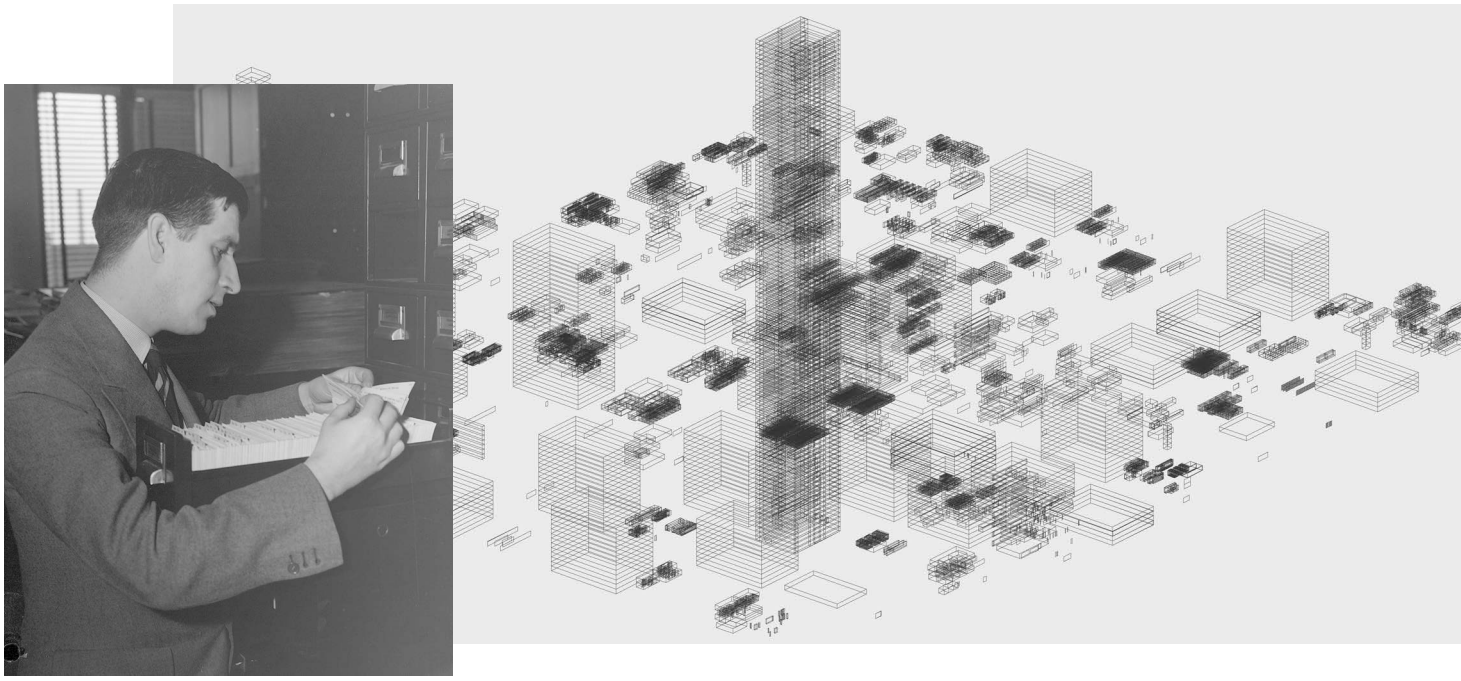
Humans in the Loop

Design Justice: Community-Led Practices to Build the Worlds We Need

“Humans in the loop” and human oversight refers to the design of AI systems so that humans are actively involved throughout their operation and decision-making. Humans can guide AI outputs, provide feedback, and maintain control over AI decisions, making them more collaborative and interactive. This approach also emphasizes transparency and explainability, allowing users to understand how AI systems produce results and provide meaningful input.

Human oversight in algorithmic systems is often treated superficially, with policymakers assuming that adding a human automatically ensures accountability or safety. In reality, such oversight often fails because it does not consider the complex interactions between humans and AI, including cognitive limitations, workflow constraints, and the risk of over-reliance on or misinterpretation of automated outputs.

By centering community voices and lived experience, human oversight challenges top-down, market-driven design models where decisions are made without accountability to affected populations. In design justice, debates focus on who controls technology, whose knowledge is valued, and how AI systems can reinforce or mitigate inequities such as racism, ableism, and cis-normativity. Embedding humans in the loop ensures oversight, meaningful participation, and social accountability, reframing design as a site of collective governance rather than purely technical innovation.



BAnQ; SNCF

DATA GOVERNANCE

Data Governance, Privacy, and Ethics

Data governance refers to the structured policies, roles, and procedures an organization uses to manage its data responsibly, securely, and ethically. It involves defining who is responsible for data, documenting how data flows within the organization, setting standards and policies for handling data, and implementing controls to protect it. Data governance also addresses ethical concerns such as obtaining consent, being transparent about data use, and preventing misuse, while ensuring compliance with legal frameworks. It requires ongoing oversight, training, and iterative improvement to maintain the integrity, privacy, and ethical use of data in organizational and societal contexts.

Governing Artificial Intelligence Means Governing Data: (re)Setting the Agenda for Data Justice

Key policy considerations about data governance focus on creating frameworks that ensure ethical, accountable, and transparent management of data. Policies should protect individual and collective rights, address bias in automated decision-making, and clarify roles and compliance mechanisms. Alternative governance approaches like data trusts, cooperatives, and recognition of Indigenous data sovereignty may also redistribute power and safeguard vulnerable communities.

CONTENT MODERATION

Behind the Screen: Content Moderation in the Shadows of Social Media

Content moderation is the commercial, human labour of evaluating and managing user-generated content on social media platforms to ensure compliance with community standards and protect brand reputation. The term Commercial Content Moderation highlights how this work, often outsourced to low-wage workers in the Global South, is essential yet largely invisible. Moderators review text, images, and videos for content such as hate speech, violence, or explicit material, requiring judgment and cultural awareness that algorithms cannot provide. Moderation can elicit emotional and psychological strain, expose workers to harmful content and labour precarity, and boost corporate control over what is visible online. In this context, content moderation is both a technological and social process that shapes public discourse while concealing the human effort sustaining it.

The Human Labour of Moderation

Scholars and practitioners debate whether social media platforms should be treated as neutral intermediaries or as publishers accountable for the content they host. This includes discussions about how moderation shapes public discourse, influences political debates, and enforces cultural norms. Market debates focus on the commercial incentives of platforms: moderation is necessary to protect advertisers, brand reputation, and user trust, but these economic pressures may conflict with broader social responsibilities. Additional debates involve transparency, accountability, bias in enforcement, and the global inequalities created by outsourcing moderation labour to low-wage regions. Overall, content moderation is framed as both a political and economic practice with wide-reaching implications for power, visibility, and speech online.

Networked Platform Governance: The Construction of the Democratic Platform

The key policy implications of content moderation involve accountability, transparency, and the distribution of power between platforms and stakeholders. Platforms are increasingly adopting a model of networked governance, involving external actors such as civil society organizations, experts, and governments in content policy-making. While this approach is framed as democratic and participatory, platforms maintain central control, shaping which voices are heard and how decisions are made. This raises critical questions for policy regarding the legitimacy of stakeholder involvement, the effectiveness of oversight, and the broader impacts on public discourse. Policymakers must ensure that governance structures are transparent, that roles of external actors are clearly defined, and that platform authority does not obscure accountability.

BAnQ; Johnathan Cooper/Unsplash

ENFORCEMENT MECHANISMS

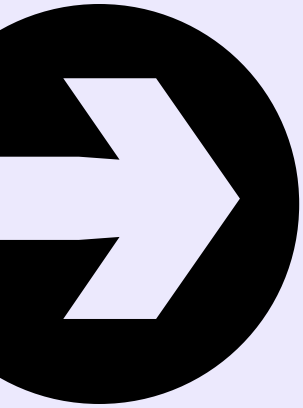
AI Regulatory Enforcement Around the World

AI enforcement mechanisms are the legal powers, institutional structures, and procedural tools that governments use to monitor, investigate, and ensure compliance with AI regulations. They include audits, investigations, fines, public disclosure orders, and, in some cases, the deletion of unlawfully trained models. In the U.S., for instance, the Federal Trade Commission applies consumer protection laws to penalize unfair or deceptive AI practices, while the EU's AI Act empowers national authorities to certify and sanction high-risk systems. While it did not pass, Canada's proposed Artificial Intelligence and Data Act (AIDA) would have authorized an AI and Data Commissioner to conduct audits and issue penalties. China enforces compliance through multiple agencies with powers to suspend or fine companies. Together, these mechanisms operationalize AI governance by translating abstract principles like accountability, fairness, and transparency into enforceable actions.

Central to debates on AI enforcement mechanisms is the balance between fostering innovation and ensuring regulation protects society. In the U.S., stakeholders worry that overly strict enforcement could slow technological development and reduce global competitiveness, while the EU prioritizes strong oversight to prevent bias, discrimination, and unsafe AI deployment. Market actors express concerns about compliance costs, liability risks, and inconsistencies across jurisdictions.

Governing with Artificial Intelligence: Are Governments Ready?

Key policy implications surrounding enforcement mechanisms include the need for clear frameworks that define accountability, oversight responsibilities, and consequences for non-compliance. Effective enforcement requires monitoring and auditing to detect errors, bias, or misuse, as well as building institutional capacity to apply regulations consistently. Transparency, stakeholder engagement, and adherence to ethical guidelines are also critical to maintaining public trust. Practical experiences from jurisdictions worldwide highlight that enforcement mechanisms must be adaptable and well-resourced to address challenges posed by rapidly evolving AI technologies.



KEY TAKEAWAYS: WHAT IS AI GOVERNANCE?

- **Policies set guiding principles, regulations make them enforceable, and standards operationalize principles.** Effective AI governance requires these tools to work together to ensure accountability, fairness, and inclusivity.
- **Risk-based, transparent, and human-centred governance is important.** Risk-based regulation prioritizes oversight based on potential harm, transparency and disclosure obligations enable monitoring and accountability, and human-in-the-loop approaches ensure meaningful oversight and social accountability.
- **Clear frameworks, institutional capacity, audits, and stakeholder engagement are essential to operationalize AI governance.** These frameworks translate ethical principles into actionable, enforceable practices that maintain public trust while balancing innovation and societal protection.

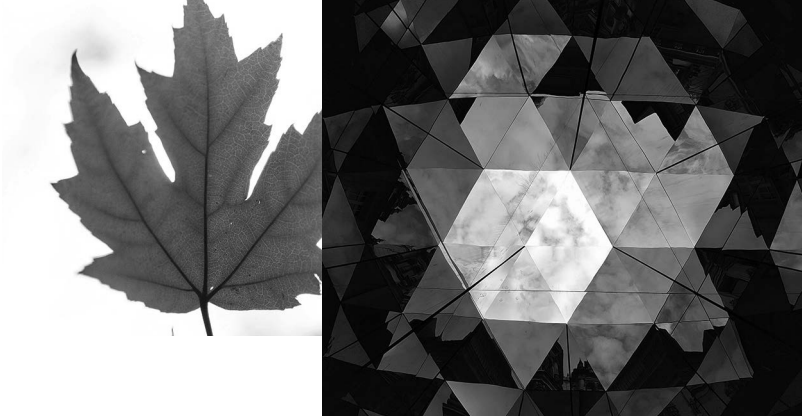
PART 4

AI at Home: Canadian Governance

Canadian Governance

Canada's approach to artificial intelligence governance has developed gradually through a mix of research investments, policy directives, and legislative proposals. The 2017 Pan-Canadian AI Strategy marked the first national effort to strengthen AI research and training through partnership with institutions like Mila, Amii, and the Vector Institute. Since then, initiatives like the Directive on Automated Decision-Making (2019) and the Digital Charter (2020) have established early frameworks for transparency, privacy, and responsible AI use in government and industry. More recent measures, including Canada's new national AI strategy, AI for All, and complementary proposed legislation in Bills C-34 and C-36, reflect an evolving attempt to manage the emerging risks that AI poses in Canada. Taken together, these policies illustrate a fragmented but growing governance landscape.

Previous page: Hamid Saiddiqi/Michael Dziedzic/Unsplash; Getty/Unsplash



Cole Keister/Neo Lee/Unsplash

CANADIAN GOVERNANCE

Canada's National Artificial Intelligence Strategy: AI for All

Launched in June 2026, AI for All is Canada's new national artificial intelligence strategy. It is organized around three guiding principles: 1) building trust; 2) creating opportunity; and 3) reinforcing Canadian sovereignty, and combines measures to accelerate AI adoption with new governance frameworks intended to address the technology's risks. To regulate AI, the strategy relies on complementary legislation, including Bill C-34, which, if passed, will establish new online safety obligations for social media platforms and AI chatbot services, and Bill C-36, which modernizes Canada's private-sector privacy law. It also commits to expanding the Canadian AI Safety Institute, introducing AI transparency measures, investing in sovereign AI infrastructure, and supporting AI adoption across the Canadian economy.

Pan-Canadian AI Strategy

Canada launched its first national AI plan in 2017—the Pan-Canadian AI Strategy—to boost AI research, training, and adoption across the country. Supported by $568 million in federal funding, the strategy connects governments, universities, and industry to grow Canada's AI ecosystem. Its three pillars, 1) commercialization, 2) standards, and 3) talent, aim to translate cutting-edge research into real-world applications. The strategy's core strength comes from its research and talent foundation, led by CIFAR, in partnership with Canada's three national AI institutes: Amii (Edmonton), Mila–Quebec AI Institute (Montreal), and the Vector Institute (Toronto).

Directive on Automated Decision-Making

Canada's first major AI policy, the Directive on Automated Decision-Making, was introduced in 2019 to guide how the federal government uses AI systems responsibly. Enforced in 2020, it ensures transparency, accountability, and fairness in automated decisions that affect Canadians. The directive requires that government departments using automated decision-making systems assess algorithmic risks, explain how AI decisions are made, test for bias, and audit those systems, especially if/when using proprietary software. Its goal is to make AI use in government both safe and trustworthy.

Digital Charter and **Digital Charter Implementation Act**

Introduced in 2019, Canada's Digital Charter laid out guiding principles for building trust in the digital age by emphasizing data privacy, online safety, universal internet access, and the ethical use of AI. To legislate these principles, the government proposed Bill C-11, the proposed Digital Charter Implementation Act, in 2022. The bill aimed to update outdated privacy rules by giving Canadians more control over their personal data, introducing tougher penalties for misuse, and creating a new tribunal to enforce digital rights.

Artificial Intelligence and Data Act

Introduced in 2022 as part of Bill C-27, the Artificial Intelligence and Data Act (AIDA) was Canada's first attempt to directly regulate AI. It aimed to set rules for how AI systems are designed, developed, and used—especially those considered "high-impact"—requiring risk assessments, transparency, and reporting of harms. The act also proposed creating an AI and Data Commissioner to oversee enforcement, though it notably excluded government use of AI from its mandate. In 2023, after public criticism, the government introduced amendments to clarify the bill's definitions, strengthen accountability, and include new rules for generative AI, but the legislation ultimately "died on the order paper" when a federal election was called in Spring 2025.

ISED's Voluntary Code of Conduct on the Responsible Development and Management of Advanced Generative AI Systems

In 2023, Canada introduced a Voluntary Code of Conduct for the responsible development and use of generative AItools like ChatGPT, DALL·E, and Midjourney. Created by Innovation, Science and Economic Development (ISED), the Code encourages companies to go beyond legal requirements by adopting strong accountability, transparency, and risk management practices. It focuses on preventing misuse, protecting against cyber threats, and documenting how AI systems are built and deployed. Organizations that sign on also commit to supporting a broader responsible AI ecosystem through collaboration, best-practice sharing, and public education.

TBS Guide on the Use of Generative AI

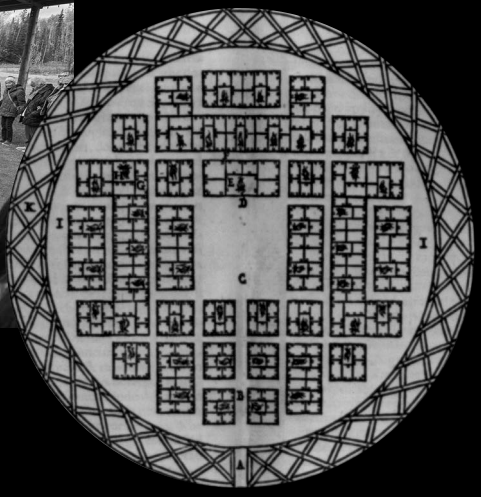
Also released in 2023, the Treasury Board of Canada's (TBS) Guide on the Use of Generative AI offers direction for how federal departments should responsibly use AI tools. Aimed at building public trust, it lays out six core principles: fairness, accountability, security, transparency, education, and relevance. The guide also shares practical advice for public servants, including how to protect information, reduce bias, ensure quality, manage legal and environmental risks, and clearly distinguish human work from machine output.

KEY TAKEAWAYS: AI AT HOME — CANADIAN GOVERNANCE

- **When it comes to AI, governments are learning alongside the public.** Even though it feels like AI is everywhere, most governments, including Canada's, are still figuring out how to regulate it responsibly. That's why recent governance initiatives are about building trust first.
- **AI Innovation is important, but it must be accompanied by guardrails that prevent harm, bias, and exploitation.** That's why most of Canada's directives and proposed policies aim to balance progress with protection. The Digital Charter and its related bills push for strong privacy rights and responsible data use, while newer initiatives like the Voluntary Code of Conduct focus on risk management for companies from powerful AI models like ChatGPT, DALL·E, and Midjourney.
- **Trust and accountability are the backbone of Canada's AI future.** Across every policy, the goal is to make AI use open, traceable, and accountable. That means testing systems for bias, explaining how decisions are made, and being honest bout how AI is used in official or public settings.

PART 5

Alternative AI Visions



Indigenous knowledge systems offer other ways of thinking about memory, corporate and political organization.

The First Nations village of Hochelaga on the site of what is now Montreal.

Decolonizing AI

As we've seen throughout this guide, AI systems—and the complex structures, physical and non-physical, that underpin them—are not neutral. They are shaped by the political, economic, and cultural systems in which they are built, and in some cases may risk carrying forward patterns of harm present in those systems. A growing body of work calls for the decolonization of AI: questioning, dismantling, renovating, or moving well beyond elements of the field. In Canada and elsewhere around the world, Indigenous scholars, technologists, organizations and community leaders are leading the charge, inviting more plural and relational approaches to the imagination of our collective technological futures. This leadership demands our attention.

Previous page: NASA/Tony Ityak/Unsplash/ Province of British Columbia; Library of Congress

Casey Horner/Unsplash

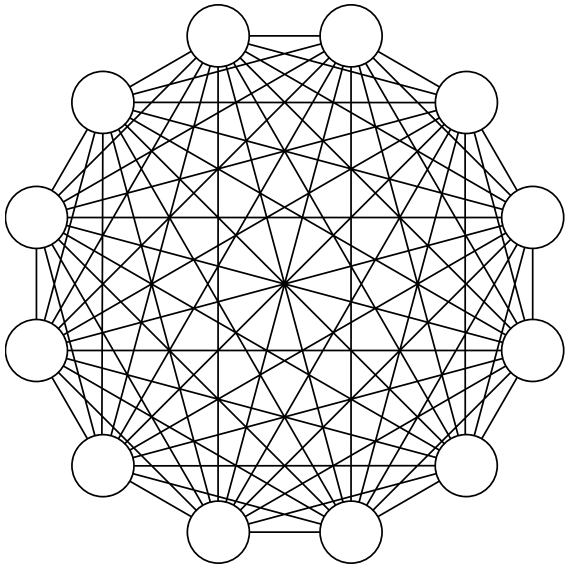
TOWARD ALTERNATIVE 'INTELLIGENCES' AND INDIGENOUS DATA SOVEREIGNTY

Why AI Must be Decolonized to Fulfill its True Potential

Artificial intelligence, largely developed in Western corporate contexts, risks reproducing the extractive and discriminatory logics of colonialism. From genomic research dominated by European data, to underpaid "ghost work" in the Global South, to biased public-sector algorithms built on incomplete statistics, AI development often mirrors older systems of exploitation and exclusion. Decolonizing AI is not about rejecting the technology altogether but is rather about diversifying perspectives and redistributing power. Such an approach is not only crucial to addressing historical injustice, but to guaranteeing a better AI future for all.

Abundant Intelligences: Placing AI Within Indigenous Knowledge Frameworks

Abundant Intelligences is an Indigenous-led, interdisciplinary research program that suggests that the failure to consider non-western rationalist approaches in existing AI research and development has resulted in the reproduction of colonial harms, including exploitation, inequality, extraction, and other forms of violence. Indigenous epistemologies and protocols to research on AI aim to cultivate a relational approach rooted in regeneration, generosity, and reciprocity, moving beyond narrow conceptions of "intelligence."

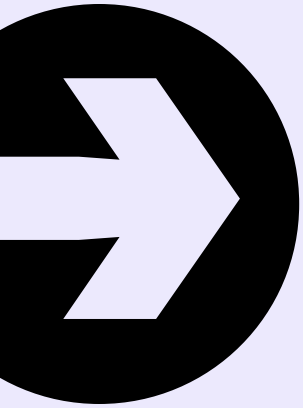


The CARE Principles for Indigenous Data Governance

The CARE Principles for Indigenous Data Governance are: Collective Benefit, Authority to Control, Responsibility, and Ethics. Intended as a complement to the "data-centric" FAIR Principles (Findable, Accessible, Interoperable, Reusable), which emphasize open data sharing without sufficient attention to contextual differences and power asymmetries, the CARE Principles foreground Indigenous self-determination and innovation, centering equitable data governance for data producers, stewards, and publishers. The joint implementation of CARE and FAIR Principles seeks to combine a focus on the open data responsibilities of producers and repositories with an understanding of historical inequities and a commitment to Indigenous rights, governance, and control.

The First Nations Principles of OCAP®

The First Nations Principles of OCAP® are: Ownership, Control, Access, and Possession. They establish that First Nations have the right to govern how information about their peoples and communities is collected, stored, protected, used, and shared. Ownership recognizes a Nation's collective relationship to its data and knowledge; Control affirms its authority over research and information-management processes; Access ensures that First Nations can obtain and make decisions about information concerning them; and Possession refers to the physical stewardship of data as a practical means of protecting those rights. These principles support First Nations data sovereignty and respond to histories of externally imposed research, extraction, and limited community benefit. Its application is not one-size-fits-all, but is determined by each Nation in accordance with its own laws, protocols, priorities, and worldview. OCAP® is a registered trademark of the First Nations Information Governance Centre (FNIGC).



KEY TAKEAWAYS: DECOLONIZING AI

- **AI risks repeating colonial patterns of harm.** From narrow research agendas, to exploitative cross-border labour practices, and datasets that overlook marginalized populations, AI's development, deployment, and supply chains can quietly reproduce older systems of extraction and inequality. In this context, serious consideration of decolonization is crucial.
- **Different worldviews open the door to a wider range of technological and social futures.** Today's AI systems reflect a specific view of the world, one grounded in probability, prediction, rationalism. Other knowledge systems, including Indigenous epistemologies, may allow us to understand 'intelligence' differently, or to imagine alternative visions of the future. Curiosity in this domain can help us answer the most critical question of all: *what do we want our machines to do for us*?
- **One area of particular importance for decolonization is data—its ownership, use, and governance.** How data is collected, shared, and controlled determines whose knowledge counts, whose interests are served, and who is represented in technological development.

Andrej Lisakov/Unsplash

Participatory AI Governance

The question of who gets a say in how AI systems are designed, deployed, and governed is growing ever more urgent. While governments and corporations often gesture toward inclusivity, genuine public participation in AI remains rare. Where it exists, it is frequently superficial or performative. Across an expanding field of research and practice, scholars and advocates are calling for a shift from consultations that treat citizens as consumers to engagement processes with real decision-making power and enforcement. As power over AI research, development, infrastructure and associated markets becomes increasingly concentrated, this call is more important than ever.

Wikipedia (Roman voting); Metropolitian Museum of Art (Congolese loudspeaker)

BRINGING THE PUBLIC ALONG: REAL ENGAGEMENT VS. PARTICIPATION-WASHING

Who are the publics engaging in AI?

When it comes to public consultation processes for AI governance, who counts as "the public" is not as straightforward as it may appear. Terms such as citizen, consumer, stakeholder, participant, and "human-in-the-loop" each carry assumptions about whose knowledge or interests matter, and therefore who is included in the process and what role they play in it. Labels meant to define participants can flatten their socio-demographic characteristics, erase power asymmetries that may exist within them, and exclude marginalized groups from engagement all together. As such, defining "the public" actually shapes who is invited into decision-making, whose voices are recognized, and how power is distributed in the governance of AI.

Democratizing AI: Principles for Meaningful Public Participation

Given AI's technical faults and biased outputs, the distance between its creators and the communities it affects, and the harms it can cause, meaningful public participation is essential to its governance. Traditional government consultation processes tend to be largely performative and unrepresentative, and corporate ethics initiatives are disconnected from the communities most affected by their technologies. To achieve meaningful participation in AI, participation should be embedded throughout its development lifecycle; equity and social justice should be centred in public engagement processes; and non-technical expertise should be valued.

Metropolitan Museum of Art (Buddhist assembly); Rijksmuseum (debate)

Beyond the Façade: Challenging and Evaluating the Meaning of Participation in AI Governance

"Participation-washing" describes engagement exercises that appear inclusive but ultimately lack real impact. These initiatives may include public consultations, audits, or stakeholder forums, often collect feedback and criticism but fail to redistribute power or influence decisions, creating only the appearance of accountability. This dynamic reflects a broader problem of corporate capture, where the private sector dominates governance processes and uses participation mainly to signal responsibility rather than to enable structural change. To move beyond this, experts argue that participation must be formally built into governance frameworks and supported by enforceable mechanisms.

The Participatory Turn in AI Design: Theoretical Foundations and the Current State of Practice

The "participatory turn" in AI refers to a growing push to involve affected communities in the design of artificial intelligence systems. Analysis of the field, however, reveals that most "participatory" efforts remain consultative rather than empowering: stakeholders are typically asked for input on interface design or preferences, not provided with any ownership or direct agency over the design process. Due to various constraints, many projects rely on proxies—including algorithmic 'stand-ins' for the public—which risk reproducing existing power imbalances.

KEY TAKEAWAYS: PARTICIPATORY AI

- **AI has a participation problem.** Even as the public hears constantly (and often in vague terms) about AI's imminent and considerable impact on peoples' lives, channels for meaningful public participation in design and governance of AI systems are limited.
- **Where it exists, participation is often performative.** Corporate and government processes alike risk reinforcing "participation-washing," where engagement is used to signal responsibility rather than share it.
- **Genuine participation can strengthen AI governance.** When designed well, participatory mechanisms can improve accountability, legitimacy, and trust. They can also surface crucial, context-specific knowledge from affected communities, helping policymakers and developers work toward more equitable technologies.
- **Promises made must be promises kept.** Participation builds trust when the aims of a process are clearly and honestly communicated, followed-up on, and measured clearly.

Glossary of Key Terms

Agentic AI

AI systems that pursue complex goals with minimal human supervision. A common example is self-driving cars, which demonstrate both the potential and the limits of agentic AI: they can operate largely independently but still rely on human oversight in high-stakes or unpredictable situations.

AI Doomerism

The belief that advanced AI poses an existential threat to humanity. Often fueled by sensationalism and promoted by Big Tech, doomerism can distract from urgent, present-day issues around responsible AI development.

AI Hallucination

Instances where large language models produce incorrect, misleading, or nonsensical outputs due to statistical errors or biases, instead of signalling uncertainty about their results.

AI Slop

A term popular in online communities for low-quality or unwanted AI-generated content (i.e., text, images, or other media) often mass-produced and found across social media and search results.

Algorithms

A set of step-by-step computational instructions designed to solve a problem or produce a specific outcome. In computer systems, algorithms provide the rules a machine follows to process input and generate results.

Alignment

The process of designing AI systems so that their goals, actions, and outcomes remain consistent with human values and intentions.

Artificial General Intelligence (AGI)

A theoretical form of AI that could understand, learn, and apply intelligence across any task that humans can perform. Often linked to notions of machine sentience, consciousness, or human-like reasoning, AGI remains largely a concept explored in science fiction.

Bias

An inherent, often subconscious, inclination or prejudice in human judgment that shapes outlooks and decisions. Since technology is created and applied by people, human bias inevitably carries over into the design, development, and use of AI.

Big Tech

A collective term for the world's most powerful technology companies (i.e., Meta, Amazon, Alphabet), whose vast computing resources and data access have shaped, and continue to shape both the direction and nature of artificial intelligence development.

Chatbots

Computer programs that simulate human conversation. Modern chatbots often rely on conversational AI and natural language processing (NLP) to understand user input and generate automated responses.

Compute

Short for 'computing power,' it refers to the combination of hardware, software, and infrastructure that powers AI systems, enabling them to process large volumes of data, perform complex calculations, and make informed decisions.

Content Moderation

The set of governance practices that structure participation in online spaces, aiming to encourage cooperation while preventing abuse. Responsibility is typically shared between platforms and government bodies.

Data Mining

The large-scale collection and analysis of personal data, often by technology companies and social media platforms, to facilitate the commodification of user activity and generate profit.

Data Poisoning

A cyberattack that manipulates the training data of machine learning models to skew their processes and outcomes.

Deepfakes

Synthetic audiovisual content generated with machine learning that makes real people appear to say or do things they never did. Deepfakes are increasingly realistic and challenging to detect.

Digital Literacy

The critical ability to understand, navigate, and participate in the digital world and its connections to everyday offline life, developed through education and practical experience.

Digital Rights

Protections, freedoms, and rules in digital environments that safeguard users' interests. They may take the form of laws, regulations, platform policies, or community standards, much like how nation-states protect rights in the physical world.

Digital Sovereignty

Encompasses the capacity for individuals, organizations, or states to control their digital assets, data, and activities, ensuring independence and self-determination in the digital sphere.

Disinformation

False content created or shared with the intent to cause harm, whether online or in person.

Explainability

The extent to which an AI system's processes and decisions can be made transparent, understandable, and traceable to humans, fostering trust, accountability, and the ability to review or contest results.

Generative AI

A branch of artificial intelligence that uses large language models and deep learning to create new content, such as images, audio, video, or code, based on user prompts.

Guardrails

Originally referring to physical barriers that prevent accidents, in AI, they are built-in safeguards designed to ensure systems operate accurately, safely, ethically, and legally.

Human in the Loop

An approach to AI and automation that deliberately integrates human input to guide, improve, or correct outputs. This framework positions AI as a complement to human judgment rather than a replacement for it.

Interoperability

Interoperability is the ability of different technological systems to communicate, share data, and work together seamlessly toward a common goal. It minimizes data friction, aligns information systems, and prevents digital fragmentation, creating the conditions for advanced tools like AI and machine learning to function effectively.

Large Language Models (LLMs)

AI foundation models trained on massive datasets, enabling them to understand and generate natural language as well as perform a broad range of other content-based tasks.

Machine Learning

A system's ability to improve its performance on programmed or routine tasks over time. Originating in computer science, the term can be misleading outside the field, as "learning" is metaphorical and does not imply sentience or agency in computers.

Misinformation

False content shared by someone unaware of its inaccuracy, whether online or in person. Unlike disinformation, it is not intentionally spread to cause harm, although it can lead to harmful effects nonetheless.

Natural Language Processing

A branch of AI that enables conversational interaction with software, powering tools such as chatbots, translation systems, virtual assistants, search engines, and spam filters.

Neural Networks

Machine learning models that are structured to mimic how neurons work together in the human brain. They process information, weigh options, and make decisions in ways that resemble human cognitive functions.

Open Source

Software whose source code is publicly available, allowing anyone to access, modify, and share it.

Personalization

Refers to AI's ability to generate user data profiles that enable tailored messaging, product experiences, and marketing content, helping organizations optimize communication and increase engagement with users and potential customers.

Predictive Analytics

A branch of data analysis, central to narrow AI applications, that uses statistical methods, machine learning, and algorithms to detect patterns in historical data to forecast future outcomes.

Privacy by Design

A framework developed by Ann Cavoukian that embeds privacy into technologies and systems from the outset rather than as an afterthought. Its seven principles include: proactive measures, privacy as default, privacy built into design, full functionality, end-to-end security, visibility and transparency, and user-centric respect for privacy.

Red Teaming

A cybersecurity practice where ethical hackers simulate authorized attacks on a network, system, or tool to identify vulnerabilities, test defences, and improve security operations.

Reinforcement Learning from Human Feedback (RLHF)

A method for training AI systems that incorporates human feedback to guide and improve model behaviour. Widely used in tools like ChatGPT, RLHF involves training a model, collecting human preference data, and refining the model to align more closely with those preferences.

Risk Assessment

The process of evaluating AI systems to identify potential harms, anticipate bias or gaps, and mitigate unintended consequences, based on the assumption that AI carries inherent risks. Independent risk assessments are most effective at promoting impartiality and accountability.

Synthetic Data

Artificially generated data that mimics the useful characteristics of real user data. It can be used to train or test machine learning models and software applications without exposing personal information, while also helping to fill gaps in datasets or replace outdated or unusable historical data.

Targeted Advertising

A marketing strategy that delivers ads to specific individuals or groups based on personal data. Widely used by digital platforms, it often aims to maximize user engagement and screen time.

Token

In large language models, a token is a small piece of text - such as a word, part of a word, a punctuation mark, or other character sequence - that the system processes when generating or interpreting language. It is worth noting that because many tokenization systems are optimized primarily for English, the same passage can require substantially more tokens in other languages, potentially increasing cost and reducing how much text can fit within a model's context window.

Turing Test

A test measuring whether a machine can exhibit human-like intelligence by engaging in conversation indistinguishable from that of a human, first proposed by Alan Turing in 1950.