



OFFICE OF THE  
INFORMATION &  
PRIVACY COMMISSIONER  
FOR BRITISH COLUMBIA



# OpenAI Investigation

## Report of Findings

The privacy commissioners for British Columbia, Canada, Quebec, and Alberta jointly investigated OpenAI's privacy compliance when developing and deploying early ChatGPT models (GPT-3.5 and GPT-4).

Among other issues, the investigation examined whether OpenAI collected, used, and disclosed the personal information of individuals in Canada:

- for an appropriate purpose and limited to this purpose; and
- with meaningful consent.

### THE OIPC

Established in 1993, the Office of the Information & Privacy Commissioner provides independent oversight of BC's access and privacy laws.

### BC LEGISLATION

The *Personal Information Protection Act* applies to any private sector organization that collects, uses and discloses the personal information of individuals in BC.

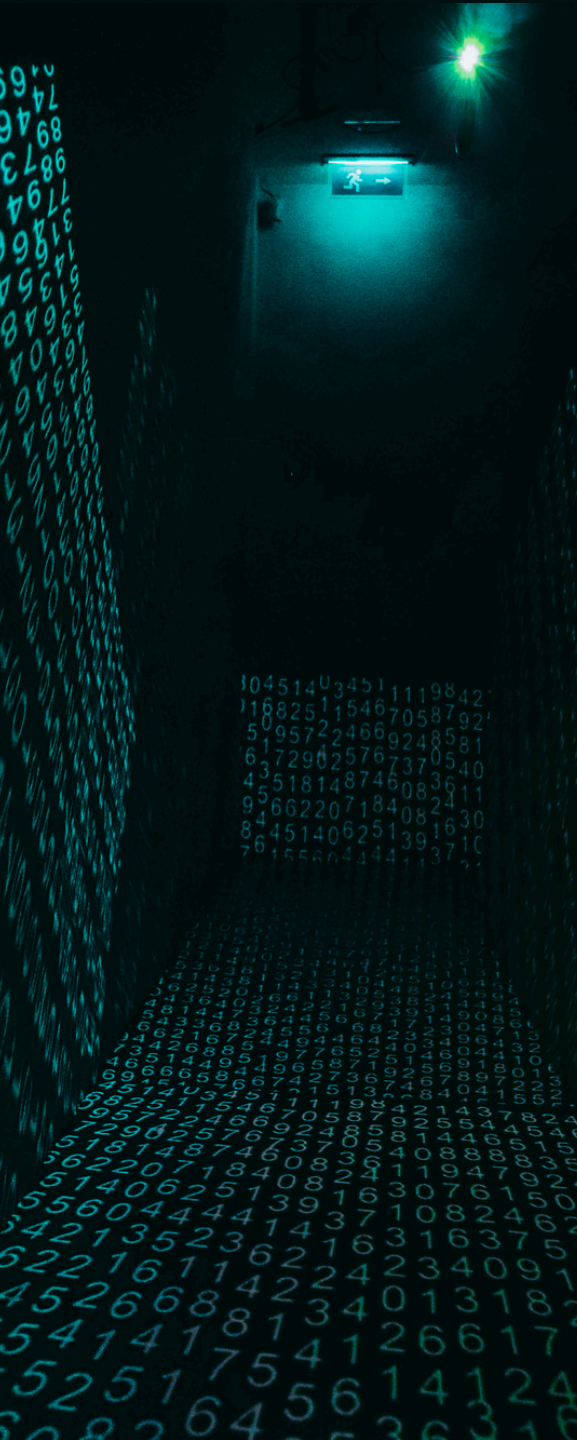
**READ MORE**



Check out *Report of Findings: Joint investigation of OpenAI OpCo, LLC* to learn more.

<https://www.oipc.bc.ca/reports/investigation-and-audit-reports/>

# Why are we concerned?



**OpenAI trained its models using personal information collected from a variety of sources, including:**

- Publicly accessible internet sources;
- licensed third party sources; and
- users' interactions with ChatGPT.

The report found that OpenAI collected this information without consent and was overly broad and inappropriate.

Further, OpenAI did not fulfill its obligations with respect to openness, accuracy, access to personal information, retention and disposal, and accountability.

# Where did ChatGPT get information from?

AI systems\* are trained by collecting massive amounts of information, including personal information you publish online, from data sources on the internet, including:

- YouTube videos and other audio or visual media content
- Forums or blog posts
- Product reviews
- User comments
- User interactions with AI, including in search engines
- Social media & professional networks
- Marketplace & e-commerce platforms
- News or other media outlets
- Essays, articles, or books
- Job boards

## Why does it matter?

- AI can use your information to identify you or predict your behaviour
- Many AI systems are “black boxes”, making their processes hard to understand
- Information is often collected without your knowledge or consent and used in ways you didn’t intend
- Collected information could be sensitive, inaccurate, and include children’s personal information

## What can you do?

Check out the infographic **Protect your privacy when using AI tools**, for simple steps to help safeguard your personal information when using AI tools. <https://www.oipc.bc.ca/resources/infographics/>

\*not limited to ChatGPT

# Was OpenAI's collection, use, and disclosure appropriate and limited?

**No.** The way OpenAI initially collected personal information from **publicly accessible websites and licensed third-party sources** to train its early models was **overly broad and inappropriate.**

OpenAI's mitigation measures were **not sufficient** to limit the scope of its collection, use, and disclosure of personal information to that which was necessary and proportional to effectively train its models.

# Did OpenAI get meaningful consent?

**No.** OpenAI failed to get meaningful consent for its collection, use, and disclosure of personal information for the purpose of developing and deploying the models.

- Information used by OpenAI to train its models was scraped from publicly available sources, such as social media and discussion forums
- Express consent is required if personal information is sensitive or beyond what a reasonable person would consider appropriate. Individuals who had their information scraped would not have reasonably expected that information about them on the internet could be collected and used by OpenAI to train its models. This practice was novel and not widely understood at the time.
- Further, users were not informed that personal information included in conversations with ChatGPT could be used for the model training.

**Web “scraping”** is the automated extraction of data from the internet, particularly from websites and social media, which can involve large volumes of personal information

# Recommendations

**The Commissioners made 10 recommendations, that focus on:**

- Limiting collection, use, and disclosure of personal information
- Obtaining valid consent from individuals before collecting, using, or disclosing personal information
- Informing users of the potential consequences of disclosing sensitive information when interacting with ChatGPT
- Publishing plain language, comprehensive, and accessible information about sources of information used to train datasets, how the models function, and how this can lead to inaccurate outputs.

**The OIPC BC and counterparts will monitor OpenAI's implementation of these commitments.**

Even if OpenAI satisfactorily implements all of these recommendations, it won't be compliant with BC's PIPA because the Act is fundamentally incompatible with how large language models work. Commissioner Harvey has written to the BC government urging them to move quickly to start a process to reform PIPA to ensure that the right protections are in place for an innovative information economy that protects privacy for those living in British Columbia.

# A MESSAGE FROM THE COMMISSIONER

The logo for the Office of the Information & Privacy Commissioner for British Columbia, featuring the lowercase letters 'oipc' in a blue, sans-serif font.

OFFICE OF THE  
INFORMATION &  
PRIVACY COMMISSIONER  
FOR BRITISH COLUMBIA

## INVESTIGATION REPORT

### Joint investigation into OpenAI

The logo for the Office of the Information & Privacy Commissioner for British Columbia, featuring the lowercase letters 'oipc' in a white, sans-serif font.

OFFICE OF THE  
INFORMATION &  
PRIVACY COMMISSIONER  
FOR BRITISH COLUMBIA

[oipc.bc.ca](https://oipc.bc.ca)

**WATCH**



Check out the full video overview:  
<https://youtu.be/nL99yxuchmo>

A small icon of a globe with latitude and longitude lines.

[oipc.bc.ca](https://oipc.bc.ca)



[@oipcbc.bsky.social](https://bsky.app/profile/oipcbc.bsky.social)