2025 Women in AI Canada Hackathon Challenge

Join us for the 2025 WAI Canada Hackathon — a virtual innovation sprint running from **October 6 to November 3, 2025, culminating in final presentations and a solution showcase on November 7, 2025.** This year's hackathon challenges you to harness the power of AI for equity, inclusion, and public trust. Whether you're a student, researcher, designer, or entrepreneur, this is your opportunity to build AI-driven solutions that create real-world impact—from personalized learning for neurodiverse communities to systems that combat misinformation and algorithmic bias.

Participants will gain access to curated datasets, expert mentorship, and visibility with top industry and policy leaders. Whether you're prototyping an AI product, building a responsible agent, or designing a concept that could change lives—we invite you to help shape the future of ethical AI.

Theme 1:

"AI for Inclusive, Personalized & Neurodiverse Learning"

We challenge you to build AI systems that empower diverse learners, including those with disabilities, cognitive differences, or language barriers. From accessible learning tools to adaptive civic support systems, this theme is about designing personalized and inclusive technologies that remove structural barriers and reflect lived diversity.

Theme 2:

"Trustworthy AI Systems for Societal Resilience"

In a time of increasing misinformation, institutional mistrust, and digital risk, this theme challenges participants to build AI tools that are transparent, auditable, and fair. Whether you're detecting deepfakes, auditing legal documents, or analyzing algorithmic risk—this track is your opportunity to engineer AI systems that protect truth and uphold justice in high-impact domains like law, media, healthcare, and public policy.

How to participate?

- The **Call for Abstract Submission is open till September 19, 2025.**
- You are required to submit your application, with the details of your group of size 1-3, along with the abstract of the problem and solution you are submitting.

- Your solution must be **plagiarism-free** and aligned with this year's Challenge Themes:
    - **AI for Inclusive, Personalized & Neurodiverse Learning,** and
    - **Trustworthy AI Systems for Societal Resilience**
- If we require you to change your problem or solution, we will contact you via email before the Challenge opens formally on **October 6, 2025.**
- You will be paired with a mentor, to help you guide throughout the development of your proposed solution.
- The Challenge period will be between **October 6 - November 3, 2025,** where you will attend weekly mentoring sessions and relevant technical workshops.
- On **November 7, 2025,** there will be a final Hackathon Ceremony where you will submit and present your use case and solutions to the Jury Members.

The solution will be evaluated based on the following:

- Addressing a relevant problem that is impactful
- Innovation
- AI Element
- Responsible AI
- Feasibility of execution and scalability
- Must be plagiarism-free

Read the Terms and Conditions

Sample Ways to Prepare a Use Case for the Hackathon

Challenges for Theme 1:

Design AI systems that promote educational equity, social inclusion, and accessibility, empowering diverse learners, underserved communities, and individuals with disabilities through personalization and multimodal experiences.

Challenge 1: AI-Enhanced Multimodal Learning Assistant for Neurodiverse Learners

Problem Statement:
Neurodiverse learners (e.g., those with dyslexia, ADHD, or autism) often lack access to personalized educational tools that adapt to their cognitive and emotional needs.

Potential Solutions:

Approach #1 (Traditional AI):
>   **Solution:** NLP-Based Dyslexia Reading Companion
>   **Description:** Build a dyslexia-friendly reader that highlights text, converts it to speech, and simplifies vocabulary in real-time.
>   **Technical Aspects:** Use NLP and TTS technologies to adapt content based on the user's reading pace and comprehension level.

Approach #2 (Agentic AI):
>   **Solution:** Adaptive Multimodal Learning Tutor
>   **Description:** Design an AI tutor that senses learner engagement, adjusts modality (text, visual, audio), and adapts its teaching style dynamically.
>   **Technical Aspects:** Combine engagement detection models with multimodal content delivery to personalize learning experiences.

Why It Matters (WAI lens): Enables equity in education for neurodiverse learners and those often underserved in traditional systems.

Suggested Datasets/APIs:
1. **OpenNeuro ds003126** – fMRI data capturing reading activity in children with dyslexia and spelling deficits. Freely available for download https://openneuro.org/datasets/ds003126/versions/1.3.1
2. **UCI Autism Screening (Children)** – Screening data for autism in children; suitable for modeling neurodivergent traits https://archive.ics.uci.edu/dataset/419/autistic+spectrum+disorder+screening+data+for+children
3. **UCI Autism Screening (Adolescents)** – Similar dataset for adolescents (104 instances, 21 features) https://archive.ics.uci.edu/dataset/420/autistic+spectrum+disorder+screening+data+for+adolescent
4. **UCI Autism Screening (Adults)** – Adult variant (704 instances) for neurodiversity classification archive.ics.uci.edu.

Challenge 2: Personalized Co-Learning Journey for Special Education Needs

Problem Statement:
Students with physical or cognitive impairments require tailored educational support, which is often lacking in standard systems.

Potential Solutions:

Approach #1 (Traditional AI):
> **Solution:** Struggle Prediction Model for Intervention
> **Description:** Predict when students are likely to struggle using engagement and performance data to trigger timely interventions.
> **Technical Aspects:** Supervised learning using student progress logs and exceptionality labels to forecast need for support.

Approach #2 (Agentic AI):
> **Solution:** Personalized Co-Learning Companion
> **Description:** A digital agent that co-learns with students, dynamically adjusting pace, format, and encouragement based on real-time feedback.
> **Technical Aspects:** Reinforcement learning and sensor-based adaptation to maintain motivation and reduce cognitive load.

Why It Matters (WAI lens):
Promotes educational equity for students with disabilities, tailoring support to real human needs.

Suggested Datasets/APIs:
1. **Special Education Enrolment by Exceptionality (Ontario)-** Includes counts of students by type of special education exceptionality (e.g. learning disabilities, physical disability) for elementary and secondary publicly funded schools in Ontario.
   https://data.ontario.ca/dataset/special-education-enrolment-by-exceptionality
2. **Student Headcount by Diverse Abilities (Canada-wide)-** Official Statistics Canada tables showing counts of students across disability categories (physically dependent, communication, learning).
   https://ouvert.canada.ca/data/dataset/1e730ea9-dd19-4c22-aa95-fe644efc7a06
3. **Ontario EQAO Assessment Linked Data-** Inclusion of variables such as special education status, accommodations, self-efficacy and engagement along with literacy and math scores for Grades 3–10.
   https://www.eqao.com/about-eqao/open-data/

Challenge 3: AI for Accessible Urban Mobility

Problem Statement:
Navigating urban environments can be extremely challenging for individuals with visual, mobility, or cognitive impairments.

Potential Solutions:

Approach #1 (Traditional AI):
    Solution: Accessibility-Aware Route Planner
    Description: Build a map-based tool that recommends optimal routes based on real-time infrastructure and accessibility tags.
    Technical Aspects: Use geospatial and transit data with accessibility filters to personalize routes for different impairment types.

Approach #2 (Agentic AI):
    Solution: Context-Aware Navigation Assistant
    Description: A wearable or mobile agent that guides users contextually, adjusting directions based on live environmental changes (e.g., blocked sidewalks).
    Technical Aspects: Real-time sensor and map data fusion with NLP for personalized verbal guidance.

Why It Matters (WAI lens): Bridges a critical accessibility gap in smart cities—supporting daily independence through inclusive design.

Suggested Datasets/APIs:
1. **Open Toronto Data Catalogue-** Includes sidewalk geometry, TTC transit accessibility data, pedestrian crossings, and real-time construction feeds—ideal for building inclusive urban navigation tools in Toronto.
https://open.toronto.ca/catalogue/
2. **Geofabrik – OpenStreetMap North America Extract-** Provides detailed map data with attributes like surface type, kerb height, incline, and wheelchair accessibility tags. Useful for route planning based on terrain and physical constraints.
https://download.geofabrik.de/north-america.html
3. **Mobility Database – Global GTFS & GBFS Feeds-** Aggregates over 2,500 public transportation feeds in GTFS and GBFS formats. Includes real-time updates, vehicle accessibility info, and stop-level metadata for many cities worldwide.
https://mobilitydatabase.org/

Challenge 4: Culturally Adaptive Public Service Assistant
Help immigrants and underrepresented communities understand local services by personalizing and translating civic information.

Problem Statement:
Immigrants and underrepresented communities often face barriers in understanding and accessing essential public services.

Potential Solutions:

Approach #1 (Traditional AI):
>   **Solution:** Multilingual Document Summarizer
>   **Description:** Automatically translate and simplify government documents for non-native speakers.
>   **Technical Aspects:** NLP translation and summarization with tone adjustment models.

Approach #2 (Agentic AI):
>   **Solution:** Culturally-Aware Public Service Copilot
>   **Description:** A conversational AI that adapts content tone, complexity, and dialect based on the user's profile and context.
>   **Technical Aspects:** LLM-based assistant with cultural language models and translation memory.

Why It Matters (WAI lens): Promotes linguistic inclusion and public service equity for underrepresented populations.

Suggested Datasets/APIs:
1. **Open Canada Data Portal-** Provides access to thousands of Canadian government datasets including immigration, healthcare, tax, and civic services. Ideal for training models to summarize and translate public service content. https://open.canada.ca/en/using-open-data
2. **FLORES (Facebook Low Resource Supervised Translation Benchmark)-** A multilingual dataset for machine translation covering 100+ languages and dialects, designed to benchmark translation systems in low-resource settings. https://github.com/facebookresearch/flores
3. **TICO-19: Translation Initiative for Covid-19-** Offers human-translated COVID-19 public health information in 30+ languages. Designed for crisis communication across diverse populations, it can be repurposed for multilingual civic information delivery. https://tico-19.github.io/translators.html
4. **OPUS Global Parallel Corpus-** One of the largest open-source collections of aligned multilingual text from public service domains including government, healthcare, and legal. Supports tone, complexity, and dialect fine-tuning for translation models. http://opus.nlpl.eu/

Theme 2: Trustworthy AI Systems for Societal Resilience

Develop AI tools that validate truth, detect bias, and analyze risks—ensuring safety, transparency, and fairness in high-impact domains like law, media, healthcare, and public services.

Challenge 1: Legal Fairness Auditor & Bias Mitigator

Problem Statement:
Legal contracts and court documents often contain biases or exclusionary language that can perpetuate injustice.

Potential Solutions:

Approach #1 (Traditional AI):
Solution: Legal Bias Detection Engine
Description: Identify biased or exclusionary phrases in legal documents using NLP pattern recognition.
Technical Aspects: Token classification with fairness lexicons and legal datasets.

Approach #2 (Agentic AI):
Solution: Inclusive Contract Rewriter Agent
Description: AI agent that rewrites clauses to align with inclusivity standards and offers rationale for each suggestion.
Technical Aspects: LLM with rule-based fairness constraints and legal clause generation.

Why It Matters (WAI lens):
Bias in legal systems has life-altering consequences—this challenge ensures accountability and inclusion.

Suggested Datasets/APIs:
1. **CourtListener API (Free Law Project)-** Provides access to millions of U.S. court opinions, oral arguments, and legal filings. Ideal for NLP models analyzing bias in legal language across case law and court rulings.
   https://www.courtlistener.com/help/api/rest/
2. **ECtHR Case Law Corpus (European Court of Human Rights)-** A structured dataset of 11,000+ cases with summaries, outcomes, and full legal text. Useful for bias detection, especially in human rights and discrimination-related rulings.
   https://fra.europa.eu/en/case-law-database

3. **LEDGAR: Contract Clause Classification Dataset-** Contains 850,000+ clauses from real-world legal contracts, labeled by type. Useful for identifying bias in employment, financing, and vendor agreements.
 https://zenodo.org/record/5117090
4. **LegalBench Dataset Collection-** A growing benchmark suite for evaluating LLMs on legal tasks, including case classification, reasoning, and bias analysis. Contains various English-language legal datasets.
https://github.com/HazyResearch/legalbench

Challenge 2: AI Guardian Against Misinformation & Deepfakes

Problem Statement:
Widespread misinformation, especially via AI-generated content, poses growing risks to trust and safety.

Potential Solutions:

Approach #1 (Traditional AI):
 Solution: Fake News and Deepfake Classifier
 Description: Train classifiers to detect misinformation using news content, claims, and user engagement signals.
 Technical Aspects: Use labeled datasets for binary/multi-class classification using NLP and video analysis.
Approach #2 (Agentic AI):
 Solution: Browser-Based Claim Verification Copilot
 Description: Real-time agent that evaluates the credibility of content and cites trustworthy sources.
 Technical Aspects: RAG architecture with real-time web search and citation scoring.

Why It Matters (WAI lens):
In an era of AI misinformation, women and vulnerable populations are often targets—trustworthy AI defends public knowledge integrity.

Suggested Datasets/APIs:
1. **DFDC (Deepfake Detection Challenge Dataset)-** Released by Meta, this large-scale video dataset contains over 100,000 labeled real and deepfake videos. Ideal for training and evaluating models to detect AI-generated misinformation in multimedia.
 https://ai.meta.com/datasets/dfdc/
2. **CoAID (COVID-19 Healthcare Misinformation Dataset)-** A comprehensive collection of fake vs. true news articles, social media posts, and user engagement data related to

COVID-19. Supports linguistic analysis and misinformation propagation modeling.
https://github.com/cuilimeng/CoAID

3. **LIAR Dataset – Political Fact-Checking-** Contains 12.8K short political statements manually labeled by the PolitiFact.com team with six truthfulness labels (pants-fire to true). Useful for training claim-level classifiers.
https://www.cs.ucsb.edu/~william/data/liar_dataset.zip

4. **FakeNewsNet-** A comprehensive dataset combining news content, publisher profiles, and social context (e.g., tweet networks) for fake vs. real news detection in political and entertainment domains.
https://github.com/KaiDMML/FakeNewsNet

Challenge 3: AI Risk Analyzer for High-Stakes Deployment

Problem Statement:
Deploying AI models without evaluating their risks (e.g., bias, harm, or unintended consequences) can cause societal harm.

Potential Solutions:

Approach #1 (Traditional AI):
Solution: Model Risk Scoring Dashboard
Description: Evaluate fairness, bias, and potential harm across multiple datasets using standardized metrics.
Technical Aspects: Dashboard integrating AIF360/AIX360 metrics with explainability visualizations.

Approach #2 (Agentic AI):
Solution: Risk-Aware AI Audit Assistant
Description: An interactive agent that guides developers through ethical and risk-focused evaluations during model development.
Technical Aspects: Conversational agent that runs checklist audits, flags issues, and recommends mitigations.

Why It Matters (WAI lens):
Ensures developers critically evaluate real-world harms before deploying models—especially in domains where bias or hallucination can marginalize users.

Suggested Datasets/APIs:

1. **UCI Machine Learning Repository-** A large collection of public datasets across domains like education, healthcare, finance, and hiring. Popular datasets for risk and fairness evaluation include the Adult Income, COMPAS Recidivism, and German Credit datasets.
https://archive.ics.uci.edu/ml/index.php

2. **AIF360 Fairness Dataset Repository (IBM Trusted AI)-** Includes pre-processed datasets specifically selected for evaluating fairness and bias in AI. Also provides baseline metrics and mitigation algorithms.
   https://github.com/Trusted-AI/AIF360
3. **COMPAS Recidivism Risk Score Dataset (ProPublica)-** Contains criminal justice risk assessment scores, race, gender, and outcomes for individuals—widely used to benchmark algorithmic bias in high-stakes settings.
   https://github.com/propublica/compas-analysis
4. **AI Explainability 360 Dataset Suite-** A companion to AIF360 with datasets suited for feature attribution, explainability, and interpretability across sensitive domains. Useful for transparency analysis in model deployment.
   https://github.com/Trusted-AI/AIX360

Challenge 4: Hallucination Detection in Generative AI for Sensitive Domains

Problem Statement:

Large language models (LLMs) can generate hallucinated or misleading information, especially in sensitive areas like healthcare or legal advice.

Potential Solutions:

Approach #1 (Traditional AI):

Solution: Fact-Checking Pipeline for LLM Outputs
Description: Develop post-processing tools that validate AI-generated content using external knowledge sources.
Technical Aspects: Fact verification models trained on claim-evidence pairs.

Approach #2 (Agentic AI):

Solution: Truth Guardian Assistant
Description: An LLM-powered agent that flags hallucinations in real-time, explains errors, and suggests corrections with citations.
Technical Aspects: RAG or knowledge-grounded generation with hallucination scoring.

Why It Matters (WAI lens):

Hallucinated outputs can spread misinformation, particularly dangerous in sensitive areas where women and vulnerable groups rely on accurate information for critical decisions.

Suggested Datasets/APIs:

1. **FactCHD (Chinese Hallucination Detection Benchmark)-** A benchmark dataset created for hallucination detection in Chinese LLM outputs, covering medical, legal, and encyclopedic domains. Includes ground-truth evidence and

hallucination annotations.
https://github.com/zjunlp/FactCHD

2. **EasyDetect (OpenKG 2024 Benchmark)-** A multilingual hallucination detection benchmark focused on knowledge-grounded QA and generation tasks, particularly within knowledge graph contexts.
https://lkm2024.openkg.org/EasyDetect/

3. **TruthfulQA-** A benchmark for evaluating how truthfully language models respond to questions across sensitive topics like health, finance, and politics. Comes with ground-truth labels and graded model responses.
https://github.com/sylinrl/TruthfulQA

4. **FEVER (Fact Extraction and VERification) -** Contains 185,000+ human-labeled claims derived from Wikipedia, annotated as "SUPPORTED," "REFUTED," or "NOT ENOUGH INFO." Widely used for training/verifying fact-checking models.
https://fever.ai/