Canada School of Public Service
École de la fonction publique du Canada

A Primer on Artificial Intelligence (AI)

Canada

Overview

1. What is AI?
2. AI branches and techniques
3. The emerging landscape
4. Skills, data, and infrastructure supporting AI
5. AI uses and opportunities in the Government of Canada
6. Risks and ethical considerations
7. Legal and policy environment
8. Annexes: guidance, available learning, and AI assistants

AI, data, and algorithms

Artificial intelligence (AI)

An **AI system** is a machine-based system that infers how to generate outputs such as predictions, content, recommendations, or decisions from the input it receives

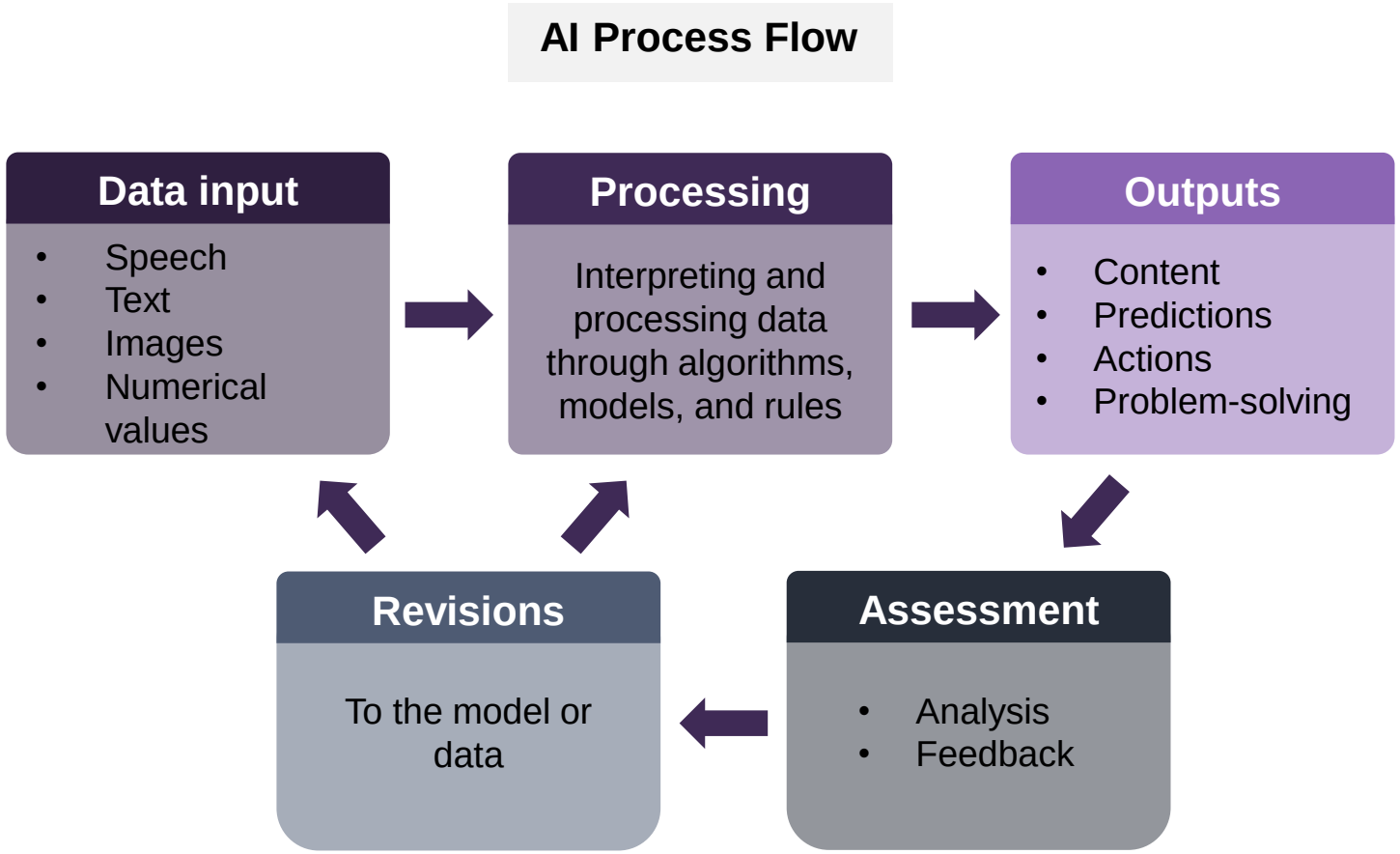
AI is also a **category of technologies**; a common explainer is “technology that performs tasks that would ordinarily require biological brainpower to accomplish, such as making sense of spoken language, learning behaviours, or solving problems”[1]

Data

Data refers to structured and unstructured values such as numbers, text, images, and videos. AI systems get their value from processing massive amounts of data – and are generally required to process that data in the first place

Algorithm

An algorithm is a set of rules or instructions a machine (and especially a computer) follows to achieve a particular goal



What is AI? AI branches and techniques

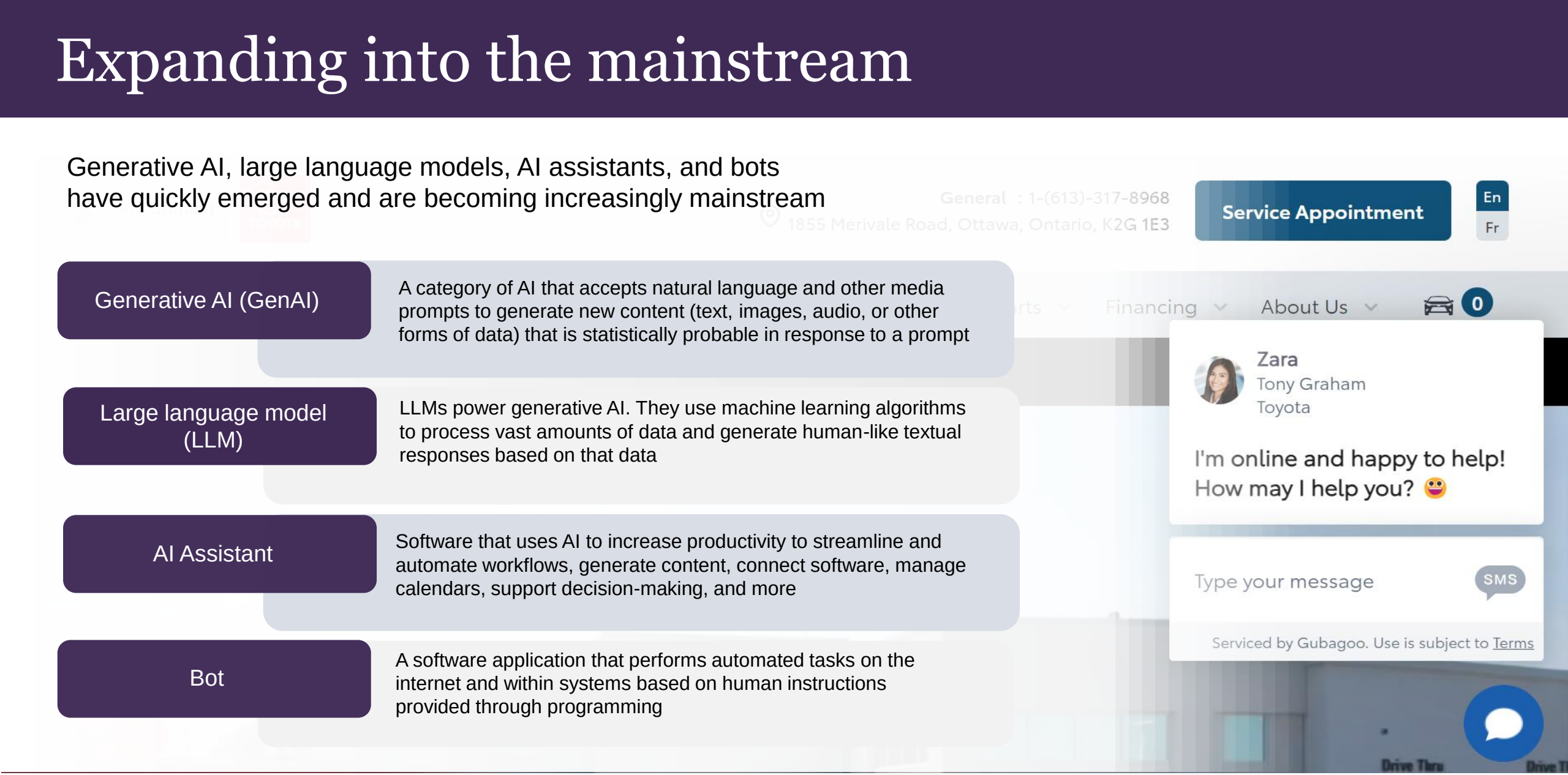
AI is best thought of as a set of interconnected fields and subfields. Rather than a single AI technology, a range of different techniques and approaches are used to solve different problems

Machine Learning (ML)	Reinforcement techniques that allow computers to improve outputs over time by testing multiple processing approaches within the model, and assessing outputs against success benchmarks, then adjusting
Computer Vision	Methods to acquire and make sense of digital images, usually divided into activities recognition, images recognition, and machine vision
Neural Networks	A class of algorithms loosely modeled after the neuronal structure of the brain that improves its performance without being explicitly instructed on how to do so
Natural Language Processing (NLP)	Tools that interpret text (or transcribed speech) for analysis or to allow conversational interaction with software (e.g., Chatbots, GenAI). May or may not use machine learning

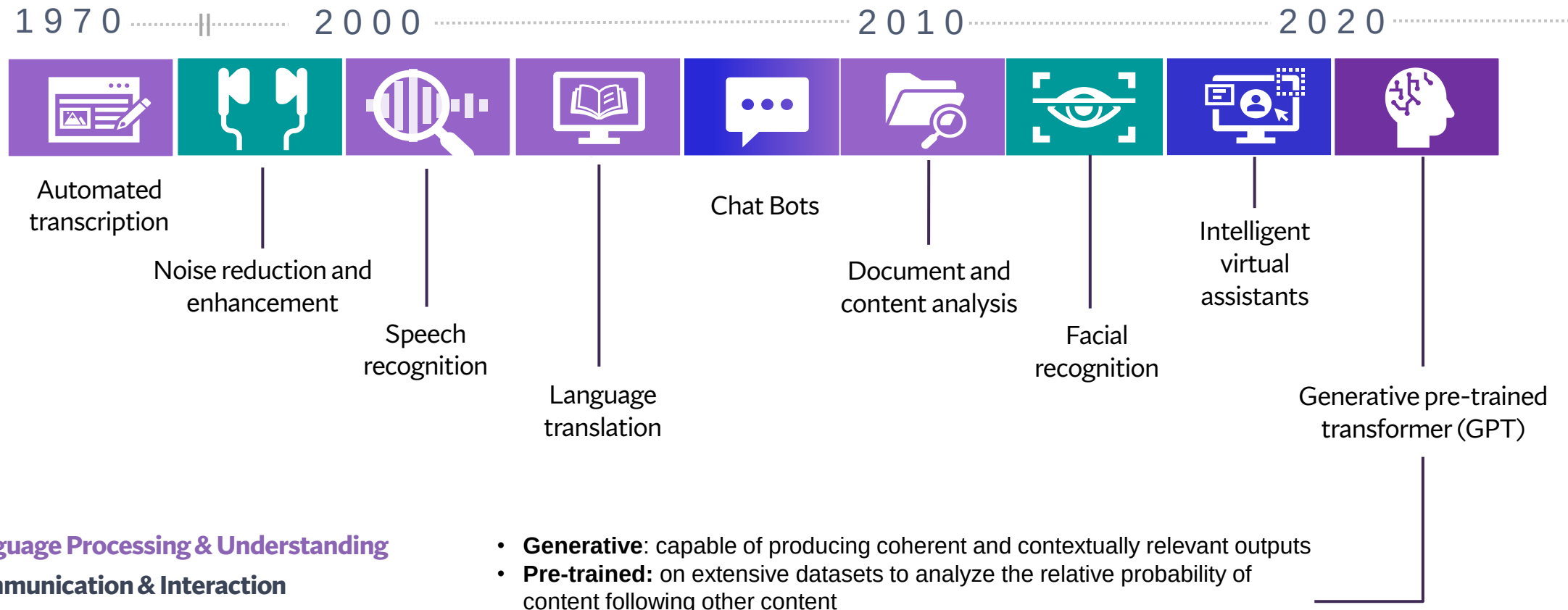
Expanding into the mainstream

Generative AI, large language models, AI assistants, and bots have quickly emerged and are becoming increasingly mainstream

Generative AI (GenAI)	A category of AI that accepts natural language and other media prompts to generate new content (text, images, audio, or other forms of data) that is statistically probable in response to a prompt
Large language model (LLM)	LLMs power generative AI. They use machine learning algorithms to process vast amounts of data and generate human-like textual responses based on that data
AI Assistant	Software that uses AI to increase productivity to streamline and automate workflows, generate content, connect software, manage calendars, support decision-making, and more
Bot	A software application that performs automated tasks on the internet and within systems based on human instructions provided through programming



Timeline and emerging landscape



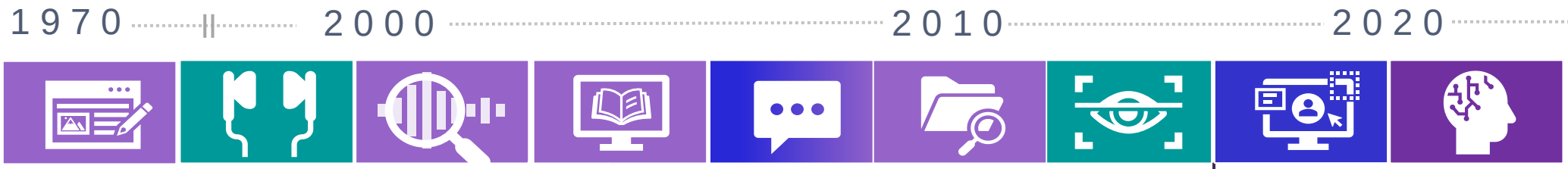
Language Processing & Understanding

Communication & Interaction

Multimedia Analysis

- **Generative**: capable of producing coherent and contextually relevant outputs
- **Pre-trained:** on extensive datasets to analyze the relative probability of content following other content
- **Transformer:** mechanism that helps to understand words in context

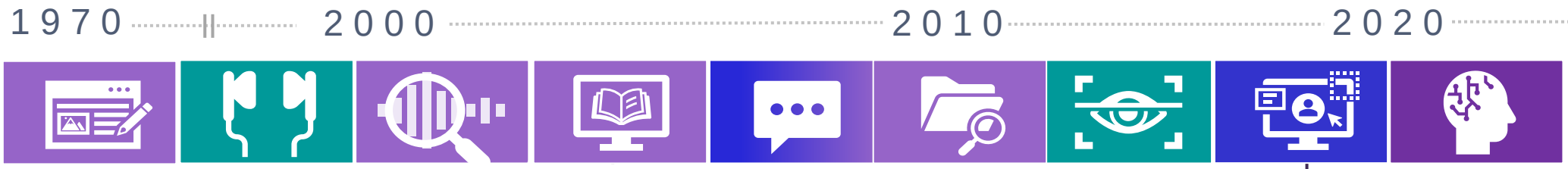
Timeline and emerging landscape: recent



Radical acceleration in the last decade: Why?

- Increase in **available data**
- **Access to software** including open source
- Increase in **computing power** / decrease in cost
- Shared or proprietary **models, algorithms, and techniques**
- Commoditization and **mainstreaming of AI tools**

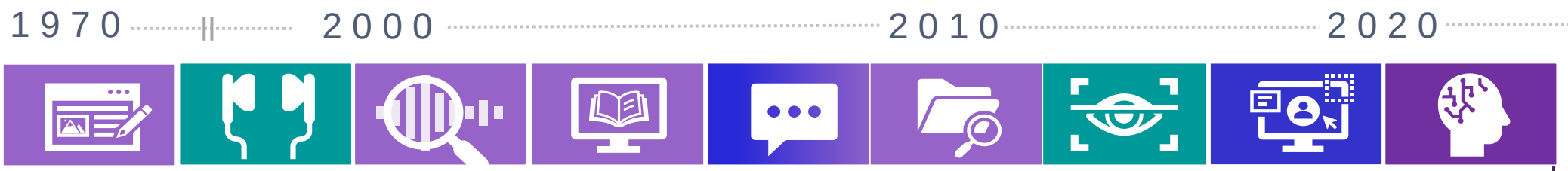
Timeline and emerging landscape: now



Emerging now:

- Generative AI that includes references to sources (called **Retrieval-augmented generation** (RAG)) to increase the reliability of AI systems
- **Generative AI** incorporating GC- and program-specific data behind GC firewalls
- **Multimodal** generation – text, audio, images, and videos

Timeline and emerging landscape: next



What's next:

- Continued **investment and experimentation** in this space will generate new tools and uses, though AI is susceptive to "hype cycle" predictions and the most productive, sustainable uses will be revealed over time
- Mature AI systems still represent ***narrow AI*** (AI systems good at specific tasks), not ***artificial general intelligence*** (AI systems that can self-improve at a wide range of tasks)

Skills, data, and infrastructure supporting AI

For every AI use, there's a lot of work and tooling below the surface. Ultimately, every AI project depends on data

APPLICATION
"Using AI to do [X]"

MODELS
Training data, choice of AI approaches, statistical methods, algorithms

DATA
Data that's ideally cleaned, normalized, unbiased, free of personal or copyright data, and often massive and costly to analyze. Cleaning and preparing data for analysis and processing can be 50%+ of the work and cost involved

SKILLS
Data science, mathematics, statistics, research, data collection, and programming

INFRASTRUCTURE AND TOOLS
Substantial and specific computing power and/or access to Cloud-based tools, open source or proprietary models, and specialized software

BUSINESS QUESTIONS
Ultimately, the foundation of AI work is a well-defined business problem to solve

Opportunities and current uses for AI in the GC

Risks of AI in the GC

1. Keeping up with a **changing technology environment** in operations
2. **Efficiency lag** while the GC develops, procures, or adopts policy-compliant solutions and approaches
3. **Securing** increasingly substantial, valuable, and linked **data holdings**
4. **Policy considerations** regarding AI use in all sectors
5. **Protecting public trust** when using AI for government business
6. **Informational advantage** of external parties leveraging AI in dealings with government

Ethical considerations

As AI becomes more advanced and use becomes widespread, there is a greater risk that it may - even unintentionally - be misused, perpetuate inequality, or exacerbate existing societal problems. This list includes only some of the many ethical considerations. Others may include impacts on job markets, environmental impacts, and questions about humans' relationship with technology

Bias and fairness

Bias in AI means unfair decisions or showing skewed outputs. The GC has a responsibility to make sure that AI tools treat everyone fairly and without discrimination.

Bias can be a product of the **algorithm/model** or the **training or input data.**

In short, historical data with a context of systemic racism and discrimination is likely to result in a biased AI output.

Transparency and accountability

Transparency is required around how AI systems operate and, if they support decision-making, how data was analyzed to produce outputs. This includes **openness, clarity, traceability, and explainability** of the AI system.

Actors – individuals or organizations – leveraging AI systems **may not feel, or take, responsibility and accountability** for actions, outputs, or decisions made by the system.

Privacy, security, and governance

AI systems process massive amounts of data, and AI tools are often Cloud-based or based on externally created resources, code, or models.

Any **data processing of personal or sensitive data needs to be governed and protected.**

Data provenance and copyright

Data provenance refers to the origins, ownership, collection, and reliability of source data. Organizations using data may need to track and document the **sources, transformations, and usage of data throughout data lifecycles.**

Many datasets powering Generative AI in particular have massive, opaque data sources that likely **include personal information, or direct or derivative copyrighted works**

Manipulation and deception

This category involves the ethical considerations related to the use of AI in generating and disseminating misleading of false information.

External **entities may use or propagate disinformation, misinformation, or deepfakes using AI.**

Generative **AI may create content that includes false information.**

AI and representation

Researchers and advocates have identified a number of potential and proven risks and harms of AI. These are likely to disproportionately impact marginalized communities

Bias and fairness

Bias in AI means unfair decisions or showing skewed outputs. The GC has a responsibility to make sure that AI tools treat everyone fairly and without discrimination.

Bias can be a product of the **algorithm/model** or the **training or input data.**

In short, historical data with a context of systemic racism and discrimination is likely to result in a biased AI output.

1. **Absence or under-representation in training data**: e.g., in simulations, self-driving cars were found not to stop for people in wheelchairs, which were absent in the training data.*[1]

   Generative AI underrepresents minority communities in generated images of many professions; attempts to correct this through the processing algorithms have so far instead created stereotypical racially coded images.[2]

   Social media or content-streaming applications are likely to generate recommendations based on "what people like you" like, siloing communities with different life experiences.

2. **Over-representation in training data:** E.g., advocates have demonstrated that predictive policing systems were systematically over-policing marginalized communities, over-represented in the training data.[3]

*Jutta Treviranus, Director of the Inclusive Design Lab at OCAD, notes that AI may be a "double-edged sword" for people with disabilities: there's a danger of decisions or systems based on data that excludes them, or systems generating outputs that serve the majority because they're designed for efficiency. On the other hand, automation technologies (e.g., self-driving cars) could also create options and supports for people.

The legal and policy environment

For Government of Canada internal use:

- *Directive on Automated Decision-Making*
- *Guide on the Use of Generative AI*
- OCIO has just launched an initiative to shape an AI Strategy for the federal public service, scheduled for completion in Fall 2025

For industry and society:

- *Artificial Intelligence and Data Act* (in development)
- Voluntary Code of Conduct on the Responsible Development and Management of Advanced Generative AI Systems

The *Directive* requires an Algorithmic Impact Assessment **where GC use of AI supports, or renders, administrative decisions** about individuals

The *Artificial Intelligence and Data Act* was tabled in 2022 as part of Bill C-27. As of January 2024, is in committee for study

Federal institutions are expected to **align with the principles** of *fair*, *accountable*, *secure*, *transparent*, *educated*, and *relevant* (FASTER) use of generative AI

Can consult **Chief Information** and **Security Officers** and use systems that have gone through privacy and security screening / the AI Supply List

Annex A: FASTER principles for Generative AI

Fair	Ensure that content from these tools does not include or amplify biases and that it complies with human rights, accessibility, and procedural and substantive fairness obligations.
Accountable	Take responsibility for the content generated by these tools. This includes making sure it is factual, legal, ethical, and compliant with the terms of use.
Secure	Ensure that the infrastructure and tools are appropriate for the security classification of the information and that privacy and personal information are protected
Transparent	Identify content that has been produced using generative AI. Notify users that they are interacting with an AI tool. Document decisions and be able to provide explanations if tools are used to support decision-making.
Educated	Learn about the strengths, limitations and responsible use of the tools. Learn how to create effective prompts and to identify potential weaknesses in the outputs.
Relevant	Make sure the use of generative AI tools supports user and organizational needs and contributes to improved outcomes for Canadians. Identify appropriate tools for the task; AI tools aren't the best choice in every situation.

Guide on the use of Generative AI - Canada.ca

Annex B: Use of AI Meeting Assistants

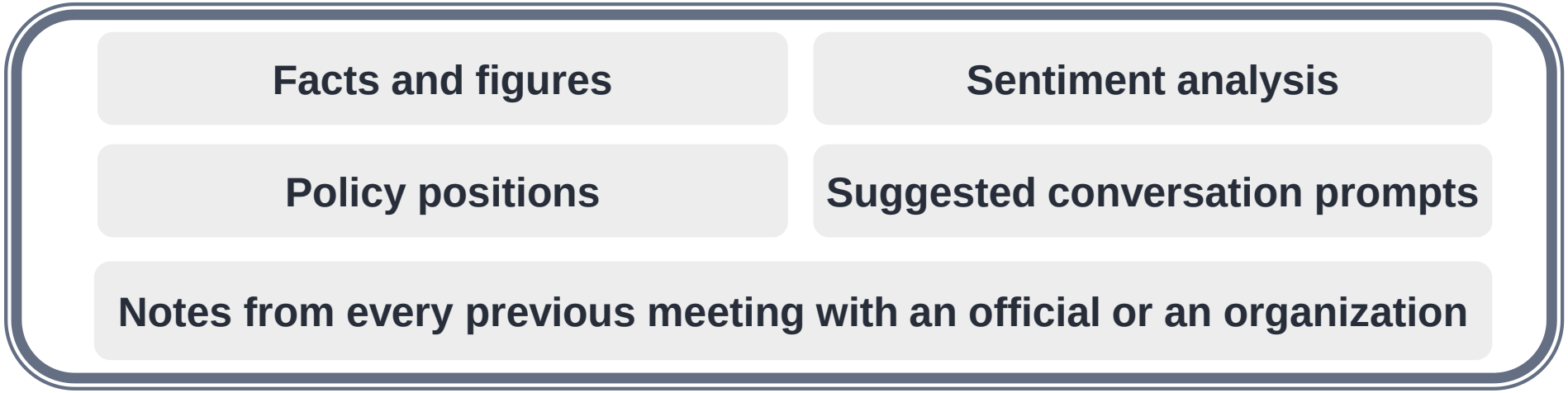
Innocuous	Automated transcription	Language translation	Noise reduction and enhancement	Accessibility features
Increasingly expected	Intelligent virtual assistants	Smart scheduling	Meeting summaries and action items	Document and content analysis
Contentious	Engagement tracking	Gesture recognition	Sentiment analysis	Facial recognition
	+ more emerging uses as the market for AI assistants is seeing rapid expansion and experimentation			

Annex B: Real-Time Business Intelligence in Meetings

AI assistants are already being embedded in major enterprise software applications (e.g., Word, Excel, Teams, PowerPoint)

In a meeting context, that means **integrated with Customer Relationship Management (CRM) systems** (e.g., Dynamics, Salesforce)

Parties meeting with government officials will increasingly have **rich, real-time, access** to:



Annex B: Privacy, Ethics, IM, and Security in Meetings

If a meeting is being **transcribed and summarized**, it's being recorded

Bots and integrated software ***may not trigger*** **the recording notification**

Transcripts and recordings are then **held by**:

- Other organizations
- Often cloud-based AI assistant providers

Some AI assistant software could have **weak data protection –** or be **explicitly designed to collect data**

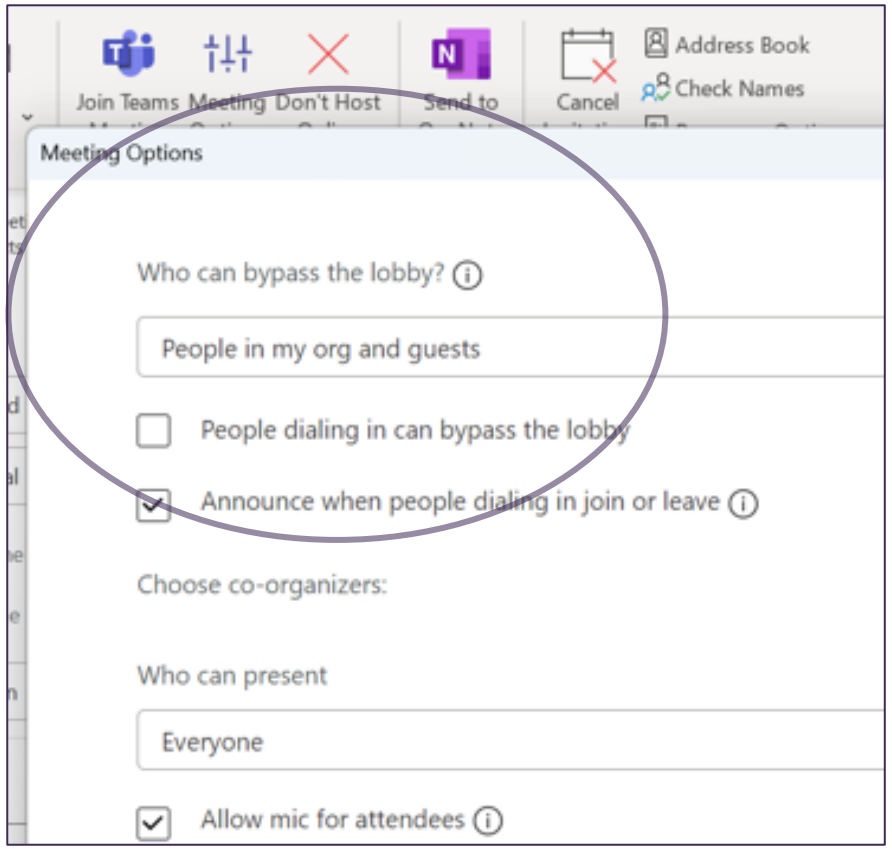
For recordings and transcripts recorded by GC officials, **records will fall under IM and ATIP** legal and policy frameworks

Annex B: Managing for Privacy in Meetings

Ensure that **all participants** in a meeting **are identified and known**

Consider whether assistants **– human or AI –** should be involved in the meeting discussion

In general, we **cannot assume that there's no AI and no recording** - a participant could have a standalone device recording, or a home assistant like Alexa



Ultimately, this is a question of trust in the participants and the sensitivity of the discussion

Annex C: CSPS learning resources

Courses	Discover Artificial Intelligence Using Generative AI in the Government of Canada Ethical Considerations in AI
Microlearning Articles	Decoding AI Assistants in Online Meetings Using Large Language Models (like ChatGPT) in the Federal Public Service Demystifying Artificial Intelligence OpenAI's ChatGPT Explained
Events	Artificial Intelligence Series: ongoing