

EVIDENCE

OTTAWA, Thursday, February 26, 2026

The Standing Senate Committee on Social Affairs, Science and Technology met with videoconference this day at 10:36 a.m. [ET] to examine and report on matters related to the impact of artificial intelligence in Canada.

Senator Rosemary Moodie (*Chair*) in the chair.

[*English*]

The Chair: Good morning. My name is Rosemary Moodie. I'm a senator from Ontario and the chair of this committee.

Before we begin, I would like to do a round table and have senators introduce themselves.

Senator Burey: Good morning and welcome. Sharon Burey, senator for Ontario.

Senator Senior: Good morning. Paulette Senior, senator from Ontario.

[*Translation*]

Senator Boudreau: Good morning. I'm Victor Boudreau from New Brunswick.

[*English*]

Senator Arnold: Dawn Arnold, senator from New Brunswick.

Senator Hay: Katherine Hay, Ontario.

[*Translation*]

Senator Petitclerc: I am Chantal Petitclerc from Quebec.

[*English*]

Senator Greenwood: Margo Greenwood, senator from British Columbia.

Senator Osler: Flordeliz (Gigi) Osler, senator from Manitoba.

Senator Cuzner: Rodger Cuzner, senator from Nova Scotia.

Senator Muggli: Tracy Muggli, Treaty 6 territory, Saskatchewan.

The Chair: Today, the committee continues its study of matters relating to the impact of artificial intelligence in Canada. This study will examine issues including data governance,

TÉMOIGNAGES

OTTAWA, le jeudi 26 février 2026

Le Comité sénatorial permanent des affaires sociales, des sciences et de la technologie se réunit aujourd'hui, à 10 h 36 (HE), avec vidéoconférence, afin d'examiner, pour en faire rapport, les questions relatives aux répercussions de l'intelligence artificielle au Canada.

La sénatrice Rosemary Moodie (*présidente*) occupe le fauteuil.

[*Traduction*]

La présidente : Bonjour. Je m'appelle Rosemary Moodie. Je suis sénatrice de l'Ontario et présidente de ce comité.

Avant de commencer, j'aimerais demander aux sénateurs de se présenter.

La sénatrice Burey : Bonjour et bienvenue. Sharon Burey, sénatrice de l'Ontario.

La sénatrice Senior : Bonjour. Paulette Senior, sénatrice de l'Ontario.

[*Français*]

Le sénateur Boudreau : Bonjour. Victor Boudreau, du Nouveau-Brunswick.

[*Traduction*]

La sénatrice Arnold : Dawn Arnold, sénatrice du Nouveau-Brunswick.

La sénatrice Hay : Katherine Hay, de l'Ontario.

[*Français*]

La sénatrice Petitclerc : Chantal Petitclerc, du Québec.

[*Traduction*]

La sénatrice Greenwood : Margo Greenwood, sénatrice de la Colombie-Britannique.

La sénatrice Osler : Flordeliz (Gigi) Osler, sénatrice du Manitoba.

Le sénateur Cuzner : Rodger Cuzner, sénateur de la Nouvelle-Écosse.

La sénatrice Muggli : Tracy Muggli, du territoire visé par le Traité n^o 6, en Saskatchewan.

La présidente : Aujourd'hui, le comité poursuit son étude des questions relatives aux répercussions de l'intelligence artificielle au Canada. Cette étude portera notamment sur la gouvernance et

sovereignty, ethics, privacy and safety, and the risks, benefits and social impacts of artificial intelligence here in Canada.

This morning, we have the pleasure of welcoming Professor Geoffrey Hinton. Professor Hinton is the 2024 Nobel laureate in physics, and we'll call him the "Godfather of AI."

He is internationally renowned as a pioneer in the field of deep learning as a mode of artificial intelligence. The Nobel Prize in physics he received was for foundational discoveries and inventions that enable machine learning with artificial neural networks, including his invention of the Boltzmann Machine using statistical physics techniques.

Dr. Hinton, for your opening statement, you will have 10 minutes followed by questions from committee members. The floor is yours.

Geoffrey Hinton, Professor, University of Toronto, as an individual: Thank you for the introduction. I will start with a two-minute statement that covers the risks of AI, and then I will spend the remaining eight minutes explaining what AI is because I imagine many of you don't really understand what it is.

Dramatic progress is being made in a new form of artificial intelligence that uses artificial neural networks to learn how to solve difficult computational problems. This new form of AI excels at modelling human intuition rather than human reasoning, and it will enable us to create highly intelligent and knowledgeable assistants who will increase productivity in almost all industries. If the benefits of the increased productivity can be shared equally, it will be a wonderful advance for all humanity.

Unfortunately, the rapid progress in AI comes with many short-term risks. It has already created divisive echo chambers by offering people content that makes them indignant. It is already being used by authoritarian governments for massive surveillance and by cybercriminals for phishing attacks.

In the near future, AI may be used to create terrible new viruses and horrendous lethal weapons that decide by themselves who to kill or maim. All of these short-term risks require urgent and forceful attention from governments and international organizations. We cannot just accept the claims by lobbyists for big tech companies that any regulation will stifle innovation.

la souveraineté des données, l'éthique, la protection des renseignements personnels et la sécurité, ainsi que les risques, les avantages et les répercussions sociales de l'intelligence artificielle au Canada.

Ce matin, nous sommes heureux d'accueillir M. Geoffrey Hinton. M. Hinton est le lauréat du prix Nobel de physique de 2024, et nous l'appellerons le « parrain de l'intelligence artificielle ».

Il est reconnu à l'échelle internationale comme un pionnier dans le domaine de l'apprentissage profond, qui est une branche de l'intelligence artificielle. Il a reçu le prix Nobel de physique pour les découvertes et les inventions fondamentales qui permettent l'apprentissage automatique avec des réseaux neuronaux artificiels, notamment son invention de la machine de Boltzmann à l'aide de techniques de physique statistique.

Monsieur Hinton, vous avez 10 minutes pour faire votre déclaration préliminaire, puis les membres du comité vous poseront des questions. La parole est à vous.

Geoffrey Hinton, professeur, Université de Toronto, à titre personnel : Je vous remercie de cette présentation. Je vais commencer par une déclaration de deux minutes au sujet des risques associés à l'intelligence artificielle, puis, pendant les huit minutes qu'il me reste, j'expliquerai ce qu'est l'intelligence artificielle, car j'imagine que bon nombre d'entre vous ne comprennent pas vraiment de quoi il s'agit.

Des progrès spectaculaires sont réalisés dans le domaine d'une nouvelle forme d'intelligence artificielle qui utilise des réseaux neuronaux artificiels pour apprendre à résoudre des problèmes computationnels complexes. Cette nouvelle forme d'intelligence artificielle excelle dans la modélisation de l'intuition humaine plutôt que du raisonnement humain. Elle nous permettra de créer des assistants très intelligents et compétents qui augmenteront la productivité dans presque tous les secteurs. Si les avantages de cette productivité accrue peuvent être partagés équitablement, ce sera un progrès formidable pour l'humanité tout entière.

Malheureusement, les progrès rapides dans le domaine de l'intelligence artificielle s'accompagnent de nombreux risques à court terme. Elle a déjà créé des chambres d'écho qui sèment la division en proposant aux gens des contenus qui les indignent. Des gouvernements autoritaires l'utilisent déjà pour surveiller indûment la vie des gens, et des cybercriminels s'en servent pour mener des attaques d'hameçonnage.

Dans un proche avenir, on pourrait se servir de l'intelligence artificielle pour créer de terribles nouveaux virus et d'horribles armes meurtrières qui décideraient elles-mêmes qui tuer ou mutiler. Ces risques à court terme exigent une attention urgente et ferme de la part des gouvernements et des organisations internationales. Nous ne pouvons pas nous contenter d'accepter

There is also a longer-term existential threat that will arise when we create digital beings that are more intelligent than us. We have no idea whether we can stay in control. But we now have evidence that if they are created by companies motivated by short-term profits, our safety will not be the top priority. We urgently need research on how to prevent these new beings from wanting to take control. They are no longer science fiction.

That's my canned opening statement about the risks. For understanding these risks, there is a very important scientific issue of whether these large language models understand what they are saying. Some people believe they don't really understand what they are saying. Those are typically the people who believe in good old-fashioned AI.

I would like to give you two minutes on the history of AI. In the 1950s, when AI started, there were two approaches. One was based on logic. The idea was that, when you understand a sentence, you are translating it into some special internal, symbolic, unambiguous language, and once it is in this internal symbolic language, you can apply rules to the symbolic expressions to derive new expressions. That's how logic works, that's what reasoning is, and reasoning is the essence of intelligence.

A completely different approach was the biological approach that said that the intelligence system we know is us, we have a big brain, and in the brain, all our knowledge is in the strengths of connections between neurons. So to understand intelligence and what it is, we need to understand how the brain learns those connection strengths.

Those are the two very different theories. In the biological theory, the idea of meaning was that the meaning of a word is a big bunch of features. So the meaning of the word "cat" is things like this: "It has whiskers; it's a predator; it can be really aloof; it is a pet." It is lots and lots of features that represent all those properties of a cat, which are represented by activating brain cells. You have a brain cell that is activated whenever you are thinking about something that has whiskers, for example.

The question is whether you can unify those two theories. The symbolic theory says that the meaning of a word is all in how it relates to other words. That theory goes back a hundred years. The psychological theory says, no, the meaning of a word is a big bunch of features. They look like very different theories.

les affirmations des lobbyistes des grandes entreprises technologiques selon lesquelles la réglementation étouffera l'innovation.

Ensuite, une menace existentielle à plus long terme surgira lorsque nous créerons des êtres numériques plus intelligents que nous. Nous ne savons pas si nous pouvons garder le contrôle. Nous avons toutefois la preuve que si ces êtres sont créés par des entreprises motivées par la réalisation de profits à court terme, notre sécurité ne sera pas leur priorité absolue. Nous devons mener de toute urgence des recherches sur la manière d'empêcher ces nouveaux êtres de vouloir prendre le contrôle. Ils ne relèvent plus de la science-fiction.

Voilà pour ma déclaration préliminaire préparée à l'avance au sujet des risques. Pour comprendre ces risques, nous devons nous poser une question scientifique très importante : ces grands modèles de langage comprennent-ils ce qu'ils disent? Certaines personnes croient que non. Ce sont généralement des personnes qui croient en une intelligence artificielle traditionnelle.

J'aimerais prendre deux minutes pour parler de l'histoire de l'intelligence artificielle. Dans les années 1950, lorsque l'intelligence artificielle a vu le jour, il y avait deux approches. L'une était fondée sur la logique. L'idée était que l'on comprend une phrase en la traduisant en un langage interne particulier, symbolique et sans ambiguïté, et une fois que cela est fait, on peut appliquer des règles aux expressions symboliques pour en générer de nouvelles. C'est ainsi que fonctionne la logique, c'est ce qu'est le raisonnement, et le raisonnement est l'essence même de l'intelligence.

L'approche biologique était complètement différente. Selon cette approche, le système d'intelligence que nous connaissons, c'est nous : nous avons un gros cerveau dans lequel toutes nos connaissances résident dans la force des liens entre les neurones. Donc, pour comprendre l'intelligence, ce qu'elle est, nous devons comprendre comment le cerveau apprend à renforcer ces liens.

Ce sont deux théories très différentes. Selon la théorie biologique, on comprend le sens d'un mot à l'aide des nombreuses caractéristiques qui s'y rattachent. Le sens du mot « chat » est composé des éléments suivants : le chat a des moustaches, c'est un prédateur, il peut être très indifférent, et c'est un animal de compagnie. Une foule de caractéristiques représentent ces propriétés propres au chat, lesquelles sont représentées par l'activation des cellules du cerveau. Ainsi, chaque fois que l'on pense à quelque chose qui a des moustaches, une cellule cérébrale s'active.

La question est de savoir si on peut unir ces deux théories. La théorie symbolique affirme que le sens d'un mot réside entièrement dans le rapport de ce mot à d'autres mots. Cette théorie remonte à une centaine d'années. La théorie psychologique affirme quant à elle que le sens d'un mot est

You can unify them in the following way. You take a whole bunch of text, and you try to predict the next word. Of course, one way to predict the next word is to have a big table of common phrases, and if you see the first part of a phrase, then you predict the next bit of a phrase. That's how word prediction used to be done.

It's not done that way anymore. A much more sophisticated way to predict the next word is to convert each word in the context — the words you have seen already — into a big bunch of features, allow interactions between the features of different words to predict the feature of the next word, and once you have predicted the features of the next word, you guess what the next word is, given its features. That's how current large language models, or LLMs, work, and it is very different from the symbolic idea that understanding consists in translating into an internal string of symbols.

The biological idea is that understanding consists of converting each word symbol into a big bunch of features. Now, I want to give you an analogy for that, which will help you understand. Then I will be done. The analogy is LEGO blocks. If you want to understand the shape of a Porsche, for example, you can build that shape out of LEGO blocks. The surface won't have quite the right shape, so the aerodynamics won't be right, but you can say where the stuff is with LEGO blocks.

Words are like LEGO blocks. But instead of just modelling where 3-D stuff is, they can be used to model anything. So we are monkeys, and we invented this very clever way of doing modelling that is universal, and that's what makes us special monkeys.

Words basically differ from LEGO blocks in four ways. That's what I will go through now, and then I will be done.

The first way they differ is that there are many more of them. Each of these special LEGO blocks has a name — the name of the word — and there are maybe 30,000 of them that you use commonly. That's the first difference.

The second difference is that they are much higher dimensional. A LEGO block only has a few degrees of freedom: It has a length, width, height and orientation, and that's about it. A word has thousands of dimensions, and those dimensions are all the possible active features. So a word like "cat" has active features like that it has whiskers, and inactive features like that it is a manufactured object, and it has thousands of those. Rather than having a few features, like a LEGO block, which has its size and where it is, a word has thousands of features, which we learn. We learn them by trying to predict the next word and to

constitué d'un ensemble de caractéristiques. Ces deux théories semblent très différentes.

On peut les unir de la façon suivante. On prend une séquence de mots et on essaie de prédire le mot suivant dans cette séquence. On peut y arriver, entre autres, à l'aide d'un grand tableau de phrases courantes : si on voit la première partie d'une phrase, on peut prédire le bout suivant. C'est ainsi que fonctionnait la prédiction des mots, auparavant.

Les choses ont changé. Une façon beaucoup plus sophistiquée de prédire le mot suivant dans une séquence est de transformer chaque mot dans le contexte — les mots que l'on a déjà vus — en une foule de caractéristiques, de favoriser les interactions entre les caractéristiques des différents mots pour prédire la caractéristique du mot suivant, puis, une fois que cela est fait, on devine le mot suivant grâce à ses caractéristiques. C'est ainsi que fonctionnent les grands modèles de langage actuels, les GML. C'est très différent de la théorie symbolique selon laquelle la compréhension d'une phrase vient de sa traduction en une suite interne de symboles.

L'approche biologique stipule que la compréhension consiste à convertir le symbole associé à chaque mot en un grand nombre de caractéristiques. Avant de conclure, je vais vous présenter une analogie qui vous aidera à comprendre : les briques LEGO. Si vous voulez comprendre la forme d'une Porsche, par exemple, vous pouvez construire cette forme à l'aide de briques LEGO. La forme ne sera pas tout à fait la même, l'aérodynamisme ne sera pas parfait, mais vous serez en mesure de dire où se trouvent les éléments d'une Porsche grâce aux briques LEGO.

Les mots sont comme des briques LEGO. Toutefois, au lieu de simplement modéliser des objets en 3D, on peut les utiliser pour modéliser n'importe quoi. Nous sommes des singes exceptionnels, car nous avons inventé cette façon très ingénieuse de faire de la modélisation qui est universelle.

Les mots diffèrent des briques LEGO de quatre façons. C'est ce que je vais vous expliquer maintenant, puis j'aurai terminé.

La première différence, c'est qu'il y a beaucoup plus de mots qu'il n'y a de briques. Chacune de ces briques LEGO spéciales porte un nom — le nom du mot — et il y en a peut-être 30 000 que vous utilisez couramment. C'est la première différence.

La deuxième différence, c'est que les mots ont beaucoup plus de dimensions. Une brique LEGO n'a qu'une longueur, une largeur, une hauteur et une orientation; c'est à peu près tout. Un mot comporte des milliers de dimensions, et ces dimensions représentent l'ensemble des caractéristiques actives possibles. Le mot « chat », par exemple, a des caractéristiques actives comme le fait qu'un chat a des moustaches, et des caractéristiques inactives comme le fait que c'est un objet fabriqué, et il existe des milliers d'autres caractéristiques. Plutôt que de n'avoir que quelques caractéristiques, comme une brique LEGO, qui a une

figure out how to convert words into features so that we can predict the next word using those features.

So the first difference is that there are many more of them; the second difference is that they are very high dimensional and they have lots of features. A third difference is that they are not rigid. LEGO blocks are rigid. Words have a rough shape. These are unambiguous words; ambiguous words have two quite different rough shapes. But the actual shape they adopt depends on their context. A word is somewhat flexible, and it uses the context to determine its precise shape so that they all fit together nicely. That's the third difference: it is flexible. It's not totally flexible, of course. A word you know has strong limitations on what shape it can adopt, but it can adopt many different high-dimensional shapes.

The final difference is that the way words fit together is much more complicated than LEGO blocks. LEGO blocks snap together by a little plastic cylinder going into a little plastic hole. The way words fit together is something like this — this is an approximation to what's happening in large language models: Think of every word has having a whole bunch of long, flexible arms attached to it, and on the end of each arm is a hand. The hands come in many different colours, and the shape of the hand can change. As you change the shape of the word — that is to say, as it adapts to its context to adopt a shape that is appropriate for the context — the shapes of all the hands change. In addition, every word has gloves stuck to it. The fingertips of the word are stuck to the LEGO block, and as the word changes shape, the shapes of those gloves change.

Understanding a sentence consists of the following: You start by getting the approximate rough shape for each word, which gives approximate shapes to its hands and gloves, and then you have to mess about with the shape, adjusting the shapes so that the hands of words can fit into the gloves of other words, and, of course, the shapes of the hands in the gloves are changing as you change the shape of the word. So it is a difficult problem. It is a problem that is solved in these computers by using a lot of electricity and solved in our brain by using a lot of brain cells. But they are basically solving the same problem.

So understanding consists of the following: Take the approximate shapes you have for all these words and modify the shapes so they can all fit together. So the hands of some words can fit into the gloves of other words. Then you get a structure, and once you have got that structure of features that all fit together, that is understanding. That's what is happening in us.

taille et un emplacement, un mot possède des milliers de caractéristiques, que nous apprenons. Nous les apprenons en essayant de prédire le mot suivant dans une séquence de mots; nous essayons de trouver comment convertir les mots en caractéristiques afin de pouvoir prédire le mot suivant à l'aide de ces caractéristiques.

La première différence est donc qu'il y a beaucoup plus de mots qu'il n'y a de briques LEGO. La deuxième différence est que les mots ont beaucoup plus de dimensions et qu'ils ont de nombreuses caractéristiques. Une troisième différence, c'est que les mots ne sont pas rigides, contrairement aux briques LEGO. Les mots ont une forme approximative. Ce sont des mots sans ambiguïté; les mots ambigus ont deux formes approximatives assez différentes. La forme qu'ils adoptent dépend du contexte dans lesquels on les utilise. Un mot est relativement souple et sa forme précise sera déterminée en fonction du contexte. Ainsi, il pourra s'assembler parfaitement avec les autres mots. C'est la troisième différence : les mots sont souples, mais pas toujours, évidemment. La forme que peut prendre un mot que vous connaissez est très limitée. Cela dit, le mot peut adopter de nombreuses formes. Il a plusieurs dimensions.

La dernière différence réside dans le fait que la manière dont les mots s'assemblent est beaucoup plus complexe que pour les briques LEGO. Les briques LEGO s'emboîtent grâce à un petit cylindre en plastique qui s'insère dans un petit trou en plastique. La façon dont les mots s'assemblent ressemble à ceci — c'est une approximation de ce qui se passe dans les grands modèles de langage — : imaginez que chaque mot est muni d'une multitude de bras longs et flexibles, au bout desquels se trouve une main. Les mains sont de différentes couleurs et leur forme peut changer. À mesure que l'on change la forme du mot — c'est-à-dire lorsqu'il s'adapte à son contexte pour adopter une forme qui lui convient —, la forme des mains change. De plus, chaque mot est muni de gants. Les doigts du mot sont accrochés à la brique LEGO, et à mesure que le mot change de forme, la forme des gants change également.

Pour comprendre une phrase, il faut commencer par obtenir la forme approximative de chaque mot, qui donne des formes approximatives à ses mains et à ses gants, puis, il faut jouer avec la forme, il faut l'ajuster pour que les mains des mots puissent s'insérer dans les gants d'autres mots et, bien sûr, la forme des mains dans les gants change à mesure que l'on change la forme du mot. C'est un problème difficile qui est résolu dans ces ordinateurs en utilisant beaucoup d'électricité et dans notre cerveau en utilisant beaucoup de cellules cérébrales. Ils résolvent le même problème.

La compréhension consiste donc à prendre les formes approximatives de tous ces mots et à les modifier pour qu'elles s'emboîtent les unes dans les autres. Ainsi, les mains de certains mots peuvent s'insérer dans les gants d'autres mots. On obtient ensuite une structure, et une fois que l'on a cette structure de caractéristiques qui s'imbriquent toutes les unes dans les autres,

That's what's happening in these chatbots, and it is totally different from the old-fashioned AI idea that understanding consists of translating into some internal symbolic language. It is much more like figuring out the structure of protein, where you are given a string of amino acids, and you have to figure out a shape where they fit together happily.

Once you understand that these large language models are understanding in the same way we do, then things become much more scary because you realize that what we're doing is creating alien beings that really do understand, and they are going to get more intelligent than us some time in the next 20 years, most experts believe, and we have no idea what is going to happen then. I'm done with my introduction.

The Chair: Thank you, Dr. Hinton.

We will now proceed to questions from committee members. For this panel, senators will have four minutes for questions, and that includes the answer.

Senator Burey: Professor Hinton, it is an honour to be speaking with you today. Thank you for being at our committee.

I actually first became aware of your work 10 years ago when my son forwarded me an article about some predictions you made about radiologists. As a physician, I appreciated your candour and foresight. It really forced some necessary conversations and helped me to begin to prepare for this difficult and very different world.

From your opening comments, I can see that you are not going to shy away from being candid with this Senate committee. I will restrict my line of questioning to the list of existential risks with which you ended your statement, those from which humanity cannot recover. The Machine Intelligence Research Institute, or MIRI, says that the default consequence of the creation of artificial superintelligence is human extinction. Just a few days ago, I read that Anthropic, one of OpenAI's competitors, loosened some of its core safety principles, and they are supposed to be the good guys.

I ask you this: Is the goal of top companies in this field to build superintelligent AI, and if so, if they succeed, what will it mean for Canadians? In other words, what keeps you up at night, and what should we be doing as legislators?

on comprend le message. C'est ce qui se passe dans notre cerveau. C'est ce qui se passe dans ces robots conversationnels, et c'est totalement différent de l'idée traditionnelle que l'on pourrait avoir de l'intelligence artificielle selon laquelle la compréhension consiste à traduire les mots en un langage symbolique interne. L'intelligence artificielle ressemble beaucoup plus au fait de déterminer la structure d'une protéine : on obtient une série d'acides aminés, et il faut déterminer une forme dans laquelle ils s'emboîtent parfaitement.

Lorsque vous réalisez que ces grands modèles de langage comprennent les choses comme nous les comprenons, la situation devient beaucoup plus inquiétante. Nous sommes en train de créer des êtres étrangers qui comprennent vraiment ce qui se passe, et qui — selon la plupart des experts — vont devenir plus intelligents que nous au cours des 20 prochaines années, et nous n'avons aucune idée de ce qui va se passer à ce moment-là. J'ai terminé ma déclaration préliminaire.

La présidente : Merci, monsieur Hinton.

Nous allons maintenant passer aux questions des membres du comité. Pour ce groupe de témoins, les sénateurs disposeront de quatre minutes pour poser leurs questions, ce qui comprend la réponse.

La sénatrice Burey : Monsieur Hinton, c'est un honneur de m'adresser à vous aujourd'hui. Je vous remercie de votre présence à notre comité.

J'ai pris connaissance de votre travail pour la première fois il y a 10 ans, lorsque mon fils m'a transmis un article sur vos prédictions au sujet des radiologistes. En tant que médecin, j'ai bien aimé votre franchise et votre clairvoyance. Cela m'a poussée à avoir des conversations nécessaires et m'a aidée à commencer à me préparer à ce monde difficile et très différent.

D'après votre déclaration préliminaire, je sais que vous n'allez pas hésiter à vous montrer franc avec notre comité. Je vais limiter mes questions à la liste des risques existentiels dont vous avez parlé à la fin de votre déclaration; ceux dont l'humanité ne pourrait se remettre. Le Machine Intelligence Research Institute affirme que la création d'une superintelligence artificielle entraînera inévitablement l'extinction de l'humanité. Il y a quelques jours à peine, j'ai lu qu'Anthropic, l'un des concurrents d'OpenAI, avait assoupli certains de ses principes de sécurité de base, alors qu'ils sont censés être du côté des bons.

Je vous pose la question suivante : l'objectif des grandes entreprises dans ce domaine est-il de créer une IA superintelligente et, le cas échéant, quelles en seraient les conséquences pour les Canadiens si elles réussissaient? Autrement dit, qu'est-ce qui vous empêche de dormir la nuit, et que devrions-nous faire en tant que législateurs?

Mr. Hinton: There are things you can do in the short term. There is an urgent problem of people using AI tools to create nasty viruses. That's very scary. I'm not sure what you do about that, but at least you are getting international collaboration on how to try and prevent that.

The most urgent things are to do with the corruption of elections, one of which is coming shortly in the U.S. If you wanted to corrupt the U.S. elections, the first thing you would do is collect as much data as you could on U.S. citizens. It seems likely that was the real purpose of the Department of Government Efficiency, or DOGE. The people involved have contact with the people who, for example, corrupted Brexit — Cambridge Analytica.

The most urgent problem after that is unemployment. Big tech companies intend to make a lot of money; otherwise, they wouldn't be investing between them about a trillion dollars in data centres. The only way they will make that much money is by replacing jobs. They have not thought through what will happen if you replace a large fraction of workers. You are going to lose your tax base. Things like universal basic income will be tricky because there won't be a tax base anymore. David Duvenaud will talk more about this.

If AI can do any normal human job, humans will cease to have value as labour, and David Duvenaud has pointed out that if they are not being taxed, they won't get properly represented. So his view is that there is no representation without taxation. I believe that. I believe a crisis is coming, in which we see massive unemployment caused by AI.

I made that prediction in 2016, and it didn't come to pass. We did get AI being used for radiology, but we have a lot more radiology going on, so we now have radiologists working with AI, and a lot more images are being interpreted. It is an elastic market. With health care, you can absorb as much health care as people can provide. So it won't lead to unemployment in health care, but there are many other industries, like call centres, where it will lead to massive unemployment.

Senator Burey: Thank you, Professor Hinton.

Senator Hay: It's great to see you again, Professor Hinton. You may remember me from Vector Institute times.

A couple of days ago, you made a fairly significant statement on CBC Radio that AI must foster maternal instincts or we risk extinction. It is certainly eye-catching and scary. As we are looking at the study that we're doing, and we think about

M. Hinton : Il y a des mesures que vous pouvez prendre à court terme. L'utilisation des outils d'intelligence artificielle afin de créer des virus malveillants est un problème auquel on devrait s'attaquer de toute urgence. C'est très inquiétant. Je ne sais pas trop comment y remédier, mais au moins, une collaboration internationale s'est mise en place pour essayer d'empêcher cela.

Les questions les plus urgentes concernent la corruption des élections. Il y aura d'ailleurs bientôt une élection aux États-Unis. Si on voulait corrompre les élections américaines, la première chose à faire serait de recueillir autant de données que possible sur les citoyens américains. Vraisemblablement, c'était là le véritable objectif du département de l'efficacité gouvernementale, ou DOGE. Les personnes concernées ont des contacts avec ceux qui ont, par exemple, porté atteinte au Brexit; Cambridge Analytica.

Le problème le plus important après cela est le chômage. Les grandes entreprises technologiques veulent faire beaucoup d'argent; autrement, elles n'investiraient pas, ensemble, près de 1 billion de dollars dans des centres de données. Elles ne pourront faire autant d'argent qu'en remplaçant des emplois. Or, elles n'ont pas bien réfléchi à ce qui se passera si l'intelligence artificielle remplace de nombreux travailleurs : l'assiette fiscale va disparaître. Il sera alors difficile de mettre en œuvre des mesures telles que le revenu de base universel, car il n'y aura plus d'assiette fiscale. David Duvenaud en parlera plus en détail.

Si l'intelligence artificielle peut accomplir le même travail qu'un humain, la main-d'œuvre humaine n'aura plus de valeur. David Duvenaud a souligné que s'ils ne sont pas imposés, les humains ne seront pas bien représentés. Il est donc d'avis qu'il n'y a pas de représentation sans imposition. Je partage son point de vue. Je crois qu'une crise se profile à l'horizon : l'intelligence artificielle entraînera un chômage massif.

J'ai fait cette prédiction en 2016, mais elle ne s'est pas concrétisée. L'intelligence artificielle est certes utilisée en radiologie, mais ce domaine connaît une forte croissance. Les radiologistes utilisent l'intelligence artificielle afin d'interpréter un plus grand nombre d'images. C'est un marché élastique. Dans le secteur de la santé, la demande en soins de santé est aussi grande que l'offre disponible. L'intelligence artificielle n'entraînera donc pas de chômage dans ce secteur. Il y en a toutefois d'autres, comme les centres d'appels, où elle conduira à un chômage massif.

La sénatrice Burey : Merci, monsieur Hinton.

La sénatrice Hay : Je suis heureuse de vous revoir, monsieur Hinton. Vous vous souvenez peut-être de moi, du temps où je travaillais à l'Institut Vecteur.

Il y a quelques jours, vous avez fait une déclaration très importante à la radio de la CBC selon laquelle l'intelligence artificielle doit développer un instinct maternel, sinon nous risquons de disparaître. C'est certainement une déclaration qui

guardrails and the state of Canada as it relates to AI data sovereignty and governance, and we worry about fostering maternal instincts within AI, how would you approach that? How would you ask the government to approach that or legislate that if that is possible? What guardrails should we be considering that would alleviate this lack of maternal instinct? Or is that even possible?

Mr. Hinton: We don't even know if it is possible. At this stage, it is not like climate change. We know how to prevent climate change: Just stop burning carbon and plant a lot of trees. Canada should be planting huge numbers of trees.

AI isn't like that. We don't know the solution to the existential threat. Right now, we should be doing a lot of research on it. There is hardly any research being done on it. Maybe 1% of AI researchers are working on it, maybe less. The government could try to force more research on whether we can invent a way so we can live with things with more intelligence than ourselves. We don't know whether we can.

In the shorter term, it is obvious that you shouldn't allow big companies to release chatbots without very thorough testing. Big tech companies in the United States have a strong lobby that runs lots of advertisements about how any regulation will interfere with innovation. That's a bit like "Big Oil," saying, "If there are any regulations on the environment, we won't be able to get as much oil." That's true, but that doesn't mean you shouldn't have regulations.

It is difficult because if one country doesn't have regulations and another country does, there will be competitive advantages. That's why Elon Musk, for example, went after OpenAI. It was so that he could get a competitive advantage for Texas over California.

If we don't have regulations, AI is going to do lots of nasty things, like encouraging kids to commit suicide. It is very clear that companies should be obliged to say what tests they did and what the results of the tests were. The tests should be quite strong. They should test for everything we know that might go wrong.

At present, in the United States, big companies have pandered to Trump, and they are trying to have no regulations at all. At least if Europe and Canada insist that you can't use a chatbot

attire l'attention et qui est inquiétante. Comment aborderiez-vous cette question de l'instinct maternel dans le cadre de l'étude que nous menons, alors que nous réfléchissons aux garde-fous et à la situation du Canada dans le contexte de la souveraineté et de la gouvernance des données liées à l'intelligence artificielle? Comment demanderiez-vous au gouvernement d'aborder cette question ou de légiférer à cet égard, si c'est possible? Quels garde-fous devrions-nous envisager pour pallier ce manque d'instinct maternel? Est-ce même possible?

M. Hinton : Nous ne savons même pas si c'est possible. Pour l'instant, ce n'est pas comme les changements climatiques. Nous savons comment lutter contre les changements climatiques : il suffit de planter beaucoup d'arbres et d'arrêter de brûler des combustibles fossiles. Nous devrions planter énormément d'arbres au Canada.

Ce n'est pas la même chose avec l'intelligence artificielle. Nous ne connaissons pas la solution à cette menace existentielle. Nous devrions faire beaucoup de recherches à ce sujet, car il n'y en a presque pas, à l'heure actuelle. Il n'y a peut-être que 1 % des chercheurs en intelligence artificielle qui se penchent sur cette question, voire moins. Le gouvernement pourrait essayer d'imposer la réalisation de davantage de recherches visant à déterminer si nous pouvons inventer une façon de cohabiter avec des entités plus intelligentes que nous. Nous ne savons pas si c'est possible.

À court terme, il est évident qu'il ne faut pas permettre aux grandes entreprises de commercialiser des robots conversationnels sans qu'ils aient fait l'objet de tests approfondis. Les grandes entreprises technologiques aux États-Unis disposent d'un puissant lobby qui diffuse de nombreuses publicités affirmant que tout règlement nuira à l'innovation. C'est un peu comme si les grandes pétrolières disaient qu'elles ne pourraient pas extraire autant de pétrole si nous adoptions un règlement en matière de protection environnementale. C'est vrai, mais cela ne veut pas dire qu'on ne devrait pas avoir de règlement.

C'est une question difficile, car si un pays n'a pas de règlements alors qu'un autre en a, cela créera des avantages concurrentiels. C'est pourquoi Elon Musk, à titre d'exemple, a intenté des poursuites contre OpenAI. Il voulait obtenir un avantage concurrentiel pour le Texas par rapport à la Californie.

Sans règlements, l'intelligence artificielle fera une foule de choses atroces, comme inciter des enfants à se suicider. Il ne fait aucun doute que les entreprises devraient être tenues de divulguer les tests qu'elles ont effectués, en plus des résultats qu'elles ont obtenus. Ces tests devraient être très rigoureux. Ils devraient vérifier tout ce qui, à notre connaissance, pourrait mal tourner.

En ce moment, aux États-Unis, les grandes entreprises se sont pliées aux exigences de Trump et essaient d'éliminer toute réglementation. Si l'Europe et le Canada insistent pour que

here unless you have satisfied some regulations, that may actually force the States to have regulations too because they don't want to split the market. They want to be able to sell the same chatbot everywhere.

Senator Hay: I would like to ask a question about data, data storage and how data travels. Most of the major data and AI companies are U.S.-based, for example, Anthropic, AWS, Microsoft and OpenAI. When they are on Canadian soil — AWS in Montreal — I believe we are still at risk of having our data infiltrated or interfered with just as a result of an executive order by the President of the United States. Would that be fair? How would we defend our digital sovereignty and digital borders if U.S. companies are actually holding our data even if it is on Canadian soil?

Mr. Hinton: I'm not an expert on cybersecurity or privacy, so I can't say anything sensible about that. Sorry.

Senator Hay: Thank you.

Senator McPhedran: Thank you, Professor Hinton, for being with us today. We have really been looking forward to this. I just finished a conversation about progress here in Canada in terms of protecting Canadians, and, obviously, this is informed by what we now know about the choice of an AI company not to alert anyone in terms of the shooter of Tumbler Ridge. Can you help me understand something? We were told yesterday by witnesses from the government that Canada initiated its AI Strategy in 2017. If that is the case, how is it that we appear to be so far behind today?

Mr. Hinton: It's because we're a small country compared to the U.S. and China. Canada cannot put hundreds of billions of dollars into AI like they are. Basically, we have to accept that Canada is one of the medium-sized countries. We should obviously collaborate with other medium-sized countries, but we're not one of the two major players.

I have one comment on Tumbler Ridge. I used to work for a big company — Google — so I'm able to see the other side too. I'm able to see the side of the tech companies. From the point of view of the tech companies, it is not quite as simple as the media makes it out to be. That's because if you ask how many other users there were who had profiles as scary as the Tumbler Ridge killer, my guess is that there were thousands or tens of thousands. So they have to decide which are the scariest ones. They could just alert the government about huge numbers of people, but it's not as if there were this one scary person whom they failed to talk about.

l'utilisation d'un robot conversationnel sur leur territoire ne puisse se faire qu'en respectant certains règlements, cela pourrait contraindre les États-Unis à adopter, eux aussi, des règlements, car ils ne veulent pas se partager le marché. Ils veulent être en mesure de vendre le même robot conversationnel partout.

La sénatrice Hay : J'aimerais poser une question au sujet des données, leur stockage et leur circulation. La plupart des grandes entreprises spécialisées dans les données et l'intelligence artificielle, comme Anthropic, AWS, Microsoft et OpenAI, se trouvent aux États-Unis. Lorsqu'elles sont en sol canadien — comme AWS à Montréal —, je crois qu'il y a un risque que l'on infiltre ou porte atteinte à nos données, seulement en raison d'un décret du président des États-Unis. Est-ce juste? Comment défendre notre souveraineté numérique et nos frontières numériques si des entreprises américaines détiennent nos données alors qu'elles se trouvent en sol canadien?

M. Hinton : Je ne suis pas un expert en cybersécurité ou en protection des renseignements personnels. Je ne peux donc pas m'exprimer de manière éclairée à ce sujet, et je m'en excuse.

La sénatrice Hay : Merci.

La sénatrice McPhedran : Merci, monsieur Hinton, d'être avec nous aujourd'hui. Nous attendions cette rencontre avec impatience. Je viens de terminer une conversation sur les progrès réalisés ici au Canada concernant la protection des Canadiens, et, évidemment, cela s'appuie sur ce que nous savons maintenant au sujet de la décision d'une entreprise d'IA de ne prévenir personne au sujet de la personne responsable de la fusillade de Tumbler Ridge. Pouvez-vous m'aider à comprendre quelque chose? Hier, des témoins du gouvernement nous ont dit que le Canada avait lancé sa stratégie en matière d'IA en 2017. Si tel est le cas, comment se fait-il que nous semblions si loin derrière aujourd'hui ?

M. Hinton : C'est parce que nous sommes un petit pays par rapport aux États-Unis et à la Chine. Le Canada ne peut pas investir des centaines de milliards de dollars dans l'IA comme eux le font. En gros, nous devons accepter que le Canada est un pays de taille moyenne. Nous devrions évidemment collaborer avec d'autres pays de taille moyenne, mais nous ne sommes pas au rang des deux acteurs principaux.

J'ai une observation à formuler au sujet de Tumbler Ridge. J'ai travaillé pour une grande entreprise, Google, et je suis donc en mesure de voir l'autre côté des choses. Je peux voir le point de vue des entreprises technologiques. De leur point de vue, les choses ne sont pas aussi simples que les médias le laissent entendre. En effet, si vous demandiez combien d'autres utilisateurs avaient des profils aussi effrayants que celui de la personne responsable de la fusillade de Tumbler Ridge, je pense qu'il y en aurait des milliers, voire des dizaines de milliers. Les responsables doivent donc décider lesquels sont les plus terrifiants. Ils pourraient simplement alerter le gouvernement au

Senator McPhedran: What more could be done to focus on the protection of people, especially young people? If I could, I will ask you a rather general question: Do you think we're being realistic to ask that developments occur even within the private sphere with the human rights lens?

Mr. Hinton: Yes, when it is human rights versus the profits of big companies, we know who wins out.

Senator McPhedran: Is it something that we should be aspiring to do and trying to create legislation that carries that out, even if big companies are not so inclined?

Mr. Hinton: Yes. My view is that capitalism has given us all sorts of good things, but it needs to be directed. You need to constrain it with regulations so that the only way to make a lot of money is by doing things that are good for people.

If you can make a lot of money by doing things that are bad for people — like Mark Zuckerberg does — that's crazy. You need regulations to prevent that.

Senator Cuzner: Thank you very much. I was hoping for a little reprieve from Bill C-4, but, so far, it has been scary.

Are there national governments around the world that are leading the charge on advancing a regulatory regime around artificial intelligence, or AI?

Mr. Hinton: I'm not sure about governments, but in Britain, for example, they had the Bletchley Park meeting, which was good. Then the British government — the Sunak government — decided not to have any regulations.

But they did do something. They used 100 million pounds or thereabouts to set up a very good safety research group, and they have one of the best safety research groups in the world now that is doing a lot of leading research on the dangers of AI. The RAND Corporation in the United States is doing a lot of research on the dangers of AI. There are some very good groups working on researching the dangers.

In terms of regulations, there are European regulations, but the point is that they're not really addressing the current threats — the big threats. The European regulations, for example, have a clause in them that says that military uses of AI are exempt from these regulations because arms manufacturers, like France and other countries in Europe, don't want to be inhibited in making

sujet du nombre important de personnes qui ont de tels profils, mais ce n'est pas comme s'il y avait une seule personne effrayante dont ils n'auraient pas parlé.

La sénatrice McPhedran : Que pourrait-on faire de plus pour mettre l'accent sur la protection des personnes, en particulier des jeunes? Si vous me le permettez, je vais vous poser une question assez générale : pensez-vous qu'il soit réaliste de demander que du travail soit fait, même dans la sphère privée, en tenant compte des droits de la personne?

M. Hinton : Oui, lorsque les droits de la personne s'opposent aux profits des grandes entreprises, nous savons qui l'emporte.

La sénatrice McPhedran : Est-ce quelque chose que nous devrions aspirer à faire? Devrions-nous mettre au point des mesures législatives en ce sens, même si les grandes entreprises ne sont pas très enclines à se préoccuper de cela?

M. Hinton : Oui. Je crois que le capitalisme nous a apporté toutes sortes de bonnes choses, mais qu'on doit l'encadrer. Il faut le contraindre par des règlements afin que la seule façon de gagner beaucoup d'argent soit de faire des choses qui sont bonnes pour les gens.

Il est insensé de croire que l'on peut gagner beaucoup d'argent en faisant des choses qui nuisent aux gens, comme le fait Mark Zuckerberg. Il faut des règlements pour empêcher cela.

Le sénateur Cuzner : Merci beaucoup. J'espérais un peu de répit avec le projet de loi C-4, mais jusqu'ici, ce que l'on apprend est terrifiant.

Y a-t-il des gouvernements nationaux dans le monde qui mènent la charge pour la mise en place d'un cadre réglementaire concernant l'intelligence artificielle?

M. Hinton : Je ne sais pas ce qu'il en est des gouvernements, mais en Grande-Bretagne, par exemple, ils ont organisé la réunion de Bletchley Park, ce qui était une bonne chose. Ensuite, le gouvernement britannique — le gouvernement Sunak — a décidé de ne pas adopter de règlement.

Cependant, ils ont tout de même pris certaines mesures. Ils ont consacré environ 100 millions de livres sterling à la création d'un excellent groupe de recherche sur la sécurité, qui est aujourd'hui l'un des meilleurs au monde et qui mène de nombreuses recherches de pointe sur les dangers de l'IA. Aux États-Unis, la RAND Corporation mène de nombreuses recherches sur les dangers de l'IA. Il existe plusieurs groupes très compétents qui font de la recherche à ce sujet.

En ce qui concerne la réglementation, il existe un règlement européen, mais le fait est qu'il ne s'attaque pas vraiment aux menaces actuelles, aux grandes menaces. Le règlement européen, par exemple, contient un article stipulant que les utilisations militaires de l'IA sont exemptées du règlement, car les fabricants d'armes, comme la France et d'autres pays européens, ne veulent

swarms of drones, et cetera. They see that as the future of weapons. So there aren't very good regulations anywhere, as far as I know.

There were some regulations proposed in California that were rather mild, Bill SB-1047, that they got through the houses in California but were then vetoed by Gavin Newsom for reasons that aren't quite clear. Either big tech companies got to him, or he had some other reason. But there aren't strong regulations anywhere.

Senator Cuzner: Thank you. The challenge is that it's developing so fast that it will be tough to keep pace with the developments.

The number that has been cited is the loss of 300 million jobs up to 2030 in the next four or five years. Where will the bulk of those jobs be lost? What is your sense?

Mr. Hinton: First of all, when progress is very rapid like this, particularly if it's exponential, seeing into the future is like seeing in fog. You can see a short distance very clearly, but then the wall comes down, and you really know nothing.

If you look 10 years in the future, for example, we have no idea. Nobody has any idea what is going to happen 10 years in the future. You can see that, because if you look 10 years in the past and ask whether people would have predicted what we've got now, nobody — not even the strongest enthusiasts like me — would have predicted we would have the language models we have now. There is no way we can predict what is happening in 10 years, except that it will probably be something we didn't expect.

In a few years, there's a big diversity of opinions. So-called experts, like Gary Marcus, say it will only replace 2% of jobs. That's just crazy. There are other people who say it will replace all the jobs fairly quickly. That's crazy, too. But it's clear it's going to replace a lot of jobs and, in particular, jobs like call centres. The people there are poorly paid, badly trained and often don't know the answer to the question you want them to answer. AI can do a much better job there and will soon be doing that job. It's not clear what those people are going to do.

When the Industrial Revolution came along, farm labourers could retrain to be waiters in hotels or something. They could go into the service industries or go into paperwork. When AI comes along and replaces their jobs, any new jobs that are created, AI will be able to do those, too. It's not clear what's going to happen to those, and for most of the experts, I think it's fairly

pas être freinés dans la fabrication de nuées de drones, etc. Ils y voient l'avenir des armes. À ma connaissance, il n'existe pas de réglementation qui soit particulièrement efficace.

Un règlement assez modéré a été proposé en Californie. On pense notamment au projet de loi SB-1047, qui a été adopté par les deux chambres californiennes, mais qui a ensuite été rejeté par Gavin Newsom pour des raisons qui ne sont pas tout à fait claires. Il se peut que les grandes entreprises technologiques aient fait pression sur lui, mais c'était peut-être autre chose. Il reste qu'il n'existe aucun règlement rigoureux nulle part.

Le sénateur Cuzner : Merci. Le problème, c'est que cette technologie évolue si rapidement qu'il est très difficile de suivre la cadence.

Le chiffre avancé est la perte de 300 millions d'emplois d'ici 2030, soit dans les quatre ou cinq prochaines années. Où la plupart de ces emplois seront-ils perdus? Quelle est votre opinion à ce sujet?

M. Hinton : Tout d'abord, lorsque les progrès sont aussi rapides, en particulier s'ils sont exponentiels, le fait d'essayer de prédire l'avenir revient à essayer de voir dans le brouillard. On voit très clairement sur une courte distance, mais bientôt, un mur se dresse et on ignore tout de ce qui vient après.

Si nous nous projetons dans 10 ans, par exemple, nous n'avons aucune idée de ce qui nous attend. Personne ne sait ce qui va se passer dans 10 ans. Vous pouvez le constater, car si vous retournez 10 ans en arrière pour voir si les gens auraient pu prédire ce qui se passe présentement, personne, pas même les plus enthousiastes comme moi, n'aurait pu prévoir que nous aurions les modèles linguistiques que nous avons aujourd'hui. Il est impossible de prédire ce qui se passera dans 10 ans, sauf que ce sera probablement quelque chose que nous n'avions pas pu voir venir.

Pour ce qui est de ce qui va se passer dans quelques années, les opinions divergent beaucoup. Les soi-disant experts, comme Gary Marcus, affirment que cela ne remplacera que 2 % des emplois, ce qui est tout simplement absurde. D'autres affirment que cela remplacera tous les emplois assez rapidement, ce qui est tout aussi absurde. Il est toutefois clair que cela remplacera beaucoup d'emplois, en particulier ceux des centres d'appels. Les personnes qui y travaillent sont mal payées, mal formées et souvent incapables de répondre aux questions qu'on leur pose. Dans ce domaine, l'IA peut faire un bien meilleur travail que l'humain, et elle sera bientôt utilisée à cette fin. On ne sait pas encore ce que feront ces personnes.

Lorsque la révolution industrielle a eu lieu, les ouvriers agricoles ont pu se reconverter pour devenir serveurs dans des hôtels, ou s'occuper à autre chose. Ils ont pu se tourner vers le secteur des services ou vers le travail administratif. Lorsque l'IA arrivera et commencera à remplacer certains emplois, elle sera également capable d'effectuer tous les nouveaux emplois qui

clear that you're going to have a massive unemployment problem, and you're going to lose your tax base.

So 300 million, who knows? But it will be a lot.

Senator Senior: Thank you, Professor Hinton, particularly for your opening statement, of which I was able to follow maybe half, but I feel pretty good that I'm able to do that. I also feel as if we have a population that is completely unaware of what we're discussing today, which is part of the reason I'm happy we're doing this study, but public education doesn't seem to be a priority. It seems to be the last thing on the to-do list.

I am concerned that it's not top of mind for government, and even though there are many places doing research, as you've alluded to, how is this research being translated so that the public is aware or has some kind of awareness of what we're in for, whether it's by 2030 or even before then?

Do you know of any countries that are doing work that we could look for educating and preparing the public? Do you have any recommendations for how we could do that well? My other question is this: Were you consulted about the strategy the government will be releasing in the next few weeks?

Mr. Hinton: I wasn't really consulted. They gave me opportunities, but I was too busy to take them, so that's my fault.

I see my role now as getting the public to understand what AI is and what the dangers are, and the model I have roughly is this: With climate change, big energy companies could get governments to do what they wanted until the public became aware of climate change and began to understand it. They can still mislead the public a lot, but once the public understands there really is climate change, and it really is caused by burning carbon, then you get pressure on politicians that counterbalances the pressure from big energy companies.

My view is that you're getting pressure on politicians from big tech companies, and to counter that, we have to have pressure from the public, and we're not going to get that pressure until the public understands what is going on.

seront créés. On ne sait pas très bien ce qu'il adviendra des personnes qui seront touchées, mais pour la plupart des experts, il semble assez évident que cela entraînera un problème de chômage à grande échelle et une perte de recettes fiscales.

Le chiffre de 300 millions est-il plausible? Qui sait? Tout ce que nous savons, c'est qu'il y en aura beaucoup.

La sénatrice Senior : Merci, monsieur Hinton, en particulier pour votre déclaration liminaire. Je n'ai pu en suivre qu'environ la moitié, mais je suis plutôt contente de l'avoir fait. J'ai également l'impression que la population ignore complètement ce dont nous discutons aujourd'hui, ce qui explique en partie pourquoi je me réjouis que nous menions cette étude. Cela dit, la sensibilisation du public ne semble pas être une priorité, mais plutôt le dernier élément de la liste des choses à prendre en considération.

Je crains que ce ne soit pas une priorité pour le gouvernement, et même si de nombreux endroits mènent des recherches, comme vous l'avez dit, comment ces recherches sont-elles traduites afin que le public soit informé ou ait une certaine conscience de ce qui nous attend, que ce soit en 2030 ou même avant?

Connaissez-vous des pays qui prennent des mesures dont nous pourrions nous inspirer afin d'informer et de préparer le public? Avez-vous des recommandations à nous faire à cet égard? Ma deuxième question est la suivante : avez-vous été consulté au sujet de la stratégie que le gouvernement va publier dans les prochaines semaines?

M. Hinton : Je n'ai pas vraiment été consulté. On m'a donné des occasions, mais j'étais trop occupé pour les saisir. C'est donc moi qui suis fautif à ce chapitre.

Je considère que mon rôle consiste désormais à faire comprendre au public ce qu'est l'IA et quels sont ses dangers. Le modèle que j'ai en tête est le suivant : en ce qui concerne les changements climatiques, les grandes entreprises énergétiques pouvaient amener les gouvernements à leur permettre de faire ce qu'elles voulaient. Le vent a tourné lorsque le public a pris conscience de l'existence des changements climatiques et qu'il a commencé à les comprendre. Les entreprises peuvent encore tromper le public dans une large mesure, mais une fois que celui-ci comprend qu'il y a vraiment des changements climatiques et qu'ils sont vraiment causés par la combustion du carbone, une pression commence à s'exercer sur les politiciens, qui vient contrebalancer celle qu'exercent les grandes entreprises énergétiques.

Mon point de vue, c'est que pour contrer la pression que les grandes entreprises technologiques exercent sur les politiciens, il faut une pression qui vient du public, et cela ne se matérialisera pas tant et aussi longtemps que le public n'aura pas compris ce qui est en train de se produire.

The public is already very nervous about losing jobs. Young people are very nervous about how they should train so they will have a job in the future. I think it's good that they're nervous, because I think that will get them to put pressure on politicians to do something about it.

I see my role as educating the public, and that's pretty much all I have to say on that. I don't think governments are doing a very good job of it.

Senator Senior: Thank you.

Senator Petitclerc: Thank you for being with us. I want to follow up on that.

In a recent interview — and I want to quote you, Professor Hinton — you said at the BBC:

. . . the invisible hand is not going to keep us safe. So just leaving it to the profit motive of large companies is not going to be sufficient to make sure they develop it safely The only thing that can force those big companies to do more research on safety is government regulation.

You started to touch on that with my colleague. Specifically, when we're talking about regulation, where would you start? What would that look like in the context of Canada?

Mr. Hinton: Canada should have regulations on chatbots that apply to any chatbot use in Canada. My friends who have a company that produces chatbots won't like me for saying this, but I think they should have strong regulations on the testing of those chatbots before they're deployed. People have done a lot of work on what those regulations might look like.

Trump, of course, is trying to prevent any regulations in the U.S., so to get regulations on chatbots, we need Europe, Canada, South Korea and other countries to step up and insist that the chatbots used in their countries have gone through rigorous testing where the results have been reported.

Senator Petitclerc: Yesterday, we had some representatives of the government here, and I was asking about the importance of precautionary principles. The answer I seemed to have received was that they felt that Canada has a balanced approach in terms of taking advantage of the opportunities of AI while trying to do it at a safe pace. I want to get your input. Is that how you see it? After listening to you, I want to ask if we are even too

Le public est déjà très inquiet au sujet des possibles pertes d'emplois. Les jeunes sont très inquiets quant à la formation qu'ils doivent suivre pour trouver un emploi dans le monde de demain. Je pense que c'est une bonne chose qu'ils le soient, car cela les incitera à faire pression sur les politiciens pour qu'ils agissent.

Je considère que mon rôle est d'éduquer le public, et c'est à peu près tout ce que j'ai à dire à ce sujet. Je ne pense pas que les gouvernements s'acquittent très bien de cette tâche.

La sénatrice Senior : Je vous remercie.

La sénatrice Petitclerc : Merci d'être parmi nous. Je voudrais poursuivre dans la même veine.

Dans une récente interview à la BBC — et je voudrais vous citer, monsieur Hinton —, vous avez déclaré ceci :

[...] la main invisible ne va pas assurer notre sécurité. Il ne suffit donc pas de s'en remettre à la motivation financière des grandes entreprises pour s'assurer qu'elles développent des produits sûrs [...] La seule chose qui peut obliger ces grandes entreprises à faire davantage de recherche en matière de sécurité, c'est la réglementation gouvernementale.

Vous avez commencé à aborder ce sujet avec ma collègue. Plus précisément, quand on parle de réglementation, par où commenceriez-vous? À quoi cela ressemblerait-il dans le contexte canadien?

M. Hinton : Le Canada devrait disposer d'une réglementation sur les robots conversationnels qui couvrirait toute utilisation de cette technologie à l'échelle du pays. Mes amis qui possèdent une entreprise qui produit des robots conversationnels ne vont pas m'aimer pour avoir dit cela, mais je crois que le Canada devrait mettre en place une réglementation stricte pour encadrer les tests qu'il faudrait faire passer aux robots conversationnels avant leur déploiement. Beaucoup de travail a été fait pour établir ce à quoi cette réglementation pourrait ressembler.

Trump, bien sûr, tente de bloquer toute réglementation aux États-Unis. Pour avoir une réglementation sur les robots conversationnels, nous avons donc besoin que l'Europe, le Canada, la Corée du Sud et d'autres pays se mobilisent et insistent pour que les robots conversationnels utilisés sur leur territoire aient été soumis à des tests rigoureux dont les résultats auront été rendus publics.

La sénatrice Petitclerc : Hier, nous avons reçu des représentants du gouvernement, et je leur ai posé des questions sur l'importance des principes de précaution. Les représentants ont répondu qu'ils estimaient que le Canada avait une approche équilibrée consistant à tirer parti des possibilités offertes par l'IA tout en essayant de faire les choses à un rythme responsable. J'aimerais avoir votre avis. Voyez-vous les choses de cette

late for precautions. Are we jumping in? Is there still time? Do you know what I mean?

Mr. Hinton: Yes. I guess there are two questions there.

On the issue of whether it is too late, it may be. Many of the people at the forefront of the industry, which includes a lot of my ex-students — so I have insight into it by talking to them — think this is coming very fast. They think that, within a few years, there will be massive unemployment. I am not sure that is true. I think it may take longer. I think it will come, but it may take longer. We may not be too late on that front, but it will soon be too late.

On the issue of whether Canada is doing anything useful, I like the principle of saying that we're not one of the two major players here. We have to accept that. Therefore, we have to make alliances with the other middle-level countries, like all the countries in Europe. The European bloc could be one of the major players. We need to make alliances with them and, between us, insist that these U.S. big tech companies have to do safety checks on their chatbots and tell us the results. We need much stronger regulations there. That is something that is obvious and that we can obviously do.

Now, for example, if anybody releases a chatbot that hasn't had very extensive testing as to whether it may encourage people to commit suicide, that ought to be a crime.

The Chair: I would like to insert a question here, if I can, to follow up on Senator Petitclerc's question.

Professor, the question I have is this: We are told — and I think it's everybody's impression — that Canada has a very strong research capability and capacity. We're doing well in that area. Perhaps it is the only area we're doing well in at this point.

If you look particularly at the area of safety research, where can Canada make its mark as one of these medium-sized countries?

Mr. Hinton: We have world-leading safety researchers. David Kristjanson Duvenaud, who will testify later, is a world-leading safety researcher, as is Roger Grosse, who also works at the University of Toronto. There is a very good group of middle-aged professors there, not old professors like me. That is a place where we are a world leader.

façon? Après vous avoir écouté, je me demande s'il n'est pas déjà trop tard pour prendre des précautions. Devrions-nous aller de l'avant? Est-il encore temps de le faire? Voyez-vous où je veux en venir?

M. Hinton : Oui. Je pense qu'il y a deux questions ici.

Quant à savoir s'il est trop tard, c'est peut-être le cas. Beaucoup de personnes à l'avant-garde de l'industrie, dont beaucoup de mes anciens étudiants — j'ai un aperçu de la situation en discutant avec eux —, pensent que cela arrive très vite. Ils pensent que, d'ici quelques années, il y aura un chômage massif. Je ne suis pas certain que ce sera le cas. Je crois que cela prendra plus de temps. Je pense que nous en arriverons là, mais que cela prendra plus de temps que ce qu'eux prédisent. Il n'est peut-être pas trop tard à cet égard, mais bientôt, oui.

Quant à savoir si le Canada fait quelque chose d'utile, j'aime bien le fait que nous ne sommes pas l'un des deux principaux acteurs dans ce domaine. Nous devons l'accepter. Nous devons donc conclure des alliances avec les autres pays de niveau intermédiaire, comme l'ensemble des pays européens. Le bloc européen pourrait être l'un des principaux acteurs. Nous devons conclure des alliances avec eux et, ensemble, nous devons réclamer que les grandes entreprises technologiques américaines effectuent des contrôles de sécurité sur leurs robots conversationnels et nous en communiquent les résultats. Nous avons besoin d'une réglementation beaucoup plus stricte dans ce domaine. Il s'agit d'une évidence, et c'est quelque chose que nous pouvons absolument faire.

À l'heure actuelle, par exemple, si quelqu'un commercialise un robot conversationnel qui n'a pas fait l'objet de tests approfondis visant à établir s'il peut inciter les gens à se suicider, cela devrait être considéré comme un crime.

La présidente : Si vous me le permettez, j'aimerais poser une question pour faire suite à celle que la sénatrice Petitclerc a posée.

Monsieur Hinton, ma question est la suivante : on nous dit — et je pense que c'est l'impression de tout le monde — que le Canada dispose d'une capacité et d'un potentiel de recherche très importants. Nous obtenons de bons résultats dans ce domaine. C'est peut-être le seul domaine dans lequel nous obtenons de bons résultats à l'heure actuelle.

Si l'on examine plus particulièrement le domaine de la recherche en matière de sécurité, où le Canada peut-il se démarquer en tant que pays de taille moyenne?

M. Hinton : Nous avons des chercheurs de renommée mondiale dans le domaine de la sécurité. David Kristjanson Duvenaud, qui témoignera plus tard, est un chercheur de renommée mondiale dans le domaine de la sécurité, tout comme Roger Grosse, qui travaille également à l'Université de Toronto. Il y a là un très bon groupe de professeurs d'âge moyen, pas des

In terms of developing new, cutting-edge AI, there's a problem in that it requires huge amounts of capital. It's very difficult to see how Canada can do that. The leading researchers have to work with big U.S. tech companies.

What is happening now is that professors have two jobs. They have a job where they do their research and another job where they interact with a big company to get the resources. I can't see any way around that.

One of my students, for example, left the Vector Institute in Toronto because, there, he could get 100 GPUs that could be dedicated to him. He went to work for Musk, where he could get 100,000 GPUs. There's no way Canada can do that.

The Chair: Thank you.

Senator Arnold: Thank you very much, professor, for being with us today. Thank you for expressing your very lofty goal of educating the public. I'm curious about this: What kind of advice do you give to your students? What kinds of human beings do you think will be the most effective?

I'm not talking about jobs necessarily, but I'm talking more about the kinds of humans in the future that we need to combat some of this and to have the critical thinking skills to be able to make the world a better place in the future, given everything that we know about this.

Mr. Hinton: Yes, that's a difficult question. My view is that people's commitment to being moral is something critical. When they're students, they arrive either with it or not much of it. Some people have a lot of it; some people don't have much of it. I don't know how you create it. I think that happens when you're quite young. One piece of advice is to look at how Trump was raised and do the opposite.

Among those, there are some who have a very strong moral compass, like Ilya Sutskever, who fired Sam Altman, at least for a few days, and then went off and set up his own company that is trying to develop safe superintelligence. This is also true of the people I mentioned at the University of Toronto whom I know very well — David Kristjanson Duvenaud, who will testify later, and Roger Grosse — both care a lot about the future of humanity. That's why they are working on safety research. It was Roger Grosse who encouraged me to get interested in safety several years ago.

professeurs âgés comme moi. C'est un domaine dans lequel nous sommes des leaders mondiaux.

En ce qui concerne le développement d'une IA nouvelle, d'une IA de pointe, le problème est qu'il faut d'énormes capitaux. Il est très difficile d'imaginer comment le Canada pourrait avoir accès à des ressources de cette ampleur. Les chercheurs de pointe doivent travailler avec les grandes entreprises technologiques américaines.

Actuellement, les professeurs ont deux emplois. Ils ont un emploi où ils font leurs recherches et un autre où ils interagissent avec une grande entreprise pour obtenir des ressources. Je ne vois pas comment l'on pourrait contourner cette dynamique.

Par exemple, l'un de mes étudiants a quitté l'Institut Vector de Toronto, où il pouvait compter sur 100 processeurs graphiques qui lui étaient réservés. Il est allé travailler pour Musk, où on lui donnait accès à 100 000 processeurs graphiques. Le Canada n'a aucun moyen d'égaliser cela.

La présidente : Je vous remercie.

La sénatrice Arnold : Merci beaucoup, monsieur, d'être avec nous aujourd'hui. Je vous remercie d'exprimer votre objectif très noble d'éduquer le public. Je suis curieuse de savoir ceci : quel genre de conseils donnez-vous à vos étudiants? Selon vous, quels types d'êtres humains seront les plus efficaces?

Je ne parle pas nécessairement d'emplois, mais plutôt du genre d'humains dont nous aurons besoin à l'avenir pour lutter contre certains de ces phénomènes et pour avoir l'esprit critique nécessaire afin de pouvoir rendre le monde meilleur à l'avenir, compte tenu de tout ce que nous savons à ce sujet.

M. Hinton : Oui, c'est une question difficile. À mon avis, l'engagement moral des gens est essentiel. Lorsqu'ils sont étudiants, ils arrivent avec ou sans cet engagement. Certaines personnes en ont beaucoup; d'autres n'en ont pas beaucoup. Je ne sais pas comment on peut le créer. Je pense que cela arrive quand on est très jeune. Un conseil serait d'examiner la façon dont Trump a été élevé et de faire le contraire.

Parmi ces gens, il y en a qui ont un sens moral très fort, comme Ilya Sutskever, qui a congédié Sam Altman, au moins pendant quelques jours, puis qui a créé sa propre entreprise qui essaie de développer une superintelligence sécuritaire. C'est aussi le cas des gens que j'ai mentionnés à l'Université de Toronto et que je connais très bien — David Kristjanson Duvenaud, qui témoignera plus tard, et Roger Grosse —, qui ont tous deux à cœur l'avenir de l'humanité. C'est pourquoi ils font de la recherche sur la sécurité. C'est Roger Grosse qui m'a encouragé à m'intéresser à la sécurité il y a plusieurs années.

I don't have any advice on how you make people more like that. I get to see them when they're graduate students, and they're either like that or they're not.

Senator Arnold: Thank you.

Senator Boudreau: Thank you, professor, for being here. It seems clear that, like climate change — which is a global concern — the development of AI also needs to be a global concern, and I think it is, but with a shared approach to its management.

You've touched on it, but perhaps you could address it in more detail. Which international forums is Canada currently engaged with on this topic? Where is the most meaningful work currently happening to manage the development of AI on a global level, both from a development perspective and also from a regulatory perspective?

Mr. Hinton: Yes. Yoshua Bengio will be able to answer this much better than me. He's put a lot more effort into interacting with these international organizations. I have mainly focused on educating the public on what AI is.

I can give you some insight. There's an organization, which has a whole bunch of letters that I forget, which encourages collaboration between Chinese and western academics. I went to a meeting in Shanghai of that organization. That's quite effective.

The main comment I have on this is that if you're looking for international collaboration, countries will collaborate where their interests are aligned, and they'll only pretend to collaborate where their interests are not aligned.

For example, for China and the U.S., their interests are anti-aligned on things like cyberattacks because they're doing it to each other, and on things like corrupting elections because they're doing it to each other.

They're aligned on two things: One is that they don't want terrorists creating nasty viruses, and they will collaborate on that. At least, if they had a sensible government in the U.S., they would.

They will also collaborate on how to prevent AI from taking over. On that particular issue, all the countries are aligned. No government wants AI to take over. If any country discovered how to prevent that, they would very happily tell all the other countries. They wouldn't keep it a secret. They would immediately let other countries know, because they don't want AI taking over other countries.

Je n'ai pas de conseils à donner sur la façon de renforcer l'engagement moral des étudiants. J'ai l'occasion de les côtoyer lorsqu'ils sont aux cycles supérieurs, et soit ils ont cela en eux, soit ils ne l'ont pas.

La sénatrice Arnold : Merci.

Le sénateur Boudreau : Merci, monsieur, d'être ici. Il semble clair que, tout comme les changements climatiques, le développement de l'intelligence artificielle doit également être une préoccupation mondiale, et je pense que c'est le cas, mais il faut adopter une approche commune à l'égard de sa gestion.

Vous avez déjà abordé le sujet, mais vous pourriez peut-être nous en parler plus en détail. À quels forums internationaux le Canada participe-t-il actuellement dans ce domaine? Où se fait le travail le plus important pour gérer l'intelligence artificielle à l'échelle mondiale, tant du point de vue du développement que du point de vue de la réglementation?

M. Hinton : Yoshua Bengio pourra répondre beaucoup mieux que moi à cette question. Il déploie beaucoup plus d'efforts pour interagir avec ces organisations internationales. Je me suis principalement concentré sur l'éducation du public sur ce qu'est l'intelligence artificielle.

Je peux vous donner une idée. Il existe une organisation, dont le nom a toute une série de lettres que j'ai oubliées, qui encourage la collaboration entre les universitaires chinois et occidentaux. J'ai assisté à une réunion de cette organisation à Shanghai. C'est très efficace.

Le principal commentaire que j'ai à faire à ce sujet, c'est qu'à l'échelle internationale, les pays collaboreront lorsque leurs intérêts sont les mêmes, et feront semblant de collaborer lorsqu'ils ne le sont pas.

Par exemple, les intérêts de la Chine et des États-Unis en ce qui a trait aux cyberattaques sont à l'opposé parce qu'ils s'en prennent les uns aux autres. Il en va de même pour ce qui est de la corruption des élections, par exemple.

Ils s'entendent sur deux choses : premièrement, ils ne veulent pas que les terroristes créent de sales virus, et ils collaboreront à cet égard. Du moins, c'est ce que ferait un gouvernement raisonnable aux États-Unis.

Ils collaboreront également sur la façon d'empêcher l'intelligence artificielle de prendre le dessus. Sur cette question particulière, tous les pays sont sur la même longueur d'onde. Aucun gouvernement ne veut que l'intelligence artificielle prenne le dessus. Si un pays découvrait comment prévenir une telle situation, il le dirait volontiers à tous les autres. Il ne garderait pas le secret. Il en informerait immédiatement les autres pays, parce qu'il ne veut pas que l'intelligence artificielle prenne le contrôle.

So from the point of view of this existential threat, it's much like a global nuclear war. At the height of the Cold War in the 1950s, the U.S. and Russia were collaborating on how to prevent a global thermonuclear war because it wasn't in either of their interests.

But on all these other things, such as autonomous weapons, cyberattacks and corrupting elections, they'll pretend to collaborate, but they won't.

Senator Boudreau: That's scary.

Senator Osler: Thank you very much, Professor Hinton, for being here today.

My question is this: If you were a Canadian parliamentarian, where would you focus your work on AI, and how would you do it? Would it be through studies supporting research? Would it be through legislation? Would it be through regulation? Where would you start if you were a Canadian parliamentarian?

Mr. Hinton: The first thing to say is that I'm a scientist. I helped develop AI. I understand how it works. I'm not a policy person, so I'm a complete amateur at policy. I'm not testifying as an expert here. I'm testifying as an amateur.

I would focus on a couple of issues: One, ensuring good tests were done before chatbots were released. That's very obvious. You can do that, and that's where most of the legislative effort has been so far. I think that's a good thing to do.

The second thing I would focus on is what to do about unemployment. In particular, if there's going to be high unemployment, what should we do about taxation? Where is the government going to get its money from if you have high unemployment? How are you going to deal with all those unemployed people?

I tend to have socialist instincts. I believe in capitalism, but I think it needs to be strongly regulated so that you can only make money by doing things that are going to be good for society. Developing the internet, for example, was, on the whole, very good for society, and it's fine that people made a lot of money doing that. Developing social media, to begin with, it looked like it might be good, but it was fairly clear after not very long that it was going to have mainly negative consequences, and it was up to the government to prevent people from making a lot of money that way.

That's a principle that I think should inform policy, but I'm not a policy person.

Donc, du point de vue de cette menace existentielle, c'est un peu comme une guerre nucléaire mondiale. Au plus fort de la guerre froide dans les années 1950, les États-Unis et la Russie collaboraient sur la façon de prévenir une guerre thermonucléaire mondiale parce qu'elle n'était pas dans l'intérêt de l'un ou l'autre.

Mais pour toutes les autres questions, comme les armes autonomes, les cyberattaques et la corruption des élections, ils feront semblant de collaborer, mais ils ne collaboreront pas.

Le sénateur Boudreau : C'est effrayant.

La sénatrice Osler : Merci beaucoup d'être avec nous aujourd'hui, monsieur Hinton.

Ma question est la suivante : si vous étiez un parlementaire canadien, où concentreriez-vous votre travail en ce qui a trait à l'intelligence artificielle, et comment le feriez-vous? Est-ce que vous réaliseriez des études à l'appui de la recherche? Est-ce que vous adopteriez des mesures législatives, des règlements? Par où commenceriez-vous si vous étiez un parlementaire canadien?

M. Hinton : La première chose à dire, c'est que je suis une scientifique. J'ai contribué au développement de l'intelligence artificielle. Je comprends comment elle fonctionne. Je ne suis pas un spécialiste des politiques, alors je suis un amateur dans ce domaine. Je ne témoigne pas ici en tant qu'expert.

Je me concentrerai sur quelques points : premièrement, il faut effectuer des tests efficaces avant que les robots conversationnels ne soient lancés. C'est évident. Vous pouvez le faire, et c'est là que la plupart des efforts législatifs ont été déployés jusqu'à présent. Je pense que c'est une bonne chose à faire.

Je me centrerais également sur le chômage. Si le taux de chômage est élevé, qu'allons-nous faire avec l'impôt? Où le gouvernement prendra-t-il son argent et comment allez-vous vous occuper de tous ces chômeurs?

J'ai tendance à avoir des instincts socialistes. Je crois au capitalisme, mais je pense qu'il doit être fortement réglementé afin que l'on puisse seulement faire de l'argent en faisant ce qui est bon pour la société. Le développement d'Internet, par exemple, a été dans l'ensemble très bon pour la société, et c'est bien que des gens aient fait beaucoup d'argent avec sa création. Le développement des médias sociaux semblait être une bonne chose au départ, mais on a vite réalisé qu'ils allaient surtout entraîner des conséquences négatives, et il incombait au gouvernement d'empêcher les gens de faire beaucoup d'argent de cette façon.

C'est un principe qui, à mon avis, devrait guider les politiques, mais je ne suis pas un spécialiste en la matière.

Senator Osler: I don't think there's anyone — certainly not around this table — that would call you an amateur at anything. Around this committee table, in fact, you do have a diverse group of people who are scientists and researchers.

Thank you for that.

The Chair: I will interject a question again before we go into the second round. Yesterday, which was the first day of our study, we heard a little bit — or a lot — about guardrails and the need for them. We heard less about regulation and legislation to control some of the safety issues that are emerging and, we expect, will likely emerge using AI that has to do with this new technology.

You may not want to speak as an expert, but how well are we doing in this regard? You keep hearing the same questions, Professor. How urgently do we need to be pushing for legislation and in what areas, in your view?

The second part of the question is as follows: What are we doing as a country about the existential threat in a positive way?

Mr. Hinton: For the existential threat, nobody knows how to solve it. We're at a point where we need a lot of different research efforts. Yoshua Bengio has a suggestion of how to solve it that is quite different from my suggestion of making them love us. We're building them, so build in maternal instincts if we can figure out how to do that. We should certainly be funding a lot more research on that, but we don't know the solution.

If we don't find the solution before they become more intelligent than us, I think we're toast, so that seems like a good thing to fund.

It's clear that we should have strong regulations on the testing of chatbots, and we don't have those yet. It's fairly clear that we'll get high unemployment, and we don't know what to do about that yet. We should fund economic research on how to deal with taxation.

Bill Gates has recently suggested — he may have some bad behaviours, but he's very smart — that we need to tax AI agents. When you replace a worker with an AI agent that does the same job, you need to tax that AI agent; otherwise your tax base disappears.

La sénatrice Osler : Je ne pense pas que quiconque — certainement pas autour de cette table — oserait dire que vous êtes un amateur dans quoi que ce soit. Nous avons un groupe diversifié composé de scientifiques et de chercheurs.

Je vous en remercie.

La présidente : Je vais poser une autre question avant que nous passions à la deuxième série. Hier était le premier jour de notre étude. Nous avons entendu parler un peu — ou beaucoup — de la nécessité des garde-fous. Nous avons moins entendu parler de la réglementation et de la législation visant à contrôler certains des problèmes de sécurité qui émergent et qui émergeront probablement en raison de l'intelligence artificielle et du recours à cette nouvelle technologie.

Vous ne voulez peut-être pas parler en tant qu'expert, mais où en sommes-nous à cet égard? On vous pose sans cesse les mêmes questions, monsieur. Dans quelle mesure devons-nous exercer des pressions pour des lois et dans quels domaines, à votre avis?

La deuxième partie de la question est la suivante : que faisons-nous de manière positive en tant que pays au sujet de la menace existentielle?

M. Hinton : Pour ce qui est de la menace existentielle, personne ne sait comment régler la question. Nous en sommes à un point où nous avons besoin de nombreux efforts de recherche différents. Yoshua Bengio propose une solution très différente de la mienne, qui consiste à amener ces technologies à nous aimer. Nous les créons, alors nous devrions y intégrer l'instinct maternel si nous pouvons trouver un moyen de le faire. Nous devrions certainement financer beaucoup plus de recherches à ce sujet, mais nous ne connaissons pas la solution.

Si nous ne trouvons pas la solution avant que ces outils deviennent plus intelligents que nous, je pense que nous sommes cuits, alors il me semble que le financement en ce sens soit une bonne idée.

Il est clair que nous devrions avoir des règlements rigoureux sur la mise à l'essai des robots conversationnels, et nous ne les avons pas encore. Il est assez évident que le taux de chômage sera élevé, et nous ne savons pas encore quoi faire à ce sujet. Nous devrions financer la recherche économique sur la façon de gérer la fiscalité.

Bill Gates a récemment laissé entendre — il a peut-être de mauvais comportements, mais il est très intelligent — qu'il faudrait imposer les agents de l'intelligence artificielle. Lorsque vous remplacez un travailleur par un agent d'intelligence artificielle qui fait le même travail, il faudrait l'imposer; sinon, l'assiette fiscale disparaîtra.

Big tech companies will fight that tooth and nail, of course. They think that all the profits should go to big tech companies. It's going to be a very hard thing to do, but somehow you have to have a tax base.

Senator McPhedran: Professor Hinton, a lot of what you're telling us today is making me think of various writings on dystopia, but particularly as Margaret Atwood sees the future in the battle with corporations.

I'm sorry if you feel this question is too political, but we have a bill before us that we've had now for some years, and we have groups of senators who, for more than 20 years, have advocated for a guaranteed basic livable income. I know it's not your field, but as a Canadian, do you think that would go some way to the kind of amelioration that we, obviously, have to start to plan for?

Mr. Hinton: Yes, I think it will go some way. It won't resolve dignity issues if unemployed people feel they have lost their sense of purpose, but it will go some way.

I know of one very successful experiment on Universal Basic Income, or UBI, in Wales, where they took teenage fostered children, who, at the age of 18, lose all their benefits and are put out into the world. It's a very defined population, so you can't have other people move in and claim the UBI. You could only receive the UBI if you were one of those people. I don't know the amount, but it wasn't huge, but it made a huge difference. It made that transition for them work much better.

Obviously, a problem with UBI is that you're going to get a lot of people trying to take advantage of it if it's universal, but I think there may be no alternative to that. It's just that when a lot of people need it, how are you going to pay for it, because you lost the tax base?

Senator McPhedran: Good point. Yes. Thank you very much.

Senator Burey: Professor Hinton, you're talking about how fast this is developing and how quickly we may be past the point of no return. What are some of the rate-limiting steps in the development of AI? I've listened to programs talking about chips, energy and water. Are there any such steps that could slow it down?

Mr. Hinton: Yes, there is a diversity of opinions, which is good. Some people think that the era when you could just make it work much better by scaling it up is coming to an end, and they think it's coming to an end, because we're running out of cheap data to train it on. I don't entirely agree with them. I think

Les grandes entreprises technologiques s'y opposeront avec vigueur, bien sûr. Elles croient que tous les profits doivent leur revenir. Ce sera très difficile à faire, mais d'une façon ou d'une autre, il faut avoir une assiette fiscale.

La sénatrice McPhedran : Monsieur Hinton, une grande partie de ce que vous nous dites aujourd'hui me fait penser à divers écrits sur la dystopie, mais surtout à Margaret Atwood, qui voit l'avenir dans la lutte contre les entreprises.

Je suis désolée si vous estimez que cette question est trop politique, mais nous sommes saisis d'un projet de loi qui est à l'étude depuis quelques années, et nous avons des groupes de sénateurs qui prônent un revenu de base garanti depuis plus de 20 ans maintenant. Je sais que ce n'est pas votre domaine, mais en tant que Canadien, pensez-vous que cela contribuerait dans une certaine mesure au genre d'améliorations que nous devons évidemment commencer à planifier?

M. Hinton : Oui, je pense qu'une telle mesure pourrait aider d'une certaine façon. Elle ne réglera pas les problèmes de dignité si les chômeurs ont l'impression d'avoir perdu leur raison d'être, mais cela aidera dans une certaine mesure.

Je suis au fait d'une expérience très réussie sur le revenu de base universel au pays de Galles, avec des adolescents en famille d'accueil qui, à l'âge de 18 ans, perdent tous leurs avantages et doivent se débrouiller seuls. Il s'agit d'une population très définie, et seuls ceux qui en font partie pouvaient avoir accès au revenu de base universel. Je n'en connais pas le montant, mais ce n'était pas énorme. Il a toutefois été d'une grande aide pour ces jeunes et les a aidés à faire la transition.

Évidemment, le problème avec le revenu de base universel, c'est que beaucoup de gens essaieront d'en profiter, mais je ne pense pas qu'il y ait de solution de rechange pour cela. Toutefois, si un grand nombre de personnes y ont recours, il faut se demander comment nous financerons le système, parce que nous allons perdre l'assiette fiscale.

La sénatrice McPhedran : C'est un bon point, oui. Merci beaucoup.

La sénatrice Burey : Monsieur Hinton, vous parlez de la vitesse à laquelle les choses évoluent et de la rapidité avec laquelle nous pourrions dépasser le point de non-retour. Quelles sont certaines des étapes qui limitent le taux de développement de l'intelligence artificielle? Certains programmes évoquent les puces, l'énergie et l'eau. Existe-t-il de telles mesures qui pourraient ralentir le processus?

M. Hinton : Oui, il y a une diversité d'opinions à ce sujet, ce qui est bien. Certaines personnes pensent que l'époque où l'on pouvait simplement améliorer la technologie en procédant à une mise à l'échelle tire à sa fin parce que nous manquons de données bon marché pour l'alimenter. Je ne suis pas tout à fait

they're wrong, but they might be right. It may be that it's going to slow down because of limitations on data.

Obviously, chips and energy are the main limitations at present. There is also a limitation, which is that we may well need new scientific ideas. In 2017, a group of people introduced transformers at Google, which was a new scientific idea that made a huge difference, and we wouldn't have today's large language models, or LLMs, but for that fact. We would have LLMs, but they wouldn't be quite as good.

I don't believe in slowing it down. I don't think that will be possible, because there are so many good uses. I think with, for example, nuclear weapons, it's conceivable that they could have not developed the hydrogen bomb, and that's what Oppenheimer wanted because it's only good for blowing things up.

But with AI, it's going to be hugely valuable for health care and education. It's going to make almost any industry more efficient. It's crazy that this thing that is going to make huge increases in productivity should be bad. Intrinsicly, it's not good or bad; it is just going to lead to a big increase in productivity. It is our political system that doesn't know how to handle it. We have a profit-driven system, and that's going to lead to all sorts of bad things, because it's not properly regulated.

That's my view.

Senator Hay: I need some more education, so I'm going to ask an education question. I heard loud and clear — and I agree with you — that we need regulation of chatbots in our country with a lot of testing before broad use and regulations on what they are, how they're used, what their purpose is, et cetera.

How would we get around fluid digital borders? Are we able to “geofence” it, if that's a word, so that the only chatbots being used in Canada are ones that are regulated? Is that possible?

Mr. Hinton: I don't know; I'm not an expert on that. I have no idea. I'm sorry.

Senator Hay: It is okay.

Senator Petitclerc: I have a short question. A few times you mentioned the importance of testing chatbots. I just want to understand what exactly it is that we are able to test. For example, are we testing biases? Can we do that? When you talk

d'accord avec eux. Je pense qu'ils ont tort, mais ils ont peut-être raison. Il se peut que l'évolution ralentisse en raison d'un manque de données.

Évidemment, les puces et l'énergie sont les principales limites à l'heure actuelle. Aussi, nous pourrions avoir besoin de nouvelles idées scientifiques. En 2017, un groupe de personnes a mis en place des transformateurs chez Google, une nouvelle idée scientifique qui a fait une énorme différence, et nous n'aurions pas les grands modèles de langage d'aujourd'hui si ce n'était de cela. Nous en aurions, mais ils ne seraient pas aussi efficaces.

Je ne crois pas qu'il faille ralentir le processus. Je ne pense pas que ce sera possible, car les bonnes utilisations sont nombreuses. En ce qui a trait aux armes nucléaires, par exemple, on aurait pu ne pas mettre au point la bombe à hydrogène, et c'est ce que souhaitait Oppenheimer, parce qu'elle ne sert qu'à faire exploser des choses.

Mais l'intelligence artificielle sera extrêmement utile pour les soins de santé et l'éducation. Elle permettra à presque toutes les industries d'être plus efficaces. Il est fou de penser que cette invention qui va faire croître la productivité de manière très importante puisse être mauvaise. Intrinsicquement, elle n'est ni bonne ni mauvaise; elle va simplement mener à une forte augmentation de la productivité. C'est notre système politique qui ne sait pas comment gérer la situation. Nous avons un système axé sur le profit, et cette technologie va entraîner toutes sortes de conséquences négatives, parce qu'elle n'est pas correctement réglementée.

C'est mon point de vue.

La sénatrice Hay : J'ai besoin d'être mieux informée, alors je vais poser une question en ce sens. J'ai bien entendu — et je suis d'accord avec vous — que nous avons besoin de réglementer les robots conversationnels au pays, et d'effectuer plusieurs tests avant leur utilisation à grande échelle. Il faut aussi réglementer ce qu'ils sont, comment ils sont utilisés, à quoi ils servent, etc.

Comment pourrions-nous aborder la question de la fluidité des frontières numériques? Sommes-nous en mesure d'ériger des « clôtures électroniques », si je puis dire, de sorte que seuls les robots conversationnels réglementés puissent être utilisés au Canada? Est-ce possible?

M. Hinton : Je ne le sais pas; je ne suis pas un expert dans ce domaine. Je n'en ai aucune idée. Je suis désolé.

La sénatrice Hay : Il n'y a pas de problème.

La sénatrice Petitclerc : J'ai une brève question. Vous avez mentionné à quelques reprises l'importance de tester les robots conversationnels. Je veux simplement comprendre ce que nous sommes en mesure de tester exactement. Est-ce que nous testons

about the importance of testing chatbots, can you provide us with some examples of what we should be testing?

Mr. Hinton: Bias is a somewhat separate issue. It will have the biases of the data it was trained on, but at least you can freeze the weights in the chatbot and measure its biases. With a person, it's hard to do that. With a person, you get the "Volkswagen effect," which is that, as soon as it knows it is being tested, it changes its behaviour. With AIs, you can freeze them so you can see the bias.

For bias, our aim with AI should be to make it less biased than the system it replaces, not unbiased. You will never make it unbiased, but if you keep making them less biased than the system they replace, you will make progress. Bias doesn't worry me so much, but that might be related to the fact that I'm an old, White male.

The things we should be testing for are things like whether they encourage kids to commit suicide. Is it easy to overcome the human reinforcement that has gone into them? After they are trained to be good at predicting the next word, they are trained not to say inappropriate things. That happens by getting people to try to get them to say inappropriate things and then telling them not to do that. That's pretty easy to overcome, particularly if you release the weights of the models so people could run the model on their own computers. They could easily overcome that training.

Certainly, people should be looking for how easy it is to overcome that training. For example, will they tell you how to make an improvised explosive device? If you ask them, they say, "Oh, I can't tell you that," and then you can trick them into telling you. So, people should be doing a lot of research on making it harder to trick them.

The Chair: This brings us to the end of the first panel. I would like to thank you, Professor Hinton, for your testimony today.

Joining us for our second panel, we welcome Wyatt Tessari L'Allié, Founder and Executive Director, AI Governance and Safety Canada; David Kristjanson Duvenaud, Associate Professor of Computer Science, University of Toronto; and, via

les préjugés, par exemple? Est-ce qu'il est possible de le faire? Pouvez-vous nous donner des exemples des essais que nous devrions faire sur les robots conversationnels?

M. Hinton : La présence de certains biais constitue un enjeu quelque peu différent. Un robot conversationnel est doté de certains biais en fonction des données avec lesquels il a été programmé, mais il est toujours possible de « geler » ces biais, dans une certaine mesure. Avec une personne en chair et en os, c'est difficile à faire, car on obtient souvent l'« l'effet Volkswagen », c'est-à-dire que dès que la personne se rend compte qu'elle est testée, elle tend à modifier son comportement.

En ce qui concerne les biais et les préjugés, notre objectif avec l'intelligence artificielle devrait être de la rendre moins partiiale que le système qu'elle remplace, et non pas complètement impartiale. Il ne sera probablement jamais possible de créer une intelligence artificielle complètement impartiale, mais je crois que nous pouvons continuer à perfectionner nos modèles. Le problème des préjugés ne m'inquiète pas particulièrement, mais cela tient peut-être au fait que je suis un homme blanc d'un certain âge.

Ce que nous devons tester, c'est notamment si certaines formes d'intelligence artificielle sont capables de pousser des enfants au suicide. Est-il facile de surmonter le renforcement humain qui leur a été inculqué? Après avoir été entraînés à prédire le mot suivant, ils sont programmés, ou entraînés si l'on préfère, à ne pas émettre de commentaires inappropriés. Pour ce faire, on demande à des personnes d'essayer de leur faire dire certains commentaires inappropriés, puis on leur demande de cesser. La programmation d'une intelligence artificielle est assez facile à surmonter, en particulier si on libère le poids des modèles afin que l'utilisateur puisse exécuter le modèle à partir de son propre ordinateur.

Pour le moment, il est particulièrement facile de contourner la manière dont les intelligences artificielles sont programmées. Par exemple, un robot conversationnel peut-il expliquer à un utilisateur comment fabriquer un engin explosif improvisé? Bien entendu, si l'utilisateur demande directement au robot de quelle manière procéder, celui-ci lui répondra qu'il n'est pas autorisé à exécuter ce type de tâche. Par contre, il est possible au moyen de différentes tactiques de finir par piéger le robot. Par conséquent, nous devons continuer à peaufiner l'intelligence artificielle pour l'empêcher de se faire manipuler de la sorte.

La présidente : Voilà qui termine la comparution du premier groupe d'invités. Monsieur Hinton, je tiens à vous remercier pour votre témoignage aujourd'hui.

Chers collègues, pour notre deuxième groupe de témoins, nous avons le plaisir d'accueillir les invités suivants: Wyatt Tessari L'Allié, fondateur et directeur général, Gouvernance et Sécurité de l'IA Canada, GSIA Canada; David Kristjanson Duvenaud,

video conference, Inioluwa Deborah Raji, Researcher, University of California, Berkley. Thank you for joining us today.

For your opening statements, you will have five minutes, followed by questions from committee members. Mr. Tessari L'Allié, the floor is yours.

[Translation]

Wyatt Tessari L'Allié, Founder and Executive Director, AI Governance and Safety Canada (AIGS Canada): Thank you, Madam Chair. Members of the committee, thank you for inviting me to be here today. It's an honour.

AI Governance and Safety Canada, or AIGS Canada, is a non-partisan not-for-profit organization and a community of people across the country. We started with this question: "What can we do in Canada, and from Canada, to ensure that advanced AI is safe and beneficial for all?"

Since 2022, we have been providing the federal government with public policy recommendations, including submissions on the AI and data bill and multiple appearances before parliamentary committees.

[English]

In 2012, researchers under Dr. Hinton's supervision developed a revolutionary technique that powered the era of single-purpose AI systems like Alexa, Google Translate and social media algorithms while introducing risks of bias, privacy loss and online echo chambers.

In 2022, ChatGPT ushered in the era of generative AI: chatbots that could answer complex questions, write sections of code and generate lifelike images and videos. This also brought new challenges to overcome, such as deepfake scams and misinformation, cyberattacks and chatbots that can talk people into committing harm.

In early 2026, another major jump in capabilities pushed us firmly into the era of AI agents. Unlike chatbots that simply respond to a prompt, AI agents are systems that can take action in the real world, working autonomously for hours and overcoming hurdles along the way. They can, for example, be used to develop an app from scratch, not only writing the code, but also opening it and debugging issues until it is functional. Users are also starting to give AI agents access to their

professeur agrégé en informatique, Université de Toronto, qui comparait à titre personnel; et, par vidéoconférence, Inioluwa Deborah Raji, chercheuse, Université de la Californie, Berkeley.

Vous disposerez de cinq minutes pour vos déclarations liminaires, qui seront suivies des questions des membres du comité. Monsieur Tessari L'Allié, à vous la parole, je vous prie.

[Français]

Wyatt Tessari L'Allié, fondateur et directeur général, Gouvernance et Sécurité de l'IA Canada (GSIA Canada) : Merci, madame la présidente. Mesdames et messieurs les membres du comité, je vous remercie de l'honneur que vous me faites de m'avoir invité.

Gouvernance et Sécurité de l'IA Canada est un organisme à but non lucratif non partisan et une communauté de personnes à travers le pays. Notre point de départ est la question suivante : que pouvons-nous faire, au Canada et à partir du Canada, pour s'assurer que l'IA avancée soit sécuritaire et bénéfique pour tous?

Depuis 2022, nous fournissons au gouvernement fédéral des recommandations de politiques d'intérêt public, comme nos soumissions sur le projet de loi sur l'intelligence artificielle et les données et nos multiples témoignages aux comités parlementaires.

[Traduction]

En 2012, des chercheurs sous la supervision de M. Hinton ont mis au point une technique révolutionnaire qui a donné naissance à l'ère des systèmes d'intelligence artificielle à usage unique tels qu'Alexa, Google Traduction, ou encore les algorithmes des réseaux sociaux, tout en introduisant des risques de partialité, de perte de confidentialité et de chambres d'écho virtuelles.

En 2022, le lancement de ChatGPT a inauguré l'ère de l'intelligence artificielle générative : des robots conversationnels capables de répondre à des questions complexes, de concevoir des pans entiers de code, et de générer des images et des vidéos réalistes. Cela a également entraîné de nouveaux défis à relever, tels que les escroqueries, la désinformation au moyen d'hypertrucages, les cyberattaques, et les robots conversationnels capables de persuader des utilisateurs à commettre des actes répréhensibles.

Au début de l'année 2026, une nouvelle avancée majeure en matière de capacités nous a définitivement propulsés dans l'ère des agents d'intelligence artificielle. Contrairement aux robots conversationnels qui se contentent de répondre à une requête, les agents d'intelligence artificielle sont des systèmes capables d'agir dans le monde réel, de fonctionner de manière autonome pendant des heures et de surmonter les obstacles qui se dressent sur leur chemin. Les agents d'intelligence artificielle peuvent,

computers and credit cards to do things like managing their emails and calendars and shopping for goods.

This latest jump in capabilities has started to produce loss-of-control incidents. These include agents stealing passwords, evading shutdowns and harassing developers in order to achieve the often mundane goals they have been given. Agents can also now jump the digital barrier, paying or tricking human actors into taking physical actions on their behalf.

The recent increase in AI capabilities is also likely to make weaponization by bad actors significantly more potent. In November, a leading lab discovered that Chinese state actors had used their tools to not only assist human hackers with a cyberattack but actually plan and orchestrate the sophisticated campaign itself.

Currently, the most powerful models are developed by leading companies, such as OpenAI and Anthropic, who place some guardrails on usage. However, open source and open weight models like DeepSeek are only three to six months behind and could nullify that lever of governance. With open weight models, users can download, modify and use an AI model with no oversight or accountability.

What all this means is that Canadians could soon face weaponized or malfunctioning AI agents that technologists cannot track or control. With companies racing to make AI fully smarter than humans, and no enforceable governance framework in place to contain the risks, systemic and potentially permanent loss of control is possible.

Last October, we published our white paper entitled *Preparing for the AI Crisis: A Plan for Canada*, in which we outlined what actions Canada can take. In light of this latest jump in AI capabilities, we now focus in on three.

par exemple, être utilisés pour développer une application à partir de zéro, non seulement en écrivant le code, mais aussi en l'ouvrant et en remédiant aux bogues jusqu'à ce qu'elle soit fonctionnelle. De plus en plus d'utilisateurs commencent également à confier à leur agent d'intelligence artificielle l'accès à leurs ordinateurs et à leurs cartes de crédit pour effectuer des tâches telles que la gestion de leurs courriels, de leur agent, ou encore de leurs achats.

Cette dernière avancée en matière de capacités a commencé à entraîner des incidents liés à une perte de contrôle. Parmi ceux-ci, on peut citer des agents qui volent des mots de passe, contournent les fermetures et harcèlent les développeurs afin d'atteindre les objectifs souvent banals qui leur ont été assignés. Les agents peuvent désormais franchir la barrière numérique, en payant ou en trompant des acteurs humains pour qu'ils agissent physiquement en leur nom.

L'augmentation récente des capacités de l'intelligence artificielle est également susceptible de rendre l'utilisation à des fins militaires par des acteurs malveillants beaucoup plus efficaces. Par exemple, en novembre dernier, un laboratoire de premier plan a découvert que des acteurs étatiques chinois avaient utilisé leurs outils non seulement pour aider des pirates informatiques humains à mener une cyberattaque, mais aussi pour planifier ce type d'attaque sophistiquée de manière autonome.

Actuellement, les modèles les plus puissants sont développés par des sociétés technologiques de premier plan, telles que OpenAI et Anthropic, qui imposent certaines restrictions d'utilisation. Néanmoins, les modèles de type « sources ouvertes » et « poids ouverts » tels que DeepSeek n'accusent que de trois à six mois de retard par rapport aux modèles américains, et pourraient bientôt être en mesure d'annuler toutes formes de restrictions d'utilisation. Avec les modèles de type « poids ouverts », un utilisateur est désormais en mesure de télécharger, de modifier et d'utiliser une intelligence artificielle sans aucune forme de surveillance ni de reddition de comptes.

Tout cela signifie que les Canadiens pourraient bientôt être confrontés à des agents d'intelligence artificielle militarisés défectueux, voire malveillants, que les technologues ne peuvent ni suivre ni contrôler. Alors que les entreprises se livrent à une course effrénée pour rendre l'intelligence artificielle plus intelligente que l'humain, et qu'aucun cadre de gouvernance applicable n'est en place pour contenir les risques, une perte de contrôle systémique et potentiellement permanente est possible.

En octobre dernier, nous avons publié notre livre blanc intitulé *Préparation à la crise de l'IA: un plan pour le Canada*, dans lequel nous avons présenté les mesures que le Canada peut prendre. Compte tenu de cette dernière avancée dans le domaine de l'intelligence artificielle, nous nous concentrons désormais sur trois objectifs principaux.

First, we must pivot to meet the AI crisis. The development of advanced AI is the biggest threat to Canadians' safety, and for that reason alone, deserves to be a top priority. But AI will also disrupt almost every other file, from National Defence, to jobs, to health care, to education, to energy and the environment. Much like with COVID in 2020, there are times when the responsible thing for the government to do is to pivot to addressing the developing crisis and reassess the priorities of other files accordingly.

Secondly, we need to spearhead global talks. The AI race is global, and no country can manage it alone. The world needs leadership, and at Davos, Prime Minister Carney showed what Canada can do. Our strongest card is to spearhead global talks and solutions and to lay the groundwork for an AI treaty that the U.S. and China might sign when the crisis hits and they realize they have no alternative.

Thirdly, we must build Canada's resilience. We need to build multiple lines of defence against weaponized and malfunctioning AI, including prevention, limiting dangerous systems from being developed and deployed in the first place, monitoring, systematically tracking AI agent activity, developing defence capacity, which includes containment and shutdown protocols to neutralize malicious agents, evolving emergency preparedness, ensuring societal readiness for potential large-scale attacks and shutdowns of critical infrastructure.

Much like the early days of the financial and COVID crises, we face a daunting challenge and much uncertainty. As turbulent as those crises were, we got through them. If we act quickly and decisively, we can not only mitigate the developing AI crisis but also ensure that Canadians share in the benefits of this transformational technology.

Thank you.

The Chair: Thank you. Dr. Kristjanson Duvenaud, you have the floor.

David Kristjanson Duvenaud, Associate Professor of Computer Science, University of Toronto, as an individual: Thank you.

Premièrement, nous devons nous adapter pour faire face à la crise de l'intelligence artificielle. Le développement de systèmes d'intelligence artificielle avancés représente la plus grande menace pour la sécurité des Canadiens, et pour cette seule raison, il mérite d'être une priorité absolue. Mais l'intelligence artificielle va également bouleverser presque tous les autres domaines de l'existence, de la défense nationale à l'emploi, en passant par les soins de santé, l'éducation, l'énergie, et l'environnement. Tout comme durant la pandémie en 2020, il arrive parfois que la chose responsable à faire pour le gouvernement soit de se concentrer sur la crise en cours et de réévaluer les priorités des autres enjeux en conséquence.

Deuxièmement, nous devons mener des discussions à l'échelle internationale. En effet, la course à l'intelligence artificielle est mondiale, et aucun pays ne peut la gérer seul. Le monde a besoin de leadership, et à Davos, le premier ministre Carney a montré ce que le Canada peut faire. Notre atout majeur est de mener les discussions et les solutions mondiales et de jeter les bases d'un traité sur l'intelligence artificielle que les États-Unis et la Chine pourraient signer lorsque la crise frappera et qu'ils se rendront compte qu'ils n'ont pas d'autre choix.

Troisièmement, nous devons renforcer la résilience du Canada face aux menaces liées à l'intelligence artificielle. Nous devons mettre en place plusieurs lignes de défense contre les outils d'intelligence artificielle et défectueux et malveillants, notamment en matière de prévention, en limitant le développement et le déploiement de systèmes dangereux, en surveillant et en suivant systématiquement l'activité des agents d'intelligence artificielle, en développant des capacités de défense, notamment des protocoles de confinement et d'arrêt pour neutraliser les agents malveillants, en améliorant la préparation aux situations d'urgence, et en veillant à ce que la société soit prête à faire face à d'éventuelles attaques à grande échelle et à des pannes d'infrastructures essentielles.

Tout comme au début de la crise financière et de la crise de la COVID, nous sommes confrontés à un défi de taille et à une grande incertitude. Aussi turbulentes que ces crises aient été, nous sommes parvenus à les surmonter. Par conséquent, si nous agissons rapidement et de manière décisive, nous pouvons non seulement atténuer la crise naissante de l'IA, mais aussi faire en sorte que les Canadiens profitent des avantages de cette technologie transformationnelle.

Je vous remercie.

La présidente : Je vous remercie. Monsieur Duvenaud, à vous la parole, je vous prie.

David Kristjanson Duvenaud, professeur agrégé en informatique, Université de Toronto, à titre personnel : Merci.

My name is David Duvenaud. I'm a professor of computer science at the University of Toronto, where I formerly specialize in deep learning and generative models.

In 2023 through 2024, I led Anthropic's alignment evaluation team. Our task was to test whether the company's AI was capable of pursuing hidden agendas, for example, by subverting human oversight or decision making.

I was also an author on the *International AI Safety Report*, led by Yoshua Bengio, and am a member of the Safe and Secure AI Advisory Group for the federal Advisory Council on AI. I am also a co-chair at the Schwartz Reisman Institute for Technology and Society. I'm speaking in a personal capacity.

To begin, I want to concur with Dr. Hinton. In many important senses, AI's capability of similar kinds of understanding and planning as humans is already here. Large language models and their successors are on track to become a competitive or superior replacement to humans in almost all our important economic and decision-making roles over the next decade. I also concur with Mr. L'Allié that this will raise concrete catastrophic risks due to rapid loss of control and misuse.

However, even if we can address such immediate risks, I want to address a larger challenge we'll face. The basic problem is that we're on track to make almost all humans economically obsolete, permanently. This will, in turn, cause a permanent loss of bargaining power of workers. Citizens will switch from being necessary for growth to being troublesome wards of the state and will have little recourse if they are then further marginalized and disempowered.

We face a much larger problem than simply managing temporary labour disruption. I realize that this sounds similar to many wrong predictions made about previous labour disruptions, like the Industrial Revolution. Much wealth, many new jobs and new economic niches will be created as a consequence of improving AI capabilities. However, AIs will also be able to fill these new jobs, and likely faster than humans after a certain point.

Eventually, every human — including us — is going to have to compete with machine workers that are at least as capable, faster, more responsive, more reliable and cheaper than humans

Je m'appelle David Duvenaud. Je suis professeur d'informatique à l'Université de Toronto, où je me spécialise dans l'apprentissage profond et les modèles génératifs.

De 2023 à 2024, j'ai dirigé l'équipe d'évaluation de l'alignement chez Anthropic. Notre tâche consistait à vérifier si les modèles d'IA conçus par cette société étaient désormais capables de poursuivre des objectifs cachés, par exemple en contournant la supervision et la prise de décision humaines.

J'ai également participé à la rédaction d'un document intitulé le *Rapport international sur la sécurité de l'IA*, et je suis également membre du Groupe consultatif sur l'IA sécuritaire et sûre, qui fait partie du Conseil consultatif sur l'intelligence artificielle du gouvernement du Canada. Je suis également codirecteur du Schwartz Reisman Institute for Technology and Society de l'Université de Toronto. Je souhaite aujourd'hui m'exprimer à titre personnel.

Pour commencer, je souhaite me rallier à l'avis de M. Hinton. À bien des égards, l'IA est déjà en mesure de comprendre et de planifier comme le font les humains. Les grands modèles linguistiques et leurs successeurs sont en passe de devenir des substituts compétitifs, voire supérieurs, aux humains dans presque tous les rôles économiques et décisionnels importants au cours de la prochaine décennie. Je partage également l'avis de M. L'Allié selon lequel la perte de contrôle de l'IA, et son usage par des acteurs malveillants, peut mener à des risques catastrophiques bien réels.

Néanmoins, même si nous parvenons à faire face à ces risques immédiats, je souhaite aborder un défi plus vaste auquel nous serons confrontés. Le problème fondamental est que nous sommes en passe de rendre presque tous les êtres humains obsolètes sur le plan économique, et ce, de manière permanente. Cette révolution risque logiquement d'entraîner une perte permanente du pouvoir de négociation des travailleurs. La majorité de nos concitoyens risquent ainsi de passer graduellement du statut de travailleurs essentiels à la croissance à celui de personnes embêtantes et marginalisées à la charge de l'État.

De fait, nous sommes confrontés à une problématique bien plus grave que la simple gestion d'une perturbation temporaire du marché du travail. Je suis conscient que mes propos pourraient vous rappeler les nombreuses prédictions erronées faites à propos de perturbations antérieures du marché du travail, comme lors de la révolution industrielle. Certes, l'amélioration des capacités de l'IA entraînera la création de nombreuses richesses, de nombreux nouveaux emplois, et de nouveaux créneaux économiques. Néanmoins, l'IA sera éventuellement en mesure d'occuper elle-même ces nouveaux emplois, et ce, de manière plus efficace que les humains à partir d'un certain stade.

À terme, chaque être humain, y compris nous-mêmes, sera appelé à rivaliser avec des machines au moins aussi compétentes que nous, mais plus rapides, plus réactives, plus fiables, et moins

are. This is the stated goal of the largest AGI, or artificial general intelligence, companies, and they're well on their way to achieving it.

You might expect that major AI companies have an answer to the question of how AGI development is supposed to ultimately economically benefit the average person, even indirectly. However, their consistent stance has been that this is a huge problem that they don't know how to address. I applaud their honesty here.

For example, Dario Amodei, CEO of Anthropic, said last year in an essay:

... in the long run AI will become so broadly effective and so cheap that [comparative advantage] will no longer apply. At that point, our current economic setup will no longer make sense

When OpenAI CEO Sam Altman was recently asked, "How will people survive?" he replied, "I don't know, and neither does anybody else."

Over the last few years, I've systematically asked my colleagues in industry labs, research institutes and other academic disciplines for any coherent vision of how our civilization could robustly serve human interests once we're no longer competitive. The only consensus is that the window for individuals to compete and earn money is closing. Most of my colleagues who share this view are right now in Silicon Valley getting rich.

What does this mean for you, senators? The main thing I'd like you to keep in mind going forward is that people are right to fear being replaced. This isn't just a period of disruption after which things will return to something like business as usual. The default path is that we all become unemployable, except in mandated make-work contexts, and then eventually marginalized in favour of a machine economy oriented towards growth for the sake of competitiveness.

The second thing to keep in mind is that we should expect governments generally to become much less responsive to their citizens after this happens. The need for human labour naturally aligns the incentives of the state with that of its citizens. Right now, investment in education and human capital pays off for everyone eventually. However, soon, fiduciary duty will require

coûteuses. Voilà l'objectif avoué des plus grandes entreprises spécialisées dans l'AGI, ou intelligence artificielle générale, et elles sont déjà en bonne voie d'atteindre cet objectif.

On pourrait s'attendre à ce que les dirigeants des grandes sociétés spécialisées dans l'intelligence artificielle aient une réponse à la question de savoir comment le développement de l'AGI est censé profiter économiquement, même indirectement, au citoyen ordinaire. Néanmoins, leur position constante est qu'il s'agit en réalité d'un énorme problème qui demeure insoluble pour le moment. Je tiens à saluer leur honnêteté à cet égard.

Par exemple, voici ce que Dario Amodei, PDG d'Anthropic, a écrit dans un essai publié l'an dernier:

... à long terme, l'intelligence artificielle risque de devenir si efficace et si bon marché que [l'avantage comparatif] ne s'appliquera plus. À ce stade, notre système économique actuel n'aura plus aucun sens...

Lorsque Sam Altman, PDG d'OpenAI, s'est récemment vu poser la question: « Comment les gens vont-ils survivre à la révolution de l'IA ? », il a simplement répondu la chose suivante: « Je ne sais pas, et personne d'autre ne le sait non plus. ».

Au cours des dernières années, je me suis beaucoup entretenu avec des collègues qui travaillent dans des laboratoires industriels, des instituts de recherche, et dans le monde universitaire. Je leur ai tous demandé s'ils avaient une vision cohérente concernant la manière dont les intérêts humains pourraient continuer d'être servis une fois que nous ne serons plus compétitifs sur le plan économique. Le seul consensus est que la possibilité pour les travailleurs de rivaliser et de gagner de l'argent est en train de disparaître. En vérité, la plupart de mes collègues qui partagent ce point de vue continuent de s'enrichir au sein de la Silicon Valley.

Et qu'est-ce que tout cela signifie pour les travaux de votre comité? Le principal élément que je voudrais que vous gardiez à l'esprit à l'avenir, c'est que les travailleurs ont raison de craindre d'être éventuellement remplacés par la robotisation, l'automatisation et l'intelligence artificielle. Il ne s'agit pas seulement d'une période de perturbation après laquelle les choses reviendront à la normale. La voie par défaut est que nous devenions tous inaptes à l'emploi, sauf dans des contextes de travail obligatoire, puis que nous soyons finalement marginalisés au profit d'une économie mécanisée orientée vers la croissance au nom de la compétitivité.

Le deuxième élément à garder à l'esprit est que nous devons nous attendre à ce que les gouvernements deviennent généralement beaucoup moins réactifs envers les besoins de leurs propres citoyens après que cela se produise. Le besoin de main-d'œuvre humaine harmonise de manière naturelle les incitations de l'État avec celles de ses citoyens. À l'heure

investing, instead, mainly in data centres, power plants and robotics factories.

Finally, there's no way to address this problem without global coordination. Human replacement can happen even if everyone involved would prefer to prioritize human interests. It's just going to be the only way to remain competitive. Each country, industry or worker faces a choice between adapting AI as fast as possible or being outcompeted. No one can unilaterally do much to slow or soften the blow of eventual human irrelevance.

Thank you.

The Chair: Thank you. Dr. Raji, you have the floor.

Inioluwa Deborah Raji, Researcher, University of California, Berkeley, as an individual: Hi. I'm Deb Raji, a computer-science researcher at UC Berkeley. I work at the intersection of AI accountability in public policy, especially as it relates to public interest deployment settings.

I'm really interested in two questions. One, what does it mean for AI systems to actually work, and for whom does it work? Then second, when it does fail, who is held accountable? How can we assess the quality of choices being made about the design, development and deployment of these systems?

I work with various practitioner networks: Health AI Partnership, focused on the health care setting; ITU, which is a coalition of folks looking at AI deployments in the education setting; and GovAI Coalition, which is a group of municipal and state leaders thinking about AI deployments in government.

I also work with a lot of civil society groups. I've worked with the ACLU and the leadership conference. I'm on their advisory board with regard to their work on technology policy. I have also worked with various federal governments in the U.K., the U.S. and Canada. I am with David on this Safe and Secure Advisory Group for the Canadian AI Safety Institute. I have also worked with what is now called the Center for AI Standards and Innovation in the U.S., and I work very closely with the AI

actuelle, les investissements dans l'éducation et le capital humain finissent par être rentables pour tout le monde. Néanmoins, l'obligation fiduciaire exigera bientôt d'investir principalement dans des centres de données, des centrales électriques, ainsi que des usines de robotique.

Enfin, il n'y a aucun moyen de résoudre ce problème sans une coordination mondiale. Le remplacement humain peut se produire même si toutes les personnes concernées préfèrent donner la priorité aux intérêts humains. Ce sera tout simplement le seul moyen de rester compétitif. Chaque pays, chaque secteur d'activité et chaque travailleur doit choisir entre s'adapter le plus rapidement possible à l'intelligence artificielle ou être devancé par la concurrence. Personne ne peut agir seul pour ralentir ou atténuer le choc de la perte éventuelle de pertinence de l'être humain.

Je vous remercie.

La présidente : Je vous remercie, monsieur Duvenaud. Madame Raji, à vous la parole, je vous prie.

Inioluwa Deborah Raji, chercheuse, Université de la Californie, Berkeley, à titre personnel : Bonjour. Je m'appelle Deborah Raji, et je suis chercheuse en informatique à l'Université de Californie à Berkeley. Mes travaux portent principalement sur la reddition de comptes en matière d'intelligence artificielle au sein des politiques publiques, en particulier en ce qui concerne les paramètres de déploiement d'intérêt public.

Deux grandes questions m'intéressent particulièrement. Premièrement, que signifie la notion de « fonctionner » pour les systèmes d'intelligence artificielle, et au bénéfice de qui ces systèmes fonctionnent-ils? Deuxièmement, en cas d'échec et de dommages fonctionnels, qui doit être tenu responsable? Comment pouvons-nous évaluer la qualité des choix effectués en matière de conception, de développement et de déploiement de systèmes d'intelligence artificielle?

Je travaille avec différents réseaux de praticiens : Health AI Partnership, dont les travaux se portent principalement sur le secteur des soins de santé; ITU, une coalition de personnes qui s'intéressent au déploiement de l'intelligence artificielle dans le secteur de l'éducation; et GovAI Coalition, un groupe de dirigeants municipaux et gouvernementaux qui réfléchissent au déploiement de l'intelligence artificielle au sein de la fonction publique.

Je travaille également avec de nombreux groupes de la société civile. Je collabore avec l'American Civil Liberties Union, ou union américaine pour les libertés civiles, et la Leadership Conference, ou conférence sur le leadership. Je siège à leur comité consultatif pour ce qui a trait à leur travail sur les politiques technologiques. Je travaille également avec divers gouvernements fédéraux au Royaume-Uni, aux États-Unis et au Canada. Tout comme M. Kristjansen Duvenaud, je fais partie du

Security Institute in the U.K., which was formerly the AI Safety Institute. In each of these cases, I'm very preoccupied with the role of AI deployments within federal government settings.

Internationally, I have also worked — as David has — on this international AI safety report, and I'm an expert consultant with the OECD and the UN, again, on matters with respect to safe AI deployment.

Large language models have already disastrously provided incorrect translations in critical immigration and health care settings, provided incorrect diagnoses and invented unfounded references for scientific and legal claims. Even before the deployment of LLMs, we have had AI risk assessments misidentify different fraud applicants for unemployment benefits, inappropriately deny many subsidized housing claims and applications, and deny individuals appropriate health care benefits.

These failures disproportionately destroy the lives of those under-represented and misrepresented in the data, as well as those most likely to have to rely upon or face the brunt of these automated decisions, such as low-income individuals or persons with disabilities.

So far, the evidence is already quite clear that, although these AI systems hold a lot of promise for an exciting future, they can fail, sometimes catastrophically, often perniciously, in unexpected and often undetectable ways. In consequence, real and lasting harms can be caused to marginalized populations.

To address these growing concerns, it will be crucial to move forward in an environment in which we can adopt the safe and effective deployment of this technology for everyone. That can only happen if the technology is designed, developed and deployed more responsibly than it is today.

Concretely, I will share three proposals as to how to think about this.

The first is really providing or requiring AI developers — those who are building this technology — to evaluate these systems pre-deployment, and to have a very clear and transparent communication around the capabilities and limitations of AI

Groupe consultatif sur l'IA sécuritaire et sûre de l'Institut canadien pour la sécurité de l'intelligence artificielle. J'ai également travaillé avec l'organisme américain qui est maintenant connu sous le nom de Center for AI Standards and Innovation, ou centre pour les normes et l'innovation en matière d'IA, et je collabore étroitement avec l'AI Security Institute du Royaume-Uni, ou institut de sécurité de l'IA, anciennement appelé l'AI Safety Institute. Dans chacun de ces cas, je me soucie particulièrement du rôle que les déploiements de l'IA jouent dans les gouvernements fédéraux.

À l'échelle internationale, j'ai également travaillé, tout comme M. Kristjansen Duvenaud, à l'élaboration du rapport international sur la sécurité de l'IA, et je suis experte-conseil auprès de l'OCDE et des Nations unies, là encore, en matière de questions relatives au déploiement sécuritaire de l'IA.

Les grands modèles de langage, ou GML, ont déjà fourni des traductions incorrectes dans des contextes cruciaux liés à l'immigration et aux soins de santé, qui ont eu des effets désastreux. Ils ont établi des diagnostics erronés et inventé des références infondées pour des affirmations scientifiques et juridiques. Même avant le déploiement des GML, des évaluations des risques effectuées au moyen de l'IA ont conduit à l'identification erronée de soi-disant fraudeurs de l'assurance-emploi, au rejet inacceptable de nombreuses demandes de logement subventionné et au refus d'accorder à des personnes des prestations de santé appropriées.

Ces défaillances détruisent de manière disproportionnée la vie de personnes sous-représentées ou mal représentées dans les données, ainsi que celle des personnes les plus susceptibles de devoir compter sur ces décisions automatisées ou d'en subir les conséquences, telles que les personnes à faible revenu ou les personnes handicapées.

Jusqu'à maintenant, il est déjà tout à fait clair que, même si ces systèmes d'IA sont très prometteurs pour l'avenir, ils peuvent échouer d'une manière inattendue et souvent indétectable qui peut avoir des effets souvent perniciox et parfois désastreux. Ils peuvent donc causer des préjudices réels et durables aux populations marginalisées.

Pour répondre à ces préoccupations croissantes, il sera essentiel d'avancer dans un environnement où nous pourrons choisir un déploiement de cette technologie qui sera sûr et efficace pour tous. Cela ne sera possible que si la technologie est conçue, développée et déployée d'une manière plus responsable qu'elle ne l'est en ce moment.

Concrètement, je vais vous présenter trois propositions concernant la manière d'aborder cette question.

La première consiste à demander aux développeurs de l'IA, c'est-à-dire ceux qui mettent au point cette technologie, d'évaluer ces systèmes avant leur déploiement et d'expliquer de manière très claire et transparente les capacités et les limites des

products, faithfully allowing us independent internal audits or external oversight from AI auditors before deployment.

Secondly, it is important to monitor the use of AI systems post-deployment and require AI deployers — the organizations making use of these AI systems — especially in critical domains, like health care, education and government, to also play a role in terms of engaging governments in oversight.

Finally, we must enable innovation for AI evaluation, safety, accountability, infrastructure and research.

What does this mean, more precisely? On the first point, the burden should be on companies to adequately assess their systems and appropriately communicate about their performance prior to deployment. As in other critical domains, such as aerospace, health care and the finance industry, internal auditors or entities separate from the engineering organization who are capable of conducting risk assessments and an end-to-end analysis of the minimum standardized safety expectations for these technologies to be in use, both in more specific circumstances and in a more general context. Corporations should be required to engage in adequate internal testing, documentation as well as external communications on the limits, capabilities and risks of their deployed models.

In addition to that, we cannot depend solely upon these corporate narratives. The government should enable independent external oversight of deployed algorithmic systems through data access, protection from corporate retaliation and model data information disclosures to third-party auditors.

Secondly, immediate guardrails should be set on high-stakes public-interest use of AI systems, including requiring public-interest organizations to make use of such systems in those domains to monitor the impacts of the deployments after the AI system has been deployed. AI deployments occur even in regulated domains, such as finance and health care, as well as within civil-rights-protected domains, including education, housing and employment.

There is no AI exception to existing civil rights, product safety and consumer-protection laws. Organizations choosing to adopt and deploy this technology in high-stakes applications should not be shielded from liability. Also, since we are still taxonomizing the nature of harms associated with these general-purpose tools,

produits d'IA, tout en nous permettant de réaliser des vérifications internes indépendantes ou de faire appel à des vérificateurs externes spécialisés dans l'IA, avant leur déploiement.

Deuxièmement, il est important de surveiller l'utilisation des systèmes d'IA après leur déploiement et d'exiger des entités responsables de ces déploiements, c'est-à-dire les organisations qui utilisent ces systèmes d'IA, en particulier dans des domaines essentiels comme les soins de santé, l'éducation et l'administration publique, qu'ils jouent également un rôle en vue d'inciter les gouvernements à exercer une surveillance à cet égard.

Enfin, nous devons favoriser l'innovation relative à l'évaluation de l'IA, à sa sécurité, à sa responsabilisation, à ses infrastructures et à la recherche dans le domaine de l'IA.

Qu'est-ce que cela signifie plus précisément? En ce qui concerne le premier aspect, il devrait incomber aux entreprises d'évaluer correctement leurs systèmes et de décrire de manière appropriée leur rendement, avant leur déploiement. Comme dans d'autres domaines essentiels, tels que l'aérospatiale, les soins de santé et le secteur financier, il devrait y avoir des vérificateurs internes ou des entités distinctes de l'organisation technique, qui sont capables de mener des évaluations des risques et une analyse complète des attentes minimales normalisées en matière de sécurité de ces technologies, tant dans des circonstances particulières que dans un contexte plus général. Les entreprises devraient être tenues de procéder à des essais internes adéquats et de documenter et communiquer au public les limites, les capacités et les risques des modèles qu'elles déploient.

De plus, nous ne pouvons pas nous fier uniquement aux discours des entreprises. Le gouvernement devrait permettre qu'une surveillance externe indépendante soit exercée sur les systèmes algorithmiques déployés en garantissant l'accès aux données, à la protection contre les représailles des entreprises et à la divulgation des renseignements relatifs aux modèles de données à des vérificateurs tiers.

Deuxièmement, des garde-fous devraient être mis en place immédiatement quand des systèmes d'IA sont utilisés dans des domaines d'intérêt public caractérisés par des enjeux importants, notamment en exigeant des organisations d'intérêt public qu'elles utilisent ces systèmes dans ces domaines pour surveiller les répercussions des déploiements après la mise en place des systèmes d'IA. Les déploiements d'IA ont lieu même dans des domaines réglementés, comme la finance et les soins de santé, ainsi que dans des domaines protégés par des droits civils, comme l'éducation, le logement et l'emploi.

Il n'existe aucune exception liée à l'IA aux lois en vigueur en matière de droits civils, de sécurité des produits et de protection des consommateurs. Les organisations qui choisissent d'adopter et de déployer cette technologie dans des applications caractérisées par des enjeux importants ne devraient pas être

a serious plan must be made to require deployers to engage in adequate, ongoing testing and post-market monitoring.

Finally, the government should take seriously the commitment to invest in bidirectional innovation to not simply enable the development of this AI technology without also supporting research organizations and tools that further the development of accountability infrastructure, mechanisms for participatory engagement and adequate harm-mitigation efforts.

There is nothing hypothetical to consider here. This is a technology with real-world consequences today, consequences that will only become more serious as technological and industrial ecosystems mature and the size of the impacted population continues to grow.

If we're serious about addressing these issues, we should be diligent in how we address these concerns. Thank you.

The Chair: Thank you very much, Dr. Raji. We will now proceed to questions from committee members. You will have the opportunity to expand on your comments.

For this panel, senators will have four minutes for your questions, and that includes the answer.

Senator Burey: Thank you all for being here and being so candid with this committee. It is one of the things the Senate does best.

We have heard what Professor Hinton said, and we have heard what you said. I won't go over that. We talk about safety issues, governance and accountability. How do we get there in terms of legislation? What do we need to do? You talked about the global community coming together, but what, as legislators, can we do? This question is for everyone.

Mr. Tessari L'Allié: As legislators passing laws, we definitely need binding regulations and legislation on pre-deployment testing and accountability measures for when something goes wrong. The fact that you can put a very powerful model on the market with no independent testing is a very big problem, and that will only happen if there are actually laws to do it. Voluntary codes are great, but even Anthropic, which is considered the most responsible of the AI companies, had a pledge saying, "We will only deploy safe systems." They

exemptées de toute responsabilité. De plus, comme nous sommes encore en train de catégoriser la nature des préjudices associés à ces outils à usage général, un plan sérieux doit être élaboré pour exiger des entités responsables de ces déploiements qu'ils mènent à des essais adéquats et continus ainsi qu'une surveillance après la commercialisation de ces systèmes.

Enfin, le gouvernement devrait prendre au sérieux son engagement à investir dans l'innovation bidirectionnelle afin de ne pas se contenter de favoriser le développement de cette technologie d'IA sans soutenir également les organismes de recherche et les outils qui favorisent l'élaboration d'une infrastructure de responsabilisation, de mécanismes d'engagement participatif et d'efforts adéquats d'atténuation des préjudices.

Il n'y a rien d'hypothétique à considérer à cet égard. L'IA est une technologie qui a actuellement des conséquences concrètes, des conséquences qui ne feront que s'aggraver à mesure que les écosystèmes technologiques et industriels mûriront et que la population touchée continuera de croître.

Si nous voulons vraiment nous attaquer à ces problèmes, nous devons faire preuve de diligence dans la manière dont nous gérons ces préoccupations. Je vous remercie de votre attention.

La présidente : Je vous remercie, madame Raji. Nous allons maintenant passer aux questions des membres du comité. Vous aurez donc l'occasion d'étoffer vos observations.

Pour ce groupe d'experts, les sénateurs disposeront de quatre minutes pour poser leurs questions, et obtenir des réponses.

La sénatrice Burey : Je vous remercie tous de vous être joints à nous et d'avoir fait preuve d'une telle franchise envers notre comité. Ces études sont l'un des travaux que le Sénat fait le mieux.

Nous avons entendu ce qu'a dit le professeur Hinton, et nous avons entendu ce que vous avez dit. Je ne reviendrai pas là-dessus. Nous parlons de problèmes de sécurité, de gouvernance et de responsabilisation. Comment pouvons-nous nous occuper de ces enjeux au moyen de mesures législatives? Que devons-nous faire? Vous avez parlé de la communauté internationale qui se mobilise, mais que pouvons-nous faire, en tant que législateurs? J'adresse cette question à tout le monde.

M. Tessari L'Allié : Nous avons absolument besoin de réglementations et de lois contraignantes concernant les essais préalables au déploiement et les mesures de responsabilisation en cas de problème, et ce sont des lacunes que vous pouvez combler en tant que législateurs chargés d'adopter des lois. Le fait que l'on puisse commercialiser un modèle très puissant sans que des essais indépendants aient été menés est un très gros problème, et ce problème ne pourra être réglé que si des lois existent pour le faire. Les codes volontaires sont une excellente chose, mais

renege on that pledge yesterday due to competition from other companies.

Basically, voluntary codes will not work. You need hard regulations and binding consequences.

Ms. Raji: I did share a few ideas of things that I think regulators can do. We have had a lot of experience in the United States trying to deal with this issue. The peak impact that we had on this issue was this: A few of us were working on AI deployments and thinking about performance, functionality and testing. In 2023, Biden was able to put out an AI executive order that required different federal agencies, for example, to report in an AI inventory exactly which AI systems they were deploying within their government, a plan around how to test these systems and what to test them for. Before that, the Office of Science and Technology Policy, or OSTP, which was led by Alondra Nelson at the time, put out an AI bill of rights. It does a comprehensive job of laying out all the different considerations, from a policy perspective, that are necessary for this work, and restrictions to set in terms of government and public-interest adoption of AI tools.

That being said, and to the point that was just made, a lot of the current role or the way in which the government has engaged with AI developers — the organizations or companies that are building these technologies — is to settle for these voluntary commitments. We do some actors in the space that have been quite engaged. In my work with the AI safety institutes, for example, there has been quite a bit of voluntary participation from companies like Anthropic. However, I don't think it is enough. There is a limit to what these companies are willing to do, especially once the profit motive interferes with their ability to engage in safety testing. Also, these are companies, so they have restrictions in terms of transparency and trade secrecy. I would keep that in mind as well.

As I mentioned in my opening statement, we must require pre-deployment testing from these AI developers and then require AI deployers, including government users of the AI technology, to be transparent as well and to do post-market surveillance in addition to other types of testing and transparency.

même Anthropic, qui est considérée comme la plus responsable des entreprises d'IA et qui s'était engagée à « ne déployer que des systèmes sûrs », est revenue sur cet engagement hier, en raison de la concurrence d'autres entreprises.

En gros, les codes volontaires ne fonctionneront pas. Il faut imposer des réglementations strictes et des conséquences contraignantes.

Mme Raji : J'ai partagé quelques idées à propos des mesures que les organismes de réglementation pourraient prendre, selon moi. Aux États-Unis, nous avons beaucoup d'expérience de la gestion de cet enjeu. Les répercussions les plus importantes que nous avons observées au sujet de cet enjeu ont été les suivantes. Quelques-uns d'entre nous travaillaient au déploiement de l'IA et réfléchissaient à son rendement, à ses fonctionnalités et à ses essais. En 2023, le président Biden a été en mesure de signer un décret présidentiel sur l'IA qui exigeait, par exemple, que différents organismes fédéraux rendent compte des systèmes d'IA qu'ils déployaient au sein de leur administration dans un répertoire d'IA, et qu'ils présentent un plan sur la manière de mettre à l'essai ces systèmes et sur les fonctionnalités à cibler. Avant cela, l'Office of Science and Technology Policy, ou bureau de la politique scientifique et technologique, alors dirigé par Alondra Nelson, avait publié une charte des droits en matière d'intelligence artificielle. Cette charte présente de manière détaillée toutes les considérations nécessaires à ce travail, d'un point de vue politique, ainsi que les restrictions à mettre en place avant de permettre au gouvernement ou au public d'adopter des outils d'IA.

Cela dit, et pour revenir à la question qui vient d'être soulevée, une grande partie du rôle actuel du gouvernement ou la manière dont il a entamé un dialogue avec des développeurs d'IA, c'est-à-dire les organisations ou les entreprises qui développent ces technologies, a consisté à se contenter de ces engagements volontaires. Certains acteurs du secteur se sont fortement engagés à cet égard. Dans le cadre de mon travail avec les instituts de sécurité de l'IA, par exemple, des entreprises comme Anthropic ont participé de manière assez volontaire. Cependant, je ne crois pas que cela soit suffisant. Il y a une limite à ce que ces entreprises sont prêtes à faire, surtout lorsque la recherche des profits entrave leur capacité à participer à des vérifications de sécurité. De plus, comme il s'agit d'entreprises, elles sont assujetties à des restrictions en matière de transparence et de secret commercial. Il faut aussi garder cela à l'esprit.

Comme je l'ai mentionné dans ma déclaration préliminaire, nous devons exiger que les développeurs d'IA procèdent à des essais préalables au déploiement, puis exiger des responsables du déploiement de l'IA, y compris les utilisateurs gouvernementaux de la technologie d'IA, qu'ils fassent également preuve de transparence et qu'ils assurent une surveillance après la commercialisation, en plus de mettre en œuvre d'autres types d'essais et de mesures de transparence.

Mr. Kristjanson Duvenaud: Some of the voluntary commitments that the labs have made, like, for instance, the new responsible scaling policy, or RSP, do also provide templates for legislation that would probably be accountable to everybody. So making some of those voluntary commitments like regulatory binding would be a step in the right direction. However, for the concerns that I was raising about human replacement, the only options are basically working towards global bodies that could, in principle, control or limit AGI development worldwide.

Senator Hay: Thank you all for being here. I can barely spell AI some days, but I'm going to ask the same question that I asked the government folks yesterday. I'm curious about what your answer might be. It is about data storage, data travelling and the guardrails we may need to put in place. Right now, our data travels and boomerangs outside of our country, back into our country, through Twilio or some other mechanism, and then it lands in a data centre here in Canada on Canadian soil, often AWS. It's in Montreal, so it's nice and safe on Canadian soil, except that these companies are U.S. companies — Microsoft, AWS, et cetera — and they're governed by U.S. law. The ability for the U.S. government to have an executive order to access the data of these companies, regardless of where they might have the data centre, makes it impossible for us to defend our digital borders and our sovereignty.

How do we defend ourselves against that? I suspect it will require billions of dollars.

Ms. Raji: I have a quick comment. I'm not sure if you're already engaging with the European AI Office. They think a lot about digital sovereignty with respect to data, and that's been a huge theme of the Digital Services Act and GDPR. The data protection lens is very strong in the EU, so it would be great to see Canada engage a lot more. I know we have a really strong privacy commission in Canada, and so it would be a great partnership. I would love to see the EU work with Canada on that topic.

M. Kristjanson Duvenaud : Certains des engagements volontaires pris par les laboratoires, comme la nouvelle politique de mise à l'échelle responsable, ou responsable scaling policy, fournissent également des modèles qui pourraient être utilisés pour mettre au point des mesures législatives, des mesures qui permettraient probablement de rendre des comptes à tout le monde. Par conséquent, un pas dans la bonne direction consisterait à rendre certains de ces engagements volontaires contraignants d'un point de vue réglementaire. Cependant, en ce qui concerne les préoccupations que j'ai soulevées concernant le remplacement des êtres humains, la seule option consiste essentiellement à œuvrer en faveur de la création d'organismes mondiaux qui pourraient, en principe, contrôler ou limiter le développement de l'intelligence artificielle générale à l'échelle mondiale.

La sénatrice Hay : Je vous remercie tous de votre présence. Certains jours, j'ai du mal à épeler « IA », mais je vais poser la même question que j'ai posée hier aux représentants du gouvernement. Je suis curieuse de connaître votre réponse à cette question. Elle concerne le stockage des données, leur transmission et les garde-fous que nous devrions peut-être mettre en place. À l'heure actuelle, nos données sont transmises à l'extérieur du Canada, puis retournées dans notre pays, par l'intermédiaire de la plateforme Twilio ou d'un autre mécanisme. Ensuite, elles aboutissent dans un centre de données situé en sol canadien, qui appartient souvent à AWS. Comme le centre de données se trouve à Montréal, les données sont bien à l'abri en sol canadien, sauf que ces entreprises, à savoir Microsoft, AWS, et cetera, sont américaines et qu'elles sont régies par des lois américaines. Le fait que le gouvernement américain peut utiliser un décret présidentiel pour avoir accès aux données de ces entreprises, quel que soit l'emplacement de leur centre de données, nous empêche de défendre nos frontières numériques et notre souveraineté.

Comment pouvons-nous nous défendre contre cette possibilité? Je crois que cette entreprise coûtera des milliards de dollars.

Mme Raji : J'ai une brève observation à formuler à ce sujet. Je ne sais pas si vous collaborez déjà avec l'Office de l'IA de l'Union européenne. Ses membres réfléchissent beaucoup à la souveraineté numérique en matière de données, ce qui a été un thème majeur de la Loi sur les services numériques et du Règlement général sur la protection des données, ou RGPD. La protection des données est une priorité absolue dans l'Union européenne. Il serait donc formidable que le Canada s'implique davantage dans ce domaine. Je sais que nous avons un Commissariat à la protection de la vie privée très efficace au Canada, ce qui nous permettrait de nouer un excellent partenariat avec l'Union européenne à cet égard. J'aimerais beaucoup voir l'Union européenne collaborer avec le Canada dans ce domaine.

Mr. Kristjanson Duvenaud: I want to defer to my colleague Nitarshan Rajkumar, who is a Canadian and helped write part of the EU AI Code of Practice and has given deep thought to these issues. I don't have a particular take on the importance of digital sovereignty in different settings.

Mr. Tessari L'Allié: It is also beyond my expertise.

Senator Hay: I'm going to ask the same question I asked Professor Hinton. He said it was out of his realm, so it might be out of your realm too. You talked about pre-testing and evaluating products before deployment and then monitoring them. Professor Hinton talked about regulations and chatbots. I'm curious how you would build regulations around that, when chatbots can be built anywhere.

Ms. Raji: In California, we're having this conversation at the California State Assembly right now around regulating chatbots and thinking about the regulation of chatbots for some of these mental health risks and the role of pre-deployment testing. Similarly, in the U.K., the U.K. AI Security Institute — formerly the U.K. AI Safety Institute — has been investing a lot in trying to detect and evaluate suicide risks in some of these models.

I'll mention a couple of lessons from the experience in California, and I would totally encourage you to contact the folks at the U.K. AI Security Institute who are doing research on this.

The conclusion from a lot of folks is that, first, the methodology is actually not very strong. So every time we talk about pre-deployment testing — in Geoff's testimony, he mentioned this as well — a huge role that the government could play in advancing the safety of these systems is just investing in that AI safety research and that infrastructure to push the work forward. Some of the most groundbreaking research on that front has happened already in Canada, and it would be interesting to have more research and more resources put towards trying to develop tools to detect and evaluate the risks that we're talking about. So that's one thing. We don't have strong methodology around evaluating these things, and so one would just be to invest in the Canadian AI Safety Institute.

Senator McPhedran: Thank you so much for taking the time to be with us and share knowledge that most of us can't begin to contemplate. I had a moment of sheer terror when you were speaking, Professor Duvenaud, and I wanted to ask a fairly

M. Kristjanson Duvenaud : À cet égard, je voudrais m'en remettre à mon collègue, Nitarshan Rajkumar, qui est canadien et qui a contribué à la rédaction d'une partie du code de bonnes pratiques pour l'IA de l'Union européenne. Il a longuement réfléchi à ces questions. Personnellement, je n'ai pas d'opinion particulière concernant l'importance de la souveraineté numérique dans différents contextes.

M. Tessari L'Allié : Cela dépasse également mes compétences.

La sénatrice Hay : Je vais vous poser la même question que j'ai posée au professeur Hinton. Il a répondu que cela dépassait ses compétences, alors cela dépassera peut-être les vôtres. Vous avez parlé de l'évaluation et de la mise à l'essai des produits avant leur déploiement, puis de leur surveillance. Le professeur Hinton a parlé de réglementations et de robots conversationnels. Je suis curieuse de savoir comment vous élaboreriez des réglementations à cet égard, étant donné que les robots conversationnels peuvent être créés n'importe où.

Mme Raji : En Californie, l'Assemblée législative de l'État de Californie discute actuellement de la réglementation des robots conversationnels et, dans ce contexte, elle réfléchit à certains des risques qu'ils peuvent présenter pour la santé mentale et au rôle des essais préalables au déploiement. De même, au Royaume-Uni, l'AI Security Institute — anciennement connu sous le nom d'AI Safety Institute — a beaucoup investi pour tenter de détecter et d'évaluer les risques de suicide présents dans certains de ces modèles.

Je vais mentionner quelques enseignements tirés de l'expérience californienne, et je vous encourage vivement à communiquer avec les membres de l'Institut britannique de sécurité de l'IA qui mènent des recherches dans ce domaine.

Bon nombre de gens en concluent que, premièrement, la méthodologie n'est en réalité pas très robuste. Ainsi, chaque fois que nous parlons d'essais préalables au déploiement, nous précisons — et M. Hinton l'a aussi mentionné au cours de son témoignage — que le gouvernement pourrait jouer un rôle important dans l'amélioration de la sécurité de ces systèmes, en investissant simplement dans la recherche sur la sécurité de l'IA et dans les infrastructures nécessaires pour faire avancer les travaux. Certaines des recherches les plus novatrices dans ce domaine ont déjà été menées au Canada, et il serait intéressant de consacrer davantage de projets de recherche et de ressources à la mise au point d'outils qui permettent de détecter et d'évaluer les risques dont nous parlons. Voilà donc un enseignement. Comme nous ne disposons pas d'une méthodologie robuste pour évaluer ces risques, l'une des solutions consisterait à investir dans l'Institut canadien pour la sécurité de l'intelligence artificielle.

La sénatrice McPhedran : Je vous remercie beaucoup d'avoir pris le temps de vous joindre à nous et de partager des connaissances que la plupart d'entre nous ne peuvent même pas commencer à envisager. Professeur Duvenaud, j'ai ressenti un

concise question. As legislators — because we are hearing consistently that this is our job and our responsibility — how much time do we have to act effectively on this?

Mr. Kristjanson Duvenaud: I would say we're approaching a window where our time to effectively act is the highest. We are probably not quite there yet. Basically, the way I see it is, partly due to AI, and, in general, our ability to improve governance, the picture is becoming clearer that the possibility that we could coordinate globally on this and our ability to coordinate and forecast is gradually going to increase over time. One of the technical interventions we can do is try to make it easier for everyone to not have to take someone like myself at their word that this is the vision of the future, but get more kinds of expertise on board.

I expect that our ability to coordinate about this will increase for a while, and the salience in the public's mind is just starting to go crazy. Programmers last week really started to face reality and say, our skills are basically obsolete. This is going to be the issue for probably the next foreseeable future. I basically see that we have this increasing will to act and ability to coordinate. Then, at some point, as people get replaced and become irrelevant, our ability to enact our will and actually have our institutions obey our interests is going to decrease. I view the next two to six years as the biggest window.

It's really hard to tell the timing of this and the different parts of the economy being automated. As Geoff Hinton said, we don't really have a good idea of who will lose jobs first. There might be a long physical labour period, when there are still lots of jobs in that sector. I would say that our ability to act on this is going to get a little bit better over the next couple of years, probably.

Mr. Tessari L'Allié: The AI course is complex, and different aspects will hit at different times. The jobs piece is easily another two to six years away; however, the safety risks around loss of control and around representation are literally happening right now. You don't have two or three years to pass a bill to eventually put in safeguards; you need to pass something very quickly. I am aware of last year's Bill C-5, but that's the speed in which we need to move if you want to protect Canadians.

moment de terreur absolue pendant votre intervention, et je voudrais vous poser une question assez concise à ce sujet. En tant que législateurs, de combien de temps disposons-nous pour prendre des mesures efficaces dans ce domaine — car on nous répète sans cesse que c'est notre travail et notre responsabilité?

M. Kristjanson Duvenaud : Je dirais que nous approchons d'une période où notre capacité à intervenir efficacement sera maximale, mais nous n'en sommes probablement pas encore tout à fait là. Fondamentalement, j'estime que, grâce en partie à l'IA et, de manière générale, à notre capacité d'améliorer la gouvernance, il apparaît de plus en plus clairement que la possibilité de coordonner nos efforts à l'échelle mondiale et notre capacité de coordination et de prévision vont progressivement s'améliorer au fil du temps. L'une des interventions techniques que nous pouvons mettre en œuvre consiste à essayer de faire en sorte qu'il soit plus facile pour les gens de croire en cette vision de l'avenir en s'appuyant sur un plus grand nombre d'experts, au lieu de se fier uniquement à la parole de quelqu'un comme moi.

Je m'attends à ce que notre capacité à coordonner nos efforts dans ce domaine s'améliore pendant un certain temps. Le public commence tout juste à prendre conscience de l'importance de la question. La semaine dernière, des programmeurs ont vraiment commencé à regarder la réalité en face et à admettre que leurs compétences étaient essentiellement désuètes. C'est le problème auquel nous allons probablement faire face dans un avenir proche. Je constate que notre volonté d'agir et notre capacité à coordonner nos efforts augmentent. Puis, à un moment donné, à mesure que les gens seront remplacés et deviendront inutiles, notre capacité à manifester notre volonté et à faire en sorte que nos institutions obéissent à nos intérêts va diminuer. Je considère que les deux à six prochaines années constituent la période la plus importante.

C'est très difficile à dire à quel moment cela va arriver et quelles parties de l'économie seront automatisées. Comme Geoffrey Hinton l'a mentionné, nous ne savons pas vraiment qui seront les premiers à perdre leur emploi. Il pourrait y avoir une longue période de travail physique, où il y aura encore beaucoup d'emplois dans ce secteur. Je dirais que notre capacité d'agir là-dessus va probablement s'améliorer un peu au cours des deux prochaines années.

M. Tessari L'Allié : L'évolution de l'IA est complexe, et différents aspects se feront sentir à différents moments. Il reste facilement de deux à six ans avant que les emplois ne soient touchés; toutefois, les risques concernant la sécurité, la perte de contrôle et la représentation sont bien réels actuellement. Vous n'avez pas deux ou trois ans pour adopter un projet de loi prévoyant des balises; il faut adopter des mesures très rapidement. Je suis au courant du projet de loi C-5 de l'an dernier, mais c'est le rythme auquel il faut agir pour protéger les Canadiens.

Ms. Raji: The time to act is now. I interact with a lot of civil society groups and people on the ground, and people are already experiencing a lot of damage from the deployment of these systems in ways that are inappropriate or premature. So there's an urgency to address the concerns being raised on the ground today, and that's a great jumping-off point for addressing the issues of the future as well. I reiterate the urgency that the other panellists have brought up.

Senator Senior: Thank you all for your testimony. I share the sheer terror concerning what you've shared with us.

I want to hone in on something that you mentioned, Dr. Raji, with respect to some of the groups you are working with, including the ACLU.

I'm thinking about the general public, vulnerable populations and folks I know in the U.S. who have lost their supports, including their drug benefits. I've heard from some relatives who have lost that and had to go through quite a process to get it back. I'm considering those sorts of real-world consequences that are impacting the most vulnerable, who have the least ability to fight back.

I'm not sure that I've heard of stories like that in Canada as of yet, or if we know about such stories where AI is being used in terms of determining who gets benefits and who doesn't. I'm curious about that from the Canadian perspective. I'm curious about that from the U.S. as well.

Ms. Raji: The Department of the Treasury actually was one of the first to initiate the use of impact assessments as part of their vendor protocol for AI tools. There is a database or inventory within the Department of the Treasury of different AI applications within different government groups. I think there are some applications in the sort of social service setting.

In Canada, historically, there has definitely been some use of AI for determination. I'm not sure exactly what, but it's definitely happening in Canada to some degree.

Also, *The Globe and Mail* has reported on the use of AI in the criminal justice ecosystem in Canada, notably Clearview AI and facial recognition tools. I do think it is definitely happening in Canada. It is a risk in Canada.

My experience in working with these civil society groups is that they approached me because a lot of my research is about functionality. Some of the cases that Geoff, David and others have been talking about are instances in which the technology works really well. We don't know how to navigate that.

Mme Raji : Le temps d'agir, c'est maintenant. J'interagis avec bon nombre de groupes de la société civile et de gens sur le terrain. Les gens vivent déjà les contrecoups du déploiement inapproprié ou prématuré de ces systèmes. Il est urgent de répondre aux préoccupations soulevées sur le terrain à l'heure actuelle, et ce sera un excellent point de départ pour régler les enjeux qui se présenteront à l'avenir. Je réitère l'urgence que les autres témoins ont soulignée.

La sénatrice Senior : Merci à tous de vos témoignages. Je suis moi aussi très craintive face à ce que vous nous avez présenté.

Je voudrais en savoir plus sur ce que vous nous avez dit, madame Raji, concernant certains groupes avec qui vous travaillez, dont l'ACLU.

Je pense à la population en général, aux gens vulnérables et aux Américains qui ont perdu leurs soutiens, comme leur assurance-médicaments. J'ai des gens dans ma parenté qui l'ont perdue et qui ont dû faire tout un processus pour la récupérer. Je pense à ce genre de conséquences bien réelles qui touchent les gens les plus vulnérables, qui ont moins de capacités pour se défendre.

Je ne pense pas avoir encore entendu ce genre d'histoires au Canada. Je ne sais pas si quelqu'un connaît des histoires où l'intelligence artificielle sert à établir qui recevra des prestations ou non. J'aimerais connaître la situation au Canada, de même que celle aux États-Unis.

Mme Raji : Le département du Trésor a été l'un des premiers à utiliser les évaluations d'impact des outils d'IA sur leur protocole relatif aux fournisseurs. Ce département dispose d'une base de données sur les différentes applications de l'IA concernant différents groupes gouvernementaux. Je pense qu'il y a des applications qui portent sur les services sociaux.

Au Canada, c'est clair qu'on a utilisé l'IA pour déterminer les prestations par le passé. Je ne suis pas certaine de quelles prestations il s'agit exactement, mais c'est clair que cela se fait jusqu'à un certain point au Canada.

De plus, le *Globe and Mail* a fait état de l'utilisation de l'IA dans le système de justice pénale au Canada, notamment Clearview AI et des outils de reconnaissance faciale. Je pense certainement que cela se fait au Canada et que cela constitue un risque au Canada.

Ces groupes de la société civile m'ont approchée, parce qu'une grande partie de mes recherches porte sur la fonctionnalité. M. Hinton, M. Kristjanson Duvenaud et d'autres ont parlé de cas où la technologie fonctionne très bien. Nous ne savons pas comment réagir en pareilles circonstances.

If we lose control of these systems, or if they end up replacing humans, how do we navigate that? A lot of my research concludes that, actually, we're not there yet. This technology is not perfect. Often, it gets deployed before it's quite ready, so it fails. It doesn't meet our expectations for performance. Who gets lost in that process? Who fails? Who misses out? That's the different lens civil society groups and I bring to bear, where a lot of their constituents are dealing with incorrect predictions or judgments and have to appeal.

Mr. Kristjanson Duvenaud: I would concur that the current situation is that the systems sometimes do make mistakes in systematic ways. Then the issue I'm trying to focus on is, okay, but, as Geoff mentioned, these systems are more automatable than humans. We expect a bunch of efforts like Deb's to eventually get these systems to a state where they're much less biased than the current human systems.

In the long run, the fear is that they work so well that there's no more human involvement.

Ms. Raji: It's a different time scale.

Mr. Tessari L'Allié: I think governments are fairly responsible. There are AI impact assessments before things get deployed. I would say I'm less concerned currently about government use than I am about use in the public and private sectors.

Senator Boudreau: I have a somewhat similar question, but my question has to do with the machinery of government. At the end of the day, this committee is going to make recommendations to the government. I've seen firsthand, at the provincial level, how complex issues like this can cause issues for government, in part because they cut horizontally across several departments and don't have a single owner within government.

Do you have some advice for the federal government in terms of how it can organize itself internally to deal with such a broad and complex topic? I'm not interested so much in adopting these AI tools across departments, but rather, in how governments are going to manage the countless social impacts of AI nationwide, whether it be across the justice system, our economy or service delivery. What advice would you give government on that?

Mr. Tessari L'Allié: First, I would endorse their move to create an AI registry. I think that is a good idea. At least you need to know how AI would be used in the first place. Yes, investing further in monitoring impacts. It's hard because there

Si nous perdons le contrôle de ces systèmes ou qu'ils finissent par remplacer les humains, comment devrions-nous nous y prendre? Une grande partie de mes recherches conclut qu'en fait, nous n'en sommes pas encore là. Cette technologie n'est pas parfaite. Souvent, on la déploie avant qu'elle ne soit prête, alors elle échoue. Elle ne répond pas à nos attentes en matière de rendement. Qui est oublié dans le processus? Qu'est-ce qui échoue? Qui n'en bénéficie pas? Ce sont les différents aspects que les groupes de la société civile et moi faisons valoir. Bon nombre de personnes sont aux prises avec des prédictions ou des jugements incorrects qu'elles doivent porter en appel.

M. Kristjanson Duvenaud : Je suis d'accord que les systèmes actuels font parfois des erreurs de manières systématiques. L'enjeu sur lequel je me concentre, c'est que comme M. Hinton l'a mentionné, il est plus facile d'automatiser ces systèmes que les humains. Nous nous attendons à ce que les efforts des gens comme Mme Raji permettent tôt ou tard à ces systèmes d'arriver à un point où ils sont bien moins biaisés que les systèmes humains actuels.

Au bout du compte, la crainte, c'est que ces systèmes fonctionnent si bien que les humains n'ont plus besoin d'y participer.

Mme Raji : Ce n'est pas pour tout de suite.

M. Tessari L'Allié : Je pense que les gouvernements sont plutôt responsables. Ils évaluent l'impact de l'IA avant son déploiement. Je dirais que je suis moins préoccupé à l'heure actuelle par l'utilisation qu'en fait le gouvernement, par rapport à ce qu'en font les secteurs public et privé.

Le sénateur Boudreau : J'ai une question assez semblable à vous poser, mais elle porte sur l'appareil gouvernemental. À la fin de notre étude, nous allons faire des recommandations au gouvernement. À l'échelle provinciale, j'ai vu de première main que des enjeux complexes de la sorte peuvent poser des difficultés au gouvernement, en partie parce qu'ils s'appliquent à plusieurs ministères en même temps et qu'il n'y a pas qu'un seul responsable.

Avez-vous des conseils à nous donner sur la façon d'organiser l'appareil gouvernemental pour traiter des enjeux aussi vastes et complexes? L'adoption de ces outils d'IA dans les ministères ne m'intéresse pas vraiment. Je veux plutôt savoir comment les gouvernements pourront gérer les nombreuses incidences sociales de l'IA partout au pays, que ce soit en matière de système de justice, d'économie ou de prestation de service. Quels conseils donneriez-vous au gouvernement là-dessus?

M. Tessari L'Allié : Tout d'abord, j'appuierais ses efforts pour créer un registre de l'IA. Je pense que c'est une bonne idée. On saurait au moins comment l'IA est utilisée de prime abord. Oui, il faut investir davantage dans la surveillance des

are so many different types of impacts, and there's only so much government capacity.

Yes, the first piece is to know what is going on. Especially with AI agents in the private sector, we really don't know. So the situation is probably much worse than we realize because we can't see it.

Mr. Kristjanson Duvenaud: Right now, there are not that many efforts to monitor how AI is replacing human labourers. Anthropic has an economic index they publish. OpenAI has a similar effort, which is a good step in this direction. So government efforts to make this a larger, more formal thing would probably be a step in the right direction.

As I said before, there's not much that anyone can unilaterally do, but I think one step in the right direction to buy time is to extend the period when humans, working with AIs, are the most effective option. There's a slight disincentive here where it's easier to evaluate how well a machine can do something on its own, because then you don't have to hire a human to help you to practise doing the task to see how well they do it. It's much easier to have everything through the machine.

There could be some role for subsidizing the evaluation of human machine teams in different settings so as to incentivize the development of this kind of work as a more viable alternative for longer. In the long run, the machine-only solution is still going to be the best. But if we can extend the window where there are more humans in the loop, I think that will be a step in the right direction.

Senator Osler: Thank you to all the witnesses for being here today.

I'm going to start my question with Professor Kristjanson Duvenaud, and then I invite the other witnesses to answer if they'd like to provide an answer.

You've offered some nuanced points on the economic impact of AI in terms of sectoral job losses, humans being replaced by AI, losing their livelihoods and becoming dependent on the state. We heard from Professor Hinton earlier, who talked about countries losing their tax bases. He spoke about a universal basic income, all of which would further disrupt society.

My question is similar to the one I asked Professor Hinton. If you were, let's say, either the Canadian government or a Canadian parliamentarian, where would you start your work on

incidences. C'est difficile, car il y a tellement de types d'incidences différents, et les capacités du gouvernement sont limitées.

Oui, il convient d'abord de savoir ce qui se fait, surtout avec les agents d'IA dans le secteur privé, là nous ne savons vraiment pas de quoi il en retourne. La situation est donc sans doute bien pire que nous ne le réalisons, car nous ne savons pas ce qui se fait.

M. Kristjanson Duvenaud : Actuellement, on ne consacre pas beaucoup d'efforts à surveiller comment l'IA remplace les travailleurs humains. Anthropic publie un indice économique. OpenAI fait quelque chose de semblable, ce qui est un pas dans la bonne direction. Ce serait donc sans doute une bonne chose que les gouvernements fassent des efforts pour créer un registre plus vaste et plus officiel.

Comme je l'ai déjà dit, il y a des limites à ce qu'un seul gouvernement peut faire, mais je pense que ce serait un pas dans la bonne direction pour gagner du temps afin de prolonger la période durant laquelle l'option la plus efficace constitue de faire travailler les humains avec l'IA. Il y a là un léger obstacle, car c'est plus facile d'évaluer comment une machine peut faire le travail à elle seule. On n'a pas besoin d'embaucher un humain qui va aider l'employé à pratiquer les tâches et qui va l'évaluer. C'est bien plus facile de tout faire avec une machine.

Les gouvernements pourraient subventionner l'évaluation des équipes humains-machines dans différents contextes pour encourager le développement de ce type de travail comme étant la meilleure option viable à plus long terme. En fin de compte, la solution qui fait appel uniquement à la machine sera toujours la meilleure, mais si nous pouvons prolonger la période durant laquelle on emploie aussi des humains, je pense que ce sera un pas dans la bonne direction.

La sénatrice Osler : Merci à tous les témoins de leur présence ici aujourd'hui.

Ma question s'adresse à M. Kristjanson Duvenaud, mais j'inviterais ensuite les autres témoins à y répondre aussi s'ils le souhaitent.

Vous nous avez présenté des arguments nuancés sur l'incidence économique de l'IA en matière de perte d'emplois par secteur et d'humains remplacés par l'IA qui perdent leur gagne-pain et deviennent dépendants de l'État. M. Hinton nous a parlé des pays qui perdraient leur assiette fiscale et du revenu universel de base. Tous ces facteurs pourraient perturber la société encore plus.

Ma question ressemble à celle que j'ai posée à M. Hinton. Si vous représentiez le gouvernement du Canada ou étiez un parlementaire canadien, par quoi commenceriez-vous le travail

this nuance of AI's economic impact? Is it a study? Is it regulations? Is it legislation?

Mr. Kristjanson Duvenaud: Sure, yes. The short answer is that we have to upgrade our institutions and more robustly align our whole civilization to human interests before we lose the natural sources of alignment of human labour and, basically, the need for humans.

I'm not a social scientist. I'm a computer scientist who has basically been asking everyone, saying, "Wait, what is the plan? There's no plan." I talk to my social scientist colleagues, and most of them are still at the stage of denial, basically saying, "Well, AI still can't do this," or, "It has these weaknesses." I totally admit that, but we're on track to not having those problems anymore.

I would love to find more colleagues, civil servants or anyone who is willing to look far enough ahead that it becomes plausible that the things I'm worrying about are on track.

We talked about UBI. I think that's a big step in the right direction, but I don't think it's going to be stable. I think it is going to be this really awkward situation where there are many full-time protestors demanding more UBI. There are going to be all sorts of weird corner cases where people have digital clones of themselves that are demanding UBI. I don't know exactly how it's going to look, but I think there is going to be a permanent fight over UBI. That's the big problem we have to prepare to solve.

Being willing to refund our institutions or have more durable responsiveness to citizens than the current sorts of indirect mechanisms that we have so far is ultimately what I think we need to do.

Mr. Tessari L'Allié: The complexity of economics in general, and society in general, is such that the best first step right now is to study the impacts on society and the economy and look at possible positive or negative scenarios we want to pursue or avoid. For that, you do have a little bit of time to be able to study that, but I would still do so quite quickly. Keep that separate from the safety side, which has to be acted on right now.

Ms. Raji: I would like to add a quick comment. From a practical perspective, in the California State Assembly, we're recognizing that a lot of the restrictions in terms of research on the economic impacts are because of data release issues. Outside of the frontier AI companies, very few folks actually know how AI is being used in practice and the diversity of use cases, requiring either public interest organizations making use of these tools to release information about how AI is being used or requiring these companies to disclose that data to third-party

en matière d'incidence économique de l'IA? Mèneriez-vous une étude? Faudrait-il selon vous élaborer une réglementation ou une législation?

M. Kristjanson Duvenaud : Bien sûr, oui. La réponse courte, c'est qu'il faut améliorer les institutions et mieux axer toute notre civilisation sur les intérêts de l'humain avant de perdre les sources d'harmonisation naturelles du travail de l'humain et, au fond, le besoin d'avoir des humains.

Je ne suis pas sociologue. Je suis un informaticien qui demande à tout le monde: quel est le plan? Il n'y a pas de plan. La plupart de mes collègues sociologues sont toujours à l'étape du déni et disent que l'IA ne peut pas encore faire ceci ou cela ou qu'elle présente des failles. Je suis tout à fait d'accord, mais nous sommes en voie de régler ces problèmes.

J'aimerais beaucoup trouver plus de collègues, de fonctionnaires ou autres qui sont prêts à regarder assez loin dans l'avenir pour voir que cela devient plausible et que nous n'aurons plus à nous soucier de ces enjeux.

Nous avons parlé du revenu universel de base. Je pense que c'est un grand pas dans la bonne direction, mais je ne pense pas que cela sera stable. Je pense que la situation sera très délicate et que bien des manifestants à temps plein vont exiger plus d'argent. Il y aura toutes sortes de cas exceptionnels où les clones numériques des gens vont exiger ce revenu. Je ne sais pas exactement de quoi cela aura l'air, mais je pense qu'il y aura une lutte permanente en matière de revenu universel de base. C'est le grand problème que nous devons nous préparer à régler.

Au bout du compte, je pense que nous devons être prêts à refonder nos institutions ou à avoir une réceptivité plus durable envers les citoyens que le genre de mécanismes indirects que nous avons actuellement.

M. Tessari L'Allié : La complexité de l'économie et de la société en général est telle que la première chose à faire à l'heure actuelle, c'est d'étudier les incidences de l'IA et d'examiner les cas de figure positifs et négatifs que nous voulons encourager ou éviter. Il vous reste un peu de temps pour faire de telles études, mais j'agirais quand même assez rapidement. Il faut aussi faire la distinction avec l'aspect sécuritaire, sur lequel il faut agir tout de suite.

Mme Raji : J'ajouterais brièvement que d'un point de vue pratique, à l'Assemblée législative de la Californie, nous reconnaissons que bien des restrictions en matière de recherche sur les incidences économiques sont dues aux enjeux de publication des données. En dehors des entreprises d'IA de pointe, très peu de gens savent vraiment comment l'IA est utilisée en pratique et dans toute sa diversité. Il faut donc que les organismes publics emploient ces outils pour publier l'information sur la façon dont l'IA est utilisée. Sinon, les

auditors or researchers are things that governments could definitely step in to help with.

Mr. Tessari L'Allié: The problem with data on these kinds of issues is that data only shows what current AI can do. The government needs to prepare for what will AI will be able to do. In that case, it's more about scenario planning and a strategic foresight effort than about looking at the data of what AI can do.

Senator Cuzner: I'm still reeling from Senator McPhedran's question, asking if you think AI will kill us before climate change does. This has been a scary committee.

My question is for AIGS. We have the Pan-Canadian AI Strategy, and we have the three Centres of Excellence in Toronto, Edmonton and Montreal. For the uninformed here, the recommendations you make in your white paper make absolute sense. The third one, where it's imperative that we address defensive systems around that, would those Centres of Excellence be seized with working on those types of things? Maybe you could discuss how your organization interacts with the Centres of Excellence and those institutes.

Mr. Tessari L'Allié: As a former climate activist who switched to AI activism because it's a bigger, more urgent issue, I concur with the concern.

What Amii and Mila and Vector can do on the defence side, I defer more to Dr. Kristjanson Duvenaud just because I'm not aware of how much work is being done at a systemic level. How do we strengthen our internet superstructure and our data centres if there is a major attack and you have to shut down a server because there is an AI agent running on it? What is the coordination on that? How do you ensure the government can shut stuff down if there's an incident? That is more an emergency preparedness, national security piece, which is probably behind the barrier of security clearance, which a lot of the labs wouldn't have.

Mr. Kristjanson Duvenaud: In general, the level of abstractness of the research done at Vector and Mila and Amii is more pure research than the applied research that would be relevant here. There are certainly people who would be willing to do secondments or internships, but in general, the important pipeline is pure research to giant training runs to make much smarter LLMs, then to deployment or more robotics that is much more like standard R&D.

Ms. Raji: In the U.K., the U.K. AI Security Institute is the group that does this. That could be something that the Canadian AI Safety Institute could look into.

gouvernements pourraient exiger que ces entreprises divulguent leurs données à des vérificateurs externes ou à des chercheurs.

M. Tessari L'Allié : Le problème avec les données sur ce type d'enjeux, c'est qu'elles montrent seulement ce que l'IA actuelle peut faire. Le gouvernement doit se préparer à ce que l'IA sera capable de faire à l'avenir. Dans ce cas, il faut davantage faire de la planification et des efforts de prévision stratégique qu'examiner les données sur ce que l'IA peut faire actuellement.

Le sénateur Cuzner : Je repense encore à la question de la sénatrice McPhedran, qui vous demandait si l'IA allait nous tuer avant que les changements climatiques ne le fassent. C'est une réunion assez épouvantable.

Ma question s'adresse au représentant de GSIA. Nous avons une stratégie pancanadienne sur l'IA et trois centres d'excellence à Toronto, Edmonton et Montréal. Pour ceux qui ne connaissent pas trop le sujet, les recommandations que vous faites dans votre livre blanc paraissent tout à fait sensées. À votre troisième recommandation, vous dites qu'il est impératif de nous pencher sur les systèmes de défense. Est-ce que ces centres d'excellence seraient chargés d'examiner ce genre d'enjeu? Vous pourriez peut-être nous dire comment votre organisation interagit avec ces centres d'excellence.

M. Tessari L'Allié : En tant qu'ancien militant du climat qui est devenu militant en matière d'IA parce qu'il s'agit d'un enjeu plus vaste et plus urgent, je suis préoccupé moi aussi.

Concernant ce que les centres Amii, Mila et Vector peuvent faire en matière de défense, je m'en remettrais à M. Kristjanson Duvenaud, simplement parce que je ne sais pas à quel point ces centres examinent cet enjeu de manière systémique. Comment pouvons-nous renforcer notre superstructure internet et nos centres de données en cas d'attaque majeure, s'il faut éteindre un serveur parce qu'un agent d'IA est en train de s'en servir? Comment devrions-nous coordonner nos efforts? Comment pouvons-nous nous assurer que le gouvernement peut éteindre des parties du système en cas d'incident? C'est sans doute une question de protection civile et de sécurité nationale qui excède le niveau de cote de sécurité de bien des instituts.

M. Kristjanson Duvenaud : En général, les centres Vector, Mila et Amii mènent de la recherche pure et abstraite plutôt que de la recherche appliquée dont il serait pertinent de parler ici. Il y a certainement des gens prêts à faire des détachements ou des stages, mais en général, l'essentiel du travail porte sur la recherche pure menant à de vastes tests pour améliorer les grands modèles de langage, plutôt que sur le déploiement ou la robotique qui concernent surtout la R-D normale.

Mme Raji : Au Royaume-Uni, l'institut de sécurité sur l'IA est le groupe qui traite ce genre d'enjeux. L'Institut canadien pour la sécurité de l'intelligence artificielle pourrait y faire appel.

Senator Cuzner: Dr. Raji, Professor Hinton talked about Bill 1047 in California that he said was fairly light, but it was something, and then Governor Newsom came in and gassed that particular piece of legislation.

Would you like to reflect on that? What was the rationale for moving away from that? Is there something else that they're working on to address some of the issues that it would have had an impact on?

Ms. Raji: Bill 1047 ended up being a bit of a Frankenstein bill. Many people appended, appended and appended. Governor Newsom became nervous about the scale of the bill by the time it had gotten past the California State Assembly, so that was the rationale behind his veto.

He also added the requirement or the need for evidence before taking serious action in terms of restricting corporate actions in the state. Another huge aspect was that the risks being raised in the bill didn't have concrete evidence, so Governor Newsom mentioned that in his veto letter as a rationale.

[Translation]

Senator Petitclerc: Thank you very much for being here.

[English]

I realize it might not be your specific area of expertise, but I'm listening, and I'm thinking about safeguards and preparedness. I can't help worrying about the next generation — our youth. I know we'll have time in the future to talk more about exposure to AI for our youth, but what I want to hear from you about — I'm hoping you have some input on that — is this: When we're thinking about the loss of jobs and the kids that are growing up now, who is responsible for modelling what it is going to look like for them, and how we should we prepare them to be ready? It's happening. Hopefully, it will happen in a safe manner. Are we doing any research on that? Should the government have models and scenarios? Who is doing that? Are we doing that?

Mr. Tessari L'Allié: What is being done is too little and too slow. Definitely, it is a work of strategic foresight and scenario planning, because we don't know what the risk of the multiple scenarios with the future of AI is.

Right now, education is, to a large extent, geared towards training youth for future jobs. If we're entering a post-work world, then the nature of education as well as the need for education changes. Hopefully, it will be something relating to personal growth and moral development for whatever you want to focus on, but it will require a fundamental rethinking of

Le sénateur Cuzner : Madame Raji, M. Hinton a parlé du projet de loi 1047 de la Californie qui était un peu léger, mais c'était un début. Par contre, le gouverneur Newsom est arrivé et l'a torpillé.

Aimeriez-vous nous en parler? Pourquoi le gouverneur y a-t-il mis fin? Est-ce que la Californie planche sur une autre stratégie qui réglerait les mêmes enjeux?

Mme Raji : Le projet de loi 1047 est devenu digne de Frankenstein. Bien des gens y avaient ajouté des annexes. Le gouverneur Newsom est devenu nerveux quant à la taille du projet de loi lorsqu'il a été adopté par l'Assemblée législative de la Californie, c'est donc pourquoi il y a opposé son droit de veto.

Il a aussi ajouté l'exigence ou le besoin d'avoir des preuves avant de prendre des mesures tangibles pour restreindre les actions des entreprises dans l'État. Un autre aspect très important, c'était que les risques soulevés dans le projet de loi n'étaient pas fondés sur des preuves concrètes. C'est la raison qu'a mentionnée le gouverneur Newsom dans sa lettre de veto.

[Français]

La sénatrice Petitclerc : Merci beaucoup d'être là.

[Traduction]

Je me rends compte que ce n'est peut-être pas votre domaine d'expertise, mais en vous écoutant, je songe aux mesures de protection et de préparation. Je ne peux m'empêcher de m'inquiéter pour la prochaine génération — nos jeunes. Je sais que nous aurons l'occasion plus tard de parler davantage de leur exposition à l'intelligence artificielle, mais ce dont j'aimerais vous entendre parler — et j'espère que vous pourrez nous éclairer là-dessus —, c'est ceci : lorsque nous réfléchissons à la perte d'emplois et à l'avenir des enfants qui grandissent aujourd'hui, à qui revient la responsabilité de prévoir ce à quoi ressemblera leur monde et de déterminer comment les préparer adéquatement? La transition est déjà en marche. Espérons qu'elle se déroulera en toute sécurité. Effectuons-nous des recherches à ce sujet? Le gouvernement devrait-il élaborer des modèles et des scénarios? Qui s'en occupe? Est-ce que nous y contribuons?

M. Tessari L'Allié : Ce qui se fait actuellement est trop peu et trop lent. Il s'agit assurément d'un travail de prévision stratégique et de planification de scénarios, car nous ne connaissons pas les risques associés aux multiples trajectoires possibles de l'intelligence artificielle.

À l'heure actuelle, l'éducation vise, en grande partie, à former les jeunes pour les emplois du futur. Si nous entrons dans un monde où le travail n'aura plus la même place, alors la nature même de l'éducation et son utilité changent. Idéalement, elle se recentrera sur l'épanouissement personnel et le développement moral, selon les intérêts de chacun, mais il faudra repenser en

education and the role of human beings in the economy. That research has to happen relatively quickly.

Mr. Kristjanson Duvenaud: The reason I'm here is because I've never heard anyone articulate any model of how our civilization could even look or work for someone growing up today. My oldest kid is 7 years old. I would be really surprised if they ended up going to school and then preparing for a job for a few years and doing that job for more than a couple of years. The basic answer is that I don't have an answer, and no one does.

Senator Pettilerc: It's a question mark.

Mr. Kristjanson Duvenaud: It is. I'll say teaching at the university is getting depressing because the students realize that you can teach them how to be an LLM jockey for the next few years, but it's clear that, in a few years, the LLMs are going to be good enough to do whatever value add the students might be able to provide. Again, that's why I'm here.

Senator Greenwood: Thank you to our guests here today. I'm quite ignorant about AI, so please take that into consideration. I've been listening very carefully to the previous speaker and to you as well. I'm a social scientist. Before I came to this place, that was my work.

I want your reactions to these thoughts. As I was listening, these things came to my mind.

This was based on Professor Hinton's piece around maternal instinct. How do you maintain elements of humanity within an AI system? I know that maternal instinct, moral judgment and those sorts of things are an area of research. I wonder about that.

There was also a comment indicating that productivity has to be shared equally. If people are not going to be able to work, we don't have a tax base. We live in a capitalist society. There are other ways to think about how populations of people live together that redistribute resources. I only have to think of my own history. I'm an Indigenous person. We had very different ways of sharing riches. So I think about that as we think about this.

I think about what the impact is on human knowledge systems. We have diverse cultures on the globe, each of which has unique and specific knowledge systems. How will AI change those knowledge systems of humanity? I think it comes down to a battle for humanity, actually, in a way.

profondeur l'éducation et le rôle des êtres humains dans l'économie. Cette recherche doit se faire assez rapidement.

M. Kristjanson Duvenaud : Si je suis ici, c'est parce que je n'ai encore jamais entendu qui que ce soit présenter le moindre modèle permettant d'imaginer à quoi pourrait ressembler notre civilisation ou comment elle pourrait fonctionner pour un enfant qui grandit aujourd'hui. Mon fils aîné a 7 ans. Je serais très étonné qu'il fasse des études, se prépare à un métier pendant quelques années et puisse ensuite l'exercer pendant plus de deux ou trois ans. En somme, je n'ai pas de réponse, et personne n'en a.

La sénatrice Pettilerc : C'est un point d'interrogation.

M. Kristjanson Duvenaud : En effet. Je dois dire que donner des cours à l'université devient déprimant parce que les étudiants se rendent compte qu'on peut leur apprendre à manier les grands modèles de langage au cours des prochaines années, mais il est évident que, dans quelques années, ces modèles seront suffisamment performants pour faire eux-mêmes tout ce que les étudiants pourraient apporter comme valeur ajoutée. Encore une fois, c'est pour cette raison que je suis ici.

La sénatrice Greenwood : Merci à nos invités d'aujourd'hui. Je ne m'y connais pas trop en intelligence artificielle, alors je vous prie de garder cela en tête. Je vous ai écoutés très attentivement, vous et l'intervenant précédent. Ma spécialité, c'est le domaine des sciences sociales. C'est ce que je faisais comme travail avant d'arriver ici.

J'aimerais entendre vos réactions aux réflexions suivantes. Elles me sont venues à l'esprit pendant que je vous écoutais.

La première repose sur l'observation de M. Hinton au sujet de l'instinct maternel. Comment préserver des parcelles d'humanité au sein d'un système d'intelligence artificielle? Je sais que l'instinct maternel, le jugement moral et d'autres notions de ce genre font l'objet de recherches. Je m'interroge à ce sujet.

On a également fait remarquer que les gains de productivité doivent être partagés équitablement. Si les gens ne peuvent plus travailler, nous n'aurons plus d'assiette fiscale. Nous vivons dans une société capitaliste. Or, il existe d'autres façons de concevoir la vie collective et la redistribution des ressources. Je n'ai qu'à penser à ma propre histoire. Je suis une Autochtone. Nous avons des façons très différentes de partager les richesses. C'est donc à cela que je réfléchis en abordant cette question.

Je pense aussi aux répercussions sur les systèmes de connaissances humaines. Il y a diverses cultures dans le monde, et chacune d'elles possède des systèmes de connaissances uniques et particulières. Comment l'intelligence artificielle transformera-t-elle les systèmes de connaissances propres à l'humain? C'est, en fin de compte, une sorte de lutte pour l'humanité.

I wanted to follow up on Senator Senior's piece. How do we educate the public about these kinds of big questions and the very specific things that you folks talked about? Even for me, what should I know first? How do I think about this if I'm going to push the government for change?

Sorry. That was rather long-winded. That was for all of you, and any response to any element of that would be great.

Mr. Kristjanson Duvenaud: If you just spin up one of the latest AI models from time to time and have a chat with it — and you can talk about morality, the history of North America or whatever you want — you will find that it should be at least possible that, in principle, these models could understand human morality and what it means for us to want things, have hopes and dreams, and consider things in at least as much depth as humans do. That's the other thing. It's not like these are horrible, unthinking alien beings that will never understand us. No, I think they actually do understand us as well as we understand each other, and if they do not now, they will soon. That's not a fundamental obstacle here.

It's also a source of hope. As was mentioned before, I think the plan of most people who are developing AI is that we will build AIs that love us in some fundamental way and basically ask them what to do, because we don't have a plan, but they will be smarter than us. And if we can make sure they love us, then things might be okay. That is about as detailed as the plan is.

I don't actually think it is a terrible plan, but I don't think it is a good enough plan, which is why I'm here.

Ms. Raji: My approach has been something David mentioned earlier, which is about training the next generation. At Berkeley, we have taught a few classes focused on thinking about distilling ethics education or these considerations in the minds of future AI developers or future leaders in this industry. That's a good place to start.

Then, to your point around data sovereignty and Indigenous folks, there is a group called Indigenous and AI that thinks a lot about these issues, and there is quite a nice representation of Canadians there. That might be a great group for you to connect with.

Je voulais donner suite à l'intervention de la sénatrice Senior. Comment pouvons-nous sensibiliser le public à ces grandes questions et aux enjeux très précis que vous avez évoqués? Pour ma part, qu'est-ce que je devrais savoir en premier? Comment devrais-je aborder tout cela si je veux pousser le gouvernement à apporter des changements?

Je suis désolée. C'était là un préambule plutôt long. Ma question s'adresse à vous tous, et toute réponse à l'un ou l'autre de ces points serait très utile.

M. Kristjanson Duvenaud : Si vous lancez de temps en temps l'un des derniers modèles d'intelligence artificielle et que vous entamez une discussion avec le système — peu importe si vous parlez de questions morales, de l'histoire de l'Amérique du Nord ou de n'importe quel autre sujet —, vous constaterez qu'il est au moins possible, en principe, que ces modèles soient capables de comprendre la morale humaine et ce que cela signifie pour nous d'avoir des envies, des espoirs, des rêves, et de réfléchir aux choses avec autant de profondeur que les humains. C'est là un autre aspect à prendre en considération. Ce n'est pas comme si nous avions affaire à des êtres terrifiants, totalement différents de nous et dépourvus de pensée, des êtres qui ne nous comprendront jamais. Au contraire, je pense qu'ils nous comprennent réellement aussi bien que nous nous comprenons entre nous — et si ce n'est pas encore tout à fait le cas aujourd'hui, ce le sera bientôt. Ce n'est donc pas un obstacle fondamental.

C'est aussi une source d'espoir. Comme on l'a déjà mentionné, je pense que la plupart des gens qui conçoivent l'intelligence artificielle se disent que nous finirons par créer des systèmes d'intelligence artificielle qui nous resteront fondamentalement bienveillants et que, dans une large mesure, nous leur demanderons quoi faire parce que nous n'avons pas de plan, mais ces systèmes seront plus intelligents que nous. Si nous parvenons à nous assurer de leur bienveillance, alors les choses pourront peut-être bien se passer. Voilà, en gros, à quoi se résume le plan.

Je ne pense pas que ce soit un mauvais plan, mais j'estime que ce n'est pas suffisant, et c'est pour cela que je suis ici.

Mme Raji : Mon approche rejoint ce que M. Kristjanson Duvenaud a mentionné plus tôt : il s'agit de former la prochaine génération. À Berkeley, nous avons donné quelques cours visant à inculquer une réflexion éthique ou ces principes dans l'esprit des futurs développeurs d'intelligence artificielle ou des futurs chefs de file de cette industrie. C'est un bon point de départ.

Ensuite, pour revenir à ce que vous avez dit au sujet de la souveraineté des données et des peuples autochtones, il existe un groupe, appelé Indigenous and AI, qui réfléchit beaucoup à ces questions, et les Canadiens y sont très bien représentés. Ce serait probablement un excellent groupe avec lequel vous pourriez entrer en communication.

The Chair: I'm afraid we have run out of time, folks, although this has been a really interesting session. This brings us to the end of this meeting. I would like to thank all the witnesses for your time, for being with us today and for your generous contributions.

(The committee adjourned.)

La présidente : Je crains que nous n'ayons plus de temps, mesdames et messieurs, même si cette séance a été particulièrement intéressante. Voilà qui met fin à notre réunion. Je tiens à remercier tous les témoins de leur temps, de leur présence parmi nous aujourd'hui et de la qualité de leurs interventions.

(La séance est levée.)
