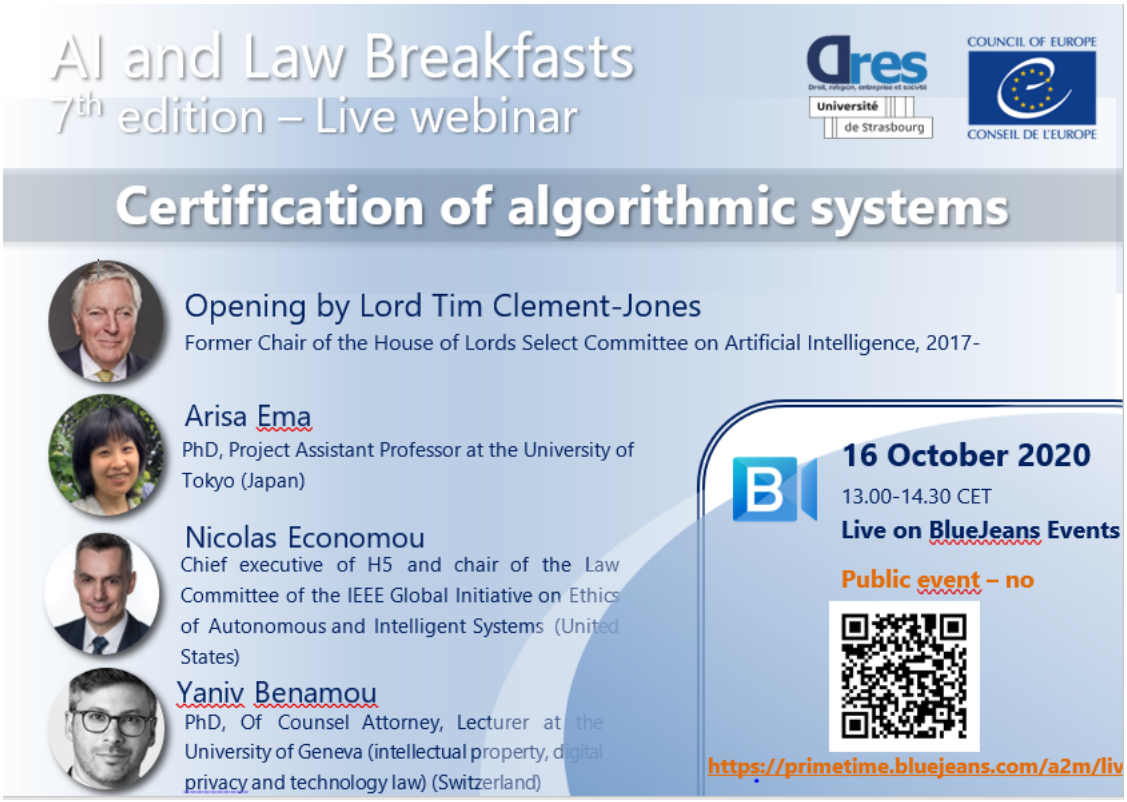
Petits déjeuners AI&Law

7eme édition : certification des systèmes algorithmiques

Résumé des interventions

(non révisé par les auteurs, seul le webinaire fait foi)

Les invités : Lord Tim Clement Jones, ancien président de la Chambre des Lords élu au Comité sur l'intelligence artificielle (2017-2018) (Royaume-Uni), **Arisa Ema,** PhD, professeure assistante de projet à l'Université de Tokyo (Japon), **Nicolas Economou,** directeur général de H5 et président du Comité juridique de l'IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (États-Unis) et **Yaniv Benamou**, PhD, Of Counsel Attorney, Maître de conférences à l'Université de Genève (Suisse).



La 7e édition a porté sur la certification des systèmes algorithmiques, y compris ceux relatifs aux derniers développements de l'intelligence artificielle (IA). A l'heure où l'opinion publique est très préoccupée par les conséquences concrètes de certaines applications de l'IA, comme la discrimination ou l'affaiblissement du contrôle humain, les régulateurs recherchent des solutions concrètes pour créer un climat de confiance pour les utilisateurs. L'idée de garantir, par l'intervention d'un tiers indépendant, la conformité d'un système algorithmique avec un certain nombre de règles, dont les principes des droits fondamentaux, a été soulevée dans la littérature académique et institutionnelle. L'objectif du webinaire sur la certification des systèmes algorithmiques était d'explorer de manière concrète les opportunités, mais aussi les enjeux pratiques d'une telle proposition.

Lord Tim Clement-Jones

Ancien président de la commission spéciale de la Chambre des Lords sur l'intelligence artificielle, 2017

Lord Clement-Jones a une longue expérience des affaires parlementaires au Royaume-Uni, dans l'UE et au niveau international, et a conseillé des gouvernements, des sociétés de premier ordre et des associations commerciales opérant dans des domaines très réglementés, notamment les services financiers, les services publics, la santé, les produits pharmaceutiques et l'environnement.

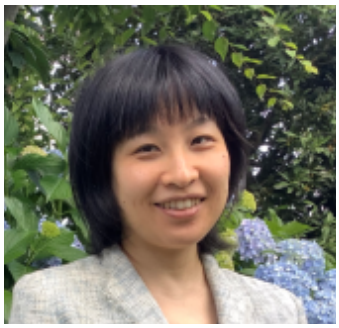
Lord Clement-Jones a ouvert cette session des petits-déjeuners de l'IA et du droit en soulignant l'importance de la confiance du public. Il a remarqué que les deux dernières années ont vu un ensemble florissant de principes éthiques au niveau national et international, et un mouvement concerté vers l'opérationnalisation de ces principes éthiques par référence aux risques, par application, par secteur. Ce type d'approche hiérarchisée des risques doit aboutir à des décisions sur la pertinence d'une approche progressive de la gouvernance, comme des codes volontaires, des normes de gouvernance d'entreprise et une réglementation obligatoire.

Cependant, Lord Clement-Jones est également préoccupé par l'idée que le principe de précaution devrait être le principe général de calibrage du risque, idée qu'il a qualifiée d'approche trop prudente.

Un cadre juridique aurait l'avantage supplémentaire d'accroître la transparence et de promouvoir la confiance, mais Lord Clément-Jones soulève également de nombreuses questions concernant les termes de ce cadre (auditeurs, normes, parties prenantes, secteur public et rôles des utilisateurs).

Lord Clement-Jones a conclu en disant qu'il pense que le label de qualité de la certification ou la marque de certification a un grand potentiel, mais qu'il fait partie d'un programme plus large visant à garantir la fiabilité et à assurer une réglementation efficace de l'IA.

Arisa Ema



Doctorat, Professeur assistant de projet à l'Université de Tokyo (Japon)

Arisa Ema est chercheuse dans le domaine des études scientifiques et technologiques (STS) avec une spécialisation sur l'éthique et la gouvernance de l'intelligence artificielle (IA). Elle dirige et participe à diverses initiatives au Japon et à l'étranger, travaillant à garantir une utilisation responsable de l'IA qui soit inclusive et bénéfique pour tous.

Arisa Ema travaille sur les moyens de construire une gouvernance de l'IA qui inclut le contrôle des données et des algorithmes et sur la façon de mettre en œuvre les principes de l'IA dans les pratiques et les implications sociales de l'IA sur le lieu de travail et les modes de vie.

Dr. Ema a partagé un aperçu des différentes discussions auxquelles elle participe. Ainsi, Dr Ema a expliqué que non seulement plusieurs institutions, dont le Conseil des principes sociaux de l'IA centrée sur l'humain au sein du bureau du Cabinet, ont publié des directives pour l'utilisation de l'IA, mais également des entreprises mondiales et des start-ups. L'Université de Tokyo souligne la nécessité d'examiner les différences régionales en matière de culture, de coutumes et d'institutions lors des discussions sur la gouvernance de l'IA, tandis que la *Japan Society for AI* souligne le fait que tant les chercheurs en IA que le système d'IA doivent respecter ces politiques. Selon le Dr. Ema, certaines entreprises considèrent l'éthique et la gouvernance de l'IA non pas comme une responsabilité sociale de l'entreprise, mais comme un cadre pour la stratégie de gestion. Le Dr Ema a également cité la *Japan Deep Learning Association*, qui organise deux tests de certification afin d'éduquer sur la façon d'utiliser et de mettre en œuvre une IA éthique qui inclut des questions d'implication éthique, légale et sociale.

Cependant, il est difficile d'estimer la qualité des services d'IA en se basant sur une seule entreprise. C'est la raison pour laquelle Dr. Ema et ses collègues de l'université de Tokyo sont en train de créer un modèle d'évaluation des risques avec une discussion entre plusieurs parties prenantes, y compris des cas d'utilisation au Japon. Dans ce modèle, le concept de base est que les risques de l'IA doivent être considérés comme relatifs à des couches de systèmes d'IA, de fournisseurs de services et d'utilisateurs.

Le Dr Ema souligne également le fait que la structure de l'industrie est également importante. Le Japon a, en effet, beaucoup d'entreprises en BtoB, impliquant une longue chaîne d'approvisionnement qui soulève des questions sur la répartition de la responsabilité partagée. C'est pourquoi le JDLA a créé un groupe d'étude sur la gouvernance de l'IA qui vise à examiner le réseau de l'écosystème impliquant des compagnies d'assurance et d'audit ainsi qu'un système de dénonciation et un comité tiers pour l'enquête sur les incidents. Jusqu'à présent, leur conclusion est que la nature de l'IA, qui est en apprentissage constant, rend difficile la production d'une évaluation de qualité. De plus, il faut tenir compte des différences d'impact entre les secteurs.

Nicolas Economou

Directeur général de H5 et président du comité juridique de l'IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (États-Unis)

Nicolas Economou a été un pionnier dans l'avancement de l'application et de la gouvernance de l'IA dans les systèmes juridiques. Il dirige les comités juridiques de *The Future Society* et de *l'IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems*. Il est membre du *Council on Extended Intelligence* (CXI).

Nicolas Economou a présenté l'organisme IEEE SA et son “Initiative globale sur l’éthique des systems autonomes intelligents ”. L'IEEE SA est l'une des plus grandes organisations techniques travaillant sur la question de l'IA digne de confiance. Le terme "systèmes autonomes intelligents", plutôt que simplement AI, est choisi en raison de sa notion plus large qui peut inclure toute la chaîne de l’IA ainsi que parce que e le travail de l'association est centré majoritairement sur les systèmes technologiques émergents.

L'IEEE a construit, en collaboration avec la CEPEJ et l'Agence de sécurité des données, un cadre, à savoir le "*Informed Trust Framework*" qui apporte une réponse à la question majeure : "comment rendre opérationnelle une IA digne de confiance" ?

La première étape à franchir consiste à se concentrer sur une définition adéquate de la fiabilité : elle doit être simple, uniforme, largement applicable et capable de s'adapter à différentes cultures et différents environnements, ainsi qu'applicable aux futures innovations en matière d'IA.

Le "cadre de confiance informé" peut aider à mettre en pratique les droits inscrits dans la CEDH et à fournir un cadre établissant les preuves dont nous avons réellement besoin pour déterminer dans quelle mesure les systèmes d'IA prévus par la loi (et au-delà) sont dignes de confiance, dans quelle mesure ils sont efficaces, dans quelle mesure les opérateurs sont compétents, dans quelle mesure les personnes sont responsables et enfin dans quelle mesure nous pouvons réellement avoir des preuves de transparence.

Nicolas a consacré les quatre éléments constitutifs (principes ou conditions de confiance) contenus dans le cadre de l'IEEE :

- L’éfficacité
- La compétence, qui ne concerne pas seulement les opérateurs, mais aussi les utilisateurs. Les utilisateurs doivent comprendre comment l'IA interagit et comment les systèmes d'IA peuvent influencer leur perception à la lumière de la décision finale.
- La responsabilité de l'identification des personnes responsables en cas de défaillance d'un système d'IA.
- La transparence, qui comprend deux éléments : 1) l'accès à l'information et 2) l'accès à une explication qui doit être adéquate pour les différentes parties prenantes.

De plus, Nicolas Economou a fourni une étude de cas américaine sur l'utilisation de la technologie TAR (*Technology assisted review*) dans le domaine civil et pénal. Plus précisément, dans ce dernier cas, l'IA est utilisée pour l'examen d'un vaste groupe de documents. Les études menées de 2008 à 2011 par le *National Institute of Standards and Technology* américain ont démontré deux choses :

A) La révision assistée par la technologie TAR aux États-Unis peut être plus efficace que l'humain

B) Ces technologies sont une méthode efficace pour mener à bien l'établissement des faits.

Ce que Nicolas Economou a montré, c'est que, dans le contexte considéré des constatations de faits, l'IA peut être qualifiée de digne de confiance lorsque les quatre conditions incluses dans le cadre de l'IEEE, sont correctement évaluées.

Il a conclu en soulignant la mission de l'IEEE : développer des instruments que les régulateurs, les tribunaux, les avocats et autres peuvent utiliser pour s'assurer que ces quatre conditions de confiance sont remplies.

Yaniv Benamou

Docteur en droit, avocat-conseil, chargé de cours à l'Université de Genève, (droit de la propriété intellectuelle, de la vie privée et des technologies numériques) (Suisse)

Dr. Benamou est expert de l'OMPI pour le droit d'auteur et les musées, Directeur exécutif de l'université d'été de droit numérique et membre du Comité d'experts de l'Initiative numérique suisse. Une de ses recherches actuelles porte sur l'autorégulation, y compris les mécanismes de certification et leur interface avec la responsabilité et la participation du public.

En tant qu'expert dans le domaine de la gouvernance numérique et de l'autorégulation, **Yaniv Benamou** a observé une augmentation importante de l'utilisation de la certification et de labelling, qui sont des instruments d'autorégulation.

Concernant cette hausse, trois recours juridiques ont été sélectionnés par Yaniv :

i) **Le premier défi concerne les différents types de modèles d'autorégulation**.
Tout d'abord, Dr. Benamou a expliqué la différence entre la corégulation, qui nécessite une intervention de l'État/approbation de l'État, comme dans le cas du RGPD ou des codes de conduite, et le label, qui est un instrument d'autorégulation purement privé ne nécessitant pas d'intervention de l'État. En tant que précédent expert universitaire impliqué dans la fondation, Yaniv a apporté l'exemple d'une étude de cas : le "*Swiss Digital Trust Label*" issu de l'initiative Swiss Digital. Le *Swiss Digital Trust Label*, comme c'est généralement le cas pour les labels, est associé à une marque de certification et à un certificateur ou un auditeur pour la vérification de la conformité.
Deuxièmement, Yaniv Benamou se demande comment assurer la légitimité démocratique lorsque ces modèles sont appliqués. En effet, ces normes sont souvent rédigées par quelques décideurs sans contrôle ni intervention de l'État et peuvent parfois servir d'outil de marketing ou même d'éthique de façade. La légitimité démocratique pourrait être mieux assurée avec des normes procédurales qui permettent la participation de toutes les parties

prenantes, y compris les organisations de consommateurs. Ce type de processus transparent a été suivi par le label suisse de confiance numérique.

ii) **Le deuxième défi est la clarification de la responsabilité et de l'obligation de rendre compte de toutes les parties concernées**.

L'obligation de rendre compte et la responsabilité doivent être assurées pour les législateurs, ainsi que pour les auditeurs et les utilisateurs certifiés. Pour les législateurs, ils restent responsables du respect des lois. En ce qui concerne les auditeurs, ils peuvent être considérés comme responsables lorsque l'audit n'est pas effectué correctement, comme l'a démontré l'étude de plusieurs cas : parmi ces cas, Yaniv Benamou a choisi une décision de la Cour américaine dans laquelle une société privée de certification est jugée responsable de l'insuffisance de l'audit des sociétés, telles que le New Times ou Apple (ordonnance finale de la FTC, n° C45-12, 12 mars 2015, Trust-e). Enfin, en ce qui concerne les utilisateurs certifiés, leur responsabilité peut être engagée même s'ils respectent le code de conduite. Par exemple, en 2011, ALSTOM (une société de transport) a été condamnée pour ne pas avoir empêché la corruption malgré la mise en œuvre du code de déontologie.

La question concernant le deuxième défi est la clarification des effets normatifs du label ou de la certification lors de l'évaluation de la responsabilité de l'entreprise labellisée. Ces effets normatifs peuvent aller d'une simple orientation d'interprétation pour les tribunaux à une véritable présomption de conformité. La suggestion faite par Yaniv Benamou est, qu'afin de protéger la sécurité juridique, la meilleure solution serait de s'appuyer sur la certification lorsque cela est possible.

iii) **Le prochain et dernier défi présenté concerne la manière d'auditer lessystèmes algorithmiques**.

M. Yaniv Benamou a souligné la nécessité que chaque certification ou label contienne différents critères vérifiables à utiliser par une entité/un cabinet d'audit. Deux approches sont possibles: a) une approche multicouche avec une couche contenant des descriptions simples compréhensibles du point de vue de l'utilisateur standard et b) une couche modifiable contenant des spécifications détaillées pour les auditeurs et les certificateurs.

Une autre façon de vérifier les systèmes algorithmiques est d'utiliser le plus possible des critères binaires ou des critères qui se réfèrent à des normes reconnues. M. Benamou reconnaît que le défi de cette dernière approche consiste à traiter des critères moins binaires, comme ceux qui sont soumis à interprétation, par exemple les meilleures pratiques ou les critères qui évoluent. Dans cette optique, les critères à prendre en considération sont par exemple l'explicabilité ou l'éthique, qui sont utilisés dans l'initiative du label suisse de confiance numérique.

Enfin, le Dr. Benamou a été confronté à la question de la nature obligatoire ou volontaire de la certification pour les technologies à haut risque. Les questions essentielles concernent les législations politiques et techniques qui définissent le champ d'application matériel des technologies à haut risque tout au long du cycle de vie des systèmes basés sur l'IA (par exemple, quel élément précis doit être certifié pendant tout le cycle de vie de l'IA). Une autre

question est : qui est le certificateur le plus adéquat pour susciter la confiance du public ? S'agit-il d'un certificateur de label international ou d'une société privée ?

En conclusion, 1) la corégulation est préférable pour assurer la légitimité démocratique et la sécurité juridique avec la présomption de conformité ; 2) la vérifiabilité des algorithmes est possible mais une action législative est nécessaire, car d'un point de vue juridique, les législations existantes sont trop limitées dans leur portée. Par exemple le RGPD a l'avantage d'être un modèle corégulé avec la légitimité démocratique renforcée et la sécurité juridique de la présomption de conformité qui lui sont associées, mais il a l'inconvénient d'être limité à la protection des données et est caractérisé par un manque de coordination entre les multiples organismes de certification.

En fin de compte, Dr Benamou estime que l'autorégulation, en particulier la corégulation, est un modèle réglementaire approprié car il offre la flexibilité nécessaire pour les technologies techniques à évolution rapide, mais seulement à condition que le contrôle démocratique et la sécurité juridique soient assurés.