

Lettre des "Lundi de la cybersécurité" n° 68

Intelligence artificielle générative géniale et dangereuse



Lundi 18 mars
18h00-20h00

Par webinaire Zoom



Image produite par Dell-E

Intelligence artificielle générative : géniale et dangereuse



Pr Jean-Paul Delahaye

Professeur émérite à l'université de Lille et
chercheur au laboratoire Cristal
Centre de recherche en informatique, signal et
automatique de Lille.



Organisateurs



Pr Ahmed Mehaoua
Université Paris Cité



Béatrice Laurent



Gérard Peliks



« Dessine-moi un mouton » demandait le Petit Prince au romancier Antoine de Saint Exupéry.

Pour cet auteur l'action était simple, même si le premier mouton qu'il avait dessiné n'était pas le mouton que le Petit Prince espérait. Par contre avec une demande beaucoup plus complexe et farfelue comme « dessine-moi une course de pères Noël à vélo dans un désert », pour Saint Exupéry, ce ne serait pas très facile, alors que pour une application d'IA générative, comme Dell-E, ça ne pose aucun problème.

Regardez le résultat : bluffant ! obtenu par des algorithmes, des calculs de statistiques sur des pixels, et nous restons dans

le merveilleux, dans celui des machines qui créent, dans les pouvoirs de l'intelligence artificielle générative !

Les méthodes d'apprentissage profond qui utilisent des réseaux de neurones pour établir des liens entre les pixels d'images, des textes, et d'autres éléments, produisent des résultats, textes, images, et même morceaux de programmes et de la musique, qui ont de quoi nous épater. Bravo pour la créativité de ce qui n'est finalement qu'une machine dopée avec des quantités

phénoménales de données, traitées par des algorithmes, et par un grand modèle de langage (LLM) et sans intervention humaine.

Vous lirez l'article du Pr Delahaye dans la revue POUR LA SCIENCE N° 557, « *Du texte à l'image : Cap franchi pour l'IA* » qui paraîtra dans quelques jours, dans le numéro de mars 2024.

Je donne la plume à l'intervenant

Avec ChatGPT, ses successeurs, et les nouveautés qui apparaissent chaque jour provenant de ces « intelligences artificielles génératives », il se produit une révolution dont les conséquences sont déjà devenues évidentes dans de nombreux secteurs d'activité : enseignement, traduction, programmation informatique, journalisme et plus généralement rédaction de textes variés, création graphique, etc.

Comprendre ce que font ces systèmes est difficile car ils évoluent très vite et se présentent sous la forme de boîtes noires dont personnes, pas même ceux qui les construisent ne sont capables d'expliquer entièrement les miracles qu'ils produisent.

L'exposé décrira cependant les principes mis en œuvre pour la fabrication de ces nouvelles IA qui utilisent massivement l'intelligence humaine. On donnera des exemples de ce qu'on en tire. La fiabilité parfois douteuse des informations produites — les fameuses hallucinations —, la possibilité de les piéger, le problème des sources d'informations utilisées lors de leur mise au point, les biais volontairement ou involontairement introduits dans leurs comportements et les textes et images que ces IA génèrent doivent nous interroger.

Il faut aujourd'hui tenter d'anticiper les conséquences que cela peut avoir sur la cyber-sécurité.

Je reprends la plume

Qui est le Pr Jean-Paul Delahaye ?



Jean-Paul Delahaye est professeur émérite, et disons-le, au mérite, à l'université de Lille. Il est chercheur au laboratoire Cristal (Centre de recherche en informatique, signal et automatique de Lille). Il travaille sur la théorie computationnelle des jeux et sur le hasard. Ses anciens thèmes de recherches ont été la programmation logique en lien avec l'intelligence artificielle.



L'intelligence artificielle générative est un domaine de l'IA qui se concentre sur la création de données nouvelles et originales, telles que du texte, des images et de la musique en se basant sur des quantités énormes de données dont elle cherche à extraire les régularités statistiques.

Jean-Paul Delahaye est également rédacteur des articles mathématiques de la chronique *Logique et calcul* de la revue *Pour la Science*.

Ce professeur écrit des articles et des livres d'information scientifique destinés au public non spécialisé, notamment sur le nombre π et sur les nombres premiers. Citons :



- Outils logiques pour l'intelligence artificielle
- Systèmes experts : organisation et programmation des bases de connaissance en calcul propositionnel
- Jeux mathématiques et mathématiques des jeux
- Information, complexité et hasard
- Merveilleux nombres premiers : Voyage au cœur de l'arithmétique
- Au-delà du Bitcoin : Dans l'univers de la blockchain et des cryptomonnaies, Paris, Dunod, 2022.



Demande d'inscription pour le Lundi de la cybersécurité du mois de mars 2024

Lundi 18 mars, à partir de 18 h 00, par visio-conférence Zoom.



Nos « Lundi de la cybersécurité » sont gratuits et veulent vous offrir une fête technologique. Demandez votre inscription, par courriel, nous vous enverrons, un peu avant le jour de l'événement, un hyperlien vers la visioconférence.

Les demandes d'inscriptions sont à adresser à :

beatricelaurent.CDE@gmail.com

Dans votre message d'inscription, nous apprécierions que vous écriviez quelques mots sur vous et sur l'intérêt que vous accordez au sujet traité. Evitez le style télégraphique, nous sommes toujours heureux de vous lire et de connaître votre intérêt pour nos événements.

Les prénoms, noms et adresses mails des inscrits seront connus des organisateurs et communiqués aux intervenants.

Si vous êtes dans la distribution de mes lettres des « Lundi de la cybersécurité » mensuels, mais vous souhaitez ne plus les recevoir, une simple demande par mail suffit. Si vous voulez être ajoutés à la liste de distribution, demandez-le nous.

Et bien entendu, si vous vous inscrivez pour assister à notre évènement, soyez connectés, avec votre prénom et votre nom, le lundi 26 février, dès 18 h 00. La salle d'attente Zoom sera ouverte aux alentours de 17 h 00. Prévenez-nous, dès que possible si vous êtes inscrits mais avez un empêchement qui ne vous permet pas d'assister à ce webinaire.

Quelques minutes avec une organisation :



Suivant la tradition de nos « Lundi de la cybersécurité », entre l'exposé de l'intervenant et la session questions / réponses, autour de 19 h 15, nous donnons pour une dizaine de minutes la parole à une organisation à but non lucratif, qui opère dans l'écosystème du numérique et dans la sécurité de l'information.



L'intervenant du Lundi de mars sera **le général d'armée (2S) Marc Watin Augouard** qui nous parlera **du FIC 2024** (renommé Forum InCyber) qui se tiendra à Lille Grand Palais de 26 au 28 mars, soit 8 jours après notre Lundi de mars. Et si vous visitez ce grand évènement, venez me voir, je tiendrai le stand de l'ARCSI avec quelques camarades.

Le coin des « Lundi de la cybersécurité » précédents.

Vous trouverez tous les détails, slides, parfois aussi les enregistrements sur le web de l'ARCSI, en <https://www.arcsi.fr/> pour les plus récents. Les événements plus anciens sont en https://www.arcsi.fr/evt_passes.php.

Et visualisez aussi les retransmissions des conférences tenues aux 15èmes Rencontres de l'ARCSI, en https://www.arcsi.fr/colloques/actes_collq_231205.php. Jean-Paul Delahaye y a traité le sujet : *Le devenir des fonctions à sens unique.*

Le coin coin de l'IA générative

Question posée à Gemini (anciennement BARD, IA générative de Google) :



« **L'Intelligence artificielle générative est-elle géniale et dangereuse ?** »

Voici une partie de la réponse reçue :

*L'intelligence artificielle générative est à la fois **géniale et dangereuse.***

Géniale car elle offre des potentialités extraordinaires dans de nombreux domaines :

Créativité :

- Générer des contenus artistiques (musique, peinture, littérature)
- Concevoir de nouveaux produits et services
- Développer des outils d'aide à la création

Productivité :

- Automatiser des tâches répétitives et fastidieuses
- Gagner du temps et de l'énergie
- Accroître l'efficacité dans divers domaines

Amélioration de la vie quotidienne :

- Personnaliser les expériences et les services
- Proposer des solutions innovantes aux problèmes du monde réel
- Améliorer la qualité de vie

Dangereuse car elle soulève des risques importants :

Désinformation et manipulation :

- Créer de faux contenus (deepfakes) pour tromper les gens
- Propager des informations erronées et de la propagande
- Influencer les opinions et les comportements

Perte de contrôle et d'autonomie :

- Dépendance accrue aux systèmes d'IA
- Risque de concentration du pouvoir entre les mains de quelques acteurs
- Diminution de la liberté et de l'autonomie individuelles

Biais et discrimination :

- Reproduction et amplification des biais présents dans les données d'apprentissage
- Exclusion et marginalisation de certains groupes
- Injustice et inégalités dans l'accès aux technologies

En résumé, l'IA générative est un outil puissant à double tranchant.

Son développement et son utilisation doivent être accompagnés d'une réflexion éthique et d'une réglementation appropriée pour maximiser ses avantages et minimiser ses risques.

L'avenir de l'IA générative dépend de notre capacité à la maîtriser et à l'utiliser de manière responsable. Enjeux importants pour l'humanité, il est crucial d'engager un dialogue ouvert et inclusif sur ces questions.

Bluffant n'est-ce pas, un peu comme *la course de pères Noël à vélo dans un désert !!!*

Gérard

Lundi de la Cybersécurité d'avril

lundi 15 avril 18h-20h

en distanciel par Zoom

Cybersécurité et aspects humains

avec Virgile Augé