Un long chemin vers une intelligence artificielle véritablement ouverte

Par Frédéric DUPEUX, Chief Information Security Officer, Banque Havilland

Le terme *open source* est devenu à la mode dans le domaine de l'intelligence artificielle (IA), avec des acteurs majeurs comme Meta et Elon Musk qui s'en défendent. Cependant, aucun consensus n'existe quant à la définition d'une IA ouverte. Cette ambiguïté permet aux entreprises leaders de manipuler le concept à leur avantage, ce qui pourrait renforcer leur position dominante.

L'essor de l'intelligence artificielle entraîne de nombreux questionnements d'ordre éthique, juridique et conceptuel au sein de la communauté *open source*. Si l'*open source* bénéficie d'une définition claire, i.e. un code source accessible, modifiable et redistribuable ; ce n'est pas le cas de l'IA ouverte. En effet, aucune définition consensuelle n'a été adoptée en raison d'intérêts divergents et de la complexité des systèmes d'IA par rapport aux logiciels traditionnels. Contrairement aux logiciels, les systèmes d'IA sont dépendants de grandes quantités de données et impliquent de nombreux composants tels que les données d'entraînement, le code de prétraitement et l'architecture du modèle.

Une des préoccupations majeures de la communauté *open source* concerne, à raison, les droits de propriété intellectuelle lorsque des algorithmes sont formés sur des quantités importantes de données sans connaissance de leur provenance. Cette incertitude décourage certains développeurs à partager leurs données, ce qui pourrait entraver le progrès dans le domaine de l'IA *open source*. Il s'agit d'une véritable bataille entre l'ensemble des acteurs du milieu, la performance des modèles actuels dépendant directement du volume de données ingurgitées.

La complexité et le manque de transparence de l'IA rendent difficile la compréhension ou la rationalité des décisions de l'IA en se basant uniquement sur le code source, remettant en question le concept d'IA ouverte. La génération de texte, d'images, de vidéos ou de code soulève donc des problèmes de licence, de sécurité et de réglementation en raison du manque de clarté sur leur origine.

Du partage au pillage

Historiquement, l'*open source* est né d'une volonté de partage et de la nécessité, pour les fournisseurs de matériel, de proposer des logiciels pour leurs machines. Aujourd'hui encore, ce mode de fonctionnement évolue constamment et encourage l'innovation, la collaboration et le partage des connaissances au sein d'une communauté diversifiée. Alors que le logiciel était au centre de l'évolution des systèmes informatiques durant les premières décennies, les données jouent un rôle central dans les avancées de l'IA depuis ces deux dernières décennies.

Les entreprises technologiques leaders dans le domaine de l'IA ont adopté diverses stratégies vis-à-vis de l'*open source*. Certains modèles IA sont partagés plus librement que d'autres. Meta a, par exemple, publié son modèle Llama 2 en tant qu'*open source*, tandis qu'OpenAI a restreint l'accès à ses modèles les plus puissants. Google propose des modèles Gemma accessibles gratuitement et conçus pour rivaliser avec les modèles de ses concurrents. De nombreux modèles qualifiés d'*open source* sont pourtant accompagnés de restrictions d'utilisation, en contradiction avec les principes mêmes de l'*open source*.

L'utilisation des données pour la création des IA est l'un des principaux points d'achoppement. Si les modèles préformés sont souvent partagés, les ensembles de données pour les former ne le sont pas, ce qui limite la possibilité de modifier et d'étudier pleinement ces modèles. Ce manque de transparence des données est un obstacle important à une véritable ouverture de l'IA.

En effet, selon Aviya Skowron, responsable des politiques et de l'éthique au sein du groupe de recherche à but non lucratif sur l'IA EleutherAI, il existe un manque de clarté quant à l'utilisation d'informations protégées par le droit d'auteur dans la formation de modèles IA. Stefano Zacchiroli, professeur à l'Institut Polytechnique de Paris et acteur majeur dans le processus de définition de l'Open Source Initiative (OSI), estime, quant à lui, qu'une description complète des données d'entrainement est essentielle pour que les modèles IA soient considérés comme *open source*.

Les grandes entreprises hésitent à partager les données d'entrainement en raison d'avantages concurrentiels et de préoccupations d'ordre réglementaire. Cette réticence nuit à l'éthique même de l'*open source* et ne peut que renforcer le pouvoir des grandes entreprises technologiques. En effet, selon le site Patronus.AI, le modèle d'openAI GPT4 ainsi que les modèles Mistral/Mixtral et Lama2 de Meta concentreraient le plus grand nombre de violations de copyright. Avec 44% de contenu protégé par des droits d'auteur, GPT4 est de loin le modèle générant le plus de reproductions exactes de contenu protégé.

L'IA à fort impact sociétal sera forcément ouverte

Une définition claire et largement acceptée de l'IA à code source ouvert est nécessaire et urgente afin d'empêcher ces entreprises puissantes de dicter des termes qui conviennent à leurs intérêts. Une IA véritablement ouverte aurait de nombreux avantages, tels que la promotion de l'innovation, de la transparence, de la responsabilité, de l'équité et des valeurs humaines, en bref une IA à fort impact sociétal et éthique. Une IA ouverte permettrait de palier aux principales menaces générées par l'IA, à savoir son utilisation malveillante ainsi que la perpétuation de préjugés et de discriminations. Une IA ouverte permettrait de générer d'importants progrès sociaux et économiques, en particulier dans des secteurs tels que les soins de santé, l'éducation et la finance.

Par exemple, dans le secteur bancaire, les promesses de l'IA sont indéniables, notamment pour les systèmes de détection de fraudes afin de mieux anticiper les activités criminelles. Il est aussi possible d'imaginer une nouvelle forme de relation clients ou de conseils financiers personnalisés qui pourraient être adaptés aux besoins individuels. Enfin, l'IA devrait permettre d'envisager une nouvelle forme de gestion des risques et de prévision des crises. Les algorithmes d'apprentissage pourraient rapidement identifier les prémices d'une crise afin de mieux en gérer les effets.

A l'heure de la prise de conscience sur les incroyables potentialités de l'intelligence artificielle, mais aussi des menaces qui l'accompagnent, il est urgent de proposer une réflexion sur une utilisation responsable et éthique de l'intelligence artificielle. L'*open source*, basé depuis ses débuts sur des valeurs de partage et de transparence, peut offrir une voie vers une intelligence artificielle en lien avec nos valeurs humaines. Cette voie nécessitera une collaboration continue entre les développeurs, les chercheurs et les régulateurs pour garantir son avenir.

PwC Cybersecurity & Privacy Day 2024:

«The AI paradox : a blessing or a curse»

Another successful edition of the PwC Cybersecurity & Privacy Day came to a close on 5 June 2024 at PwC Luxembourg's Crystal Park premises. A total of 200 attendees and speakers took part in a day filled with keynote speeches and workshops, discussing "The AI paradox: a blessing or a curse." Furthermore, both the Jury's and the People's Choice Award were presented at the end of the day, with Contrast Security winning the Jury's Choice Award and Data & More taking home the People's Choice Award. During the event, this year's "Out of the shadows: CISOs and DPOs in the spotlight 2024" market survey was discussed.

Morning Speaker highlights

The exceptional capabilities of Artificial Intelligence (AI) have the potential to revolutionise cybersecurity. However, it is crucial to safeguard AI systems themselves from attacks to prevent harm to individuals, processes, and technology. Throughout the day, this year's theme - "The AI Paradox: A Blessing or a Curse," - shone a spotlight on the dual-edged nature of AI in the realm of cybersecurity and privacy.

The morning kicked off with welcome words from Koen Maris, Advisory Partner, Cybersecurity & Privacy Leader at PwC Luxembourg and host of the event. "AI is here, and it will never go away. Is it a weapon or a tool? Is it friend or foe?" He was joined in the greeting by Grant Waterfall, Partner, Cyber Security & Privacy Leader, PwC EMEA who praised the mix of people in the audience, ranging from tech experts to start-ups, regulators, and cyber and privacy specialists, as being part of the true value of the day.

Mika Lauhde, Senior Fellow at Maastricht University, gave an impressive talk entitled, "AI and sign of the times", in which he discussed the geopolitical context of our times. As Head of Technology, Delegation for CyberSpace, International Committee of Red Cross, he leads the R&D unit that develops new digital technologies for global humanitarian use with cybersecurity in mind. He pointed out three reasons why there is currently a need to "double down on security": Geopolitical fragmentation (which increases the risk of espionage, sabotage and cyberattacks), the emergence of AI itself (which poses new challenges and opportunities) and the new era of cyber transparency, with the pressure from regulators (DORA, CS (Cyber Security) Act) demanding more mandatory reporting.

Mika also warned against the monopolisation of technology, with only a few companies dominating everything from mobile technology to the internet.

Dr. Donia Elkateb. Senior IT-Security Engineer, European Investment Bank (EIB), gave a talk on "Application Security and AI," saying notably that "The bad guys are also using AI," and punctuating this comment with an example of how hackers are using ChatGPT to create polymorphic malware.

This was followed up by Stan Scharnigg, Co-founder, Chunk Works, who told the crowd, "The AI genie is out of the bottle," comparing our current situation with the dawn of nuclear power, with both great applications but also the danger of bad actors. He made a fascinating point about Quantum Computing and AI and how breakthroughs could come much faster than we ever imagined. He advised that agility holds the key in dealing with the future, but are we ready for the future?

In his talk "Security in the era of Generative AI", Nico Sienaert, Sr. GTM Lead Security at Microsoft proposed a startling analogy. "If cybercrime, which currently is estimated at around US 8 trillion, was a country's economy, it would be the third biggest in the world." In fact, estimates vary on the scale of cybercrime, but it gave the audience a vivid image of what the world is up against.

Herwig C. H. Hofmann, Professor of European and Transnational Public Law, Head of the Department of Law Univer-sity of Luxembourg, FEDEF provided key insights in relation to "Information Management in the Regulation of Privacy and AI". He explained that information management is emerging as the central focus of EU regulation in digitalisation package, imposing obligations on individual actors which can be only complied with by means of an increasingly granular collection of information sourcing.

The afternoon was dominated by a series of workshops where attendees could attend up to two on a variety of topics. The late afternoon keynote was a real treat with Geoff White, Investigative Journalist, Speaker and Author who spoke about, "Cybercrime's Enablers- Inside the Network of Financial Crooks that Keep Hackers in Business. With his new book "Rinsed" hot off the presses, Geoff, an incredible public speaker, wowed audience with his stories of how technology can greatly enable financial crime using real life examples, including the dark web. Illuminating, scary and entertaining all at the same time.

The Pitching Competition 2024

As every year, one of the most exciting aspects of the conference was the pitching competition. After an international call for submissions, PwC Luxembourg selected five cybersecurity and privacy companies with relevant solutions for the Luxembourgish market. On 5 June, the Jury's and the People's Choice Awards were presented at the end of the day, with 'Data & More' winning the People's Choice Award and 'Contrast Security' taking home the Jury's Award.

Jury member, Roman Borisovich, International Financier, who has been coming to the event for five years, spoke publicly about the rising quality of the pitches over time and how difficult this year's decision was, saying that it is a testimony to the recognition of this event that rather than just start-ups, but is now more established companies that are travelling from across the world for a chance to pitch at this event.

Cybersecurity, Privacy and Regulatory expert insights

The 2024 edition of the Out of the shadows: CISOs and DPOs in the spotlight! Market survey was at the heart of the roundtable discussion, moderated by Maxime Pallez Cybersecurity Director, Cybersecurity Governance, Risk and Compliance Leader and Antonin Jakubse; Privacy Senior Manager, PwC Luxembourg with the support of the Club de la Sécurité de l'Information (CLUSIL), the Commission Nationale pour la Protection des Données (CNPD), the Commission de Surveillance du Secteur Financier (CSSF), and the Institut Luxembourgeois de Régulation (ILR). This allowed for some valuable insight both regulatory and from cybersecurity professionals.

This roundtable was followed by, "The latest trends in cyberattacks, technology watch from PwC LU's Ethical Hackers" with Maxime Clementz, Ethical hacker & Cybersecurity Senior Manager, PwC Luxembourg.

Finally, at the very end of the day, Koen Maris, Advisory Partner, and current Cybersecurity & Privacy Leader at PwC Luxembourg, announced that he would be leaving Luxembourg, although he will remain in the PwC network. At the same time, Thierry Kremser, Advisory Partner, Deputy Advisor and Technology Leader, PwC Luxembourg, announced that replacing Koen would be not one but two people, Maxime Pallez, Director at PwC Luxembourg currently leading the Cybersecurity Governance, Risk and Compliance practice and Simon Petitjean , Cybersecurity Director, Offensive Security & Red Team Leader, PwC Luxembourg.

The PwC Cybersecurity & Privacy Day is PwC's annual flagship event, the mission of which is to help CISOs, DPOs and CEOs ensure they keep their organisation secure in a digital society. The organisers extend a warm thank you to all the participants and speakers for another successful get together.