Rapport de synthèse 18e Security Talk : L'IA générative - outil de production performant ou risque fondamental pour la sécurité ?

Exposé de Katharina Fulterer

Opportunités, défis et possibilités d'utilisation de l'intelligence artificielle (IA)

Quelles sont les conséquences réelles de l'IA pour l'homme ? Il y a des déclarations, comme celle de Bill Gates, qui affirme que l'IA est notre nouveau téléphone portable. Ou Sundar Pichai de Google qui la compare à des innovations comme le feu et l'électricité. Mais il y a aussi des voix comme celle de Jeff Bezos, qui disent que nous n'en sommes qu'au début de cette évolution. Quels sont donc les domaines d'application possibles, où pouvons-nous utiliser l'IA de manière utile en tant qu'entreprise ou institution publique ? Le thématique de l'optimisation des itinéraires est certainement de grande importance. Nos applications de navigation, nos trackers de fitness nous aident dans différents domaines de notre vie. De grandes entreprises comme la Poste suisse utilisent également l'IA pour optimiser les itinéraires lors de la livraison de colis. Différentes données sont prises en compte, comme les données météorologiques, le volume de trafic ou le volume des colis, afin de garantir une utilisation efficace de leurs ressources. L'IA apporte également une contribution précieuse à la lutte contre la fraude et est utilisée par les banques pour vérifier chaque jour des milliards de transactions.

L'IA générative en tant que nouvel acteur

Il existe aujourd'hui de nombreux exemples, par exemple dans le domaine de la génération d'images, où **l'IA s'approche effectivement des performances humaines, voire les dépasse.** Dans le contexte de l'IA générative, il faut différencier le sujet. Il s'agit en fin de compte de la thématique des machines intelligentes ou des programmes informatiques intelligents - l'intelligence artificielle. Depuis les années 1950, différents experts font des recherches à ce sujet. Que ce soit dans le domaine du Machine Learning ou, ces dernières années, de plus en plus dans le domaine des réseaux neuronaux, qui relèvent du Deep Learning.

Si nous examinons le Deep Learning d'un peu plus près, nous pouvons constater qu'il y a deux domaines qui présentent aujourd'hui un potentiel de développement. Il s'agit d'une part du domaine de l'IA prédictive, l'IA classique. Il s'agit là de reconnaître des modèles sur la base

de données historiques et de prédire des événements futurs. Les prévisions météorologiques ou l'optimisation des rouages en sont des exemples. En revanche, depuis le lancement de ChatGPT, nous voyons des modèles qui relèvent du domaine de l'intelligence artificielle générative. Il s'agit de générer des résultats entièrement nouveaux sur la base de données d'apprentissage. Ceux-ci peuvent être générés sous forme de texte, de voix, d'images et de bien d'autres modalités. **L'intelligence artificielle générative est un moyen qui nous permet d'entrer en dialogue avec l'intelligence artificielle.** C'est quasiment une interface de dialogue avec laquelle nous, les humains, pouvons tenir une conversation dans notre vie quotidienne.

Par ailleurs, il existe des applications qui, combinées à l'IA prédictive et générative, fonctionnent aujourd'hui beaucoup mieux qu'il y a quelques années. Il s'agit par exemple de la traduction de langues, du test, de la correction et de la génération de code logiciel, ainsi que de la création automatisée de documents. Et puis il y a des domaines où l'IA générative nous donne des résultats qui étaient impossibles auparavant. Par exemple, une interaction personnalisée avec le client ou le sujet des deepfakes.

Risques liés à l'utilisation de l'IA

Outre une multitude d'opportunités, l'utilisation de l'IA comporte également de nombreux risques qui doivent être abordés. Ceux-ci peuvent être divisés en trois domaines. Il y a les risques qui nous concernent en tant qu'êtres humains, par exemple lorsqu'une IA produit des résultats qui comportent un certain biais ou qui révèlent des données personnelles. Il y a des risques qui concernent les entreprises, par exemple lorsque des données sont réutilisées et soumises à des obligations légales ou lorsque des employés sont mécontents parce qu'ils doivent confier certaines tâches à une IA. Il y a également des risques qui concernent la société, par exemple l'influence de l'intelligence artificielle sur la sécurité de l'approvisionnement en Suisse. Enfin, il s'agit de développer des solutions d'IA fiables et de les mettre en œuvre.

Applications de l'IA générative dans les entreprises suisses

Où les applications d'IA générative sont-elles utilisées aujourd'hui par les entreprises suisses sur le marché suisse ? Dans le discours public, la difficulté de l'utilisation de l'IA dans le contexte de l'entreprise n'est pas toujours communiquée de manière transparente. Selon l'OCDE, seules 8% des entreprises de la zone OCDE ont utilisé l'IA en 2023. Cela s'explique par la complexité de la mise en œuvre. Souvent, après un PoC (Proof of Concept) réussi, les entreprises ne parviennent pas à faire passer le projet à grande échelle et à le rendre productif. Dans le monde des entreprises, de nouveaux facteurs d'influence entrent en jeu et les nouvelles solutions d'IA doivent être connectées aux systèmes existants. Souvent, de tels projets échouent aussi parce que certains thèmes de sécurité et de gouvernance n'ont pas été suffisamment éclairés et pris en compte dans un premier PoC allégé. C'est pourquoi nous essayons, en tant qu'entreprise, d'éclairer suffisamment ces sujets et de réfléchir à la manière dont nous pouvons dès le départ concevoir des systèmes sûrs et fiables. Un exemple d'une telle solution est l'enregistrement et la saisie d'informations pertinentes.

L'IA générative dans le domaine de la gestion des connaissances

L'IA générative a de très larges champs d'application, mais celui qui est actuellement suivi le plus activement par de nombreuses entreprises sur le marché suisse est dans le contexte de la gestion des connaissances. Comment réussir à mettre à disposition de nos collaborateurs, et peut-être dans un deuxième temps de nos clients, les données disponibles de la manière la plus efficace et la plus pertinente possible ? Derrière tout cela, il y a la capacité d'analyser des documents, de les synthétiser et d'en faire de bons résumés. Dans le travail de la police, il existe des cas d'utilisation très similaires, notamment lorsqu'il s'agit d'enregistrer des incidents. Il s'agit d'un processus relativement laborieux que nous essayons actuellement d'automatiser dans le cadre d'un PoC, en faisant en sorte qu'une IA enregistre les paroles prononcées, établisse un procès-verbal et résume à nouveau les points importants avant de les transférer dans les systèmes existants.

L'IA générative pour les chatbots et les voicebots

Les chatbots et les voicebots constituent un autre cas d'application de l'IA. Dans leur version actuelle, ils sont souvent un véritable fléau pour les client.e.s. L'IA générative est désormais capable de comprendre le language parlé, non seulement l'allemand standard, mais aussi les différents dialectes suisses, de reconnaître l'intention des client.e.s et d'en déduire les actions correspondantes.

L'IA générative dans le contexte militaireGenerative

Nous voyons également de champs d'application de l'IA générative dans le contexte militaire. Il pourrait s'agir par exemple de la simulation de scénarios d'entraînement. Aujourd'hui, les articles d'actualité, les posts sur les réseaux sociaux, les nouvelles sur Twitter sont laborieusement analysés à la main afin de rendre les simulations et les formations aussi réalistes que possible. C'est précisément là que l'IA générative peut apporter sa contribution grâce à sa capacité créative.

Exposé de Stefan Preuss

L'IA générative à des fins de marketing

Le cas d'application de l'IA générative le plus simple est l'utilisation à des fins de marketing. La technologie fonctionne très bien. On peut faire tellement de choses avec. L'IA est partout. Actuellement, la prochaine vague de la génération text-to-video ou text-to-image est en cours. Il n'y a pas de règle du type « guardrail » qui dicte ce qui est autorisé et ce qui ne l'est pas. On a donc aussi la possibilité de générer des vidéos

problématiques. Il se peut que quelqu'un connaisse l'outil Flux de Black Forest, une entreprise allemande. C'est l'un des générateurs d'images les plus rapides actuellement sur le marché. Quand on utilise Flux pour générer un docteur, on obtient un docteur blanc de mon âge. Il y a donc un biais dans cet outil, car il ne peut pas du tout interpréter cette tâche. En revanche, si l'on compare cela à une recherche Google et que l'on pose la même question, le résultat était le même il y a cinq ans, mais il est aujourd'hui plus divers. Ces outils ont donc leurs limites.

Échec des applications de l'IA

« The best AI fails 2020-2024 » : nous avons commencé à rassembler des échecs majeurs de l'IA en 2020. Un bel exemple est un robot capable de jouer aux échecs et d'utiliser simultanément plusieurs échiquiers. Celui-ci n'était toutefois pas entraîné à ce que le petit garçon contre lequel il jouait s'immisce simplement dans l'aire de jeu. Le bras du robot aurait alors dû s'arrêter et sortir. Mais il ne s'est pas rendu compte de la situation, a agi et a cassé le doigt du garçon. Un cas un peu plus récent : McDonalds a lancé un projet relativement important mais a dû l'abandonner. McDonalds voulait utiliser l'IA générative dans un drive-through pour automatiser le processus de commande. Cela a conduit à la commande de 260 thés glacés et à la combination de différentes commandes prises parallèlement dans d'autres sites. Nous devons apprendre de ces exemples comment les erreurs se produisent et peuvent être évitées.

La confiance en l'IA, une composante essentielle

Pourquoi devrions-nous faire confiance à l'IA ? Après avoir vu le mur de la honte, on commence par douter de cette technologie. Nous utilisons la technologie des milliers de fois dans notre vie quotidienne et lui faisons donc confiance. Nous prenons le train, nous faisons confiance au conducteur et aux CFF pour que le train arrive à destination. Nous faisons confiance au pilote qui pilote notre avion pour qu'il atterrisse sans encombre. En fin de compte, l'IA est aussi une question de confiance. Un article très intéressant de Bruce Schneier, un spécialiste de la cybersécurité, a été publié à ce sujet. Celui-ci a défini quatre dimensions pour savoir quand nous pouvons faire confiance à la technologie. D'une part, lorsque la technologie en elle-même implique des mécanismes de sécurité et est arrivée à maturité. Ensuite, lorsqu'il existe des valeurs au sein de l'entreprise qui propose la technologie, valeurs qui ne résident pas uniquement dans la maximisation des bénéfices, mais aussi dans la mise à disposition d'un produit qui fonctionne ou encore la conscience de sa propre réputation, qui est liée à ce produit. Et enfin, une réglementation judicieuse qui tienne également compte d'un système de sanctions en cas de mauvais comportement et d'échec.

La question est de savoir si je fais confiance à la technologie et si je l'utiliserais personnellement. Schneier a dit qu'en ce qui concerne la Secure Technology et la réglementation, les caractéristiques dites de confiance sociale sont pertinentes et qu'en ce qui concerne la réputation et la valeur, ce sont les caractéristiques de confiance interpersonnelle qui sont essentielles. Et lorsque nous parlons d'IA générative, nous reconnaissons beaucoup d'éléments humains dans cette technologie. L'IA répond de la même manière que nous parlons. Et c'est

pourquoi nous mettons davantage l'accent sur la confiance interpersonnelle et négligeons un peu trop la confiance sociale.

La conduite autonome comme exemple

Je prends la conduite autonome comme exemple, mais ce n'est pas un sujet spécifique à l'IA génétique. Les constructeurs automobiles disent tous que la technologie est prête et qu'elle fonctionne. En tant qu'observateur critique, je dirais que si la conduite autonome est prête, le nombre de chauffeur.euse.s de poids lourds devrait également diminuer, ce qui n'est pas le cas selon les données disponibles aux Etats-Unis et en Europe. C'est donc en totale contradiction avec la technologie et la promesse technologique. Pourquoi ? La technologie existe. Nous avons dans les véhicules modernes des radars, des GPS, des 4G, des capteurs qui mesurent toutes les valeurs possibles de 2 cm à 250 m et qui calculent chaque scénario de conduite 10 secondes à l'avance. Les véhicules modernes d'aujourd'hui sont des centres de données sur roues. La technologie est donc prête. Est-ce une question de réglementation ? En Suisse, il existe désormais au moins un projet de réglementation de la conduite autonome. En Allemagne, la conduite autonome est possible au niveau 5, jusqu'à 60 km/h sur autoroute (!). Le problème n'est ni la technologie, ni la réglementation, ni les utilisateurs. Nous pouvons regardez l'exemple de San Francisco, où la plus grande expérience de conduite autonome est actuellement en cours. Environ 1600 véhicules sont en service. Si l'on regarde les rapports de conduite, il s'agit statistiquement du mode de conduite le plus sûr. Depuis l'accident d'Uber il y a six ans, il n'y a plus d'accidents mortels. La technologie est arrivée à maturité, mais des défis subsistent.

Voyons ce qui peut se passer. Première possibilité : le réseau téléphonique tombe en panne. Deuxième possibilité : la police arrête un véhicule autonome, le véhicule freine brièvement, la police sort, court, le véhicule accélère et repart pour s'arrêter à nouveau 200 mètres plus loin. Numéro trois, des gangs de jeunes s'amusent à couvrir les capteurs avec des cônes de signalisation et créent ainsi un embouteillage artificiel. Numéro quatre, un véhicule conduit par un être humain renverse une femme, celle-ci est projetée sous un véhicule autonome, le véhicule autonome n'est pas préparé à cette situation. Finalement, quelqu'un s'est amusé à mettre un T-shirt avec un panneau stop et s'est placé sur le bord de la route. Le véhicule autonome s'est arrêté parce qu'il avait détecté un panneau stop.

L'IA fonctionne à 95-99%, jamais à 100%.

Il y a des situations qu'on ne peut pas prévoir. Je l'ai appelé le « dilemme de la vache et du vol », car personne ne s'attend à ce qu'une vache vole dans les environs et que le véhicule doive l'éviter. L'IA fonctionnera donc à 95 ou 99 %, mais elle ne fonctionnera jamais à 100 %. C'est pourquoi nous devons être en mesure d'amortir ce 1 % de manière à ce qu'il n'y ait pas de dommages et surtout pas de personnes blessées.

En résumé, la réalité écrit ses propres lois, même dans le contexte de l'IA. Même si nous essayons de mettre en œuvre et de maîtriser au mieux toutes les dimensions. Nous avons le sentiment qu'aujourd'hui que nous devons tout faire avec l'IA, mais tous les problèmes ne peuvent pas être résolus par l'IA. Ils peuvent peut-être être résolu autrement. Il est

néanmoins important de se demander quand l'IA et quand la puissance humaine ou des solutions alternatives sont appropriées.

Referat Dr. Thomas Rothacher

Développement technologique au cours du siècle dernier

Au cours des 100 dernières années, on constate une augmentation exponentielle des trans ferts de technologie dans notre environnement. Autrefois, l'armée était encore un moteur de technologie. Aujourd'hui, c'est au contraire le monde civil qui est le moteur de technologie. Les conditions financières ont changé. De plus, le changement technologique a beaucoup plus d'influence sur notre culture que nous ne le réalisons. L'iPhone, qui est utilisé depuis 14 ans, a changé le monde de manière disruptive. Le savoir a pris une autre valeur grâce à Google et l'intelligence artificielle va nous le faire comprendre encore plus clairement. Un autre phénomène est la pénétration de cette technologie, qui se fait de plus en plus rapidement et à grande échelle. Il a fallu trois ans et demi à Netflix pour atteindre un million d'abonné.e.s/utilisateur.trice.s. ChatGPT, en revanche, n'a eu besoin que de cinq jours pour atteindre cette étape. Un autre graphique montre combien de temps il a fallu à un algorithme pour réaliser une certaine activité mieux que l'être humain moyen. Dans le cas de la reconnaissance de l'écriture manuscrite, cela a pris plus de dix ans. Si l'on se place en 2018, il n'aura fallu que deux ans pour la reconnaissance et la compréhension de la parole. Il y a donc eu une énorme accélération de l'évolution de ces capacités.

Influence de l'IA générative sur la cybersécurité

Si nous parlons d'armasuisse, nous avons un grand avantage sur les entreprises de l'économie privée. Si celles-ci veulent vendre un produit, toutes les éventualités doivent pouvoir être prises en compte. Vous avez entendu tout à l'heure qu'il pourrait être difficile d'atteindre le dernier pour cent jusqu'au contrôle total de l'IA. C'est un peu plus facile pour S&T (division Sciences et technologies d'armasuisse). Le S&T est préoccupé par la question de savoir comment ils peuvent apporter de nouveaux gadgets à la troupe ou à l'application, et ce encore au stade de l'essai ou du test.

Nous nous sommes demandé où nous devions utiliser cette technologie. Dans l'armée, il y a certaines compétitions où le bien affronte le mal. La Suisse a participé à un concours

international et s'est retrouvée dans le dernier tiers en 2017. Nous avons alors collecté des données sur plusieurs années et les avons utilisées pour programmer et entraîner des algorithmes. Lors du concours Block Shields 19, nous nous sommes retrouvés dans le premier tiers grâce à l'aide de ces algorithmes. Cette forme d'IA est déjà utilisée de manière opérationnelle dans les systèmes de l'armée depuis 2020.

Capacité d'apprentissage de l'IA

Il existe déjà aujourd'hui des jeux de guerre qui peuvent être joués à l'aide de l'IA. Nous pouvons par exemple faire jouer deux intelligences artificielles l'une contre l'autre. Il s'agit ici d'équiper des troupes avec différentes technologies et de les positionner l'une contre l'autre. Ce qui est surprenant, c'est que l'espace de solution proposé par une intelligence artificielle est différent de celui que nous choisirions. Bien que l'intelligence artificielle puisse avoir un biais, le biais de l'homme est en général encore plus prononcé. Nous avons appliqué cette technologie au nouveau simulateur tactique de l'armée de l'air. L'IA calcule, à l'aide d'un grand nombre de données, quel serait le chemin optimal pour l'adversaire afin d'attaquer nos positions. Nous pouvons en déduire ce que cela signifie pour nos systèmes - comment ils doivent être conçus et comment nous pouvons les disposer au mieux.

Drones contrôlés par l'IA

Les drones sont l'un des grands sujets qui ont été alimentés ou accélérés par l'IA. Un officier ukrainien qui dirige ces opérations de drones nous a dit qu'en Ukraine, ils utilisent actuellement un million de drones tous les trois mois environ. En outre, ils détruiraient 95% des véhicules protégés avec ces drones, car ils n'ont pas suffisamment de munitions d'artillerie. Les drones sont lancés et la règle est alors « Fire and Forget ». Ils ne sont donc plus sous contrôle humain, mais choisissent eux-mêmes leurs cibles grâce à l'IA et sont constamment perfectionnés. L'Ukraine est le premier pays à avoir introduit une unité sans pilote en tant qu'unité à part entière. C'est un sujet dont nous discutons déjà aujourd'hui dans certains secteurs de l'armée.

La Suisse est également un pays en tête dans le domaine de l'IA et des drones. Nos hautes écoles sont considérées comme faisant partie de la plus grande pépinière de talents du monde. Nous sommes également numéro un de l'indice mondial de l'innovation depuis quelques années. Armasuisse a œuvré à la création de la Task Force Drones au cours des deux derniers mois. L'objectif est de mettre en réseau les différents acteurs en Suisse et d'apporter une contribution à notre sécurité.

« Ce n'est pas l'espèce la plus intelligente qui survivra, mais celle qui saura le mieux s'adapter».

Ma conclusion : la vitesse des évolutions évoquées augmente de manière exponentielle, notamment dans le domaine de la sécurité. Et nous assistons aujourd'hui à des révolutions dans la conduite de la guerre, qui vont au-delà d'éventuelles disruptions. De plus, les nations qui nous entourent s'arment toutes énormément. Ce n'est pas mon avis en tant que directeur de S&T, mais en tant que chef adjoint de l'armement. Les Suisse.sse.s se trouvent dans une bulle

et ne perçoivent pas ce qui se les entourent. Nous devons pourtant considérer ce qui se passe autour de nous, car, ce n'est pas l'espèce la plus intelligente qui survivra, mais celle qui saura le mieux s'adapter.

Table ronde

Thomas Rothacher, d'autres experts ont participé au panel : Jennifer Scurrell, doctorante au Center for Security Studies de l'EPF de Zurich, Patrick Fontana, Digital & App Innovation Specialist chez Microsoft et Dr. Peter Friedli, partenaire chez Eraneos Switzerland. Le panel a été animé par Fredy Müller, directeur du FORUM SÉCURITÉ SUISSE.

L'IA comme changement fondamental

Fredy Müller : Stefan Preuss, la révolution technologique que nous vivons avec l'intelligence artificielle est-elle un développement aussi importante que la lecture, l'écriture ou le calcul ?

Stefan Preuss : A mon avis, oui, sans aucun doute. Si l'on considère l'évolution depuis la révolution industrielle, en commençant par la machine à vapeur, en passant par l'électricité, la technologie de l'information, et maintenant aussi le thème de l'intelligence artificielle, c'est définitivement un pas révolutionnaire, car il y a des données derrière. Et nous disposons désormais de ces données de manière exploitable, nous pouvons les utiliser. Si l'on ne considère pas seulement l'IA générative, mais aussi les procédés de reconnaissance d'images dans le domaine médical, cela nous donne un coup de fouet.

Freddy Müller : Patrick Fontana, vous travaillez depuis 20 ans dans le domaine de la technologie. Qu'est-ce qui a changé chez Microsoft pendant cette période ?

Patrick Fontana : Beaucoup de choses. D'une part, on est passé d'une simple entreprise de vente et de licence à une entreprise de développement et d'adaptation de technologies. L'intelligence artificielle est l'étape suivante. On ne parle plus simplement de sortir un produit, mais nous mettons la plupart des technologies à disposition sous forme de services. Une

seule entreprise n'a souvent plus les ressources pour tout faire elle-même. Cela apporte toujours de la nouveauté au sein de l'entreprise, car il faut bien réfléchir au cas d'utilisation des applications d'IA. Toutes les tentatives de mise en œuvre qui ont échoué n'avaient pas de cas d'utilisation clairement défini.

Freddy Müller : L'objectif n'est donc plus de mettre sur le marché un produit en soi, mais un service ou un système global ?

Patrick Fontana : D'une part, certainement un service, mais d'autre part aussi un écosystème.

Freddy Müller : Peter Friedli, tu travailles depuis de nombreuses années dans le domaine du conseil et de la sécurité. Comment as-tu vécu ce changement technologique ?

Dr Peter Friedli : Je détecte un grand besoin de conseil dans ce domaine. Comme l'a dit Patrick Fontana, il faudrait en fait réfléchir aux problèmes que nous avons effectivement et à la manière dont nous pouvons les résoudre. Reste à savoir si l'IA ou l'IA générative est la solution.

Freddy Müller : Jennifer, tu fais partie de la nouvelle génération et tu es en plein milieu de tes recherches. Comment vis-tu ce hype pour l'IA et l'IA générative ?

Jennifer Scurrell : Je ne qualifierais pas l'IA de hype, mais de révolution. Surtout si l'on considère mon expérience en matière d'interaction homme-IA. On voit bien l'impact de l'IA aujourd'hui. Que ce soit les chatbots, qui deviennent de plus en plus intelligents, mais aussi dans le domaine de la médecine ou de la psychothérapie, ces bots peuvent faire beaucoup de bien. Cela fait environ huit ans que je m'intéresse au Machine Learning, au Deep Learning et à l'IA, et je tiens donc à répéter que cette technologie est absolument révolutionnaire.

Comprendre le fonctionnement de l'IA - condition optionnelle pour garantir la confiance

Freddy Müller : Dans quelle mesure les gens connaissent-ils l'IA et l'IA générative ?

Jennifer Scurrell : Étonnamment, beaucoup de gens utilisent ChatGPT, y compris ma mère. Je ne dirais pas qu'elle sait comment cela fonctionne. Mais doit-elle vraiment le savoir ? Il est plus important d'acquérir des compétences générales dans l'utilisation de technologies telles qu'Internet, les réseaux sociaux et l'IA. La pensée critique est essentielle dans l'utilisation de toute technologie.

Freddy Müller : Stefan Preuss, du point de vue d'une entreprise, que doit-on savoir à propos de cette technologie ? Faut-il simplement l'utiliser et ne pas trop se poser de questions en tant qu'utilisateur ?

Stefan Preuss : Du point de vue de l'assurance, ce n'est peut-être pas la meilleure solution, car on joue alors beaucoup avec la confiance du client.e. Je recommanderais d'utiliser ces technologies avec prudence et en connaissance de cause, de faire des essais et des

comparaisons. On voit bien à quel point une technologie, par exemple dans le développement d'un champ de bataille, bouleverse complètement un comportement habituel et change les choses. Mais cela ne convient évidemment pas à tous les scénarios d'entreprise.

Les six principes pour une utilisation responsable de l'IA

Freddy Müller : Patrick Fontana, nous connaissons les six principes nécessaires pour utiliser l'IA de manière responsable. Chez Microsoft, l'utilisation de ces principes est certainement un thème permanent.

Patrick Fontana : Les six principes pour l'utilisation de l'IA responsable sont définitivement un sujet, mais ils reviennent finalement exactement à la thématique que nous avons entendue tout à l'heure. Car lorsque nous utilisons une technologie, nous devons avoir confiance dans le résultat. En ce qui concerne les principes, une IA doit être transparente, sûre et compréhensible afin d'éviter les biais. Si une entreprise peut s'y identifier, cela crée de la confiance. Reste à savoir si cela doit être inscrit dans la loi pour que tout le monde s'y conforme, ou s'il suffit d'une auto-déclaration qui n'entraîne pas de conséquences. C'est la question qui se cache derrière les six principes de Responsible AI.

Régulation de l'IA

Freddy Müller : Une auto-déclaration suffit-elle ou avons-nous besoin de dispositions légales pour l'IA afin d'éviter les abus ?

Dr Thomas Rothacher : Dans le monde militaire, la réglementation est un peu difficile. Je suis quelqu'un qui ne croit pas vraiment à cette réglementation. Car toutes les évolutions et les fonctions qui sont possibles se produiront tôt ou tard. Nous devons réfléchir davantage à la manière dont nous voulons les gérer et ne pas réglementer quelque chose qui va de toute façon arriver. L'approche de l'UE vis-à-vis d'une telle réglementation n'est définitivement pas la bonne.

Jennifer Scurrell : Je suis également d'accord d'un point de vue scientifique. Une réglementation judicieuse est acceptable, mais je trouve la manière dont l'UE aborde le sujet problématique. Le progrès et la créativité sont entravés, ce qui freine également l'Europe en tant que leader potentiel dans le domaine de l'IA. C'est pourquoi il faut une réglementation judicieuse, mais limitée.

Freddy Müller : C'est un exercice d'équilibre. Stefan Preuss, tu l'as montré sur une de tes slides, il faut une régulation.

Stefan Preuss : Je suis personnellement tiraillé entre deux sentiments. D'une part, je suis d'avis qu'il faut impérativement une réglementation pour une technologie disruptive comme l'IA. Avec internet et les médias sociaux, nous avons raté l'occasion de réglementer et aujourd'hui, on essaie péniblement de refermer la boîte de Pandore. En ce sens, j'ai été heureux que l'EU AI Act existe. Mais il faudra encore quelques années avant que l'on définisse les objets que l'on souhaite réglementer et que l'on comprenne de quoi l'on parle. Je suis donc très partagé à ce sujet. De mon point de vue, je suis bien sûr content d'avoir un objectif auquel je peux m'orienter. Mais il faut encore beaucoup de temps pour la mise en œuvre.

Freddy Müller : Patrick Fontana, faudrait-il une réglementation globale ou locale ? Il a été dit que l'on avait probablement raté des choses dans le domaine des médias sociaux. Microsoft est souvent pointé du doigt dans ce contexte.

Patrick Fontana : Pour nous, ce serait bien sûr beaucoup plus simple si tout était réglementé quelque part au niveau mondial. Nous pourrions alors harmoniser l'ensemble de nos processus avec cette réglementation et nous aurions donc moins de travail. Le fait est qu'actuellement, chaque État et chaque organisation tente d'introduire de telles réglementations. Nous devons à chaque fois nous adapter à ces réglementations et les respecter. Nous le faisons volontiers, car cela favorise la confiance. Nous sommes un peu partagés : Oui à la réglementation, mais seulement d'une manière qui n'entrave pas l'innovation et le développement.

Changements sociaux et politiques induits par l'IA (générative)

Freddy Müller : Ensuite, nous parlons des changements sociaux et politiques générés par l'IA ou l'IA générative. C'est surtout le domaine de recherche de Jennifer Scurrell.

Jennifer Scurrell : Je m'intéresse surtout à la manière dont les chatbots peuvent influencer l'opinion politique sur les réseaux sociaux. Le sujet a été largement médiatisé pour la première fois en 2016, lorsque l'on a soupçonné des bots russes d'avoir influencé les élections américaines. Dans ce cas, les scientifiques n'étaient pas tous d'accord. Certaines études ont montré que les bots avaient effectivement influencé les résultats en raison de leur grand nombre. Mais d'autres études prouvent que ces bots n'étaient pas assez intelligents pour pouvoir influencer efficacement l'opinion politique. Avec ChatGPT, nous sommes dans une situation totalement différente, car les bots deviennent de plus en plus performants et intelligents. Les gens ne font parfois plus la différence entre les bots et les humains. Ces derniers mois, de nombreuses études ont été publiées à ce sujet. Il a été démontré que ces bots peuvent influencer d'autres personnes au moins aussi bien que les humains avec de la propagande. C'est déjà là que nous avons un gros problème.

Gestion des bots

Freddy Müller : Patrick Fontana, comment une entreprise comme Microsoft gère-t-elle le sujet des bots ?

Patrick Fontana : Chez Microsoft, nous ne proposons pas un bot qui peut tout faire. Nous proposons plutôt des architectures de référence claires et soutenons directement les client.e.s. Lorsque nous utilisons OpenAI, nous avons un système de sécurité à deux niveaux. Nous avons une reconnaissance automatisée de la grille de questions qui ne doivent pas être posées, comme par exemple sur le sujet du suicide ou les questions à caractère manipulatoire. Pour des raisons de protection des données, cette composante humaine peut également être retirée. Dans ce cas, nous devons toutefois nous protéger en tant qu'entreprise, car nous avons la responsabilité de ne pas répondre à certaines questions. Un exemple à ce sujet est une preuve de concept que nous avons réalisée avec l'armée.

L'IA et la nouvelle génération

Freddy Müller : Peter Friedli, comment une entreprise de conseil fait-elle face à ce défi ?

Dr Peter Friedli : Nous sommes presque tous des natifs du numérique, ou du moins nous le sommes devenus. Ma fille de 9 ans arrive toutefois dans un monde où le numérique est la nouvelle normalité. Comment faire en sorte que cette génération sache s'en servir ? C'est en quelque sorte un principe de « train-the-trainer », comme on le connaît dans l'armée. En tant que formateurs, sommes-nous aussi prêts à transmettre cela? Les enseignants sont-ils prêts à transmettre cela ? Je pense que ce sont aussi des questions qui peuvent être transposées dans les entreprises.

Jennifer Scurrell : De nombreuses personnes, qu'il s'agisse d'élèves ou d'adultes**, sont concernées, mais ne sont pas préparées à cela**. Nous sommes quotidiennement influencés par les réseaux sociaux et bombardés d'informations. Lorsque ChatGPT est sorti, on a voulu l'interdire dans les écoles et les universités. Mais ce n'est pas la bonne approche. Nous devons apprendre aux enfants, dès leur plus jeune âge, à utiliser ces technologies. Dans une matière scolaire traitant du sujet de technologie, on pourrait en parler une fois par semaine. Il s'agirait d'expérimenter de manière ludique quelles questions génèrent quelles réponses, quelles données sont utilisées et comment celles-ci sont combinées ?

Freddy Müller : La société, en particulier les entreprises, devrait-elle être formée et sensibilisée à ce qui nous attend ?

Stefan Preuss : Il est presque impossible d'y parvenir si l'on se tient à la perspective des parents. Quand on voit ce qui est transmis aux enfants par les réseaux sociaux et que l'IA renforce encore beaucoup de choses, c'est un combat très difficile. D'un autre côté, l'évaluation de l'impact de la technologie est passionnante. Car c'est la première fois que nous sommes confrontés à une technologie qui nous oblige à réfléchir à nos positions éthiques fondamentales. Voulons-nous conserver la composante humaine dans la boucle pour les systèmes d'armement ? Ou que signifie l'intégration de systèmes de scoring social dans la société, non pas pour les appliquer en Suisse, mais pour les vendre ?

Freddy Müller : Thomas Rothacher, que pouvons-nous faire pour ne pas nous laisser berner par les deepfakes, que ce soit dans la société civile ou dans le monde militaire ?

Dr Thomas Rothacher : Je ne comprends pas vraiment la technologie derrière l'IA, mais ce que j'ai appris, c'est de réaliser certaines réflexions de plausibilité. Dans le cas des vidéos deepfake, il est utile de se demander dans quelle mesure ce qui est montré est réaliste. De mon point de vue, cette pensée critique et cette discussion critique devraient être encouragées de manière plus ciblée.

Les entreprises doivent s'adapter à l'IA.

Freddy Müller : Patrick Fontana, je suppose que Microsoft est très demandé dans les écoles en raison de son expertise, etc. Avez-vous l'occasion d'y intervenir ?

Patrick Fontana : Nous nous mettons activement à disposition. La demande est toutefois très faible, surtout au niveau de l'école primaire, ainsi que dans le domaine de l'enseignement de base ou des huitième et neuvième années. En revanche, les établissements d'enseignement supérieur s'occupent activement de ce sujet. Même lorsque nous nous rendons dans les entreprises, nous disons toujours qu'il faut un programme d'adaptation. La première étape consiste à apprendre à utiliser cette technologie et à valider correctement les informations fournies.

Freddy Müller : Devrions-nous être plus professionnels dans l'utilisation d'internet, des réseaux sociaux et de l'IA ?

Dr Peter Friedli : De mon point de vue, là où il n'y a pas encore de besoin, il devrait obligatoirement en avoir un. Pour les entreprises qui sont sorties de la phase d'idéation, nous rédigeons souvent des stratégies d'IA. Les stratégies d'IA contiennent déjà une partie technique. Mais il s'agit plutôt de l'opérationnel, des processus et des réflexions sur le flux de données. Un cas d'utilisation typique : je dois écrire une réponse à un e-mail. Pour cela, je copie le contenu, j'utiliser ChatGPT et je lui fais rédiger des réponses. Lors de ce processus, une grande quantité de données s'écoule. Aucun pare-feu ne peut y remédier. Pour finir, il y a une sorte de gestion de portefeuille et une priorisation de l'activité de l'IA. La technique n'est donc qu'une petite partie du processus.

Application réussie de l'IA dans les entreprises et les pouvoirs publics

Freddy Müller : Passons au troisième bloc thématique, l'application de l'IA dans les entreprises et les autorités. Stefan Preuss, tu as expliqué clairement ce qui peut mal tourner lors de l'utilisation de l'IA. Quels sont les trois éléments les plus importants pour que l'utilisation de l'IA soit une réussite ?

Stefan Preuss : Comme nous l'avons déjà mentionné, l'IA se base sur les données de l'entreprise. C'est la seule façon de générer une valeur ajoutée dans le contexte spécifique de l'entreprise. Les informations doivent être disponibles dans une qualité et une quantité appropriées. C'est probablement le plus grand défi pour la plupart des entreprises. En outre, dans un scénario stratégique, il faudrait autoriser le plus grand nombre possible de petits scénarios de jeu avant de faire le grand saut. Car nous aussi, à La Mobilière, nous ne pouvons pas

nous permettre d'être précis à 99%, nous avons besoin de 100%. Avant de franchir ce pas, il faut donc encore beaucoup de données empiriques.

Freddy Müller : Nous aimerions demander au public : utilisez-vous l'IA et avez-vous déjà des expériences ? Qui utilise un chatbot ?

Georg Kaufmann [question du public]: Je suis chef d'état-major dans le domaine du personnel de l'armée. Nous sommes en train de mettre en place un chatbot. Actuellement, nous avons une hotline qui répond à 30.000-40.000 demandes par an. 60 % par téléphone et 40 % par e-mail. Le chef de l'armée souhaite maintenant intégrer un nouveau projet, le support de premier niveau. Cela nécessiterait 1,6 million de personnel supplémentaire, ce qui n'est pas envisageable. C'est pourquoi nous avons opté pour le chatbot et maintenant pour le MVP. Actuellement, nous effectuons des tests dans les troupes. Ensuite, je présenterai le projet au commandement de l'armée. Celui-ci décidera alors si le chatbot peut être utilisé, car cela nécessite également des ressources financières et humaines.

Patrick Fontana : Le commandement de l'instruction a un projet en cours depuis longtemps. Nous entrons souvent en contact avec l'IA par le canal du support. Dans ce domaine, ChatGPT a introduit la composante de la compréhension texte-parole, ce qui nous permet de communiquer activement avec nos données. Il manque actuellement à l'IA générative une composante relativement importante, celle d'avoir une vue d'ensemble du cas et de proposer différentes variantes. Nous sommes pour l'instant limités dans ce domaine. Ce serait la prochaine étape de développement. Aujourd'hui, on stabilise toute la phase de l'IA. A-t-on vraiment envie de passer à l'étape suivante ? Il faut d'abord clarifier les questions éthiques à ce sujet, avant que l'IA n'agisse de manière quasi autonome, qu'elle ait une vue d'ensemble de la situation et qu'elle trouve des variantes.

Freddy Müller : Nous parlons d'AGI, Artificial General Intelligence. Quelle est ton estimation, quand va-t-on commencer à utiliser ?

Patrick Fontana : Comme toujours, Microsoft se projette au maximum deux ans dans l'avenir avec ses feuilles de route et regarde ensuite plus loin. Bill Gates a dit dans son dernier podcast, The Next Big Thing, qu'il prévoyait ce développent dans10 à 15 ans. Je me rangerais à cet avis.

L'IA dans le contexte militaire

Freddy Müller : Nous arrivons au quatrième bloc thématique, celui du contexte militaire. Thomas Rothacher, tu as dit que la Suisse peut jouer dans la cour des grands. Que devons-nous faire pour maintenir notre position derrière les Etats-Unis et la Chine ?

Thomas Rothacher : Nous occupons la troisième place dans le domaine de la recherche et de l'innovation, mais la mise en œuvre est un peu plus difficile. Comment pouvons-nous parvenir à fabriquer des produits importants en Suisse dans des conditions difficiles, c'est-à-dire lorsque les frontières ne sont plus aussi ouvertes, comme en cas de conflit par exemple ? C'est, à mon avis, le plus grand défi.

Dr Peter Friedli : Il est amusant de constater que deux études d'armasuisse ont été publiées aujourd'hui sur la manière de promouvoir la place technologique suisse. L'un des objectifs principaux était de créer de la transparence. Je pense qu'armasuisse ou Swiss Innovation Forces s'efforcent de rapprocher l'écosystème des start-ups en Suisse de l'industrie de l'armement et d'y briser les résistances. L'utilité pour l'armée et la défense doit être mise en évidence.

Patrick Fontana : En tant qu'officier supérieur chez Microsoft Suisse, je peux également prendre position à ce sujet. L'écosystème de la défense fonctionne. La question est toujours de savoir à quel niveau régional une action doit être menée. Est-ce que je fais par exemple confiance à une entreprise américaine lorsque je suis dans un conflit ou est-ce que tout doit se faire à l'intérieur des frontières nationales ? Nous échangeons activement avec nos collègues américains de la DARPA (Defense Advanced Research Projects Agency), mais nous mettons également en œuvre des cas d'utilisation correspondants ici en Suisse et soutenons ainsi à notre tour l'écosystème local.

Freddy Müller : Comment pouvons-nous utiliser l'IA et l'IA générative de manière optimale dans le domaine de la police, de la justice et des services de renseignement ?

Dr Thomas Rothacher : Il existe déjà quelques exemples à ce sujet. Nous utilisons par exemple l'IA dans le domaine des exercices d'état-major. Comment puis-je alimenter le système avec mes propres données ? Nous faisons désormais appel à l'IA pour le développement de scénarios, ainsi que pour l'analyse des données lors de la fusion des images de la situation. L'analyse de textes et d'images fait également appel à l'IA. Les premières analyses sont déjà souvent effectuées par l'IA et l'IA aide ainsi les responsables à prendre des décisions. L'homme est donc toujours dans la boucle.

Le questionnement critique remplace le savoir

Nathalie Gratzer [question du public] : L'intelligence artificielle ne risque-t-elle pas de nous rendre stupide ? Comme cela a été mentionné plus tôt, l'IA élargit mes compétences dans une certaine mesure et me fournit rapidement des solutions sans que je doive réfléchir moi-même, raisonner logiquement et faire travailler mon cerveau. Si ChatGPT avait déjà existé à l'époque où j'étais à l'école, je n'aurais probablement fait que très peu de tâches moi-même et je n'aurais appris que la moitié.

Thomas Rothacher : Le rapport a complètement changé et nous devons aujourd'hui gérer certaines choses différemment. J'ai beaucoup appris par cœur à l'école. Pour moi, cela avait

une valeur de performance. Aujourd'hui, il n'est pas difficile d'accéder au savoir. La difficulté est d'évaluer ce savoir et de le rendre plausible ? Je n'ai aucune idée de comment apprendre à mes enfants à distinguer les vérités des contre-vérités. Traiter ce sujet est un défi.

Stefan Preuss : Je comprends ce que tu veux dire, la capacité à remettre les choses en question de manière critique gagne en importance. Si nous considérons un système d'assistance médicale, un médecin expérimenté est en mesure de juger si les informations communiquées par l'IA sont pertinentes et correctes. Le questionnement critique est une compétence importante pour nous, les humains. La question est de savoir si nous pouvons conserver cette compétence.

Dans mon cas, cela se présente ainsi au quotidien. J'ai trois chatbots différents ouverts devant moi et je leur pose la même question. Je compare ensuite leurs réponses et continue à travailler avec celle qui me semble la plus plausible. J'essaie de porter un regard critique sur les réponses et je demande toujours les sources et les liens pour pouvoir vérifier les informations. C'est une question de méthode, de compétence méthodologique.

Freddy Müller : Jennifer Scurrell, dans quelle mesure t'appuies-tu sur l'IA et l'IA générative ?

Jennifer Scurrell : Je ne les utilise pas dans mes recherches et je n'utilise pas non plus une réponse de ChatGPT pour répondre à ma question de recherche. Pour ne pas devenir stupide, il est important d'apprendre à utiliser de telles informations et de remettre en question le savoir recraché. Nous devons également apprendre à poser les bonnes questions. Si nous y parvenons, le travail avec l'IA a un grand potentiel.

Patrick Fontana : Nous savons désormais que nous avons besoin de cours pour évaluer activement les informations correspondantes. Un chatbot et un assistant IA sont utiles pour faire des affirmations et acquérir certaines connaissances. Mais il faut aussi être conscient des conséquences qui en découlent. C'est une compétence que nous devrons développer à l'avenir.

Mot de la fin: Lisa Kontratieva

En résumé, nous avons vu que l'IA, et en particulier l'IA générative, offre un très grand potentiel. Il y a une infinité d'applications possibles et pas seulement des chatbots, mais aussi des LLM multimodaux, par exemple pour la génération d'images ou de vidéos. Nous pouvons automatiser certains processus. Les possibilités d'application sont très diverses. Et pourtant, le risque existe que

quelque chose ne fonctionne pas ou que nous soyons trompés par des contenus très réalistes. Il y a donc un risque sécuritaire, un risque éthique, un risque politique et des défis technologiques. La Suisse a maintenant la possibilité de jouer un rôle de premier plan dans le domaine de l'IA. Nous pouvons y parvenir si nous gérons de manière active ces risques et ces défis.