Table ronde NARALO – Abus DNS et IA : combattre et favoriser les menaces

ICANN82 – 10 mars 2025

Greg Shatan - NARALO / Moses Singer LLP: Merci à tous, et merci d'être venus à notre table ronde sur l'abus de DNS et l'intelligence artificielle. Un sujet très actuel, une combinaison de deux sujets très actuels, bien sûr. Et sans plus tarder, puisque nous avons un panel de sommités, permettez-moi de les présenter d'abord directement à ma droite.

Jeff Bedser est un entrepreneur chevronné et un leader dans l'investigation en cybersécurité, le renseignement sur les menaces et la mitigation des abus DNS. Il est le PDG de CleanDNS, une entreprise dont la mission est de nettoyer Internet pour de bon, pour le meilleur ou pour le pire, en détectant et en stoppant les activités malveillantes liées au DNS. Dans une vie antérieure, il était détective privé, et il utilise cette expertise pour protéger les utilisateurs finaux d'Internet, les personnes qui nous tiennent à cœur ici à At-Large. Il siège à divers conseils et comités sur la gouvernance, la stabilité et la juridiction d'Internet, et possède 25 ans d'expérience dans le domaine. Il a fondé iThreat, une entreprise de renseignement sur les menaces, et il siège au SSAC de l'ICANN depuis 2007, et a auparavant été membre du conseil d'administration de PIR. Alors, Jeff, merci beaucoup d'être avec nous.

Tim Maurer est à sa droite. Tim est actuellement le directeur de la politique mondiale de cybersécurité chez Microsoft, où il collabore étroitement avec les équipes internes de Microsoft et les parties prenantes externes. Avant Microsoft, Tim travaillait pour le gouvernement des États-Unis, d'abord en tant que conseiller principal pour la cybersécurité et les technologies émergentes au sein de la Sécurité intérieure, puis au Conseil de sécurité nationale de la Maison-Blanche. Je crois que c'était sous le président Biden, et il a passé une décennie à travailler sur la cybersécurité et la politique technologique, et est l'auteur de Cyber Mercenaries, the State, Hackers, and Power.

À sa droite, nous avons Ana Neves. Beaucoup d'entre vous connaissent probablement Ana en tant que représentante du GAC du Portugal. Ana porte plusieurs casquettes. Elle est membre de l'actuel MAG de l'IGF, un conseil de poids lourds qui maintient l'IGF en fonctionnement. Elle est la présidente du conseil consultatif de DNS.PT et dirige le Bureau de la Gouvernance de l'Internet à la FCT, FCCN et UMIC au Portugal. Elle est également vice-présidente de la Commission des Nations Unies pour la science, la technologie et le développement. Elle a plus de 30 ans d'expérience dans le développement de la science et de la technologie numériques et les politiques publiques, et a été reconnue en 2018 comme CIO du secteur public portugais de l'année et leader numérique européen.

Donc, nous avons un panel qui me fait me sentir obscur et peu accompli. Alors, sans plus tarder, j'aimerais lancer la discussion.

Donc, si nous pouvons aller ici, nous sommes à la première diapositive. Avant d'aller plus loin, j'aimerais mentionner que toute cette présentation a été rédigée par un copilote Microsoft.

Je ne plaisante pas.

Et pour une raison quelconque, c'est là. Donc, au lieu que Dieu soit mon copilote, Microsoft est mon Copilot. J'espère que tu apprécies, Tim.

Tim Maurer - Microsoft: Je suis prêt. Je suis prêt. Le backup humain est là.

Greg Shatan - NARALO / Moses Singer LLP: Voilà. Selon Copilot, l'abus de DNS fait référence à l'utilisation abusive du système de noms de domaine à des fins malveillantes, telles que le phishing, la distribution de logiciels malveillants et d'autres menaces cybernétiques. La définition officielle de l'ICANN de l'abus de DNS comporte cinq catégories, peut-être une pour chaque région géographique de l'ICANN. Les logiciels malveillants, les botnets, l'utilisation de botnets, le phishing et le farming, et le spam, mais seulement lorsque le spam sert de mécanisme de livraison pour l'une des quatre autres catégories. Donc, le spam en soi n'est pas considéré comme un abus de DNS.

Ceux d'entre vous qui sont chez ICANN depuis un certain temps peuvent se rappeler que nous avons eu plusieurs années de discussions, c'est un bon terme pour le dire, sur les limites et la définition de l'abus de DNS. Et plutôt que d'en discuter davantage, nous vivons au moins avec cette définition de l'abus de DNS à mesure que nous avançons, car cela nous pose déjà suffisamment de problèmes.

Comme

vous savez, et je sais, et Microsoft Copilot sait, que l'IA et l'apprentissage automatique sont de plus en plus utilisés pour lutter contre les abus de DNS en détectant et en atténuant ces effets de manière plus efficace.

Je vous laisse lire tout ça à votre rythme. Nous n'avons pas besoin de l'entendre. Je préfère entendre ce que nos panélistes ont à dire plutôt que ce que Microsoft Copilot a à dire.

Et le design, bien sûr, a été choisi pour moi par Microsoft PowerPoint. Et l'image sur la couverture a également été choisie par Copilot. Je suis en fait une personne très créative, mais maintenant j'ai décidé de laisser Copilot faire tout cela.

Donc, en tout cas, j'aimerais d'abord me tourner vers Tim. Je vais juste dire encore une minute, bien sûr, l'abus de DNS se manifeste également par l'intelligence artificielle. Et il y a plusieurs exemples dans ce PowerPoint des façons dont l'IA crée de nouvelles formes et mute les anciennes formes d'abus de DNS. Donc, nous avons en quelque sorte une bataille de l'intelligence artificielle. C'est pour le bien et pour le mal. Et nos trois panélistes discuteront à la fois de la cybersécurité en général

et l'interface IA, comment cela affecte le DNS, et tout ce qui leur passe par la tête.

Donc, pour donner un aperçu général, Tim, qui est le seul non-initié de l'ICANN sur notre panel, va commencer.

Tim Maurer - Microsoft: Merci beaucoup, Greg. C'est un plaisir d'être avec vous tous ici. Je parlais justement avec Ana, je pense que la dernière fois que nous avons eu l'occasion de nous rencontrer était avant la pandémie, ce qui est, je crois, aussi la dernière fois que j'ai eu la chance d'assister à une réunion de l'ICANN. Et c'est un plaisir particulier d'être ici avec vous tous aujourd'hui parce que c'est une communauté qui nous enthousiasme beaucoup chez Microsoft, dans l'espoir de nous aider à utiliser l'IA et à maximiser son potentiel pour la défense d'Internet et s'assurer qu'elle est utilisée pour réellement atteindre ce que nous espérons que la technologie débloquera, à savoir un progrès humain supplémentaire, et aussi s'assurer que nous l'utilisons de manière à nous aider à avancer les objectifs et les buts existants que nous partageons.

Pour rebondir sur ce que Greg a dit, j'ai rejoint Microsoft il y a un peu plus d'un an, donc c'est une expérience fascinante pour moi d'être maintenant dans l'entreprise et à l'avant-garde de l'innovation en matière d'IA. Et la manière dont nous envisageons l'IA de manière plus générale, mais aussi spécifiquement en ce qui concerne l'abus de DNS, c'est que nous sommes conscients du fait que l'IA existe depuis un certain temps, selon la façon dont on la définit, et que l'apprentissage automatique et l'IA prédictive, ainsi que l'IA analytique, ont déjà été utilisés de diverses manières pour atténuer les abus de DNS.

Donc, comme Greg l'a mentionné avec Copilot, l'IA générative va, espérons-le, nous aider à débloquer davantage de potentiel de plusieurs manières et rendre également la façon dont nous pouvons aider à atténuer les abus DNS plus accessible à un plus grand nombre de défenseurs en éliminant les barrières qui pouvaient exister auparavant concernant le niveau de compétences et en utilisant l'IA générative, en particulier Copilot, et d'autres outils que nous avons pour faciliter l'utilisation des outils par le défenseur pour atténuer les abus.

Une chose que je vais dire d'emblée, et ensuite j'espère rendre cela aussi interactif que possible avec les panélistes et vous tous dans la salle, c'est que l'entreprise examine l'IA dans le contexte de la sécurité et dans le contexte de l'abus de DNS à travers trois perspectives.

La première question est : comment maximiser l'utilisation de l'IA pour la défense ? Et cela signifie, comment mobiliser les clients et la communauté comme la vôtre et la communauté ICANN pour adopter rapidement la technologie ? Et la raison pour laquelle je souligne la rapidité est que je peux vous garantir que ceux qui ont des intentions malveillantes seront parmi les plus rapides à utiliser les nouveaux outils pour découvrir comment ils peuvent les utiliser pour faire avancer leurs activités malveillantes existantes. Donc, plus nous pourrons avancer rapidement dans l'IA et l'utiliser pour des techniques de défense et d'atténuation, plus nous pourrons nous assurer de non seulement rattraper, mais aussi dépasser les acteurs malveillants qui frappent les organisations du monde entier chaque jour.

Le deuxième point que je voudrais souligner est que l'entreprise se concentre très tôt sur la détermination des mesures à prendre pour que les acteurs malveillants ne puissent pas tirer parti de nos produits et services d'IA. Ainsi, il y a un an, l'entreprise a publié une politique qui stipulait clairement que si nous détectons qu'un acteur malveillant suivi par l'entreprise, dont vous pouvez imaginer qu'il y en a beaucoup au fil des ans qui ont par le passé effectué un accès non autorisé ou autrement attiré notre attention, et que nous remarquons qu'ils utilisent des produits ou services d'IA, nous prendrons des mesures pour résilier leur compte. Peu importe s'ils utilisent l'IA de manière productive. Nous ne voulons pas que les criminels deviennent plus productifs. Nous prendrons donc des mesures contre cela. Et c'est une politique qui est venue de la direction, non en réponse à une réglementation ou à des demandes gouvernementales, mais qui reflète l'engagement de s'assurer que nous donnons du pouvoir aux

défenseurs et essayons de rendre la tâche aussi difficile que possible pour les acteurs malveillants impliqués.

Et puis, le dernier point, en termes de cadrage global, est que l'entreprise se concentre sur ce que nous devons faire de notre côté pour nous assurer que la technologie n'est pas utilisée de manière inappropriée, en pensant à Azure et à nos services cloud, et comment nous pouvons également protéger nos produits et services d'IA, qui, comme vous pouvez l'imaginer, sont ciblés par d'autres acteurs malveillants qui aimeraient connaître les détails de notre utilisation. Donc, permettez-moi d'utiliser cela comme un cadrage général en trois dimensions. Je sais qu'il y a beaucoup à discuter lorsque nous entrerons dans les détails de l'abus de DNS.

Greg Shatan - NARALO / Moses Singer LLP: Merci beaucoup, Tim. Et j'ai remarqué, bien sûr, que vous avez mentionné la défense DNS ou la défense contre les abus DNS à plusieurs reprises. Nous avons la chance d'avoir avec nous Jeff Bedzer, dont la vie est entièrement définie par la défense et le nettoyage du DNS. Je ne vais pas vous appeler le nouveau shérif en ville parce que c'est probablement la marque déposée de quelqu'un d'autre à ce stade. Mais en tout cas, Jeff, je pense... Oh, voilà. Vous avez un badge. Nous n'avons pas besoin de badges puants. Donc, en tout cas, Jeff, j'aimerais que vous creusiez un peu plus. Et je sais que vous avez une présentation et probablement qu'elle n'a pas été écrite par un copilote Microsoft. Elle a peut-être même été écrite par vous. Donc, sans plus tarder, je vais laisser la présentation et vous parler d'eux-mêmes.

Jeff Bedser - SSAC / CleanDNS: Merci, Greg. Et non, celui-ci n'a pas été fait par l'IA, mais je ne peux pas nier que j'ai utilisé l'IA avec succès pour créer de nombreuses présentations. Donc

Je vais profiter du fait que j'ai parlé à ce groupe, il y a déjà une semaine, hier, avant-hier, dimanche, de l'abus de DNS en général. Et je pense que cette présentation que j'ai préparée porte sur la manière dont CleanDNS utilise les processus d'IA pour améliorer cette lutte, comme l'a dit Tim.

Et j'ai mentionné, je pense, dans notre présentation, ma présentation l'autre jour, que si vous y pensez maintenant, ils y ont déjà pensé, l'ont fait, et sont probablement passés à quelque chose de plus rapide parce qu'ils sont tellement en avance sur nous. Et si nous n'utilisons pas les technologies émergentes au même rythme et à la même vitesse pour la défense, nous allons être tellement en retard si rapidement que cela va donner le tournis à tout le monde.

Donc, ce dont je vais parler, c'est comment nous avons activement déployé des modèles d'API dans ce que nous faisons parce que, encore une fois, comme je l'ai dit l'autre jour, il y a un volume significatif d'abus dont nous ne sommes pas au courant parce qu'ils ne sont pas signalés. Ce que nous voyons dans tous ces rapports actifs qui sont faits provient d'environ cinq à sept sources sur lesquelles la plupart des gens

basent leurs rapports, mais ce n'est que la partie émergée de l'iceberg. Et littéralement, si vous ne le voyez pas, ce n'est pas signalé. Et la majorité des gens ne signalent pas le phishing ou le smishing ou tout autre chose qu'ils reçoivent. Donc, nous n'en sommes pas au courant. Nous ne pouvons pas agir en conséquence.

Donc, la meilleure solution est de le détecter rapidement. Et quand vous le détectez rapidement, vous pouvez le mitiger rapidement. Et c'est dans cet espace que CleanDNS se situe. Nous nous trouvons entre les entreprises qui détectent les abus et ensuite l'industrie ou notre industrie entre les registres, les bureaux d'enregistrement et les sociétés d'hébergement qui doivent les mitiger. Et notre objectif est constamment d'accélérer non seulement le signalement, mais aussi la détection. Ainsi, nous pouvons arriver à un point où le processus entre la détection et la mitigation se fait en quelques secondes.

Alors, je ne sais pas si quelqu'un a lu ce tome dont j'étais le président du groupe de travail de l'SSAC, SAC 115 sur l'interopérabilité et la gestion des abus DNS. Mais nous avons découvert dans cette étude il y a plusieurs années que le temps moyen entre le signalement et la mitigation était de 96 heures. Et je pense que nous pouvons tous deviner combien de personnes cliquent sur un lien de phishing en 96 heures, combien de victimes cela représente et combien il y en aurait moins si nous pouvions réduire ce délai à quelques secondes au lieu de 96 heures. Alors, laissez-moi passer à la présentation.

Donc, le SAC 115 a été le modèle à partir duquel j'ai décidé de construire CleanDNS parce que de nombreuses personnes dans l'industrie m'ont fait remarquer que, bien que ce soit un très bon article académique, à moins que quelqu'un ne construise les solutions évoquées dans l'article, cela n'allait pas exister. Et CleanDNS a été construit pour correspondre à cela. J'aimerais appeler cela un cycle vertueux et je donnerai crédit là où il est dû. Byron Holland de CIRA a utilisé ce terme avec moi lorsque nous en parlions. Nous essayons de non seulement résoudre un problème, mais aussi d'améliorer les rapporteurs et de rendre l'industrie meilleure en résolvant le problème au milieu de vous l'avez trouvé. Maintenant, quel type de preuve est nécessaire pour agir et ensuite raccourcir le cycle pour l'éliminer ? C'est un modèle d'agrégation pour essentiellement accélérer ce processus.

Alors, d'accord, laissez-moi vous expliquer ce que nous faisons avec l'IA et tout ça.

Donc, le premier est facile et c'est l'un de ces domaines où l'on peut rapidement voir que l'IA est meilleure que l'humain. Tout le monde ici a déjà été victime de phishing. Si quelqu'un ne l'a pas été, veuillez expliquer pourquoi vous êtes à une conférence technologique sans avoir de téléphone. Non ? D'accord.

Le phishing repose sur des appâts. Un appât est une entreprise qui vous fait croire que vous allez vers quelque chose qui n'est pas réel. Que ce soit une banque, une plateforme de commerce électronique, votre opérateur télécom, ou votre médecin

local, il y a un appât impliqué. Et cet appât est presque toujours une image qui vous fait croire que c'est cette banque, cette plateforme de commerce électronique, etc.

Eh bien, pour les humains qui analysent le phishing, ils ne connaissent pas toutes les banques de la planète. Ils ne connaissent pas toutes les entreprises de la planète. Ils n'ont pas cette connaissance. Ils le font depuis 30 ans. Ils ne sauront jamais tout. Les systèmes d'IA peuvent puiser dans des bases de données et des systèmes qui possèdent ces fichiers d'images pour des comparaisons rapides.

L'analyse d'image peut donc vous donner la capacité de valider le logo. L'analyse d'image peut examiner et voir s'il y a une tentative de récupérer des identifiants. Y a-t-il une tentative de récupérer des informations personnellement identifiables ? Que contient d'autre cette image ? Et quand je parle d'analyse d'image, il s'agit essentiellement d'une capture d'écran. Donc, la page de destination d'un phishing, d'un pharming, d'une fausse boutique ou d'un smishing. Je vais arrêter de passer en revue tous les acronymes. Mais vous êtes essentiellement conçu pour faire cette capture d'écran, puis vous pouvez l'analyser pour le contenu. Et ensuite, l'IA est très douée pour dire, oui, c'est structuré d'une manière qui est probablement un phishing.

L'analyse de sentiment repose sur l'examen du texte et d'autres composants pour dire, oh là là, ce texte est petit à l'écran. Je m'excuse. Nous pouvons partager la présentation après si tout le monde la veut.

Peut-être que je peux le lire sur mon écran pour vous. Attendez.

Analyse des rapports d'e-mails entrants liés au contenu et à l'intention du rapporteur. Ainsi, l'industrie, au cours des 20 dernières années, a essentiellement utilisé un mécanisme de rapport basé sur les e-mails. Et bien sûr, le problème avec cela est que ce sont des données non structurées. Donc, l'analyse de sentiment vous permet de prendre ces données d'e-mails non structurées et d'essayer de comprendre ce qui se passe.

Les humains sont par défaut des conteurs. Et quand quelque chose vous arrive, vous voulez raconter l'histoire de comment j'ai été victime d'une arnaque sentimentale et pourquoi je pensais que cette personne me trouvait si attirant(e) et qu'elle voulait que je lui envoie 50 $ pour qu'elle puisse venir me rendre visite. Et les histoires sont fascinantes à lire, mais prennent beaucoup de temps. Et ce que vous cherchez vraiment dans ce texte, c'est quelle est l'URL ou le domaine impliqué ? Et de quoi l'accusez-vous ? Ah, vous avez été victime de phishing et c'est l'URL. C'est tout ce qui m'importe. Le reste de l'histoire est sans importance pour un effort de mitigation. Donc l'analyse de sentiment nous aide à y parvenir.

Et puis une relation contextuelle, encore une fois, je mets mes lunettes, désolé, c'est l'intégration vectorielle des similarités entre les blocs de texte. Nous recherchons des motifs où, par exemple, le spam ou le phishing, est-ce la même campagne du même

cybercriminel parce qu'ils utilisent la même structure, le même format, etc. Ainsi, nous pouvons développer une compréhension de groupes plus larges. Nous ne les éliminons pas un par un. Pouvons-nous utiliser les relations contextuelles pour comprendre, oh, il y en a 10 000. Prenons-les tous en même temps parce qu'ils sont tous liés au même cybercrime basé sur le timestamp, les relations contextuelles entre les données.

Donc, je m'avance un peu, mais la détection de logos est une partie de l'analyse d'image. La détection des informations personnelles identifiables (PII) est encore une autre clé. Beaucoup de domaines qui vont être utilisés pour des abus ou qui ont été utilisés pour des abus sont redirigés vers une page de stationnement avant d'être utilisés pour des abus. Parfois, ils les utilisent pour des abus, puis les remettent sur une page de stationnement. Ainsi, quand les "gentils" les regardent, ils se disent : "Oh, c'est juste une page de stationnement. Il n'y a rien là." Ensuite, ils attendent que ce trafic disparaisse et ils la réactivent. Et maintenant, c'est de nouveau une page de phishing. Donc, la détection de si c'est une page de stationnement ou une page de phishing est un composant important de l'utilisation de l'analyse d'image, car vous n'avez pas besoin d'un humain pour le faire, ce qui raccourcit encore le délai et accélère le processus.

Donc, certaines pistes futures potentielles que nous envisageons consistent simplement à affiner tout ce processus. C'est excellent pour le phishing, mais quels autres préjudices pouvons-nous traiter ici ? Car encore une fois, l'abus de DNS tel que défini dans le cadre des parties contractantes et de leurs obligations est celui dont Greg a parlé. Mais tout le monde dans cette salle sait qu'il existe de nombreux autres préjudices en ligne qui doivent être atténués. Et la plupart des parties qui représentent ceux de la maison des parties contractantes et du CCNSO agissent sur de nombreux autres types de préjudices qu'ils ne sont pas tenus de traiter, mais ils le font parce que c'est la bonne chose à faire. Alors, comment affiner un dossier de preuves qui indique que ce type de préjudice, lorsqu'il est présenté de cette manière, est un préjudice afin qu'il puisse être atténué ?

La détection de code malveillant, n'est-ce pas ? C'est difficile parce que si vous devez traiter le code pour déterminer s'il est malveillant, vous mettez vos systèmes en danger d'être compromis. Donc, utiliser l'IA pour examiner les modèles et identifier à quelle famille de logiciels malveillants cela appartient, quelles parties sont impliquées, comment cela est ciblé, est un autre domaine où l'IA peut être utilisée pour lutter contre cela.

Et enfin, le composant d'IA générative de tout cela, nous l'avons utilisé pour structurer les communications sortantes vers les parties concernées. Si c'est un registre qui rapporte à un registraire, nous l'utilisons pour générer le texte expliquant ce que nous avons observé, ce qui s'est passé, mais nous l'utilisons pour créer des modèles, pas pour créer les communications réelles parce que quiconque a joué avec l'IA, comme Greg l'a souligné, et je lui ai raconté l'histoire, doit être très prudent parce que si l'IA générative sans restriction commence à fantasmer, et j'ai en fait demandé à l'IA d'écrire un article pour moi sur toutes les entités travaillant sur les abus de DNS, et elle a

inventé des entités formidables que j'ai recherchées, et j'ai vérifié les notes de bas de page, et aucune de ces entités n'existait. Les sites web qu'elle a référencés n'existaient pas.

Donc, il faut être très prudent avec l'IA générative pour contrôler l'environnement à partir duquel vous lui imposez des restrictions.

Donc, les écueils. La transparence. Personne ne travaille avec moi sans comprendre que nous utilisons ces technologies, et nous ne voulons pas prétendre le contraire. Il faut surveiller les biais de données, et c'est aussi une préoccupation légitime en ce qui concerne les ensembles de données que vous appliquez. Si je commence à intégrer des ensembles de données de journalistes contenant beaucoup de faux positifs, mais que je ne sais pas qu'ils sont faux, je peux corrompre tout mon ensemble de données de validation.

Par exemple, s'il y a tout un tas de fichiers image qui ne sont en fait pas du commerce électronique ou des banques, et qu'ils se retrouvent dans cet ensemble de données, et qu'ils sont maintenant confirmés comme tels, ce biais peut être problématique. Les faux positifs et négatifs nécessitent toujours, à quel moment les soumettez-vous à une révision humaine ? Apparemment, c'est Tim. Donc j'aurai besoin de votre adresse e-mail pour cela.

Mais bien sûr, la gestion des cas sensibles concernant les données et les contrôles de ces données, en particulier pour des raisons de confidentialité, etc. Donc je pense que c'est

la dernière diapositive. Je te rends la parole, Greg.

Greg Shatan - NARALO / Moses Singer LLP: Merci beaucoup, Jeff. Très intéressant et opportun.

Ensuite, nous allons passer à Ana Neves. Ana, vous pouvez dire ce que vous voulez.

Ana Cristina Amoroso da Neves - GAC / Portugal: D'accord, merci beaucoup. Donc, ce que je veux, eh bien, c'est très agréable d'être ici à Seattle, et en particulier ici à At-Large. C'est toujours un grand honneur d'être ici à At-Large, car depuis le début de ma vie au GAC, j'ai toujours défendu que le GAC et At-Large devraient vraiment travailler ensemble, car nous devrions renforcer nos efforts dans le secteur public et du côté des utilisateurs finaux.

Que puis-je dire après que Microsoft et Jeffrey ont parlé et expliqué l'abus de DNS ? Ce que je peux dire à propos de Microsoft, c'est qu'il est très bon de savoir que vous avez une politique pour atténuer l'abus de DNS. Mais qu'en est-il des autres entreprises qui n'ont pas cette politique ? Et cette politique

est envisagée par qui ? Et pouvons-nous compter sur les entreprises pour avoir une telle politique ? Pouvons-nous compter ? Pourquoi ?

Donc, du point de vue du secteur public, c'est ce que je pense que j'ai été invité à faire partie de cette table ronde.

pour renforcer la collaboration entre les secteurs public et privé, établir davantage de partenariats et voir avec les organismes de réglementation ce qui peut être fait, tout en gardant toujours à l'esprit la cybersécurité. Les gouvernements, lorsqu'ils régulent ou réfléchissent à des politiques publiques, doivent toujours avoir en tête l'abus du DNS. Ainsi, les efforts multipartites sont vraiment importants pour garantir une utilisation responsable de l'IA dans la sécurité du DNS.

Un autre point important que j'aimerais souligner ici est le développement de cadres de gouvernance de l'IA. Je pense donc qu'il est très bon de discuter de cela ici à l'ICANN et avec AtLarge, qui pourrait explorer des politiques garantissant que la sécurité pilotée par l'IA soit en accord avec les principes de gouvernance mondiale de l'internet. C'est une très bonne année pour le faire, car nous réfléchissons à la révision de WSIS plus 20, comment rationaliser ce processus avec le pacte numérique mondial et ce que nous devrions vraiment envisager dans tous ces processus et négociations.

Et la dernière chose que je voudrais souligner dans ce premier commentaire est de donner du pouvoir aux utilisateurs finaux. La protection basée sur l'IA doit être conçue avec transparence et permettre aux individus de faire des choix éclairés concernant les paramètres de sécurité et de confidentialité. Il s'agit donc toujours d'avoir du bon sens mais aussi d'informer correctement les utilisateurs finaux. Ainsi, le renforcement des capacités est toujours important, la confidentialité est toujours importante à souligner dans tous nos efforts.

Je dirais donc que renforcer la collaboration, avoir plus de partenariats public-privé, développer des cadres de gouvernance de l'IA et autonomiser les utilisateurs finaux sont mes trois premiers points pour cette discussion et pour cette table ronde.

Merci beaucoup.

Greg Shatan - NARALO / Moses Singer LLP: Merci Ana. Pouvons-nous revenir à la diapositive où je me suis arrêté, c'est-à-dire la première vraie diapositive ici. Donc

passons à la diapositive suivante s'il vous plaît, juste pour couvrir certaines des choses que nos bons amis de Copilot ont trouvées, juste pour mettre quelques faits supplémentaires sur la table.

Nous voyons donc des attaques de phishing alimentées par l'IA, vous savez, qui utilisent l'IA pour personnaliser le phishing, le rendant presque un peu plus semblable à du

spearphishing, imitant des contacts de confiance, contournant les filtres de sécurité, améliorant la syntaxe et la grammaire, à moins que les erreurs ne soient là pour s'assurer que les personnes intelligentes ne les lisent pas.

Ensuite, l'IA peut également être utilisée pour pré-juger les identifiants, tester des milliers d'identifiants volés en quelques secondes, permettant des attaques par force brute et, en gros, la nature répétitive de ces attaques. C'est une affaire de volume, et donc l'IA peut renforcer la puissance du volume dans les menaces des acteurs malveillants.

Le prochain malware peut également être amélioré par l'IA avec un code qui s'adapte dynamiquement pour échapper à la détection et, vous savez, l'aider à se cacher des logiciels antivirus qui, bien sûr, recherchent des motifs qu'ils connaissent et apprennent. Donc, bien sûr, en utilisant l'IA, c'est encore une sorte de bataille des bots.

Nous pouvons également observer des opérations de reconnaissance utilisant des outils de scan de réseau avec l'IA, à grande vitesse et à grande échelle, pour identifier les points faibles de l'infrastructure de sécurité, vous savez, des attaques de pénétration automatisées. De plus, l'ingénierie sociale peut devenir plus dynamique, plus réaliste mais toujours fausse, pour contourner les filtres traditionnels, encore une fois en utilisant l'IA pour ajuster les éléments afin qu'ils paraissent moins comme des menaces évidentes.

Cela montre clairement pourquoi nous devons utiliser des fonctionnalités de défense sophistiquées et en constante amélioration pour suivre tout cela.

Donc, certaines de ces solutions sont bien sûr des outils de sécurité DNS alimentés par l'IA utilisant une architecture de confiance zéro avec une vérification plus stricte. Je pense que nous avons tous remarqué une utilisation accrue de l'authentification multi-facteurs, plus d'utilisation de différents types d'authentification et de couches, même au cours des six derniers mois à un an, juste pour essayer de suivre les niveaux de menace.

Donc, vous savez, les plateformes de renseignement sur les menaces doivent continuer à s'améliorer et vous savez, Tim a déjà ou plutôt Jeff a déjà fait allusion à cela. Bien sûr, il y a toujours le facteur humain. Vous savez, il faut encore former de vraies personnes, les employés et les utilisateurs finaux, le grand public, à reconnaître le phishing et les attaques d'ingénierie sociale, vous savez, la sensibilisation doit être maintenue en permanence. Je viens de faire la formation annuelle de cybersécurité de mon cabinet et elle est probablement déjà obsolète.

Vous savez, en tant qu'avocat, je peux apprécier la nécessité de continuer à améliorer les cadres juridiques. La loi ne progresse jamais aussi rapidement que l'IA, bien sûr, mais il est important d'essayer de maintenir des règles juridiques et éthiques qui régissent l'IA de manière plus responsable. Une des choses qui m'intrigue est de savoir

si les entreprises qui travaillent sur l'IA sont suffisamment conscientes de l'univers des abus de DNS, par opposition à tous les autres univers qui utilisent l'IA. Vous savez, notre domaine, bien qu'important, n'est pas nécessairement une priorité. Donc, est-ce que les entreprises d'IA font suffisamment d'efforts pour collaborer avec le reste de l'écosystème ?

Enfin, bien sûr, continuer à adopter des mesures de sécurité plus proactives, une paranoïa accrue est également cruciale. Diapositive suivante, s'il vous plaît.

Voici la biographie de Jeff, du moins celle qui m'a été fournie par Microsoft Copilot. Vous pouvez la lire à votre guise. Elle semble assez précise, ils n'ont pas trop inventé de choses.

J'ai d'abord demandé celle de Tim. Il y avait un point que vous devriez souligner. Oh, non. Qu'est-ce que c'était ?

Jeff Bedser - SSAC / CleanDNS: Revenez à la diapositive précédente. Revenez à la diapositive précédente.

Greg Shatan - NARALO / Moses Singer LLP: Regardez l'année en bas du deuxième paragraphe.

Jeff Bedser - SSAC / CleanDNS: Oh, l'année 2071.

Greg Shatan - NARALO / Moses Singer LLP: Eh bien, en fait, c'est parce qu'il y avait une note de bas de page que je n'ai pas supprimée. Donc, c'est une erreur humaine.

C'était la première tentative d'utiliser l'IA pour obtenir la biographie de Tim. Elle ne mentionnait pas du tout Microsoft. Encore une fois, espérons que ce soit plus ou moins précis, mais encore une fois, pas à jour.

J'ai ajouté Microsoft dans l'invite.

J'ai ajouté Microsoft à l'invite, et voilà, nous découvrons qu'il est le Directeur Principal de la Politique Mondiale de Cybersécurité chez Microsoft. Mais je ne l'aurais pas su si je n'avais pas mis cela dans l'invite. Donc, Microsoft Copilot ne connaît pas encore bien Tim. Mais je le vois prendre des notes, donc cela sera probablement corrigé.

Je voulais en savoir plus sur Tim et ses passe-temps. Donc la randonnée, le cyclisme, la photographie, la lecture. Il a des intérêts variés et s'engage dans le développement personnel et le bien-être. Du moins, c'est ce que dit Microsoft Copilot.

Voici la biographie d'Ana. Encore une fois, cela ne tient pas vraiment la comparaison avec la biographie qu'elle nous a donnée, mais cela ne semble pas complètement hors

sujet. Passons à la diapositive suivante. Ma première tentative a abouti à une Ana Neves complètement différente, Ana Luisa Neves. Je n'ai pas utilisé son deuxième prénom ici, mais cette personne est également très présente dans notre domaine d'une certaine manière. Elle est Directrice de la Santé Numérique Globale à l'Imperial College de Londres. J'espère que vous vous êtes rencontrées à un moment donné. Sinon, peut-être devriez-vous vous contacter. Vous avez probablement beaucoup de choses intéressantes à discuter sur notre avenir numérique. Diapositive suivante, s'il vous plaît.

C'est moi, si vous voulez savoir quelque chose sur moi. C'est un peu vieux. Comme je l'ai dit, j'ai assisté à plus de 10 réunions de l'ICANN. La vraie réponse est plus de 35 réunions de l'ICANN. Et je ne suis plus le président de l'ISOC New York. Je me suis rétrogradé au poste de secrétaire, toujours au conseil d'administration. Tout le reste est plus ou moins vrai. Diapositive suivante, s'il vous plaît.

C'est tout. Voici une autre photo de Seattle dans le brouillard, choisie par Microsoft Copilot. Revenons donc aux diapositives. Vous pouvez laisser celle-ci affichée. Elle est plutôt jolie, bien que quelque peu apocalyptique.

Passons à la discussion sur l'attaque et la défense de l'IA. Et je me demande, je vais commencer avec toi, Jeff. Qu'anticipes-tu pour les prochaines étapes ? Et quand tu regardes comment l'IA évolue, quel impact cela aura-t-il sur le domaine des abus de DNS dans lequel nous nous trouvons ?

Jeff: C'est une excellente question, Greg. J'espère vraiment que personne n'écoute ça et ne trouve des idées grâce à moi.

Je pense donc que l'espace où nous allons le plus le voir est dans la gestion des identités et les fausses identités. La capacité de l'IA à créer des cartes d'identité nationales et locales presque parfaites avec les identités de vraies personnes, mais peut-être une photo différente, etc., rendant presque impossible de distinguer une vraie personne d'une fausse, est une grande préoccupation.

Je pense donc que l'espace où nous allons en voir le plus est la gestion des identités et les fausses identités. La capacité de l'IA à créer des cartes d'identité nationales et locales presque parfaites avec les identités de vraies personnes, mais peut-être une photo alternative, etc., fait qu'il devient presque impossible de distinguer une vraie personne d'une fausse personne, ce qui est une grande préoccupation.

Et puis, quand vous pensez au fait que vous pouvez maintenant dire, d'accord, IA, pas son IA, parce qu'elle arrête ces choses, et je ne suis pas sarcastique. La sienne arrête ces choses. IA, je veux un numéro de carte de crédit qui n'a pas encore été utilisé et qui a été volé lors d'une violation. Je veux le faire correspondre à une pièce d'identité de cette même personne avec leur adresse personnelle réelle. Et s'ils résident dans l'État de Pennsylvanie aux États-Unis, je veux un permis de conduire qui correspond au modèle de permis de conduire de Pennsylvanie. Et je veux mettre ma photo dessus de manière

à ce qu'elle ressemble à une photo de permis de conduire. Donc, je suppose pas de sourire. Et puis générer cette image pour que lorsque je veux acheter un domaine, acheter des services avec cette carte de crédit volée, elle passe toutes les vérifications de connaissance du client. La carte de crédit n'a pas été signalée auparavant. Il y a une pièce d'identité parfaite pour la correspondre. Et tout cela a été généré en, je ne sais pas ce que vous en pensez, quatre ou cinq secondes.

Lorsqu'il s'agit de cybercriminalité, celle-ci doit être alimentée par quelque chose. Elle est principalement alimentée par des cartes de crédit volées et des comptes bancaires compromis. Et vous accédez à cet argent avec des identifiants volés modifiés. Et la manière d'obtenir ces identifiants est essentiellement ces identifiants solides qui valident l'identité d'une vraie personne. Je pense que c'est l'un des plus grands aspects que nous allons voir émerger et qui va vraiment changer rapidement le paysage.

Greg Shatan - NARALO / Moses Singer LLP: Merci, Jeff. Tim, je vais te poser à peu près la même question, mais je vais aussi te demander de réfléchir à la question précédente que j'ai posée, à savoir ce que font les développeurs d'IA en particulier. Sont-ils conscients de notre domaine de l'abus de DNS par rapport à tous les autres usages potentiels de l'IA, bons et mauvais ?

Tim: Oui, la réponse est oui, en partie parce que c'est dans l'intérêt de Microsoft, étant donné que nous gérons également une infrastructure cloud importante et que nous devons nous assurer que l'internet fonctionne globalement de manière à ne pas être abusé par des acteurs malveillants.

Pour revenir sur l'exemple que nous avons entendu sur la manière dont les attaquants l'utilisent, je pense qu'il y a de bonnes nouvelles. Depuis l'avènement de l'IA générative et des différents modèles qui ont été mis en place, certains prédisaient un scénario beaucoup plus apocalyptique et ce qui pourrait se dérouler. Nous n'avons pas vu la résurgence des vers que nous avons vus au début des années 2000, que l'on aurait pu imaginer avec l'IA générative écrivant de nouveaux codes qui auraient empêché d'éviter les mécanismes de détection existants ou simplement les filtres. Ce que nous avons principalement vu, c'est que les acteurs de la menace existants adoptent progressivement l'IA pour des TTP connus et avancent leurs objectifs. Ce n'est évidemment pas idéal, mais en termes de spectre des possibilités, je pense qu'il y a de bonnes nouvelles ici en termes de ce que nous avons vu.

Je pense donc qu'il y a déjà des leçons intéressantes à tirer en regardant les derniers mois sur la manière dont les défenseurs utilisent cette technologie. Comme vous l'avez démontré avec vos invites, l'IA peut être amusante en termes de ce qu'elle produit, mais cela souligne également l'importance de la manière dont vous utilisez réellement la technologie, quelles invites vous utilisez, à quel point elles sont spécifiques, et de traduire cela maintenant en techniques défensives. Si vous utilisez, par exemple, un Copilot ou un Copilot pour la sécurité ou tout autre produit et service d'IA pour produire des rapports spécifiques lorsqu'une anomalie est détectée et pour simplement

automatiser davantage, par exemple, signaler qu'une anomalie spécifique a été détectée, produire un résumé exécutif.

Nous voulons également nous assurer que nous instruisons le modèle à privilégier l'inclusion de plus d'anomalies potentielles plutôt que moins, afin de capturer les faux positifs. Les humains devraient ensuite examiner ce qui est signalé, et nous préférons capturer plus de faux positifs que nous pouvons ensuite écarter, plutôt que de réduire le nombre de signalements, ce qui pourrait inclure plus de faux négatifs. Donc, je pense que, premièrement, l'humain reste important. La manière dont nous pouvons tirer parti de l'IA est idéalement utilisée de manière à nous aider à utiliser les humains plus efficacement, et ce qui me rend quelque peu optimiste, malgré les préoccupations que vous avez soulevées, notamment sur la question de l'identité, avec laquelle je suis entièrement d'accord, c'est que l'un des principaux défis que nous avons rencontrés du côté défensif a été le manque de compétences et de personnes suffisamment formées pour accomplir certaines tâches. Si nous pouvons utiliser l'IA à grande échelle pour réduire ces barrières d'entrée et produire des rapports plus facilement compréhensibles même pour des personnes moins qualifiées en cybersécurité, alors nous pourrions avoir un effet positif global en termes d'évaluation nette de savoir si l'IA aide à faire avancer la défense ou l'attaque.

Juste une petite remarque sur le point d'Ana que je ne voulais pas laisser sans commentaire concernant l'importance de l'engagement entre le secteur privé et le secteur public. En regardant les trois dernières années, il a été fascinant de voir comment, après la sortie de Chat GPT et la une du New York Times, les régulateurs ont rapidement réagi avec l'AI Act de l'UE ou, dans le cas du gouvernement américain, les diverses politiques en place. Il y a eu une mise en œuvre très rapide de nouvelles réglementations, lois et politiques, à tel point que les entreprises ont dû s'assurer de suivre le rythme. Et avec l'émergence des instituts de sécurité de l'IA, qui sont censés aider à mettre en place des mesures d'atténuation pour éviter des abus comme taper "oh, donne-moi ce numéro de carte de crédit", j'ai été assez impressionné par la rapidité avec laquelle les gouvernements et les régulateurs ont mis en place des cadres en très peu de temps et comment ils ont collaboré avec le secteur privé, que ce soit Microsoft ou de nombreuses autres entreprises, pour essayer de mettre en place des cadres basés sur les risques et se concentrer sur ce à quoi nous devrions prêter attention.

Greg Shatan - NARALO / Moses Singer LLP: Merci Ana. J'aimerais avoir votre point de vue et en particulier, je demanderai une perspective européenne parce que nous trois, parmi d'autres choses que nous avons en commun, nous sommes tous très souvent en Amérique, ce qui bien sûr change énormément aussi. Mais nous devons, je serais intéressé d'ajouter quelques perspectives différentes dans le mélange ici.

Ana Cristina Amoroso da Neves - GAC / Portugal: Oui, absolument. Donc, en Europe, je pense que nous comprenons que le paysage des menaces pour les abus de DNS évolue rapidement avec l'IA et l'apprentissage automatique. Donc,

nous essayons de développer des lignes directrices pour l'utilisation éthique de l'IA dans la prévention des abus DNS tout en abordant les préoccupations de confidentialité liées à l'analyse du trafic DNS pilotée par l'IA.

Donc, nous avons ce rapport réalisé dans l'Union européenne. L'étude exhaustive de l'Union européenne sur le DNS

abusif, qui fournit des recommandations pour les registres et les bureaux d'enregistrement. Il date, je crois, de décembre 2022. Et il est intéressant de voir que des domaines comme .de et .eu

sont les domaines qui sont moins abusifs. Donc, là où l'abus de DNS n'est pas

si solide. Cela prouve que certains efforts et la mise en œuvre technique, ainsi que ce type de collaboration entre les secteurs public et privé, sont très importants.

Il est également très important de collaborer avec les chercheurs en IA, car ce sont eux qui peuvent nous aider à prévenir. Ainsi, avoir ces composantes de prévention en amont de ce qui peut se produire est très important. Encore une fois, le multi-parties prenantes est crucial dans tous ces processus numériques et l'abus du DNS en est un autre. Donc, l'importance d'inclure le monde académique, les chercheurs et les développeurs d'IA dans tous ces processus est

extrêmement important. Et donc, ce n'est qu'avec ce type d'implication et en écoutant toutes ces personnes et entités et différents acteurs impliqués que nous pouvons atteindre différents niveaux de politiques, qu'il s'agisse de politiques pour les entreprises ou de politiques publiques pour lutter contre l'abus du DNS. C'est mon avis pour l'instant.

Greg Shatan - NARALO / Moses Singer LLP: Merci, Ana. Nous avons beaucoup de monde dans la salle et virtuellement également. Donc, je voudrais passer aux questions et réponses. Si vous êtes à distance, veuillez mettre vos questions dans le module de Q&R et elles seront lues à haute voix par Michelle DeSmyter. Et si vous êtes dans la salle et que vous êtes sur Zoom, veuillez lever la main et je vois que la main de Joanna est levée. Et si vous n'êtes pas sur Zoom, levez simplement la main. Et si vous n'êtes pas à la table, nous avons un micro mobile qui viendra à vous. Donc, ne soyez pas timide si vous n'êtes pas assis près d'un microphone. Le microphone viendra à vous. Nous allons donc commencer avec Joanna Kulesza.

Joanna Kulesza - ALAC / University of Lodz: Merci. Merci beaucoup, Greg. C'est Joanna Kulesza pour les besoins de la transcription. Merci pour toutes les présentations. C'est merveilleux d'entendre parler du lien entre l'IA et le DNS.

J'ai une question qui fait suite à l'intervention d'Ana et à ce dernier commentaire. Je pense qu'il serait logique de l'adresser à Tim. Merci beaucoup pour cette revue très complète. Vous avez en quelque sorte abordé ce sujet. Je ne suis pas sûr que vous soyez à l'aise pour répondre avec votre casquette Microsoft. Donc, n'hésitez pas à choisir comment y répondre. Nous avons un passé de recherche en commun. Donc, si vous voulez parler des politiques mondiales de cybersécurité, c'est parfaitement acceptable aussi.

Je vais commencer par un petit contexte pour cette question, mais j'aimerais ensuite avoir votre avis. D'une part, Microsoft a été essentiel pour soutenir l'Ukraine après l'agression russe sans précédent. L'intelligence, la cybersécurité et l'analyse des menaces de Microsoft ont été cruciales pour que les Ukrainiens préparent leurs défenses. La situation pourrait changer bientôt. Il semble que le partage d'informations ait été interrompu. Maintenant, cela pourrait revenir. Je pense que c'est une indication très claire de la manière dont la réglementation nationale pourrait cibler les tentatives d'assurer la cybersécurité, que ce soit au niveau de l'abus de DNS ou plus généralement en ce qui concerne le partage d'informations.

D'autre part, Ana a mentionné, et vous l'avez également indiqué, qu'il y a une réglementation renforcée sur l'IA, y compris l'AI Act. Le sénateur Cruz, en décembre de l'année dernière, a déclaré que l'AI Act freine l'innovation et pourrait être considéré comme une ingérence étrangère dans le développement de l'IA aux États-Unis. Cela m'amène à mon dernier point et à la question elle-même. Notre bon ami, Marietje Schaake, a publié un livre, le Tech Coup, vous êtes peut-être au courant du concept. Et récemment, en en parlant, elle a mentionné que cela pourrait être le moment quantique pour les chercheurs qui ne se sentent pas très à l'aise de faire des recherches aux États-Unis de déménager dans l'UE.

L'UE, a-t-elle dit, devrait dérouler le tapis rouge, un peu comme dans un scénario inverse de ce que nous avons vu au milieu des années 1930. Donc, avec votre casquette de chercheur, je ne suis pas sûr que ce soit une question confortable pour Microsoft à répondre, et je ne la poserais pas dans ce contexte. Quel est, selon vous, le rôle de la réglementation dans le développement de l'IA pour la cybersécurité ? Et je vais transformer cela en une question liée à la politique avant que Jonathan ne me remette à ma place. Que peut faire la communauté des utilisateurs finaux, peut-être en travaillant avec le comité consultatif gouvernemental, pour s'assurer que nous utilisons l'IA pour protéger les individus contre les cybercriminels ? Je crois que c'est une manière diplomatique de poser une question. Et je sais que vous pouvez clairement identifier le défi géopolitique que j'essaie de soulever. Toutes les réflexions que vous pourriez partager seront appréciées. Merci beaucoup.

Tim Maurer - Microsoft: Merci, Joanna. Et c'est aussi un plaisir de te revoir, ce qui remonte également à l'époque avant la pandémie et avant mon passage au gouvernement.

Je suis heureux de répondre à cela en tant que représentant de Microsoft, car l'entreprise a été très claire concernant certains des défis géopolitiques auxquels nous avons été confrontés et s'engage fermement à assurer la cybersécurité, tant pour l'entreprise que pour les clients. Il y a l'initiative Secure Future, qui a remis la sécurité au premier plan pour l'ensemble de l'entreprise.

Permettez-moi d'aborder d'abord le dernier point sur ce que cette communauté peut faire. Je pense que nous sommes à un moment vraiment crucial de l'innovation technologique où les progrès que nous réalisons avec l'intelligence artificielle et Microsoft ont régulièrement lancé de nouveaux produits et services et prévoient de continuer à le faire. Il s'agit pour cette communauté de nous aider à déterminer comment utiliser au mieux la technologie pour le bien et, dans ce contexte, particulièrement pour la défense. Et cela dépend vraiment de vous, car vous pouvez parier que les personnes qui se concentrent sur la monétisation d'actions néfastes seront rapides à tester comment ces nouveaux outils peuvent aider à faire avancer leur objectif. Nous devons donc nous assurer d'adopter la technologie aussi rapidement. Et si vous identifiez des domaines où vous voyez des vulnérabilités, où vous considérez qu'il y a des questions de recherche sur lesquelles une entreprise comme Microsoft devrait se concentrer, alors veuillez les partager avec nous, car nous sommes tous dans le même bateau pour essayer de maximiser le potentiel de la technologie pour le bien. Nous accueillons donc certainement, que ce soit par le biais du GAC ou d'autres mécanismes, les contributions de cette communauté et nous voulons entendre chacun d'entre vous.

Sur le dernier point, sur le premier point concernant simplement l'UE et la réglementation, l'entreprise a été très claire sur le fait que nous pensons qu'il y a certains domaines où la réglementation sera utile. Je pense que là où l'entreprise et Microsoft commencent à s'inquiéter davantage, c'est si la réglementation est trop prescriptive et non axée sur les résultats, étant donné la rapidité de l'évolution technologique. Et lorsque nous commençons à voir différents cadres réglementaires apparaître dans de nombreux pays et juridictions qui sont parfois divergents ou même en contradiction les uns avec les autres, cela devient rapidement un défi pour nous de naviguer et de mettre en œuvre. Nous croyons donc qu'en ce moment, nous sommes à l'aube de réaliser des avancées significatives où, pour revenir à mon commentaire précédent sur le talent, nous pourrions être en mesure de faire pencher la balance au niveau systémique en faveur de la défense. Nous voulons donc nous assurer d'atteindre ce potentiel.

Greg Shatan - NARALO / Moses Singer LLP: Merci beaucoup, Tim. Et merci, Joanna, pour la question très complète. Si nous pouvions passer à la section de questions-réponses ensuite, et après cela, nous irons vers Hadia.

Michelle DeSmyter - ICANN: Merci, Craig. Et ici Michelle pour The Record. Nous avons une question de Saeed Najeeb. Quelle méthode utilisez-vous pour trouver l'IA dans CleanDNS ?

Jeff Bedser - SSAC / CleanDNS: Eh bien, les méthodes que nous utilisons sont dans la présentation. Je sais que nous utilisons Copilot en interne. Mais sur la plateforme, je pense que c'est probablement propriétaire. Nous utilisons une solution commerciale disponible, mais je ne suis pas autorisé à divulguer laquelle.

Michelle DeSmyter - ICANN: Très bien. Notre prochaine question vient de Siva Subramanian. D'un autre côté, l'IA ne peut-elle pas être utilisée pour concevoir des sites de phishing qui paraissent plus authentiques, au point que le site de phishing soit exempt des motifs et éléments typiques ?

Jeff Bedser - SSAC / CleanDNS: Je vais prendre cette question.

Donc, oui. Mais un site de phishing n'est pas seulement la page d'accueil. Il y a aussi d'autres détails concernant l'infrastructure entourant ce site. Par exemple, la plupart des sites de phishing ont été créés dans le but de phishing. Je veux dire, il y a des hôtes compromis, mais la majorité, c'est une inscription malveillante pour créer ce site. Ainsi, l'âge du domaine est un facteur clé, ainsi que la structure, les serveurs de noms sous-jacents, les IP, leur réputation. Souvent, vous constaterez que l'adresse IP sous-jacente et les hôtes sont des hôtes à l'épreuve des balles, et les hôtes à l'épreuve des balles sont ceux qui ne répondent pas aux assignations ou à tout type de procédure et ne journalisent rien. Ce n'est vraiment pas une pratique commerciale légitime. Donc, si vous faites cela, cela signifie en gros que vous facilitez une activité criminelle. Il y a donc beaucoup de facteurs à prendre en compte en dehors de la simple page d'accueil. Mais oui, les pages d'accueil deviennent de plus en plus sophistiquées grâce à l'IA. Il n'y a aucun doute là-dessus.

Greg Shatan - NARALO / Moses Singer LLP: Merci, Jeff. Maintenant, je vais donner la parole à Hadia Elminiawi dans la salle.

Hadia s'approche du micro.

Hadia Elminiawi - SSAC / AFRALO: Merci, Greg. Et pour information, je suis Hadia. Ma question s'adresse à Ana. Ana, tu as parlé de transparence dans l'IA. Ma question est donc : qu'est-ce que la transparence signifie concrètement ? Est-ce que cela veut dire

L'IA open source ? Nous savons, par exemple, que l'IA open source ne démystifie toujours pas comment l'IA prend ses décisions. Et les systèmes en boîte noire restent des boîtes noires, même avec l'open source. Donc, en termes pratiques, qu'est-ce que cela signifie vraiment ? Merci.

Ana Cristina Amoroso da Neves - GAC / Portugal: Merci pour cette question. Très intéressante. Je pense que lorsque j'ai mentionné la transparence de l'IA, il s'agissait de transparence et de responsabilité. Donc, les modèles d'IA devraient être explicables. Cela permettrait aux opérateurs NES et aux professionnels de la cybersécurité de

comprendre les processus de prise de décision. Si vous ne comprenez pas ce pour quoi vous vous battez, c'est très difficile. Donc, si les modèles d'IA peuvent être explicables, il est beaucoup plus facile de traiter ce qui pourrait être incorrect. Donc,

c'est tout en résumé. Merci.

Greg Shatan - NARALO / Moses Singer LLP: Merci. Jonathan, aviez-vous une question ?

Jonathan Zuck - ALAC: Oui, Jonathan Zuck pour le compte rendu.

Je n'ai que ma propre IA, que je suppose appeler intelligence antique, pour détecter les tentatives de phishing et ce genre de choses. Mais il semble que la plupart des emails de phishing soient liés d'une manière ou d'une autre à des entités déposées ou à de grands noms reconnaissables. Et quelque part dans cet email se trouve un lien vers un site de farming. Et je me demande dans quelle mesure un gestionnaire d'emails pourrait examiner et détecter de qui cela était censé provenir et examiner les liens intégrés dans l'email et les vérifier pour voir s'ils appartiennent à l'entité dont l'email semble provenir. Cela semble être une tâche assez simple pour une IA, mais je n'ai rien vu de tel, je suppose, à ce jour. Je me souviens avoir été vraiment surpris un jour quand j'ai cliqué sur envoyer dans Gmail et qu'il a dit, hé, il semble que vous vouliez joindre un fichier. J'étais vraiment excité et ensuite un peu effrayé, mais bon, ils lisent le message. Je veux dire, à l'arrivée, il semble que ce genre d'analyse dise, hé, cela semble provenir de Norton Utilities, mais les liens intégrés dedans, alors faites attention ou quelque chose comme ça.

Tim Maurer - Microsoft: Je pense que cela revient à ce que nous avons discuté au début à propos de l'IA, beaucoup de cela n'est pas nouveau et nous avons en fait déjà un bon nombre de systèmes en place basés sur l'apprentissage automatique et des algorithmes qui, pour répondre à votre point, sont les raisons pour lesquelles de nombreux e-mails vont directement dans le dossier spam, car il y a maintenant des filtres en place comparé à il y a 15, 20 ans qui rendent votre boîte de réception plus propre qu'elle ne l'était. C'est là que je pense, pour répondre à votre point et aussi à ce que vous avez soulevé, que le véritable potentiel que nous espérons voir avec l'IA à l'avenir est de mieux connecter certains des points qui n'ont pas pu être connectés et de rassembler différentes sources de données et des insights afin que vous puissiez détecter si un site a été enregistré d'une manière qui nous inquiète quant à l'identité en raison de la longueur de la page. Et c'est là que nous pensons que le véritable potentiel peut se trouver, en plus de le rendre plus convivial et consommable et, en ce qui concerne l'explicabilité, de comprendre ce que l'humain a réellement besoin de savoir et de lire pour ensuite agir en conséquence. J'espère que cela répond à votre question.

Greg Shatan - NARALO / Moses Singer LLP: Nous avons dépassé le temps, mais je vais prendre une dernière question. Amrita, s'il vous plaît.

Amrita Choudhry - CCAOI: Merci beaucoup, Greg. Amrita, pour mémoire, j'ai deux questions. La première est que j'aimerais comprendre de Tim, toi et Jeffrey, nous avons parlé de l'abus de la langue anglaise, et comment l'IA peut l'utiliser ou non, ou l'aggraver. Qu'en est-il des non-anglophones ? Parce que c'est la majorité mondiale, que se passe-t-il là-bas ? La deuxième est que nous voyons que cet abus de noms de domaine est en augmentation. Le contenu, évidemment, mais aussi son effet financier et sur la réputation partout. Et surtout venant des pays de la majorité mondiale où vous avez de nouveaux utilisateurs qui ne sont pas très lettrés en termes de numérique ou même lettrés à ce point. Alors, que faites-vous pour ces utilisateurs finaux, avec des vidéos ou des tutoriels de formation, etc. ? Comment faites-vous, surtout quand Microsoft, il y a environ 10 ans, avait des cours d'alphabétisation numérique pour les gens. Donc, pour ce genre de sécurité, que faites-vous ? Parce que les gens perdent aussi de l'argent. Merci.

Tim Maurer - Microsoft: Je vais commencer et ensuite, après...

Donc, Microsoft, avec ses activités dans le monde entier, déploie des produits, des services et des modèles dans plusieurs langues, y compris en matière de sécurité, en ce qui concerne nos efforts de red teaming qui ne se déroulent pas seulement en anglais, mais aussi dans d'autres langues. Nous nous engageons donc fermement à rendre la technologie accessible aux personnes du monde entier et à ce que les mesures d'atténuation de la sécurité et de sûreté soient adaptées aux utilisateurs qui les utilisent.

En ce qui concerne le renforcement des capacités et la formation, nous avons fait une grande annonce l'année dernière concernant le Kenya et un accent particulier sur le Sud global, car nous sommes très conscients que nous voulons utiliser la technologie pour amener de nouveaux utilisateurs en ligne et nous assurer qu'ils savent comment utiliser la technologie. Nous avons également beaucoup de formations internes où nous sommes tous encouragés à l'utiliser nous-mêmes. Donc, ce n'est pas seulement axé sur des groupes d'utilisateurs spécifiques ou des pays. Je pense que c'est un effort global pour s'assurer que nous fournissons la formation en même temps que nous déployons de nouveaux produits et services afin de maximiser l'impact de la technologie.

Jeff Bedser - SSAC / CleanDNS: Merci pour cette précision, car j'ai omis de mentionner dans ma présentation que ce n'était pas uniquement en anglais. Mais oui, les outils utilisés sont conçus pour n'importe quelle langue dans laquelle la menace se présente, que ce soit pour l'analyse d'images ou de textes. Et c'est encore une des raisons pour lesquelles l'IA est si efficace pour nous, car je n'ai pas besoin d'avoir quelqu'un dans l'équipe qui parle et lit toutes les langues du monde. L'IA peut évaluer la langue, déterminer ce qui se passe et extraire les composants pertinents.

Donc, oui, c'est un élément très important de s'assurer que nous traitons non seulement les langues majoritaires, mais toutes les langues, car comme nous fournissons des services à travers le monde, nous voyons du phishing, des logiciels

malveillants, etc., être livrés dans toutes les langues. Nous utilisons donc l'IA de manière efficace pour nous aider, car je ne pourrais pas me permettre une équipe aussi grande qui gère toutes ces langues.

Greg Shatan - NARALO / Moses Singer LLP: Merci, Jeff. Nous arrivons maintenant à la fin de notre session. Malheureusement, nous avons encore quelques questions intéressantes. Désolé, Dana et Atif. Mais bien sûr, c'est une discussion continue. Je voudrais remercier tous nos panélistes, Jeff Bedser, Tim Maurer, Ana Neves, d'avoir été présents. Je tiens à vous remercier tous pour votre écoute et pour avoir fait de cette table ronde NARALO un moment mémorable. Nous reviendrons dans cette salle à la demi-heure pour l'assemblée générale de NARALO. Merci. Et cette réunion est maintenant levée.