IA GÉNÉRATIVE

# DE NOUVELLES MENACES POUR LES CABINETS

**L'intelligence artificielle générative (ou IA Gen) n'est plus une promesse technologique ni un concept futuriste : elle est déjà un outil quotidien. Conversations en langage naturel, solutions de rédaction automatisée de rapports, génération de codes, traitement automatisé de documents…**
**Les usages sont multiples et souvent séduisants. Mais, comme toute avancée technologique, l'IA Gen apporte aussi son lot de menaces. Certaines sont connues, d'autres totalement nouvelles. Elles concernent autant la sécurité des données que la fiabilité des informations fournies et la réputation des cabinets.**

Par **Dominique Périer**, en charge du pilotage des Grands projets numériques de la mandature, **Fabrice Faivre**, directeur Veille, Innovation et Normalisation, Conseil national, **Sophie Lambert**, lieutenant-colonel, chef de l'anticipation et de la gestion de crise cyber au ministère de l'Intérieur, & **Jérôme Notin**, directeur général de Cybermalveillance.gouv.fr

## L'IA GÉNÉRATIVE, ACCÉLÉRATEUR… DES CYBERMENACES AUSSI !

L'IA Gen est capable de produire très rapidement, textes, images, sons, vidéos ou lignes de code à partir de simples instructions. Toutefois, elle ne remplace pas seulement le travail humain… elle peut aussi industrialiser la fraude !

L'IA Gen impose aux cabinets une vigilance accrue. En 2024, plus de 348 000 atteintes numériques ont été recensées en France, soit + 74 % en cinq ans[1]. Cette croissance reflète l'industrialisation des cyberattaques permise par l'IA : campagnes de *phishing* hyper-personnalisées, *deepfakes* crédibles, malwares polymorphes capables d'échapper aux antivirus traditionnels.

### La désinformation automatisée

Grâce à l'IA Gen, il est désormais possible de générer rapidement des contenus convaincants, imitant le style d'un interlocuteur, d'une entreprise ou d'un média. On parle de *deepfakes* textuels ou visuels.

Ces faux contenus peuvent :
- diffuser de fausses informations financières ou fiscales ;
- imiter un dirigeant ou un collaborateur dans un échange d'emails, oral ou visio ;
- tromper un client ou un partenaire sur la santé d'une entreprise.

De nombreux cas d'escroquerie avérée par *deepfakes* ont été recensés ces deux dernières années. Les modes opératoires ont souvent été similaires : les voix des personnes imitées ont été récupérées sur internet, et il suffit désormais aux IA génératives de disposer de 30 secondes à 1 minute de voix pour leur permettre de générer un faux message vocal. Coupler ces messages aux techniques de génération vidéo sur la base de photos permet d'obtenir une vidéo réaliste d'un interlocuteur connu présentant un message inventé de toutes pièces.

1. Cf. le rapport annuel relatif à la cybercriminalité, publié sur le site du ministère de l'Intérieur

**Le *phishing* nouvelle génération**
Le *phishing* classique se reconnaissait souvent par ses fautes ou son style maladroit. Avec l'IA Gen, un cybercriminel peut désormais produire :
- des messages parfaitement rédigés, contextualisés, et personnalisés ;
- des appels téléphoniques synthétiques imitant la voix d'une personne connue (cf. *supra*) ;
- des faux sites web générés automatiquement pour récolter des données.

L'objectif clairement identifié de ces *phishings* améliorés et diversifiés auprès du plus grand nombre et sur le plus de canaux (mails, SMS, voix…) est de récupérer des informations de connexion à des sites, des informations personnelles et bancaires, voire faire réaliser par les victimes des paiements directement sur des faux sites.

La massification de campagnes de *phishing* ciblant des petits montants permet dès lors malheureusement la recrudescence de ce type d'attaques.

**La compromission des données par mauvaise utilisation**
Les solutions d'IA Gen fonctionnant en ligne requièrent souvent l'envoi de données sur des serveurs distants.

Si un collaborateur soumet des informations confidentielles à un outil grand public, ces données peuvent :
- être stockées et analysées par le fournisseur de l'IA ;
- fuir accidentellement lors d'une mise à jour ou d'un incident de sécurité ;
- être exploitées pour entraîner des modèles futurs.

La plupart des outils d'IA Gen proposent des versions gratuites afin d'inciter le plus grand nombre à les utiliser dans des versions limitées (à l'usage) mais surtout en termes de protection des informations fournies. Les versions gratuites ne garantissent pratiquement jamais la confidentialité des données qui leur sont partagées.

**L'automatisation des cyberattaques**
Des modèles spécialisés et entraînés grâce à l'IA peuvent générer du code malveillant ou trouver des failles plus rapidement dans un système. Ainsi, l'IA Gen permet la production rapide de scripts et codes malveillants qui facilitent dès lors aux cyberattaquants :
- le lancement d'attaques par force brute plus rapides : celles-ci consistent à tester de façon systématique un grand nombre de possibilités (mots de passe, clés, codes PIN, etc.) jusqu'à trouver celle qui fonctionne. L'attaquant automatise les essais (scripts, bots, cartes GPU) et peut tester des millions, voire des milliards de combinaisons par seconde selon son matériel et l'objectif ;
- la création de logiciels espions adaptés à des environnements précis. La massification d'informations ingérées par les modèles d'IA Gen permet de réaliser ces actions très rapidement et d'accéder à des informations très ciblées selon les contextes ;
- l'adaptation d'attaques à chaque cible en temps réel.

**La dépendance technologique et les biais**
L'usage massif de l'IA Gen peut amener à déléguer des tâches critiques à un système opaque, ce qui pose deux risques :
- Les biais algorithmiques : réponses erronées ou discriminatoires. Parfois, l'IA Gen fournit une réponse ou une prédiction qui favorise ou défavorise certaines catégories de manière injuste ou non représentative de la réalité. Il ne s'agit pas d'une « volonté » de l'IA car elle n'a pas d'intention, mais d'un effet lié aux données, aux modèles ou à la façon dont on utilise les algorithmes ;
- La perte de savoir-faire : moins de contrôle et de vigilance humaine. Naturellement, plus les humains auront tendance à se reporter aux IA Gen pour produire des contenus à leur place, plus ils risqueront de perdre un savoir-faire.

Mais les modèles d'IA générative risqueront également à l'avenir de s'épuiser en termes de connaissances (bouclage sur des contenus produits par des IA) et les biais algorithmiques risquent de se retrouver automatiquement décuplés.

## LES MENACES SPÉCIFIQUES POUR LES CABINETS D'EXPERTISE COMPTABLE

Face à cette accélération, les cabinets doivent d'abord encadrer l'usage interne de ces outils.
Toute saisie de données fiscales, sociales ou financières dans une solution d'IA externe peut entraîner une fuite ou une compromission, avec des conséquences graves en matière de responsabilité professionnelle ou de conformité (RGPD, secret professionnel).

Une charte interne précisant les outils autorisés, les données exclues et la nécessité de validations humaines est indispensable[2].

**Atteinte à la confidentialité des dossiers**
Les cabinets traitent des données fiscales, sociales et financières hautement sensibles. Une fuite, même partielle, peut avoir des conséquences lourdes : perte de confiance des clients, sanctions RGPD, mise en cause de la responsabilité civile.

De nombreux cas de transferts de données financières confidentielles à des IA Gen ont déjà été constatés dans le monde ; elles ont naturellement été exploitées immédiatement ensuite par des concurrents ou cybercriminels.

De plus, lorsque la fuite est constatée, il n'est malheureusement plus possible par la suite de revenir en arrière et d'effacer cette diffusion…
Les cas avérés ont conduit toutefois à des condamnations juridiques des responsables, mais trop tard, la donnée avait fuité.

2. Cf.« Rédiger et déployer une charte IA en cabinet », *SIC mag* n° 449, juin 2025, pp. 36-38.

**Escroqueries à l'ordre de virement**
Avec la voix ou le style d'écriture d'un dirigeant reproduit parl'intelligence artificielle, un cybercriminel peut ordonner un transfert d'argent en se faisant passer pour un client.
La précision et la crédibilité de ces imitations rendent la détection beaucoup plus difficile.
Les informations propres aux dirigeants d'entreprises clientes des cabinets d'expertise comptable sont facilement disponibles et récupérables sur internet.

Comme évoqué *supra*, il est facilement possible pour un cyberattaquant de trouver des informations vocales sur ces personnes (réseaux sociaux) et de réaliser des messages vocaux demandant au cabinet comptable de réaliser un ordre de virement.

Le cyberattaquant appelle le collaborateur du cabinet comptable et passe la voix du dirigeant pour demander l'ordre de virement.
Si le cabinet dispose du mandat de règlement et que le collaborateur n'a pas le réflexe de contre-vérifier la demande, l'escroquerie peut réussir !

**Falsification de documents comptables ou fiscaux**
Aujourd'hui, les IA Gen peuvent créer de fausses notes de frais, de faux justificatifs bancaires, de fausses factures ou des faux bulletins de paie parfaitement crédibles.
Ces faux peuvent tromper un collaborateur non averti et compromettre la qualité des missions.

**Altération involontaire de la qualité des travaux**
En se reposant sur des outils d'IA Gen non vérifiés, un collaborateur peut intégrer dans un rapport des données erronées ou non sourcées, engageant ainsi la responsabilité du cabinet.

## COMMENT S'EN PRÉMUNIR PAR QUELLES BONNES PRATIQUES ?

Face à ces nouvelles menaces, la réponse ne peut pas être uniquement technique. Elle doit combiner d'autres composantes toutes aussi importantes telles que la gouvernance, la formation et la culture de la sécurité.

**Mettre en place une charte interne d'utilisation de l'IA**
Mettre en place une charte IA en entreprise, c'est un peu comme définir un code de conduite pour l'usage de cette technologie.

Cette charte doit permettre d'encadrer les usages et de réduire les risques, de sécuriser les données, de garantir conformité et éthique et de donner un cadre clair aux collaborateurs.
Ainsi cette charte doit clairement préciser :
- les outils autorisés (encadrés contractuellement ou non) et ceux qui sont proscrits ;
- les types de données pouvant être traitées par IA selon les outils autorisés ;
- les validations humaines obligatoires avant diffusion d'un contenu généré.

**Utiliser des IA sécurisées et souveraines**
Comme indiqué précédemment, toutes les IA génératives ne sont pas sécurisées de la même manière. Il convient donc de privilégier :
- des solutions hébergées sur des serveurs internes ou dans des environnements certifiés ;
- des outils conformes au RGPD et offrant des garanties contractuelles sur la non-utilisation des données à des fins d'entraînement (versions payantes notamment).

Ainsi, il est extrêmement important d'encadrer et organiser les outils que les collaborateurs peuvent utiliser en contractualisant notamment avec ces outils afin de disposer des bonnes clauses de protection.

**Former et sensibiliser les équipes**
Au-delà de la gouvernance, la culture cyber reste la clé. Former régulièrement les collaborateurs aux menaces spécifiques de l'IA (*deepfakes*, fraudes au président, faux justificatifs) permet de développer les bons réflexes.
Les collaborateurs doivent être capables de :
- identifier un contenu potentiellement falsifié ;
- reconnaître les signaux faibles d'une tentative de fraude ;
- comprendre les limites techniques et juridiques de l'IA.

**Renforcer les procédures de vérification**
Il convient de définir, voire de renforcer les contrôles existants tels que :
- la double validation pour les ordres de virement ;
- le contrôle systématique des justificatifs et documents clients ;
- la vérification des sources des données générées et la traçabilité des échanges sensibles.

**Simuler des attaques internes**
Des exercices de simulation (*phishing* test, faux appels audio) entretiennent la vigilance et révèlent les failles organisationnelles. Il convient de mettre régulièrement en place ce type d'exercices dans les cabinets pour identifier les failles et y remédier.
Il convient d'autant plus de régulariser ces exercices dans le cadre d'une rotation régulière des équipes et ressources.

## QUELS OUTILS DISPONIBLES POUR LES CABINETS ?

**Dossiers thématiques du Conseil national**
Le site privé de l'Ordre présente un dossier thématique Cybersécurité qui regroupe de nombreuses ressources dont notamment :

- les 11 commandements de la cybercriminalité qui décrivent les règles de base que chaque cabinet doit *a minima* mettre en œuvre pour se sécuriser ;

- le lien vers l'assistance 17Cyber.gouv.fr ;
- un kit mission permettant au cabinet de s'autodiagnostiquer sur les aspects cyber et également de proposer la mission à ses clients ;
- un plan de reprise et de continuité d'activité permettant de mettre en œuvre les réflexes de base pour anticiper une crise et mieux y répondre.

Enfin, le dossier thématique Parlons IA contient également des ressources telles que :

- les 12 piliers pour un usage efficace et sûr de l'IA
- IA et IA Gen adaptées à la profession du chiffre
- un modèle de charte d'utilisation de l'IA générative au sein d'un cabinet à adapter à chaque cabinet

**Les outils Cybermalveillance.gouv.fr**
Ce dispositif national de prévention et d'assistance aux risques numériques, propose différents contenus accessibles et simples à destination des cabinets, mais aussi de leurs clients :

- Comment piloter sa cybersécurité ? (guide pour les dirigeants)
- Mémento de cybersécurité
- Guide cybersécurité (avec Bpifrance)
- Séparation des usages pro-perso
- Que faire en cas de cyberattaque ? (guide pour les dirigeants)
- Que faire en cas de cyberattaque ? (consignes d'urgence aux collaborateurs)

Il est important de les consulter et de les faire connaître auprès de l'ensemble des entreprises en France. En complément des actions de prévention à destination des collaborateurs des entreprises, il est essentiel de protéger son organisation en amont contre le risque cyber. C'est pourquoi Cybermalveillance.gouv.fr a créé sur sa plateforme Mon ExpertCyber, un service dédié à la sécurisation des systèmes d'information. Ce service permet aux publics professionnels de bénéficier d'une mise en relation directe avec un prestataire de confiance labellisé ExpertCyber qualifié pour sécuriser des systèmes d'information professionnels.

## VERS UNE IA GÉNÉRATIVE ÉTHIQUE ET MAÎTRISÉE

L'IA Gen n'est pas uniquement un risque : elle peut aussi devenir un levier positif si elle est intégrée de manière encadrée et responsable. Elle peut également renforcer la sécurité, par exemple en détectant plus vite les fraudes ou en automatisant la veille réglementaire. L'enjeu est donc de maîtriser l'outil, non de le bannir.

La profession comptable, déjà rompue à la confidentialité, dispose d'atouts pour relever ce défi : culture de la rigueur, respect des normes, sens de la responsabilité. En intégrant l'IA Gen dans un cadre sécurisé et éthique, les cabinets peuvent transformer une menace potentielle en une véritable opportunité maîtrisée.

En conclusion, l'IA générative marque un tournant : jamais la création de contenus crédibles n'a été aussi rapide et accessible… pour le meilleur comme pour le pire. Les cabinets d'expertise comptable doivent anticiper ces menaces, adapter leurs procédures et former leurs équipes.

De plus, toutes ces problématiques ne sont pas propres qu'aux cabinets et la profession peut, et doit, également informer et accompagner dans la mesure du possible ses clients sur ces sujets.

Car dans ce nouveau paysage numérique, la vigilance humaine reste – et restera – la meilleure défense.

### POUR EN SAVOIR PLUS

Consulter sur www.experts-comptables.org (site privé de l'Ordre) :
- les dossiers thématiques Cybersécurité ;
- le dossier thématique Parlons IA.

Visionnez sur Fuz'experts.tv le replay du webinaire « Cybersécurité et intelligence artificielle : nouvelle donne pour les cabinets d'expertise comptable », organisé en juin 2025.