

Note d'analyse — septembre 2026

Le risque de fausse instruction

Pourquoi la voix et l'image d'un dirigeant ne sont plus une preuve d'identité, et ce qu'une organisation peut mettre en place sans attendre un outil de détection.

Une attaque n'a plus besoin de franchir un pare-feu. Il lui suffit d'imiter une voix connue, d'invoquer l'urgence et la confidentialité, et de demander à une personne habilitée de faire ce qu'elle fait tous les jours : un virement, un changement de coordonnées bancaires, l'envoi d'un document. Le système d'information n'est pas attaqué. La chaîne de décision, elle, l'est.

Cette note décrit ce mécanisme à partir de sources publiques et institutionnelles, explique pourquoi les défenses techniques habituelles ne le couvrent pas, et propose un protocole de vérification qui ne coûte rien à installer.

Ce que contient cette note. Les chiffres français disponibles sur la fraude par manipulation, l'anatomie d'une fausse instruction, les raisons pour lesquelles l'entourage cède, six règles de vérification applicables immédiatement, le cadre juridique, et une auto-évaluation en dix questions.

Ce qu'elle n'est pas. Ni une consultation juridique, ni un audit de sécurité informatique, ni une évaluation d'un cas particulier. Elle ne contient aucune information issue d'un dossier client et ne recommande aucun produit.

En résumé

- En France, la fraude par manipulation de la personne a atteint **516 millions d'euros en 2025**, soit un peu plus de 40 % du montant total de la fraude aux moyens de paiement, et sa part a doublé en quatre ans.
- Le point faible n'est pas l'outil, c'est la **règle de décision** : qui a le droit de dire oui, sur quel canal, et à quelles conditions.
- Une voix ou un visage ne constituent plus une authentification. Les seules preuves qui résistent sont **le canal de retour et le temps**.
- La jurisprudence française récente rappelle qu'une opération **autorisée** par une personne habilitée laisse la charge du contrôle à l'organisation, pas à sa banque.
- Un protocole de vérification écrit, connu d'un petit nombre de personnes, retire à l'attaque ses deux leviers — l'urgence et l'interdiction de vérifier — sans investissement technique.

1. Ce que disent les chiffres

Les données françaises de référence sont publiées par l'Observatoire de la sécurité des moyens de paiement, adossé à la Banque de France. Elles ne mesurent pas les hypertrucages en tant que tels, mais la catégorie qui les englobe : la fraude obtenue en manipulant une personne habilitée plutôt qu'en contournant un dispositif technique.

Indicateur	Valeur	Période
Fraude aux moyens de paiement, tous canaux	1,2 milliard d'euros, +3,8 %	Année 2025
Dont fraude par manipulation de la personne	516 millions d'euros , un peu plus de 40 % du total	Année 2025
Part de la manipulation dans la fraude totale	A doublé en quatre ans	2021 → 2025
Fraude par manipulation	245 millions d'euros, +37 %	1 ^{er} semestre 2025
Dont virements émis depuis la banque en ligne	174 millions d'euros, +54 % sur un an	1 ^{er} semestre 2025
Fraude au virement, valeur totale	230,4 millions d'euros, +43,8 %	1 ^{er} semestre 2025
Nombre de virements frauduleux	109 519 opérations, +106,7 %	1 ^{er} semestre 2025

Indicateur	Valeur	Période
Montant moyen d'un virement frauduleux	2 104 euros, contre 3 023 euros un an plus tôt	1 ^{er} semestre 2025

Sources : Banque de France, rapport annuel de l'Observatoire de la sécurité des moyens de paiement 2025 — banque-france.fr/fr/publications-et-statistiques/publications/rapport-de-lobservatoire-de-la-securite-des-moyens-de-paiement-2025 ; note statistique du premier semestre 2025, publiée le 27 janvier 2026.

Deux lectures de ces chiffres

La première est rassurante : le taux de fraude au virement reste très faible, 0,0014 % des flux. La seconde l'est moins : le nombre d'opérations frauduleuses a plus que doublé en un an tandis que le montant moyen baissait d'un tiers. Autrement dit, les tentatives se multiplient et se banalisent. Une organisation n'est pas exposée à un taux, elle est exposée à un événement.

Côté entreprises, une enquête menée du 13 février au 2 avril 2025 auprès de 145 entreprises françaises de plus de quinze secteurs indique que 85 % d'entre elles ont été visées par une tentative de fraude, que le virement est le canal d'exfiltration dans 57 % des fraudes abouties, que 64 % ne récupèrent jamais les fonds, et que **seules 44 % disposent d'une procédure formalisée**. Dans cette même enquête, 83 % des répondants citent l'intelligence artificielle et les hypertrucages comme la principale menace émergente. Source : enquête Memo Bank et DFCG, juin 2025 — memo.bank/communique-de-presse/enquete-fraude-dfcg-2025.

L'écart entre 83 % qui identifient la menace et 44 % qui ont une procédure écrite est l'espace exact dans lequel opère ce type d'attaque.

2. Anatomie d'une fausse instruction

Le schéma est stable. Il ne dépend ni de la taille de l'organisation ni de son secteur, et il se déroule en cinq temps.

Temps 1 — La collecte

Tout ce qui est nécessaire est public : un organigramme, un communiqué de presse, une interview vidéo où la voix du dirigeant est audible pendant quelques secondes, un profil professionnel qui nomme l'assistante et le directeur financier, une publication qui annonce un déplacement ou une opération en cours. Aucune intrusion n'a eu lieu.

Temps 2 — Le prétexte

L'attaque choisit un moment où l'inhabituel devient plausible : une acquisition confidentielle, un contentieux, un déplacement à l'étranger, une clôture comptable, une absence annoncée publiquement.

Temps 3 — Le contact

Un message, puis un appel. La voix ressemble. Le numéro affiché peut ressembler aussi. Le FBI décrit précisément cette séquence dans une alerte publique : des messages textes et des messages vocaux générés par intelligence artificielle usurpant l'identité de hauts responsables, avec une caractéristique récurrente — l'interlocuteur demande presque immédiatement de basculer la conversation vers une application de messagerie chiffrée, hors des canaux traçables de l'organisation. Sources : FBI, Internet Crime Complaint Center, alertes du 15 mai 2025 et du 19 décembre 2025 — ic3.gov/PSA/2025/PSA250515 et ic3.gov/PSA/2025/PSA251219.

Temps 4 — La contrainte

Trois éléments arrivent ensemble et ce sont eux, bien plus que la qualité de l'imitation, qui font céder : l'urgence, la confidentialité, et l'interdiction implicite de vérifier. La personne visée n'est pas naïve. Elle est placée dans une situation où vérifier semble être une faute professionnelle.

Temps 5 — L'exécution

Le virement part, le RIB est modifié, le document est transmis. L'opération est techniquement régulière : elle a été ordonnée par une personne habilitée, depuis un poste légitime, avec une authentification valide.

Le point à retenir. Aucun de ces cinq temps ne met en cause un outil informatique. Ils mettent en cause une règle de décision absente ou contournable.

3. Pourquoi les défenses habituelles ne couvrent pas ce risque

Défense en place	Ce qu'elle arrête	Pourquoi elle ne suffit pas ici
Filtrage des courriels et antivirus	Les pièces jointes et liens malveillants	L'attaque passe par la voix, le téléphone ou une messagerie personnelle, hors du périmètre filtré
Authentification forte	L'usurpation d'un compte	Ce n'est pas le compte qui est usurpé, c'est l'intention. L'utilisateur légitime agit lui-même
Double validation des paiements	L'erreur isolée d'une seule personne	Inopérante si les deux valideurs reçoivent la même instruction crédible, ou si le second est court-circuité par l'urgence
Formation annuelle à la vigilance	La reconnaissance des schémas déjà connus	Elle apprend à se méfier, mais ne dit pas quoi faire, ni à qui désobéir, dans les trente secondes qui comptent
Détection d'hypertrucages	Certains contenus synthétiques	Aucun détecteur n'est infaillible, et la personne au téléphone n'a pas le temps de soumettre la voix à un outil pendant l'appel

Europol, dans le premier rapport de son observatoire consacré au sujet, estime que la technologie d'hypertrucage pourrait devenir un outil courant de la criminalité organisée. Source : Europol Innovation Lab, Facing reality? Law enforcement and the challenge of deepfakes — europol.europa.eu/publications-events/publications/facing-reality-law-enforcement-and-challenge-of-deepfakes.

La conséquence est contre-intuitive. Face à un risque né de l'intelligence artificielle, la contre-mesure la plus solide n'est pas technologique : c'est une règle écrite, connue à l'avance, que personne n'a le droit de faire sauter, pas même le dirigeant lui-même.

4. Six règles qui tiennent sans outil

Ces six règles ne demandent aucun logiciel, aucun budget et aucune compétence technique. Elles demandent une décision, écrite, et connue d'un petit nombre de personnes identifiées.

1. Le canal de retour

Aucune instruction sensible ne se vérifie sur le canal par lequel elle est arrivée. On rappelle sur un numéro connu à l'avance, enregistré avant l'incident. Jamais le numéro affiché, jamais celui donné dans le message, jamais un lien de visioconférence transmis dans l'échange.

2. Le délai de latence

Un délai minimal et non négociable s'applique à toute instruction inhabituelle : quinze minutes suffisent. L'urgence est l'outil de travail de l'attaquant ; lui retirer le temps, c'est lui retirer son levier principal.

3. Le mot de rappel

Une phrase convenue à l'avance, connue seulement du dirigeant et de trois ou quatre personnes de confiance, jamais écrite dans un courriel ni stockée sur un téléphone. Elle ne prouve pas l'identité à elle seule, mais son absence suffit à suspendre l'exécution.

4. Le droit de dire non

La règle la plus importante et la plus rarement écrite : la personne qui refuse d'exécuter une instruction non vérifiée est couverte, par écrit, même si l'instruction était authentique et même si le refus a coûté une opportunité. Sans cette phrase, tout le reste s'effondre au premier appel insécurisant.

5. La personne de recours

Une seule personne désignée, joignable, vers qui remonter un doute sans passer par la voie hiérarchique habituelle — puisque c'est précisément cette voie que l'attaque imite.

6. Le journal

Toute tentative, même déjouée, même douteuse, est consignée : date, heure, canal, numéro, contenu, personne visée. C'est ce qui permet de repérer une campagne qui reviendra, et c'est ce qui sert de preuve en cas de plainte.

Test simple. Si votre organisation subissait un appel de ce type demain, la personne visée saurait-elle, en moins de trente secondes et sans demander à quiconque, quoi faire et qui appeler ? Si la réponse n'est pas un oui immédiat, le protocole n'existe pas encore.

5. Ce que dit le droit français et européen

Le droit pénal sanctionne, après coup

L'article 226-8 du code pénal sanctionne la publication d'un montage réalisé avec les paroles ou l'image d'une personne sans son consentement, d'un an d'emprisonnement et de 15 000 euros d'amende, avec des peines aggravées lorsque le montage est à caractère sexuel ou diffusé par un service de communication en ligne. Le texte a été étendu aux contenus générés par intelligence artificielle par la loi du 21 mai 2024 dite SREN.

L'escroquerie est par ailleurs punie de cinq ans d'emprisonnement et de 375 000 euros d'amende par l'article 313-1. Source : CNIL, Hypertrucage (deepfake) : comment se protéger et signaler, mise à jour du 3 février 2026 — cnil.fr/fr/hypertrucage-deepfake.

Le droit européen impose une transparence, pas une protection

Les obligations de transparence de l'article 50 du règlement européen sur l'intelligence artificielle s'appliquent depuis le 2 août 2026 : les contenus synthétiques doivent être marqués de manière lisible par machine, et les hypertrucages signalés comme tels par ceux qui les déploient. Ces obligations pèsent sur les fournisseurs et les déployeurs de systèmes légitimes. Un attaquant ne marquera pas son contenu. La transparence organise le marché, elle ne défend pas une organisation. Source : Commission européenne, Quick facts: transparency rules for AI systems — digital-strategy.ec.europa.eu/en/factpages/quick-facts-transparency-rules-ai-systems.

La charge du contrôle reste sur l'organisation

C'est le point le plus important pour un dirigeant. Dans deux arrêts du 12 juin 2025, la chambre commerciale de la Cour de cassation a jugé que, lorsque les opérations de paiement ont été **autorisées** par une personne habilitée, seule la responsabilité contractuelle de la banque peut être recherchée au titre d'un éventuel manquement à son devoir de vigilance — et elle a cassé une décision qui avait retenu cette responsabilité sur le seul constat d'anomalies apparentes. Références : Cass. com., 12 juin 2025, n° 24-13.697 et n° 24-10.168. Source : Eurojuris, analyse du 21 août 2025 — eurojuris.fr/actualites-juridiques/actualites-juridiques-entreprises/droit-finances-entreprise/droit-bancaire-financier/articles/fraude-president-responsabilite-banque-43490.htm.

En pratique : une instruction exécutée par un collaborateur habilité reste une opération autorisée, même si elle a été obtenue par tromperie. L'existence d'un processus de contrôle interne documenté n'est donc pas une précaution de confort : c'est ce qui détermine où s'arrête la perte. Cette note ne constitue pas une consultation juridique ; l'analyse d'une situation précise relève d'un avocat.

À noter du côté des dispositifs collectifs : un service de vérification du bénéficiaire a été instauré en octobre 2025, et un fichier national des comptes signalés pour risque de fraude fonctionne depuis mai 2026. Ces mécanismes réduisent la fenêtre d'exécution, ils ne remplacent pas la vérification humaine. Source : Banque de France, rapport de l'Observatoire 2025.

6. Auto-évaluation en dix questions

À passer en revue avec la personne qui répond au téléphone, pas seulement avec la direction. Chaque réponse négative est un point d'entrée.

Question	Oui / Non
1 Toute instruction financière urgente reçue par téléphone est-elle vérifiée par un rappel sur un numéro connu à l'avance ?	
2 La liste des numéros de rappel de référence existe-t-elle, hors téléphone et hors boîte de courriels ?	

	Question	Oui / Non
3	Un délai minimal s'applique-t-il à toute demande inhabituelle, sans exception pour la direction ?	
4	Un mot de rappel existe-t-il entre le dirigeant et les personnes habilitées aux paiements ?	
5	Est-il écrit, et signé, que refuser une instruction non vérifiée ne peut pas être reproché ?	
6	Une personne de recours est-elle désignée et joignable hors voie hiérarchique ?	
7	Un changement de coordonnées bancaires d'un fournisseur déclenche-t-il une vérification sur un contact antérieur au changement ?	
8	Les tentatives déjouées sont-elles consignées quelque part ?	
9	Les proches et l'assistanat de direction, souvent visés en premier, connaissent-ils le protocole ?	
10	Le protocole a-t-il été testé au moins une fois en conditions réelles, sans prévenir ?	

Les questions 5 et 9 méritent une attention particulière. Une procédure que l'entourage immédiat ne connaît pas, ou qu'un collaborateur n'ose pas appliquer à son dirigeant, n'existe pas.

Méthode, limites et portée de cette note

- **Sources.** Uniquement des publications publiques et institutionnelles, citées avec leur adresse complète. Aucune donnée issue d'un dossier, d'un client ou d'un tiers n'a été utilisée.
- **Ce que les chiffres ne disent pas.** Les statistiques de la Banque de France mesurent la fraude aux moyens de paiement et la manipulation de la personne, pas l'usage d'hypertrucages en particulier. Aucune statistique publique française ne permet aujourd'hui d'isoler rigoureusement la part des attaques recourant à une voix ou à une image synthétique. Les chiffres qui circulent sur ce point précis proviennent le plus souvent de fournisseurs de solutions et n'ont pas été retenus ici.
- **Ce que cette note n'est pas.** Ni une consultation juridique, ni un audit de sécurité des systèmes d'information, ni une évaluation d'une situation particulière. Elle ne comporte aucune recherche sur une personne et aucune activité relevant du livre VI du code de la sécurité intérieure.
- **Réutilisation.** Ce document peut être diffusé et cité librement, à condition de ne pas en modifier le contenu et d'en mentionner l'origine.

À propos. Sentinel Reputation Office est une activité indépendante de conseil et de coordination en matière de réputation, de confidentialité et de préparation aux situations sensibles, basée à La Réunion et intervenant en français et en anglais. Les missions portent sur l'organisation, les procédures et la préparation — jamais sur la recherche d'informations concernant un tiers.

Contact : sentinel.reputation@gmail.com
Note rédigée par Alexandre Potonie, septembre 2026.