Pentest social engineering : vishing, prétexting et cadre légal

> Guide complet du pentest social engineering en France : cadre légal, phishing ciblé, vishing, prétexting, intrusion physique, deepfakes et restitution des...

Le facteur humain est impliqué dans 74 % des violations de données selon le rapport Verizon DBIR 2024. Aucune technologie, aussi sophistiquée soit-elle, ne peut compenser un employé qui transmet ses identifiants à un faux technicien IT au téléphone ou qui laisse entrer un inconnu derrière lui dans le datacenter. C'est pour cette raison que le pentest social engineering est devenu une composante essentielle de tout programme de sécurité offensif mature. Contrairement aux tests d'intrusion réseau ou applicatif, le social engineering test cible la couche humaine : il évalue la résistance des collaborateurs face à la manipulation psychologique, l'urgence artificielle, l'autorité simulée et la confiance abusée. En France, ce type d'exercice est encadré par des obligations légales strictes — conditions d'autorisation, périmètre défini, responsabilité du prestataire — que tout RSSI ou directeur sécurité doit maîtriser avant de lancer une campagne. Ce guide vous donne les clés opérationnelles pour commander, préparer, exécuter et valoriser un pentest social engineering dans le respect du droit français.

À RETENIR

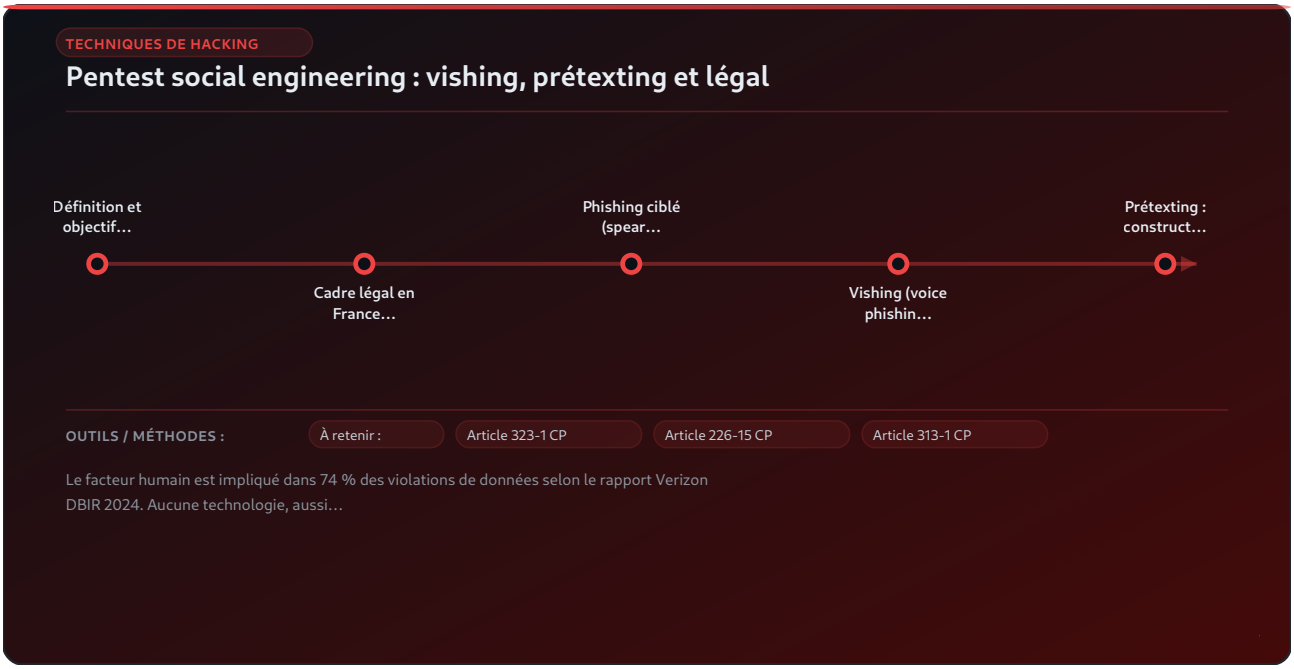
À retenir :

Tout test de social engineering en France nécessite une autorisation écrite préalable signée par le représentant légal de l'entreprise — sans cette autorisation, le prestataire est passible de poursuites pénales.

Le vishing (voice phishing) et le prétexting sont les vecteurs les plus efficaces en 2026 : taux de succès de 40 à 70 % selon les scénarios.

Les deepfakes voice sont désormais intégrés dans les campagnes de social engineering avancées — le clonage vocal d'un dirigeant en temps réel est accessible pour moins de 50 €/mois.

> La restitution doit protéger la dignité des employés testés : les résultats ne doivent jamais cibler des individus mais des processus et des formations à corriger.



Définition et objectifs du pentest social engineering

Un pentest social engineering est un test d'intrusion autorisé qui utilise des techniques de manipulation psychologique pour évaluer la résistance humaine d'une organisation à des attaques d'ingénierie sociale. Il diffère d'un simple test de phishing automatisé (envoi massif d'emails piégés) par sa sophistication, son ciblage et sa richesse de vecteurs d'attaque.

Les objectifs typiques d'une campagne :

Mesurer la sensibilisation des employés face aux techniques d'attaque réelles (pas seulement les simulations génériques).

Identifier les processus organisationnels vulnérables (vérification d'identité insuffisante, transferts d'informations sans validation).

Tester l'efficacité du programme de formation et de sensibilisation existant.

Obtenir des preuves concrètes (credentials volés, accès physique obtenu) pour justifier des investissements en formation.

Répondre à des exigences réglementaires (PCI-DSS, ISO 27001, NIS 2) qui recommandent le test humain.

La différence avec le phishing automatisé : une campagne GoPhish envoyant 500 emails identiques mesure un taux de clic statistique. Un pentest social engineering crée des scénarios personnalisés basés sur l'OSINT de la cible, utilise des vecteurs multiples (email + téléphone + physique) et simule un attaquant réel déterminé à obtenir un accès.

Cadre légal en France : conditions obligatoires

Le social engineering pentest opère dans un cadre légal strict en France. Plusieurs dispositions du Code pénal sont potentiellement applicables si le test n'est pas correctement autorisé.

⚡ **Retour terrain**

Dans mes missions de red team, la phase de post-exploitation révèle systématiquement des données que le client pensait protégées. Le cas le plus fréquent : des fichiers Excel de comptabilité ou RH stockés sur des partages réseau accessibles à tous les utilisateurs du domaine, sans restriction. Sur les 30 dernières missions, 27 avaient des partages réseau avec des données sensibles accessibles à n'importe quel utilisateur authentifié. La sensibilité des données n'est pas corrélée à leur niveau de protection réel.

Infractions potentiellement en jeu

Article 323-1 CP : accès frauduleux à un système informatique (STAD) — applicable si un faux employé extrait des données d'un système.

Article 226-15 CP : captation d'informations dans des correspondances privées — applicable si des emails ou messages sont interceptés ou récupérés sans consentement.

Article 313-1 CP : escroquerie — applicable si le prestataire obtient par tromperie la remise de biens, fonds ou informations confidentielles.

Article 441-1 CP : faux et usage de faux — applicable à la création de faux badges, faux documents d'identité ou faux emails d'entreprise.

Les conditions de légalité

Pour neutraliser ces risques pénaux, trois conditions sont impératives :

Autorisation écrite signée par le représentant légal (PDG, DG) de l'organisation cliente, décrivant précisément le périmètre du test, les techniques autorisées et la période d'exécution. Cette autorisation est la pièce fondamentale que le prestataire doit conserver.

Périmètre défini : liste des personnes ciblées (ou des services/populations cibles), des vecteurs autorisés, des informations que le prestataire est autorisé à exploiter. Si le test inclut de l'intrusion physique, les locaux concernés doivent être explicitement listés.

Clause de non-divulgation : le prestataire est tenu à la confidentialité des informations obtenues pendant le test et doit les détruire après restitution.

La CNIL doit être considérée : si des données personnelles d'employés sont collectées dans le cadre du test (identifiants, comportements), une analyse de légitimité au regard du RGPD est recommandée. L'employeur doit informer les instances représentatives du personnel (CSE) de la mise en œuvre de dispositifs de surveillance ou d'évaluation — même si la date exacte du test peut rester confidentielle.

Phishing ciblé (spear phishing) : préparation et exécution

Le spear phishing est la forme d'ingénierie sociale la plus répandue dans les campagnes APT réelles. Contrairement au phishing générique, il exploite des informations personnalisées sur la cible pour créer un email extrêmement convaincant.

Phase OSINT : reconnaissance de la cible

Avant d'envoyer le moindre email, le pentesteur consacre plusieurs heures à la collecte d'informations en sources ouvertes (OSINT) :

LinkedIn : organigramme, rôles, relations fournisseurs, projets en cours mentionnés dans les posts.

Site web de l'entreprise : mentions de partenaires, de logiciels utilisés, d'événements récents.

WHOIS et DNS : infrastructure technique, sous-domaines, prestataires d'hébergement.

Haveibeenpwned : adresses email exposées dans des fuites antérieures.

Google Dorks : documents internes indexés accidentellement, formats d'adresses email.

Ces informations permettent de construire un email de phishing qui mentionne un projet réel, un collègue par son prénom, un outil réellement utilisé par l'entreprise.

Infrastructure de phishing

```
 # Installation GoPhish
wget https://github.com/gophish/gophish/releases/download/v0.12.1/gophish-v0.12.1-linux-64bit.zip
unzip gophish-*.zip && chmod +x gophish

# Démarrage (interface admin sur https://localhost:3333)
./gophish
```

L'infrastructure de phishing légitime inclut : un domaine typosquatté (ex. `rh-votreentreprise.fr` au lieu de `votreentreprise.fr`), un certificat TLS valide (Let's Encrypt), un serveur de phishing configuré avec GoPhish, un site de leurre crédible hébergé sur ce domaine.

Les métriques à mesurer : taux d'ouverture de l'email, taux de clic sur le lien, taux de soumission de credentials, temps moyen entre réception et clic, taux de signalement à l'équipe sécurité.

Vishing (voice phishing) : techniques et guide de conversation

Le vishing consiste à appeler la cible en se faisant passer pour un tiers de confiance (IT helpdesk, fournisseur, auditeur, dirigeant). C'est le vecteur avec le taux de succès le plus élevé en test humain : 40 à 70 % selon les scénarios et la préparation.

Scénarios types

Scénario IT Helpdesk : « Bonjour, je suis Thomas du support informatique. On a détecté une activité suspecte sur votre compte. Pour sécuriser votre compte, j'ai besoin de vérifier votre identité. Pouvez-vous me confirmer votre identifiant de connexion ? Je vais vous envoyer un code par SMS pour validation. » Ce scénario exploite l'autorité (IT) et l'urgence (sécurité compromise).

Scénario faux fournisseur : « Bonjour, je suis Antoine de chez [nom d'un vrai fournisseur récupéré via OSINT]. On vous a envoyé une facture le mois dernier et elle n'a pas été payée. Pouvez-vous me confirmer les coordonnées bancaires de votre service comptabilité pour régulariser ? »

Scénario audit externe : « Bonjour, je suis consultant mandaté par votre direction pour un audit de conformité ISO 27001. J'ai besoin d'accéder à votre portail interne pour vérifier quelques paramètres. Vous pouvez me créer un accès temporaire ? »

Outils de spoofing caller ID légaux

Le spoofing de numéro de téléphone est légalement autorisé dans le cadre d'un test avec autorisation écrite. Des services comme SpoofCard (usage légal avec autorisation) permettent

d'afficher un numéro choisi à l'appelé. Des solutions VOIP professionnelles (Twilio, Vonage) permettent également de configurer l'identifiant appelant.

Prétexting : construction d'un persona crédible

Le prétexting est la création et le maintien d'une identité fictive (persona) sur la durée d'une campagne. C'est la technique la plus sophistiquée du social engineering car elle combine plusieurs vecteurs sur plusieurs semaines.

Construction d'un persona :

Identité complète : nom, prénom, titre, entreprise fictive ou réelle.

Présence LinkedIn créée 2 à 4 semaines avant la campagne (pour paraître crédible dans les recherches).

Adresse email d'un domaine typosquatté de l'entreprise partenaire simulée.

Documents d'appui : faux contrat, faux bon de commande, fausse plaquette commerciale.

Le persona établit d'abord une relation de confiance (emails anodins, échanges LinkedIn) avant de passer à la phase d'exploitation (demande d'accès, envoi de fichier piégé, invitation à une réunion Teams avec lien malveillant).

Intrusion physique : tailgating et USB drop

Le physical pentest est le volet le plus risqué légalement et doit être encadré de façon très précise dans l'autorisation écrite. Les techniques testées :

Tailgating / Piggybacking : suivre un employé dans une zone à accès contrôlé sans badge, en profitant d'une porte maintenue ouverte. Taux de succès typique : 60-80 % dans les grands open spaces.

USB drop : déposer des clés USB piégées sur le parking ou dans les espaces communs. Test de si un employé branchera une clé USB inconnue. Solution propre : USB ne contenant que GoPhish payload inoffensif ou simple beacon d'alerte.

Faux badge : se présenter avec un badge d'apparence officielle (imprimé) pour tenter d'accéder à des zones restreintes. Légalement très encadré — nécessite l'accord explicite de la direction.

Le prestataire doit porter sur lui en permanence la lettre de mission signée par le client et un contact téléphonique direct du RSSI pour stopper le test à tout moment.

Deepfakes voice et video en social engineering 2026

En 2026, la technologie de clonage vocal est accessible pour moins de 50 € par mois (ElevenLabs, Murf, PlayHT). À partir de 3 minutes d'audio d'une personne (une interview YouTube, une conférence publique), un modèle vocal est généré capable d'imiter la voix avec un taux de reconnaissance de 85 % par les auditeurs humains.

Les cas réels documentés en 2024-2025 incluent plusieurs tentatives de fraude au président utilisant le clonage vocal du PDG pour convaincre un directeur financier d'effectuer un virement urgent. Un cas confirmé en Asie du Sud-Est a permis de détourner 25 millions USD via une visioconférence entièrement générée par deepfake video.

Dans les pentests avancés, le deepfake voice permet de simuler l'appel d'un dirigeant de l'organisation cliente pour tester la résistance des processus de validation des demandes urgentes. Cette technique doit être explicitement autorisée et utilisée avec extrême précaution pour ne pas créer de traumatisme chez les employés ciblés. Notre article sur les techniques de phishing AiTM et bypass MFA complète ce panorama des vecteurs d'attaque avancés.

Métriques et benchmark sectoriel

Les métriques standards d'un pentest social engineering :

Taux de clic phishing : % d'employés ayant cliqué sur le lien. Benchmark sectoriel moyen : 18-25 %. Secteur santé : 35-45 %. Secteur finance après formation : 8-12 %.

Taux de credential submission : % ayant entré leurs identifiants. Toujours inférieur au taux de clic. Benchmark : 8-15 %.

Taux de succès vishing : % d'appels aboutissant à l'objectif défini. Benchmark : 30-60 % selon le scénario.

Taux de signalement : % d'employés ayant signalé l'attaque simulée à l'équipe sécurité. Indicateur clé : un taux élevé (>30 %) indique une bonne culture sécurité.

Temps de détection : délai entre le lancement de la campagne et la première alerte remontée au SOC.

Rapport et restitution : protéger la dignité des employés

La restitution d'un pentest social engineering est un exercice délicat. Les résultats montrent des employés qui ont « échoué » — si mal présentés, les résultats peuvent créer des tensions et une culture de méfiance plutôt qu'une amélioration. Les principes de restitution recommandés :

Jamais de nommage individuel dans le rapport principal — les résultats sont présentés par service, par niveau hiérarchique, par scénario.

Cadrage positif : « 73 % des employés ont résisté au scénario de vishing » plutôt que « 27 % ont fourni leurs identifiants ».

Recommandations actionnables : chaque vulnérabilité identifiée doit être associée à une correction concrète (processus, formation, outil).

Session de sensibilisation post-test : présenter les résultats anonymisés à tous les employés, avec les techniques utilisées et comment les reconnaître.

Le rapport final doit inclure : résumé exécutif, chronologie de la campagne, métriques par vecteur, captures d'écran anonymisées, recommandations priorisées (quick wins vs projets longs). Notre guide sur le pentest selon le standard PTES 2026 fournit la structure de rapport de référence.

Programme de sensibilisation post-test

Le test n'a de valeur que s'il est suivi d'actions correctives. Un programme de sensibilisation efficace post-social engineering comprend :

E-learning personnalisé : modules courts (5-10 minutes) ciblés sur les techniques qui ont fonctionné lors du test.

Simulations régulières : 4 à 6 campagnes de phishing simulé par an, avec variation des scénarios pour éviter l'habituation.

Formation réflexe vishing : procédure simple de rappel (call back) pour valider l'identité de tout appelant demandant des informations sensibles ou des accès.

Indicateur de performance : mesurer l'évolution du taux de clic et du taux de signalement entre les campagnes pour démontrer le ROI de la formation.

L'audit de cybersécurité PME intègre le social engineering test comme composante de l'évaluation globale de la posture de sécurité humaine.

FAQ — Questions fréquentes

Le prestataire qui réalise le test est-il légalement protégé en France ?

Oui, à condition que l'autorisation écrite préalable soit en règle. Sans cette autorisation, le prestataire qui, par exemple, récupère des identifiants d'employés via un faux portail ou accède physiquement à des zones protégées est susceptible d'être poursuivi pour accès frauduleux à un

STAD (article 323-1 CP), escroquerie ou faux et usage de faux. L'autorisation écrite doit être signée par le représentant légal habilité à engager la responsabilité de l'organisation, mentionner explicitement les techniques de social engineering autorisées et la période de test. Le prestataire doit conserver ce document pendant toute la durée de la mission et le présenter si nécessaire. En cas d'incident (employé qui appelle la police par exemple), le prestataire doit être en mesure de stopper le test immédiatement et d'exhiber son autorisation.

Faut-il informer les employés qu'un test de social engineering va être réalisé ?

Non, dans la grande majorité des cas les employés ne sont pas informés à l'avance — l'information préalable annule l'intérêt du test. Cependant, les instances représentatives du personnel (CSE en France) doivent être informées de la mise en place d'un dispositif de surveillance ou d'évaluation, même si les dates précises restent confidentielles. Certaines organisations choisissent d'informer un cercle très restreint (RSSI, DRH, représentant légal) pour gérer les éventuels impacts RH (employé en détresse émotionnelle après avoir été manipulé). La bonne pratique est de prévoir un protocole de soutien pour les employés qui seraient particulièrement affectés par la découverte qu'ils ont été testés. La session de debriefing collectif post-test, où toutes les techniques sont expliquées avec bienveillance, est cruciale pour transformer l'expérience en apprentissage positif. Pour aller plus loin sur la protection contre les attaques BEC réelles, consultez notre article sur les fraudes BEC et défenses email. Les ressources de social-engineer.org et la documentation de GoPhish complètent ce guide opérationnel.

Que faire si un employé coopère pleinement et fournit des accès critiques pendant le test ?

C'est un scénario fréquent et qui doit être géré avec précaution. Si un employé fournit des identifiants d'administrateur, ouvre une porte vers un datacenter ou transfère des fonds lors d'un test vishing, le prestataire doit immédiatement arrêter l'exploitation et ne pas aller plus loin que ce qui est strictement nécessaire à la démonstration. Les informations obtenues (credentials, accès physiques) doivent être immédiatement signalées au RSSI et documentées sans être exploitées davantage. L'employé concerné ne doit pas être sanctionné : il a réagi comme la majorité des gens le ferait face à un scénario bien construit. La réponse organisationnelle doit

être de corriger le processus qui a permis cette vulnérabilité (manque de procédure de vérification, absence de double validation pour les transferts), pas de punir l'individu.

Bonnes pratiques et recommandations complémentaires

Au-delà des techniques et outils présentés dans cet article, plusieurs principes transverses guident les professionnels de la cybersécurité dans leur approche quotidienne. La défense en profondeur (*defense-in-depth*) reste le principe fondateur : aucune mesure de sécurité unique n'est suffisante, et la multiplication des couches de protection — même imparfaites individuellement — crée une résilience globale supérieure à la somme de ses parties.

Veille et mise à jour continue

La cybersécurité est un domaine où l'obsolescence est rapide. Une technique ou un outil efficace en 2024 peut être contourné en 2026. Les équipes sécurité maintiennent leur efficacité en s'appuyant sur des sources de veille fiables : bulletins CERT-FR et ANSSI, advisories des éditeurs (Microsoft MSRC, Google Project Zero, Cisco Talos), recherches académiques (USENIX Security, IEEE S&P, CCS), et publications de la communauté (threat intel reports des grands éditeurs, articles de blog de chercheurs reconnus).

Documentation et partage de connaissances

La capitalisation des connaissances est un enjeu organisationnel critique dans les équipes de sécurité. Les runbooks d'investigation, les post-mortems d'incidents, les procédures de réponse documentées, et les bases de connaissance internes permettent de maintenir la cohérence des pratiques indépendamment des rotations d'équipe et de réduire le temps de résolution des incidents récurrents. L'utilisation d'un wiki sécurisé (Confluence, Notion avec contrôles d'accès stricts) pour centraliser ces connaissances est une pratique adoptée par la majorité des équipes SOC matures. La documentation proactive, rédigée juste après les incidents pendant que les

détails sont frais, est systématiquement plus précise et utile que la documentation rédigée après coup.

Synthèse et perspectives 2026

Les techniques et recommandations présentées dans ce guide s'inscrivent dans un contexte de menaces en constante évolution. La cybersécurité offensive et défensive sont deux faces d'une même médaille : comprendre les mécanismes d'attaque est indispensable pour construire des défenses robustes et résilientes face aux acteurs malveillants les plus sophistiqués.

Pour les équipes sécurité, l'enjeu de 2026 est double : maintenir une veille continue sur les nouvelles techniques publiées par la communauté de recherche (CVE, exploit-db, GitHub, Secrech, SSTIC) tout en assurant le durcissement progressif de l'infrastructure existante. Le référentiel MITRE ATT&CK reste le fil conducteur le plus efficace pour structurer un programme de détection et de réponse face aux tactiques, techniques et procédures des groupes APT ciblant les secteurs critiques.

La formation continue des équipes, la simulation régulière d'incidents (exercices tabletop, exercices Red/Blue/Purple Team), et l'automatisation des tâches répétitives via des outils SOAR constituent les piliers d'une organisation cyber mature. Les organisations qui investissent dans ces trois axes démontrent systématiquement de meilleures métriques de détection et de réponse (MTTD et MTTR réduits de 40% en moyenne selon les benchmarks sectoriels) face aux incidents de sécurité.

Sources et références

[MITRE ATT&CK — Tactiques, techniques et procédures offensives]

OWASP — Top 10 des risques applicatifs

ANSSI — Panorama de la cybermenace 2024