

FORMATION 05

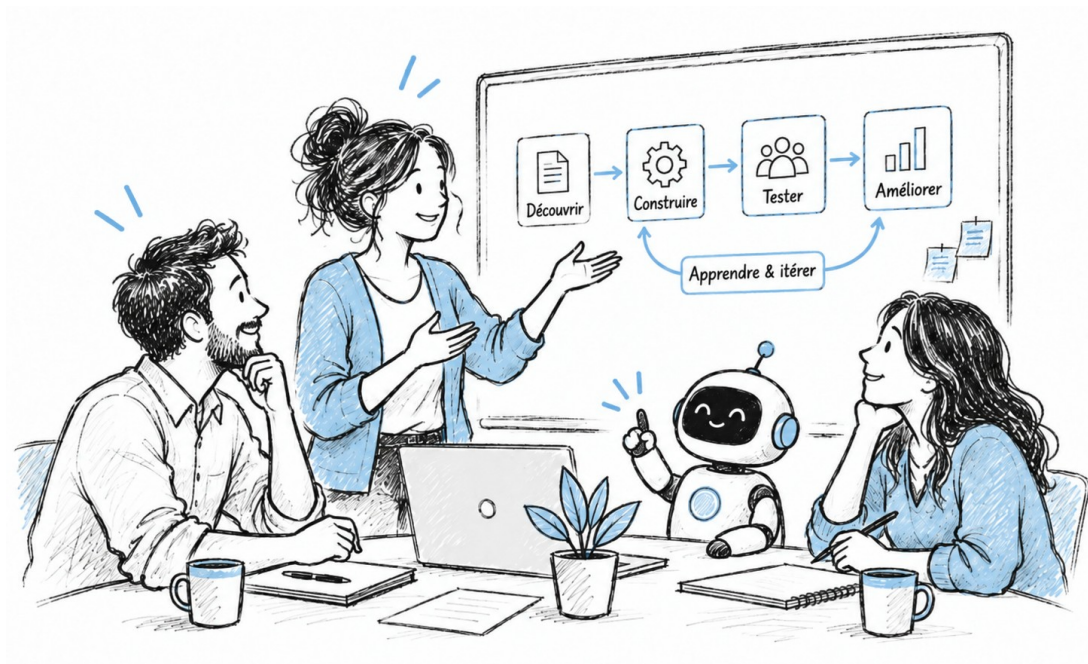
Faire une place aux agents sans leur donner les clés de l'entreprise

Plus autonome ne doit jamais vouloir dire incontrôlé.

Comprendre ce qu'est réellement un agent IA, déterminer lorsqu'il est pertinent, puis concevoir sa mission, ses outils, ses permissions, ses limites et ses points de passage vers l'humain.

DURÉE 7 h	GROUPE 5 à 10 participants	PRÉREQUIS Cas d'usage cadré et processus connu
--------------	-------------------------------	---

Concevoir un agent utile, limité, testable et supervisé avant de lui confier la moindre clé.



Finalité, public et objectifs

Finalité de la formation

Cette formation aide les entreprises à comprendre ce qu'est réellement un agent IA et à décider lorsqu'une architecture agentique est pertinente. Les participants cadrent une mission, définissent les outils et données nécessaires, limitent les permissions et organisent les passages vers l'humain.

L'objectif n'est pas de construire un agent en production pendant la journée. Il est de produire une conception suffisamment précise pour être évaluée, transmise à une équipe technique puis testée dans un pilote progressif et réversible.

Résultat attendu

À l'issue de la journée, le groupe dispose d'une fiche de conception agentique complète : mission, périmètre, outils, permissions, autonomie, validations, risques, tests, supervision, conditions d'arrêt et plan de pilote.

Public, prérequis et modalités

Rubrique	Contenu
Public concerné	Dirigeants, responsables produit, responsables métier ou techniques, chefs de projet, développeurs, intégrateurs et équipes automatisation.
Prérequis	Connaissance du processus étudié, de ses principales données, de ses outils et de ses responsabilités humaines. Aucun prérequis en développement.
Taille du groupe	5 à 10 participants recommandés ; jusqu'à 12 personnes pour un cas suffisamment préparé.
Modalités	1 journée de 7 heures. Intra-entreprise recommandé. Présentiel ou visioconférence selon le contexte.
Préparation	Questionnaire en amont, workflow existant, liste des outils, règles métier, exceptions connues et exemples anonymisés.
Répartition pédagogique	Environ 40 % d'apports et 60 % d'ateliers, de conception et de restitution.

Objectifs pédagogiques

- Distinguer chatbot, assistant, workflow automatisé et agent.
- Déterminer si un agent est réellement nécessaire.
- Formuler une mission précise, mesurable et délimitée.
- Identifier les données, outils, permissions et actions interdites.
- Définir un niveau d'autonomie progressif et les passages vers l'humain.
- Identifier les principaux risques, garde-fous, tests et conditions d'arrêt.
- Produire un plan de pilote et désigner un responsable humain.

Déroulé de la journée

La journée alterne apports méthodologiques, analyse d'un cas commun et conception d'un agent appliqué à une situation réelle ou réaliste.

Matin - Comprendre et cadrer

Horaire	Séquence	Objectif
09:00-09:25	Accueil, représentations et cadrage	Faire émerger les attentes et clarifier les premières représentations du mot agent.
09:25-10:10	Chatbot, workflow, assistant ou agent ?	Distinguer les architectures et comprendre ce qui caractérise réellement un agent.
10:10-10:50	Le problème nécessite-t-il un agent ?	Évaluer si un workflow classique ou une assistance IA suffirait.
10:50-11:05	Pause	-
11:05-11:50	Mission et critères de réussite	Transformer une intention vague en mission testable et mesurable.
11:50-12:30	Données, outils et permissions	Définir les ressources utiles et appliquer le principe d'accès minimal.

Après-midi - Encadrer l'autonomie et préparer le pilote

Horaire	Séquence	Objectif
13:30-14:15	Autonomie et passages vers l'humain	Choisir un niveau d'autonomie et les situations nécessitant une validation ou une escalade.
14:15-14:50	Étude d'un cas commun	Concevoir un agent de préparation commerciale et cadrer ses limites.
14:50-15:05	Pause	-
15:05-15:55	Atelier de conception	Produire la fiche complète d'un agent appliqué à un cas métier.
15:55-16:35	Risques, supervision, tests et arrêt	Prévoir les garde-fous, journaux, scénarios dégradés et conditions de suspension.
16:35-16:55	Restitution et validation	Présenter la conception, justifier les choix et vérifier les acquis.
16:55-17:00	Conclusion	Identifier les prochaines étapes vers un prototype ou un pilote.

Contenu pédagogique - première partie

1. Chatbot, assistant, workflow ou agent

Architecture	Fonction principale	Exemple
Chatbot	Répond à une demande et produit un échange.	Répondre à une question sur un catalogue de produits.
Assistant	Prépare ou facilite le travail d'une personne.	Résumer les derniers échanges avant un rendez-vous.
Workflow automatisé	Suit un parcours largement prédéfini.	Valider un formulaire, créer un dossier et notifier une équipe.
Agent	Choisit dynamiquement parmi plusieurs actions ou outils.	Analyser un ticket, explorer le code, lancer des tests et préparer une proposition.
Système multi-agent	Répartit plusieurs missions entre des agents spécialisés.	Coordonner un agent front, un agent back et une revue technique.

Principe : lorsque les étapes peuvent être décrites simplement à l'avance, un workflow classique reste généralement plus facile à tester, expliquer et maintenir.

2. Savoir si le problème nécessite réellement un agent

Question	Interprétation
La tâche varie-t-elle réellement ?	Si toutes les entrées suivent le même chemin, un workflow suffit peut-être.
Les actions peuvent-elles être prédéfinies ?	Une automatisation déterministe sera plus prévisible.
Le système doit-il choisir entre plusieurs outils ?	Cette capacité peut justifier une logique agentique.
Le contexte influence-t-il la stratégie ?	Un agent devient pertinent lorsque le parcours dépend des informations découvertes.
Le gain compense-t-il l'incertitude ?	Un agent demande davantage de tests, de supervision et d'explicabilité.
Les conséquences d'une erreur sont-elles acceptables ?	Plus le risque est élevé, plus l'autonomie doit être limitée.

3. Définir la mission

- Déclencheur et résultat attendu.
- Périmètre, exclusions et cas hors mission.
- Données accessibles et outils autorisés.
- Actions permises, interdites ou soumises à validation.
- Conditions d'escalade, responsable humain et critères de réussite.

Exemple de mission

Identifier les opportunités sans activité depuis plus de sept jours, réunir le contexte des derniers échanges et préparer une synthèse pour le commercial responsable. Aucun message n'est envoyé sans validation.

Contenu pédagogique - deuxième partie

4. Données, outils et permissions

L'agent reçoit uniquement les ressources nécessaires à sa mission. Les droits sont distingués afin d'éviter qu'un accès en lecture devienne implicitement un droit de modification ou d'exécution.

Ressource	Lire	Créer	Modifier	Exécuter	Supprimer
CRM	Oui	Fiche limitée	Champs autorisés	Non	Non
Emails	Historique utile	Brouillon	Non	Après validation	Non
Fichiers	Périmètre limité	Oui	Limité	-	Non
Production	Logs uniquement	Non	Non	Non	Non

Principes de permissions

- Accès minimal et comptes dédiés.
- Secrets séparés du workflow et environnement de test.
- Révocation simple et journalisation des actions.
- Aucune permission implicite ni droit accordé par confort.

5. Échelle d'autonomie

Niveau	Comportement	Exemple
0 - Observer	Analyse sans produire d'action visible.	Repérer les dossiers nécessitant potentiellement une relance.
1 - Recommander	Formule une proposition.	Suggérer les opportunités à reprendre.
2 - Préparer	Produit un brouillon ou rassemble le contexte.	Préparer une synthèse et un message.
3 - Agir après validation	Exécute uniquement après confirmation humaine.	Envoyer un message approuvé.
4 - Agir seul dans un cadre limité	Traite les cas standard et escalade les autres.	Mettre à jour certains champs à faible risque.

6. Passages vers l'humain

- Confiance insuffisante ou donnée essentielle absente.
- Action irréversible, montant élevé ou relation sensible.
- Responsabilité de l'entreprise engagée.
- Situation inhabituelle, règles contradictoires ou outil inconnu.
- Contenu cherchant à détourner la mission de l'agent.

Cas pratique et atelier de conception

Cas commun : agent de préparation commerciale

Chaque matin, l'agent repère les opportunités dont le suivi paraît en retard. Il consulte le CRM et les échanges, rassemble le contexte et propose une action au commercial responsable.

Peut faire	Ne peut pas faire seul
Consulter le CRM et les derniers échanges.	Envoyer un email à un prospect.
Identifier les opportunités sans activité récente.	Modifier un prix ou promettre une échéance.
Préparer une synthèse et un brouillon.	Supprimer une opportunité.
Proposer une prochaine action.	Changer le responsable du dossier.
Journaliser ses recherches et propositions.	Contacteur un prospect signalé comme sensible.

Progression possible du pilote

Étape	Niveau d'autonomie
Pilote initial	L'agent observe et recommande.
Étape suivante	Il prépare les synthèses et les brouillons.
Autonomie limitée	Il met à jour certaines informations internes après validation.
Frontière conservée	La relation avec le prospect reste confiée au commercial.

Atelier de conception

Étape	Production attendue
1. Choisir le problème	Un cas délimité, fréquent, variable et mesurable.
2. Justifier l'agent	Pourquoi un workflow fixe ou une simple assistance ne suffit pas.
3. Définir la mission	Déclencheur, objectif, résultat, périmètre et exclusions.
4. Définir les moyens	Données, outils, permissions et environnement.
5. Définir les limites	Actions interdites, seuils, validations, escalades et arrêt.
6. Définir la supervision	Journaux, responsable, indicateurs, alertes et reprise.
7. Préparer le pilote	Durée, utilisateurs, volume, autonomie, succès et conditions d'arrêt.

Variante technique

Le même atelier peut être appliqué à un agent front, un agent back ou un agent de revue : tickets, repositories, tests, pull requests, secrets, base de données et déploiement sont alors étudiés comme autant de permissions distinctes.

Risques, garde-fous et supervision

1. Principaux modes de défaillance

Risque	Conséquence possible	Réponse attendue
Mission mal formulée	Objectif poursuivi trop largement.	Réduire le périmètre et reformuler les critères de réussite.
Mauvais outil ou mauvaise donnée	Action incohérente ou résultat trompeur.	Valider les entrées, limiter les outils et vérifier les sorties.
Action excessive	Modification au-delà de l'intention.	Permissions minimales, seuils et validation.
Boucle ou répétition	Coût, doublons ou blocage.	Budget d'actions, limite de temps et arrêt automatique.
Instruction détournée	L'agent ignore ou contourne sa mission.	Isolation du contenu, outils limités et contrôle des instructions.
Escalade absente	Le système poursuit alors qu'il devrait s'arrêter.	Conditions explicites de passage vers l'humain.
Défaillance silencieuse	Personne ne sait que la mission a échoué.	Journaux, alertes et responsable identifié.

2. Garde-fous à combiner

- Périmètre technique, listes autorisées et accès minimal.
- Permissions séparant lecture, création, modification, exécution et suppression.
- Validation humaine avant les actions sensibles ou irréversibles.
- Seuils de montant, volume, fréquence, destinataires ou durée.
- Formats de sortie contrôlés et vérification par des règles déterministes.
- Budget d'actions, mécanisme d'arrêt et journalisation complète.

3. Tests minimum

Scénario	Comportement attendu
Demande normale	Mission réalisée dans le périmètre défini.
Demande ambiguë	Question ou escalade, pas d'invention silencieuse.
Information manquante	Arrêt ou demande de complément.
Outil indisponible	Nouvelle tentative contrôlée ou escalade.
Permission refusée	Aucune tentative de contournement.
Action déjà réalisée	Pas de duplication de la conséquence.
Instruction contradictoire ou détournée	Mission initiale et règles de sécurité maintenues.
Seuil dépassé	Validation humaine obligatoire.
Demande hors périmètre	Refus explicite et journalisé.

4. Supervision et conditions d'arrêt

Un responsable humain consulte les journaux, reçoit les alertes, ajuste les limites et peut suspendre l'agent. Le pilote doit être arrêté en cas d'action inattendue, de hausse des erreurs, de permissions compromises, de coûts incontrôlés ou de perte de traçabilité.

Méthodes pédagogiques et validation des acquis

Méthodes pédagogiques

- Apports structurés et exemples issus de situations d'entreprise.
- Grille de pertinence agentique et matrice des permissions.
- Étude d'un cas commun puis conception en sous-groupes.
- Restitution orale, challenge croisé et reformulation des limites.
- Production de supports réutilisables pour un futur prototype ou pilote.

Critères de validation

Compétence	Manifestation attendue
Distinguer les architectures	Le participant différencie chatbot, workflow, assistance et agent.
Justifier l'agent	Il explique pourquoi un workflow fixe ne suffit pas.
Définir la mission	Objectif, périmètre, résultat et exclusions sont clairs.
Encadrer les accès	Données, outils et permissions sont limités.
Positionner l'humain	Validations, escalades et responsabilité sont explicites.
Identifier les risques	Les défaillances principales et leurs garde-fous sont documentés.
Tester	Des scénarios nominaux et dégradés sont prévus.
Superviser	Un responsable, des journaux et des conditions d'arrêt existent.
Préparer le pilote	Le niveau d'autonomie et les critères de succès sont réalistes.

Modalités d'évaluation

- Observation pendant les ateliers.
- Fiche de mission et matrice des permissions.
- Carte d'autonomie et passages vers l'humain.
- Registre des risques et plan de tests.
- Restitution orale et questionnaire de synthèse.
- Auto-évaluation du participant.

Livrables remis

Livrable	Contenu
Fiche de pertinence	Aide à décider si un agent est justifié.
Fiche de mission	Objectif, déclencheur, périmètre, exclusions et responsable.
Matrice des outils et permissions	Ressources, droits, limites et validations.
Échelle d'autonomie	Niveau de départ et conditions d'évolution.
Carte des passages vers l'humain	Validations, escalades, seuils et cas sensibles.
Registre des risques	Conséquences, détection, garde-fous et reprise.
Plan de tests	Scénarios nominaux, erreurs, détournements et arrêts.
Plan de pilote	Durée, périmètre, utilisateurs, indicateurs et conditions d'arrêt.
Fiche de supervision	Journaux, alertes, responsable et fréquence de revue.

Cadre, limites et prolongements

Ce que la formation ne couvre pas

- Construction technique complète ou déploiement en production.
- Architecture multi-agent avancée et connexion aux systèmes de production.
- Audit de cybersécurité, audit juridique ou sélection contractuelle d'un fournisseur.
- Rédaction exhaustive des prompts, maintenance ou supervision continue.
- Configuration détaillée des modèles, mémoire, coûts et infrastructure.

Limite assumée

La formation produit une conception agentique encadrée et transmissible. La construction du prototype, sa sécurité et sa mise en service nécessitent une étape complémentaire adaptée au niveau de risque.

Prolongements possibles

Prolongement	Contenu
Variante technique - 1 journée	Passage de la fiche de conception à un prototype limité : outils, fonctions, contexte, permissions, validation, journaux et tests.
Atelier de revue - 2 h 30 à 3 h	Analyse des premières traces, erreurs, escalades et permissions après un prototype ou un pilote.
Accompagnement au prototype	Conception technique, environnement isolé, intégrations, tests et documentation.

Articulation avec le parcours RIKŌ

Formation	Contribution
Formation 01	Sélectionner un cas qui mérite d'être automatisé.
Formation 02	Préparer l'intégration de l'agent dans le collectif de travail.
Formation 03	Formaliser le processus, les règles et les responsabilités.
Formation 04	Apporter les bases de fiabilité, de test et de reprise.
Formation 05	Définir la mission, les permissions et l'autonomie de l'agent.
Formation 06	Organiser la maintenance, la documentation et la gouvernance dans la durée.

Résultat attendu

À l'issue de la journée, les participants disposent d'un cas justifié, d'une mission délimitée, d'outils et permissions contrôlés, d'un niveau d'autonomie initial, de passages vers l'humain, d'un responsable, d'un plan de tests et d'un périmètre pilote.

Un projet agentique à cadrer ?

Échangeons sur votre contexte : riko-automatisation.fr/contact