

Les demandes clients récurrentes

LE WORKFLOW, TEL QU'IL SE FAIT AUJOURD'HUI

≈ 3 000 demandes par mois · ≈ 60 % portent sur trois motifs · ≈ 1 jour de délai de première réponse (ordres de grandeur illustratifs)

● point de friction



AVANT DE COMMENCER

Préparer les sources.

Valider les réponses types, résoudre les contradictions, attribuer un propriétaire à chaque document et définir les données accessibles. L'outil vient ensuite.

ET TECHNIQUEMENT ?

Une architecture mixte : données de commande lues par API, en lecture seule et limitées au client authentifié ; recherche hybride (vecteurs + BM25) avec reranking dans les conditions, la FAQ et les modes d'emploi ; LLM instruit de s'appuyer sur ces éléments et d'en citer la source ; garde-fous d'entrée et de sortie avec routage vers les conseillers. Boucle d'apprentissage : journal, revue métier, jeu d'évaluation. Schéma et contrôles en page 2.

ÉTAPE 1 L'HUMAIN DANS LA BOUCLE

L'assistant propose, le conseiller décide

- 1 La demande arrive — dans la même file qu'aujourd'hui.
- 2 L'assistant prépare un projet de réponse — avec la source ou la commande citée.
- 3 Un conseiller décide — il relit, corrige si besoin, puis envoie.
- 4 Chaque correction est notée — motif par motif.

Ce que l'étape produit : une mesure, motif par motif, de la qualité des réponses proposées. Durée indicative : 8 à 12 semaines.

ÉTAPE 2 L'ASSISTANT PREND LE STANDARD

Il répond sur les motifs autorisés, les conseillers prennent les exceptions

RÉPOND LORSQUE

- La demande relève d'un motif autorisé
- Le client est authentifié et la donnée est à jour
- La réponse peut citer sa source ou la commande

TRANSMET AU CONSEILLER LORSQUE

- Il s'agit d'une réclamation, d'un litige ou d'un cas ambigu
- La source ou la donnée est absente, contradictoire ou ancienne
- La demande touche notamment un remboursement, un geste commercial, un dommage ou la sécurité d'un produit

Chaque semaine, le service client relit un échantillon de réponses et tient les sources à jour.

POURQUOI ICI, ET À QUELLE CONDITION

Répété souvent · Porteur clair · Livrable net · Irritant ressent

- Des demandes d'information statut de commande, conditions, mode d'emploi.
- Des données de commande consultées en lecture seule, client authentifié.
- Aucune décision commerciale ni remboursement ni geste confié à l'IA.
- Les réclamations, litiges et cas sensibles transmis à un conseiller.

Le risque ne disparaît pas : il devient visible, mesurable et gouvernable.

PASSER DE L'ÉTAPE 1 À L'ÉTAPE 2

Décidé motif par motif, jamais à l'avance, sur ce que le service client observe durant l'étape 1 :

- Exactitude de la réponse et de la source citée
- Corrections nécessaires, par motif
- Transmission correcte des exceptions
- Pas d'incident significatif observé sur le périmètre testé
- Des sources à jour, chacune avec un propriétaire

Le statut de commande peut passer à l'étape 2 pendant que les modes d'emploi restent à l'étape 1.

CE QUE L'ON CHERCHE, ET COMMENT ON LE MESURE

Non pas remplacer les conseillers, mais réduire le délai sur les demandes d'information, éviter les réponses réécrites et rendre du temps pour les cas qui demandent un jugement.

Délai de première réponse sur une demande standard

≈ 1 jour → ≈ 10 min

aujourd'hui (illustratif) → cible à l'étape 2 · suivi aussi : la part des demandes traitées sans conseiller, 40 à 50 % (cible illustrative)

Architecture de référence : comment le dispositif reste contrôlé



CONTRÔLE ET SÉCURITÉ

Évaluation

Jeu de référence d'environ 150 demandes réelles anonymisées. Métriques : exactitude, fidélité aux sources et aux données, citation correcte, taux de détection et de bonne transmission des exceptions (rappel du routage). Rejoué à chaque changement de modèle, de prompt ou de sources.

Observabilité

Pour chaque requête, journal des seuls éléments nécessaires : demande minimisée, références des données et passages récupérés, réponse, signaux de contrôle, décision de routage. Accessible aux seules personnes autorisées et conservé pour une durée définie.

Identité et droits

Client authentifié avant toute donnée de commande. Accès limité à ses propres commandes, appliqué dès la lecture, pas après la génération.

Données

Hébergement et transferts de données vérifiés selon la solution retenue ; conditions contractuelles de non-utilisation des données pour l'entraînement ; chiffrement ; durée de conservation et accès au journal définis ; analyse nLPD documentée pour le traitement réellement déployé.

Sécurité applicative

Protection contre l'injection de prompt, y compris dans les e-mails reçus. L'orchestration limite les actions de sortie à deux : répondre ou transmettre.

Trois voies de mise en œuvre

Trois voies : le module d'IA de l'outil de service client en place, une plateforme RAG managée reliée aux commandes, ou un développement sur mesure (orchestration, base vectorielle, API de modèle). Le choix se fait après l'audit des sources, pas avant.