Implémentation — OCR ML Kit + voix whisper.rn (offline-first)

Le scan de reçu et la dictée sont **offline-first** dès que les moteurs locaux sont disponibles. Après photo (caméra / galerie), l'utilisateur **recadre** la zone utile, puis ML Kit lit le ticket. Pour la voix : compte à rebours → WAV PCM 16 kHz → whisper.rn (`tiny-q5` / `base-q5`). Les modèles **ne sont pas** téléchargés au boot. **Sans** licence cloud : download **explicite** (~31 Mo, barre de %) avant le micro. **Avec** `receipt_online_access` + réseau et modèle absent : dictée **Grok online** tout de suite, Whisper en **fond** ; dès que le cache est prêt, retour au flux offline + **seuil**. Un parseur local extrait montant, commerçant, date et catégorie (lexicon FR + matching catégories), calcule un score (montant 40 %, merchant / date / catégorie 20 % chacun) ; si score ≥ `EXPO_PUBLIC_RECEIPT_OFFLINE_SCORE_THRESHOLD` (défaut 0,7), pas d'appel cloud. Grok sinon en secours (licence + online + score bas, ou voix sans modèle local). Hors ligne avec licence : warning non bloquant. Images jointes : JPEG 0,72 sans resize. Saisie manuelle toujours libre.

État : livré côté app.
i18n FR/EN (`lang/fr.js` , `lang/en.js`) : crop, countdown, loader, offline, gate Whisper, cloud-while-download.

Ops : `docs/RECEIPT-OFFLINE-OPS.md` · Outils : `docs/OFFLINE-OCR-VOICE.md` · Licence : `docs/LICENSE.md`

---

1. Décisions

Besoin	Solution livrée
Scan offline	`@react-native-ml-kit/text-recognition`
Dictée offline	`whisper.rn` + ggml `base-q5_1` / `tiny-q5_1`
Audio	WAV PCM 16 kHz mono
Online Grok	`receipts/scan`, `receipts/dictate` — escalade **ou** voix sans modèle si licence
Gate Whisper	Explicite sans licence ; fond + online si licence
Recadrage	`ReceiptCropModal` + `expo-image-manipulator@~14.0.8`
Catégorie	`receiptCategoryMatch`
Manuel	Toujours libre

Principe : offline d'abord quand le modèle / OCR local suffit. Online = coût → licence `receipt_online_access` + réseau (+ score sous seuil, sauf dictée sans modèle encore en cache).

---

2. Matrice licence × réseau

2.1 Scan photo (OCR)

Licence	Internet	Comportement

Non	—	Offline seul (ML Kit)
Oui	Non	Offline + warning
Oui	Oui	Offline → score ≥ seuil : garder · score < seuil : Grok

2.2 Dictée (voix) — gate modèle

Licence	Réseau	Modèle Whisper en cache	Comportement
Non	*	Non	CTA download + % ; **micro bloqué**
Non	*	Oui	Offline-first (jamais Grok)
Oui	Non	Non	Micro bloqué / message modèle (pas de cloud)
Oui	Non	Oui	Offline + warning
Oui	Oui	Non	Micro **OK** → **Grok direct** + download Whisper **en fond** (bandeau %)
Oui	Oui	Oui	Offline-first + seuil → Grok si score bas

Slug : `FEATURES.RECEIPT_ONLINE` → `"receipt_online_access"` .

Permissions OS : caméra / photos / micro **avant** capture (≠ licence).

---

3. Scoring & seuil

`helpers/receiptScore.js` + `RECEIPT_OFFLINE_SCORE_THRESHOLD`
(`EXPO_PUBLIC_RECEIPT_OFFLINE_SCORE_THRESHOLD` , défaut **0.7**).

Champ	Points
`amount` > 0	0.40
`merchant`	0.20
`date` `YYYY-MM-DD`	0.20
`category_id`	0.20

OCR confidence optionnelle : `score *= 0.7 + 0.3 * ocrConfidence` .
Multi-dépenses voix : moyenne (`scoreExpenses`).

Score	Licence + online	Sinon
`>= seuil`	Garder offline	Offline
`< seuil`	Escalade Grok	« À compléter »

Sans réseau : jamais d’appel online.
Exception voix : pas de modèle local + licence + online → Grok **sans** passer par le score offline (path `online_grok_model_pending`).

---

4. Warning hors-ligne

`warnOfflineIfNeeded` si `hasOnlineReceipt && !isOnline` : alert `receipt.offlineWarning*` → continuer offline. Pas de blocage abo.

---

5. Flux réel (as-built)

```
caméra | galerie
    │
    └─ crop (ReceiptCropModal) → ML Kit → parse → score
         ├─ score OK              → formulaire
         ├─ score bas + licence   → Grok scan
         └─ sinon                 → à compléter

voix — gate (ExpenseSourceModal + useWhisperModelGate)
    │
    ├─ sans licence / offline, pas de modèle → CTA download (bloquant)
    ├─ licence + online, pas de modèle
    │     → micro OK → record → Grok dictate
    │     → startWhisperDownload({ quiet: true }) en fond
    └─ modèle en cache
          → record → whisper.rn → parse → score
               ├─ score OK            → formulaire
               ├─ score bas + licence → Grok dictate
               └─ sinon               → à compléter
```

Entrées : `TransactionTypeModal`, `NewExpenseBudgetBtn`
Flow : `useReceiptScanFlow`
Gate : `receiptWhisperGate` + `useWhisperModelGate`
UI gate : `ExpenseSourceModal`, `WhisperModelBanner`
Loader : `ReceiptScanLoader` · Debug : `ReceiptScanDebugPanel`

---

6. Modules livrés

6.1 Photo / OCR

Fichier	Rôle
`ReceiptCropModal.js`	Recadrage avant OCR
`receiptOcr.js`	ML Kit → parse + score
`receiptOfflineParse.js`	Montant / commerçant / date
`receiptCategoryMatch.js`	Catégorie locale
`receiptScan.js`	Persist JPEG 0.72, même résolution

`expo-image-manipulator@~14.0.8` (pas 58.x — casse Gradle Expo 54).

6.2 Voix / STT / gate

Fichier	Rôle
`VoiceRecorderModal.js`	Countdown 2 s + micro
`receiptPcmWavRecorder.js` / `receiptWavEncode.js`	PCM → WAV
`receiptWhisperModel.js`	Cache, download CDN→HF→miroir, upgrade Wi-Fi fond
`receiptWhisperGate.js`	Singleton état ready / downloading / %
`useWhisperModelGate.js`	`canDictate` = modèle prêt **ou** licence+online
`WhisperModelBanner.js`	Bandeau flottant % (download fond ou forcé)
`ExpenseSourceModal.js`	CTA download **ou** micro cloud selon licence
`receiptVoiceLexicon.js`	Homophones FR
`receiptStt.js`	Whisper lazy ; `allowDownload: false` à l'analyse
`receiptOfflineConstants.js`	Seuil, prompt, catalogues

Patch : `patches/whisper.rn+0.7.4.patch`.

6.3 Online Grok

`receipt.services.js` — `scanReceipt` / `dictateExpense`.
Voix sans modèle + licence : appel **direct** (pas seulement escalade sous seuil).

6.4 UX / i18n

- Loader étapes, countdown, crop
- Gate : `whisperDownload*`, `whisperReady`, `whisperCloudAvailable`, `whisperCloudWhileDownload`
- Offline / modèle manquant

---

7. Détails receipt offline

7.1–7.2 Parse OCR / voix

Montant (préférence max), commerçant, date, `matchCategorySmart`.
Voix : lexicon → split multi-dépenses → `WHISPER_DICTATION_PROMPT`.

7.3 Modèles Whisper

Priorité	Fichier	~Taille
Qualité	`ggml-base-q5_1.bin`	~57 Mo (Wi-Fi / fond)
Secours	`ggml-tiny-q5_1.bin`	~31 Mo (download bloquant / 1er)

Legacy	`ggml-base.bin` / `ggml-tiny.bin`	reuse cache seulement

~140 Mo = full `base` **non** utilisé (≠ 31+57).
URL : CDN → HF → `hf-mirror.com` . Cache : `documentDirectory/whisper/` .

`ensureWhisperModel({ allowDownload: false })` dans le chemin analyse — le gate UI / `startWhisperDownload` gère le download.

7.4–7.6 Audio, crop, persist

PCM 16 kHz mono WAV · crop natif · JPEG 0.72 sans resize.

7.7 Env

Variable	Défaut	Rôle
`EXPO_PUBLIC_RECEIPT_OFFLINE_SCORE_THRESHOLD`	`0.7`	Escalade sous seuil
`EXPO_PUBLIC_WHISPER_MODEL_BASE_URL`	vide	CDN ; sinon HF

---

8. Hors app

Équipe	Travail
Billing	`receipt_online_access` sur offres Grok
Infra	CDN des 2 `.bin`
QA	§2 + §11

---

9. Critères de done (app)

- [x] Offline-first OCR + score / seuil
- [x] Escalade Grok sous seuil (scan + voix avec modèle)
- [x] Gate Whisper sans licence (CTA + micro bloqué)
- [x] Licence + online + pas de modèle → Grok direct + download fond
- [x] Modèle prêt → retour offline + seuil
- [x] Warning hors-ligne licence
- [x] Crop, PCM WAV, category match, JPEG 0.72, i18n FR/EN
- [ ] CDN prod + feature billing

---

10. Notes build

- Whisper / ML Kit : rebuild natif (`expo run:android` / `run:ios`)
- `expo-image-manipulator@~14.0.8`
- `useFrameworks: "static"` (iOS) · patch whisper.rn

---

11. QA manuelle (device)

#	Setup	Attendu
1	Sans licence	OCR/voix offline ; 0 Grok ; dictée exige download
2	Licence, airplane	Warning ; offline si modèle ; sinon pas de cloud
3	Licence + online, ticket lisible	Offline, pas Grok
4	Licence + online, photo floue	`receipts/scan`
5	Crop	Zone / entière / annuler
6	Dictée, modèle en cache	Offline + éventuelle escalade seuil
7	**Sans licence**, pas de modèle	CTA % ; micro bloqué
8	**Avec licence**, pas de modèle, online	Micro OK → Grok ; bandeau download fond
9	Après fin download (8)	Prochaine dictée → offline + seuil
10	Multi-photos / manuel	Crop un à un / inchangé