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 avec **un seul modèle** `ggml-base-q5_1.bin` (~~57 Mo~~). Pas de tiny. Les modèles ~~**ne sont pas** téléchargés au boot. **Sans** licence cloud : download **explicite** (~~57 Mo, barre de % / bulle) 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**. Parse local (montant, commerçant, date, catégorie), score (montant 40 %, reste 20 % chacun) ; si score ≥ `EXPO_PUBLIC_RECEIPT_OFFLINE_SCORE_THRESHOLD` (défaut 0,7), pas d'appel cloud. Hors ligne avec licence : warning non bloquant. Images : JPEG 0,72 sans resize. Saisie manuelle toujours libre.

État : livré côté app.
i18n FR/EN : crop, countdown, loader, offline, gate Whisper (~57 Mo), 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` + **uniquement** `ggml-base-q5_1.bin` (~57 Mo)
Audio	WAV PCM 16 kHz mono
Online Grok	Escalade sous seuil **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

---

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 `base-q5` en cache	Comportement
Non	*	Non	CTA download + % ; **micro bloqué**
Non	*	Oui	Offline-first (jamais Grok)
Oui	Non	Non	Micro bloqué / message modèle
Oui	Non	Oui	Offline + warning
Oui	Oui	Non	Micro **OK** → **Grok direct** + download base-q5 **en fond** (bulle %)
Oui	Oui	Oui	Offline-first + seuil → Grok si score bas

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

---

3. Scoring & seuil

`helpers/receiptScore.js` · env `EXPO_PUBLIC_RECEIPT_OFFLINE_SCORE_THRESHOLD` (défaut **0.7**).

Champ	Points
`amount` > 0	0.40
`merchant`	0.20
`date`	0.20
`category_id`	0.20

Exception voix : pas de modèle + licence + online → Grok direct (`online_grok_model_pending`), sans score offline.

---

4. Warning hors-ligne

`warnOfflineIfNeeded` si licence cloud + offline → alert non bloquante.

---

5. Flux réel

```
photo → crop → ML Kit → parse → score → [Grok si bas + licence]

voix — gate
  ├─ sans licence, pas de modèle → CTA download base-q5 (~57 Mo)
  ├─ licence + online, pas de modèle → Grok + download fond
  └─ modèle en cache → whisper base-q5 → parse → score → [Grok si bas]
```

Modules clés : `useReceiptScanFlow`, `receiptWhisperGate`, `useWhisperModelGate`, `ExpenseSourceModal`, `WhisperModelBanner` (bulle bas-droite).

---

6. Modules voix / modèle

Fichier	Rôle
`receiptOfflineConstants.js`	Seuil + **WHISPER_MODEL_PREFERRED** = base-q5 seul
`receiptWhisperModel.js`	Cache / download base-q5 (CDN→HF→miroir)
`receiptWhisperGate.js`	État ready / downloading / %
`receiptStt.js`	Transcribe `language: "fr"` ; `allowDownload: false` à l'analyse

Plus de `WHISPER_MODEL_FALLBACK` (tiny). Legacy full `ggml-base.bin` peut encore être réutilisé s'il est déjà en cache.

---

7. Modèle Whisper

Fichier	~Taille	Rôle
`ggml-base-q5_1.bin`	~57 Mo	**Seul** modèle téléchargé / utilisé
`ggml-base.bin` (legacy)	~140 Mo	Reuse cache seulement, **non** téléchargé

Tiny (`ggml-tiny*`) : **abandonné** (ni download, ni sélection).
URL : CDN → HF → `hf-mirror.com`. Cache : `documentDirectory/whisper/`.

---

8. Hors app

Équipe	Travail
Billing	`receipt_online_access`
Infra	CDN : **`ggml-base-q5_1.bin` uniquement** (tiny optionnel à retirer)
QA	§2 + §11

---

9. Critères de done (app)

- [x] Offline-first OCR + seuil
- [x] Gate Whisper sans licence
- [x] Licence + online sans modèle → Grok + download fond
- [x] **Uniquement base-q5** (~57 Mo), plus de tiny
- [x] Crop, PCM WAV, i18n FR/EN, bulle download
- [ ] CDN prod + feature billing

---

10. Notes build

`expo-image-manipulator@~14.0.8` · patch whisper.rn · rebuild natif.

---

11. QA manuelle

#	Setup	Attendu
1	Sans licence, pas de modèle	CTA ~57 Mo ; micro bloqué
2	Licence + online, pas de modèle	Grok OK ; bulle download base-q5
3	Modèle prêt	Offline + seuil
4	Ticket / crop / multi-photos	Inchangé
5	Aucun download de `ggml-tiny*`	Vérifier logs / cache