Offline OCR & voix — comparaison d’outils

App majoritairement offline. Aujourd’hui scan / dictée = API (`receipts/scan`, `receipts/dictate`) → inutilisables sans réseau.

OCR et STT sortent du **texte** : il reste une extraction structurée (montant, commerçant, date, catégorie).

Outil	Lien	Rôle	Où ça tourne
ML Kit Text Recognition	developers.google.com/ml-kit/…	OCR	**On-device**
Tesseract	github.com/tesseract-ocr/tesseract	OCR	On-device ou serveur
whisper.rn	npmjs.com/package/whisper.rn	STT	**On-device** (RN)
faster-whisper	github.com/SYSTRAN/faster-whisper	STT	**Serveur**

1. ML Kit Text Recognition (OCR) — préféré

Liens

• https://developers.google.com/ml-kit/vision/text-recognition/v2

• Android : https://developers.google.com/ml-kit/vision/text-recognition/v2/android

• iOS : https://developers.google.com/ml-kit/vision/text-recognition/v2/ios

Avantages

• Meilleure qualité terrain que Tesseract

• Offline, gratuit, pensé mobile

• Blocs / lignes / bounding boxes utiles au parseur

• Privacy (image sur device)

Inconvénients

• Dépendance Google

• Module natif (prebuild / dev client)

• Pas de champs dépense prêts à l’emploi

• Poids si modèle bundled

2. Tesseract (OCR) — plan B OSS

Liens

• https://github.com/tesseract-ocr/tesseract

• https://tesseract-ocr.github.io/

• tessdata : https://github.com/tesseract-ocr/tessdata

Avantages

• Offline, gratuit, 100 % open source

• Pas de dépendance Google

Inconvénients

• Qualité souvent inférieure à ML Kit sur tickets photo

• Texte brut uniquement

• Wrapper RN à maintenir + packs langue

3. whisper.rn (STT on-device)

Liens

• https://www.npmjs.com/package/whisper.rn

• https://github.com/mybigday/whisper.rn

• whisper.cpp : https://github.com/ggerganov/whisper.cpp

Avantages

• Dictée offline réelle, gratuit

• Audio reste sur le device

• Modèles quantifiés (tiny / base / small)

Inconvénients

• Poids modèle + perf / batterie selon device

• Qualité FR / bruit sous un gros modèle serveur

• Intégration native (NDK, assets `.bin`)

• Texte brut → parse ensuite ; éviter les modèles `*.en`

4. faster-whisper (STT serveur)

Liens

• https://github.com/SYSTRAN/faster-whisper

• CTranslate2 : https://github.com/OpenNMT/CTranslate2

Avantages

• Rapide / léger côté serveur, bonne qualité

• Self-host, MAJ sans republier l’app

• Brancheable sur `receipts/dictate`

Inconvénients

• Pas offline mobile

• Coût infra, upload, privacy

Synthèse

Critère	ML Kit	Tesseract	whisper.rn	faster-whisper
Offline mobile	Oui (OCR)	Oui (OCR)	Oui (voix)	Non
Coût API externe	Non	Non	Non	Non (self-host)
Qualité	Bonne	Moyenne	Variable	Bonne
Privacy	Forte	Forte	Forte	Faible
Aligne offline-first	Oui	Oui	Oui	Non

Recommandation pragmatique

• **OCR** → ML Kit on-device (Tesseract = plan B OSS)

• **Voix** → whisper.rn (`tiny` / `base` multilingual)

• **Serveur** → faster-whisper / API actuelle en complément online seulement

```
Offline : photo → ML Kit → parse  |  audio → whisper.rn → parse
Online  : fallback API si besoin
```