

◆ DOCUMENT STRATÉGIQUE PRIVÉ

Stratégies de Monétisation

6 stratégies de monétisation + deep dive AI fine-tunée pour professionnels (avocats, médecins, experts-comptables). Workflows détaillés, business models et projections financières.

HARDWARE

RTX 5090

SOFTWARE

Claude Code Max 20x

HORIZON

12 mois

POTENTIEL

40k+/mois

Sommaire

01 Agence dev AI-augmentée

Cash flow immédiat, marge 90%+, scalable

02 Templates & micro-produits

Revenus semi-passifs, marketplaces

03 SaaS bootstrappé à la chaîne

Stratégie shotgun, MRR croissant

04 Services AI sur mesure

Inference locale, coût marginal nul

05 Formation & contenu

Audience, cours, coaching, communauté

06 Produit AI vertical avec moat

Fine-tuning, données propriétaires, moonshot

★ Workflow complet — Agence dev AI-augmentée

Phase 0 à 4, de la préparation au scale — le plan d'action détaillé

◆ Deep Dive — AI fine-tunée pour professionnels

8 niches avec PRD, features PO/P1/P2, conversion, projections 24 mois

⌚ Architecture technique — RAG-first, fine-tuning later

Pipeline RAG, pgvector, Légifrance API, vLLM sur RTX 5090

▶ Guide de démarrage — Coder JurisAI avec Opus 4.6

Plan d'exécution 4 semaines, prompts Claude Code jour par jour

N1 Assistant juridique
Avocats – TOP PICK

N2 Codage médical
CCAM / Facturation

N3 Rédaction médicale
Dictée → Compte-rendu

N4 Saisie comptable
OCR + Catégorisation

N5 Assistant notarial
Actes + Conformité

N6 Veille réglementaire
Cross-niche

N7 Due diligence
M&A / Data rooms

N8 Gestion RH / Paie
Cabinets comptables

Stratégies de monétisation – RTX 5090 + Claude Code Max 20x

Document de stratégie personnelle. Combinaison : RTX 5090 (inference locale, rendu, fine-tuning) + Claude Code Max 20x (développement AI-augmenté massif).

// Table des matières

1. [Agence dev AI-augmentée](#)
2. [Templates & micro-produits](#)
3. [SaaS bootstrappé à la chaîne](#)
4. [Services AI sur mesure](#)
5. [Formation & contenu](#)
6. [Produit AI vertical avec moat](#)
7. [WORKFLOW COMPLET – Agence dev AI-augmentée](#)

// 1. Agence dev AI-augmentée

Concept

Tu vends du développement web/SaaS au tarif marché, mais tu livres 10–20× plus vite grâce à Claude Code. Le client paie pour la valeur livrée, pas pour le temps passé.

Workflow précis

PHASE 1 – Prospection (jour 1-7)

- Créer un profil Malt, Upwork, LinkedIn optimisé
 - └ Positionnement : "Développeur fullstack – livraison rapide, qualité premium"
 - └ PAS de mention "AI" – tu vends du résultat, pas l'outil
- Publier 3-5 projets portfolio (faits avec Claude Code en 1-2 jours chacun)
 - └ Landing page SaaS
 - └ Dashboard admin
 - └ App mobile-first
 - └ Intégration API complexe
 - └ Site e-commerce headless
- Fixer les tarifs
 - └ Landing page : 2 000-4 000 €
 - └ MVP SaaS : 5 000-15 000 €
 - └ Dashboard/app : 3 000-8 000 €
 - └ Intégration API : 1 500-5 000 €
- Répondre à 5-10 appels d'offres/jour

PHASE 2 – Exécution (par projet)

- Jour 0 : Appel découverte (30 min)
 - └ Comprendre le besoin exact
 - └ Identifier les contraintes techniques
 - └ Proposer un devis + délai (diviser ton estimation réelle par 2)
- Jour 1 : Setup + Architecture
 - └ Claude Code : scaffold du projet complet
 - └ Claude Code : composants UI de base
 - └ Claude Code : routes API + schéma DB
 - └ RTX 5090 : si features AI → setup inference locale pour prototype
- Jour 2-3 : Développement core
 - └ Claude Code : implémenter toutes les features
 - └ Claude Code : tests + error handling
 - └ Claude Code : responsive + polish
- Jour 4 : QA + Livraison
 - └ Review manuelle du code généré
 - └ Tests manuels des parcours critiques
 - └ Déploiement (Vercel/Railway/VPS)
 - └ Livraison client avec démo
- Jour 5-7 : Buffer retours client + ajustements

PHASE 3 – Scale (mois 2+)

- Augmenter les prix progressivement (+20% tous les 2 mois)
- Créer des templates internes réutilisables
 - └ Starter SaaS (auth + billing + dashboard)
 - └ Starter landing page
 - └ Starter e-commerce
- Embaucher un VA pour la prospection/admin
- Viser 3-4 projets simultanés

Chiffres estimés

MOIS	PROJETS/MOIS	CA MOYEN/PROJET	CA MENSUEL	COÛT (CLAUDE + INFRA)	NET
1	2	3 000 €	6 000 €	200 €	5 800 €
3	4	4 000 €	16 000 €	200 €	15 800 €
6	5	5 000 €	25 000 €	300 €	24 700 €
12	6	7 000 €	42 000 €	400 €	41 600 €

Outils

OUTIL	USAGE
Claude Code Max 20x	Développement principal
Vercel / Railway	Déploiement rapide
Stripe	Facturation
Malt / Upwork / LinkedIn	Prospection
Notion	Gestion projets clients
Loom	Démos async aux clients

// 2. Templates & micro-produits

Concept

Créer des templates/starters premium vendus en one-shot sur des marketplaces. Revenus semi-passifs après la création initiale.

Workflow précis

PHASE 1 – Étude de marché (jour 1-3)

- Analyser les bestsellers sur :
 - Gumroad (filtrer "SaaS template", "Next.js starter")
 - Lemon Squeez
 - ThemeForest (tendances)
 - IndieHackers (demandes récurrentes)
- Identifier 5 niches sous-servies
 - Exemple : "SaaS starter avec multi-tenancy"
 - Exemple : "Dashboard analytics temps réel"
 - Exemple : "Template email marketing avec Resend"
 - Exemple : "Landing page AI SaaS"
 - Exemple : "Admin panel avec RBAC"
- Valider la demande (recherche keywords, posts Reddit/Twitter)

PHASE 2 – Production (jour 4-14, par template)

- Jour 1 : Architecture + Design system
 - Claude Code : scaffold complet Next.js/Tailwind
 - Définir la proposition de valeur unique
- Jour 2-3 : Développement
 - Claude Code : toutes les features
 - Claude Code : documentation inline
 - Claude Code : README détaillé
- Jour 4 : Polish
 - Review manuelle qualité
 - Screenshots + vidéo démo (RTX 5090 pour rendu rapide)
 - Écrire la page de vente
- Jour 5 : Publication
 - Publier sur Gumroad/Lemon Squeez
 - Post Twitter/Reddit/IndieHackers
 - Envoyer à 3-5 newsletters dev

PHASE 3 – Itération continue

- Publier 1 nouveau template par semaine
- Mettre à jour les templates populaires
- Créer des bundles (3-5 templates = prix réduit)
- Collecter les emails → newsletter → upsell

Pricing

PRODUIT	PRIX	VOLUME ESTIMÉ/MOIS
Template landing page	29-49 €	20-50 ventes
SaaS starter kit	79-149 €	10-30 ventes
Bundle complet	199-299 €	5-15 ventes
Template + setup call	499 €	2-5 ventes

Chiffres estimés

MOIS	TEMPLATES PUBLIÉS	CA MENSUEL	NET (APRÈS FRAIS PLATEFORME)
1	4	500–1500 €	400–1200 €
3	12	2 000–5 000 €	1 600–4 000 €
6	20+	4 000–10 000 €	3 200–8 000 €
12	30+	6 000–15 000 €	4 800–12 000 €

// 3. SaaS bootstrappé à la chaîne

Concept

Lancer rapidement plusieurs micro-SaaS (1–2 semaines par produit), valider avec du trafic réel, doubler sur ceux qui marchent, abandonner les autres.

Workflow précis

PHASE 1 – Idéation (jour 1-2)

- Trouver 10 idées de micro-SaaS
 - Sources : Reddit (r/SaaS, r/Entrepreneur), Twitter, ProductHunt
 - Pattern : "Je fais X manuellement, ça me prend Y heures"
 - Pattern : "Il n'existe pas de bon outil pour Z"
 - Pattern : wrapper autour d'une API existante
- Filtrer avec la grille ICE (Impact / Confidence / Ease)
 - Impact : le problème est-il douloureux ? (1-10)
 - Confidence : les gens paieraient-ils ? (1-10)
 - Ease : faisable en 1-2 semaines avec Claude Code ? (1-10)
- Sélectionner les 3 meilleures idées

PHASE 2 – Build rapide (jour 3-12, par SaaS)

- Jour 1 : Landing page + waitlist
 - Claude Code : landing page complète
 - Formulaire email (Resend / ConvertKit)
 - Partager sur Twitter/Reddit/IndieHackers
 - Objectif : 50+ signups en 48h = feu vert
- Jour 2-3 : MVP core feature
 - Claude Code : auth (NextAuth/Clerk)
 - Claude Code : feature principale (1 seule, bien faite)
 - Claude Code : billing (Stripe, plan gratuit + payant)
- Jour 4-5 : Polish + Launch
 - Claude Code : onboarding flow
 - Claude Code : emails transactionnels
 - Lancer sur ProductHunt
 - Posts Reddit/Twitter/HackerNews
- Jour 6-10 : Itération
 - Écouter les retours utilisateurs
 - Claude Code : fix bugs + top 3 feature requests
 - A/B test pricing
- Jour 11-14 : Décision
 - > 10 paying users → DOUBLER (passer en mode scale)
 - > 50 free users, 0 payants → pivoter pricing/feature
 - < 10 free users → ABANDONNER, passer au suivant

PHASE 3 – Scale (pour les SaaS validés)

- Claude Code : features avancées
- RTX 5090 : ajouter des features AI (le différentiateur)
 - Inference locale → pas de coût API pour toi
 - Exemples : résumé automatique, classification, extraction
- SEO + Content marketing (Claude Code pour générer les articles)
- Intégrations (Zapier, Slack, etc.)
- Viser 100+ MRR avant d'investir plus

Exemples de micro-SaaS rapides à builder

IDÉE	FEATURE CORE	AI ANGLE (RTX 5090)	PRICING
Invoice generator	Créer/envoyer des factures	OCR extraction automatique	9 €/mois
Meeting notes	Enregistrer + résumer	Whisper transcription locale	12 €/mois
Social media scheduler	Planifier des posts	Génération de captions AI	15 €/mois
Competitor tracker	Surveiller des sites	Analyse sémantique des changements	19 €/mois
Email warmup tool	Améliorer délivrabilité	Score de qualité AI	29 €/mois

Chiffres estimés (pour 1 SaaS qui marche sur 5 lancés)

MOIS	MRR	COÛT INFRA	NET
1	0-200 €	50 €	-50 à 150 €
3	500-2 000 €	100 €	400-1900 €
6	2 000-5 000 €	200 €	1800-4 800 €
12	5 000-15 000 €	500 €	4 500-14 500 €

// 4. Services AI sur mesure

Concept

Utiliser la RTX 5090 pour faire tourner des modèles AI en local et vendre l'accès comme service (API, chatbot, transcription, etc.). Coût marginal quasi nul.

Workflow précis

PHASE 1 – Setup infrastructure (jour 1-5)

- RTX 5090 : installer les modèles
 - LLM : Llama 3.1 70B (Q4 quantized) ou Mistral Large
 - STT : Whisper Large v3
 - TTS : Bark / XTTs v2
 - Vision : LLaVA / Florence 2
 - Images : Stable Diffusion XL / Flux
- Exposer via API
 - Option 1 : vLLM + FastAPI (plus performant)
 - Option 2 : Ollama + API wrapper
 - Tunnel : Cloudflare Tunnel ou Tailscale
- Claude Code : construire le dashboard de monitoring
 - Usage par client
 - Quotas
 - Billing automatique
- Claude Code : documentation API (auto-générée)

PHASE 2 – Packager les services (jour 6-14)

- Service 1 : Transcription audio/vidéo
 - Upload fichier → Whisper → texte + timestamps
 - Pricing : 0.01 €/minute (vs 0.006 \$/min OpenAI)
 - Avantage : données restent en France, RGPD
- Service 2 : Chatbot personnalisé
 - RAG sur les documents du client
 - Llama/Mistral en backend
 - Widget intégrable
 - Pricing : 49-199 €/mois selon volume
- Service 3 : Génération d'images
 - API simple : prompt → image
 - Fine-tuning sur le branding du client (LoRA)
 - Pricing : 0.02 €/image
- Service 4 : Extraction de données
 - PDF/images → données structurées
 - Florence 2 + LLM pour comprendre le contexte
 - Pricing : 0.05 €/document
- Service 5 : API LLM privée
 - Pour les entreprises qui veulent de l'AI sans envoyer leurs données chez OpenAI
 - Llama/Mistral en backend
 - Pricing : 199-999 €/mois (accès dédié)

PHASE 3 – Commercialisation

- Site vitrine (Claude Code, 1 jour)
- Démarchage direct : agences web, PME, startups
 - Angle : "AI privée, données en France, RGPD compliant"
- Partenariats : intégrateurs, ESN, agences marketing
- Content marketing : articles sur l'AI privée/souveraine
- Upsell : du service API → au SaaS clé en main

Capacité RTX 5090

MODÈLE	TOKENS/SEC (ESTIMÉ)	REQUÊTES/JOUR
Llama 3.170B Q4	~40–60 tok/s	~5 000
Mistral 7B	~120–150 tok/s	~20 000
Whisper Large v3	~30× temps réel	~700 heures audio/jour
SDXL	~2–3 sec/image	~30 000 images/jour

Chiffres estimés

MOIS	CLIENTS	CA MENSUEL	COÛT (ÉLECTRICITÉ + INFRA)	NET
1	2–5	500–2 000 €	100 €	400–1 900 €
3	10–20	3 000–8 000 €	150 €	2 850–7 850 €
6	20–50	8 000–20 000 €	200 €	7 800–19 800 €
12	50–100	20 000–50 000 €	300 €	19 700–49 700 €

// 5. Formation & contenu

Concept

Créer des cours et du contenu éducatif sur le développement AI-augmenté. Monétiser via des plateformes de cours, YouTube, et du consulting.

Workflow précis

PHASE 1 – Construire l'audience (mois 1-3)

- YouTube (2 vidéos/semaine)
 - Format : "Je build X en Y minutes avec AI"
 - RTX 5090 : rendu vidéo rapide (DaVinci Resolve / Remotion)
 - Claude Code : scripts, code des tutos
 - Exemples de vidéos :
 - "Je build un SaaS complet en 4h avec Claude Code"
 - "Landing page qui convertit en 30 min"
 - "De 0 à deployed : app full-stack en live"
 - "RTX 5090 : faire tourner Llama 70B en local"
 - "Je remplace 5 outils SaaS avec du self-hosted AI"
- Twitter/X (1-3 posts/jour)
 - Tips dev, threads "how I built X"
 - Screenshots de code, before/after
 - Claude Code : générer les threads
- Newsletter (1/semaine)
 - Recap des meilleurs tips
 - Projet de la semaine
 - Outils recommandés (affiliation)
- Blog/SEO (2 articles/mois)
 - Claude Code : rédiger + optimiser SEO
 - Tutoriels longs format

PHASE 2 – Monétiser (mois 3-6)

- Cours en ligne
 - Cours 1 : "Développement AI-augmenté : de junior à 10x dev" (199-499 €)
 - Cours 2 : "Lancer un micro-SaaS en 2 semaines" (149-299 €)
 - Cours 3 : "Self-hosted AI avec une RTX" (99-199 €)
 - Plateforme : Gumroad ou propre site (Claude Code)
- Coaching 1:1
 - 150-300 €/h
 - 5-10 sessions/mois
- Affiliation
 - Vercel, Railway, Clerk, Resend, etc.
 - 10-30% de commission récurrente
- Sponsors YouTube
 - À partir de 10k abonnés : 500-2 000 €/vidéo
- Communauté payante (Discord/Slack)
 - 19-49 €/mois
 - Accès aux projets sources, Q&A, workshops mensuels

PHASE 3 – Scale (mois 6-12)

- Cohort-based course (live, 4-6 semaines, 999-2 499 €)
- Embaucher un monteur vidéo
- Repurpose content : YouTube → TikTok, shorts, clips
- Écrire un ebook/guide complet

Chiffres estimés

MOIS	ABONNÉS YT	CA MENSUEL	SOURCES
1	100-500	0-500 €	Affiliation
3	1000-3 000	1 000-3 000 €	Cours + affiliation
6	5 000-10 000	5 000-15 000 €	Cours + sponsors + coaching
12	15 000-30 000	15 000-40 000 €	Tout combiné

// 6. Produit AI vertical avec moat

Concept

Construire un produit SaaS spécialisé dans une niche verticale, avec un avantage compétitif AI difficile à reproduire (données propriétaires, modèles fine-tunés, workflow spécifique).

Workflow précis

PHASE 1 – Recherche de niche (semaine 1-2)

- Critères de sélection
 - Marché B2B (willingness to pay plus élevée)
 - Processus actuellement manuel ou mal outillé
 - Données spécifiques au domaine (= moat)
 - Réglementation qui favorise le local/privé (RGPD)
- Niches prometteuses
 - Comptabilité freelance AI (OCR factures + catégorisation auto)
 - Réécriture juridique assistée (contrats, CGV, RGPD)
 - Analyse immobilière (estimation prix, détection opportunités)
 - Recrutement technique (screening CV + matching AI)
 - Gestion de chantier BTP (planning AI + suivi photo)
 - Veille réglementaire (extraction + résumé des JO/textes)
- Validation
 - 20-30 interviews utilisateurs potentiels
 - "Combien payez-vous aujourd'hui pour résoudre ce problème ?"
 - "Que feriez-vous avec 10h de plus par semaine ?"
- Sélectionner LA niche

PHASE 2 – MVP (semaine 3-6)

- Semaine 3 : Architecture
 - Claude Code : setup projet complet
 - Définir le data model
 - Choisir les modèles AI (Llama, Whisper, Florence selon besoin)
 - RTX 5090 : tester les modèles sur des données réelles du domaine
- Semaine 4-5 : Build
 - Claude Code : app complète (Next.js + Supabase)
 - Claude Code : intégrations métier (API comptable, API immo, etc.)
 - RTX 5090 : pipeline AI (extraction, classification, génération)
 - Claude Code : tests + monitoring
- Semaine 6 : Beta
 - 10-20 beta testers (issus des interviews)
 - Feedback loop serré (Slack/Discord dédié)
 - Itérer quotidiennement avec Claude Code
- Gate : 5+ utilisateurs qui reviennent sans qu'on les relance = GO

PHASE 3 – Fine-tuning & moat (mois 2-3)

- RTX 5090 : fine-tuner les modèles sur les données du domaine
 - LoRA/QLoRA pour efficacité mémoire
 - Données collectées via les beta testers (avec consentement)
 - Objectif : surpasser les modèles généralistes sur TA niche
- Construire le dataset propriétaire
 - Chaque utilisation enrichit le modèle
 - Effet réseau : plus d'utilisateurs = meilleur modèle = plus d'utilisateurs
- Claude Code : features avancées basées sur le feedback
- Documenter les métriques (précision, temps gagné, satisfaction)

PHASE 4 – Go-to-market (mois 3-6)

- Pricing
 - Freemium : usage limité gratuit
 - Pro : 29-99 €/mois (selon la niche)
 - Enterprise : 299-999 €/mois (features avancées, support, SLA)
- Distribution
 - SEO (Claude Code : contenu optimisé)
 - Partenariats avec des acteurs du domaine
 - Présence sur les salons/événements de la niche
 - Bouche à oreille (le produit est le marketing)
- Sales
 - Démo personnalisée (Loom ou live)

- └ Onboarding white-glove pour les 50 premiers clients
- └ Case studies avec les early adopters
- └ Objectif mois 6 : 50-100 clients payants, 5-15k MRR

PHASE 5 – Scale (mois 6-12)

- └ Lever si nécessaire (ou rester bootstrappé)
- └ Migrer l'inference vers le cloud si la demande dépasse la RTX
 - └ Mais garder le fine-tuning en local (contrôle)
- └ Embaucher : 1 dev, 1 sales/customer success
- └ Intégrations (Zapier, API ouverte, webhooks)
- └ Expansion : features adjacentes, nouveau marché géographique

Chiffres estimés

MOIS	CLIENTS PAYANTS	MRR	COÛT	NET
1-2	0 (beta)	0 €	200 €	-200 €
3	10-20	500-2000 €	300 €	200-1700 €
6	50-100	5 000-15 000 €	1000 €	4 000-14 000 €
12	200-500	20 000-50 000 €	3 000 €	17 000-47 000 €
24	500-2000	50 000-200 000 €	10 000 €	40 000-190 000 €

WORKFLOW COMPLET — Agence dev AI-augmentée

C'est la stratégie la plus rentable à court terme car elle génère du cash immédiat avec un investissement minimal. Elle finance ensuite les autres stratégies (SaaS, produit AI).

// Pourquoi c'est le meilleur point de départ

CRITÈRE	SCORE
Temps jusqu'au premier euro	1–2 semaines
Investissement initial	~200 €/mois (Claude Code)
Risque	Très faible (pas de produit à maintenir)
Scalabilité	Moyenne (limitée par ton temps, mais marge élevée)
Cash flow	Immédiat et prévisible
Apprentissage	Maximum (tu découvres plein de stacks/domaines)

// Phase 0 — Préparation (jour 1-3)

0.1 Stack de base à maîtriser

Tu n'as pas besoin de tout connaître — Claude Code comble les lacunes. Mais tu dois être fluent sur :

MUST KNOW (tu guides Claude Code) :

- └─ Next.js (App Router) – 80% des projets clients
- └─ Tailwind CSS – styling rapide
- └─ Supabase ou PostgreSQL – base de données
- └─ Stripe – paiements
- └─ Vercel – déploiement
- └─ Git – versionning

NICE TO KNOW (Claude Code gère, tu supervises) :

- └─ React Native / Expo – apps mobiles
- └─ tRPC / GraphQL – API complexes
- └─ Redis / queues – background jobs
- └─ Docker – déploiement custom
- └─ AWS / GCP – infra lourde

0.2 Crée ton environnement de travail

Templates internes (à construire une fois, réutiliser toujours) :

- └─ starter-saas/
 - └─ Next.js 16 + Tailwind 4
 - └─ Auth (NextAuth ou Clerk)
 - └─ Stripe billing
 - └─ Dashboard layout
 - └─ Email transactionnel (Resend)
 - └─ Landing page template
- └─ starter-landing/
 - └─ Next.js + Tailwind
 - └─ Sections modulaires
 - └─ Formulaire contact
 - └─ SEO optimisé
 - └─ Analytics (GA4 / Plausible)
- └─ starter-api/
 - └─ Next.js API routes ou Hono
 - └─ Validation Zod
 - └─ Rate limiting
 - └─ Auth middleware
 - └─ Error handling standardisé
- └─ starter-admin/
 - └─ Dashboard avec sidebar
 - └─ CRUD générique
 - └─ Filtres / recherche / pagination
 - └─ Charts (Recharts)
 - └─ Export CSV/PDF

Action: Utiliser Claude Code pour générer ces 4 starters en 2–3 jours. Ils te feront gagner 1–2 jours par projet client.

0.3 Outils business

CATÉGORIE	OUTIL	COÛT
Facturation	Stripe / Tiime	Gratuit - 15 €/mois
Contrats	Bonsai / template maison	Gratuit - 19 €/mois
Gestion projet	Notion	Gratuit
Communication client	Slack / Discord	Gratuit
Démos async	Loom	Gratuit
Prospection	Malt + LinkedIn	Gratuit - 30 €/mois
Email pro	Google Workspace	6 €/mois
Domaine	ton-agence.com	10 €/an

// Phase 1 – Positionnement & présence (jour 3-7)

1.1 Positionnement

NE PAS dire : "Je suis développeur freelance" **DIRE** : "Je suis un studio de développement spécialisé en [niche]"

Niches rentables pour le dev AI-augmenté :

NICHE	TJM MOYEN	DEMANDE	COMPLEXITÉ PERÇUE (= MARGE)
SaaS MVP pour startups	600–1000 €/j	Très haute	Haute
E-commerce headless	500–800 €/j	Haute	Moyenne
Dashboards / admin panels	500–700 €/j	Haute	Moyenne
Landing pages premium	400–600 €/j	Très haute	Basse (mais volume)
Intégrations API/automations	500–900 €/j	Haute	Haute
Apps avec AI intégrée	700–1200 €/j	Croissante	Très haute

Recommandation : commencer par "SaaS MVP pour startups" + "Landing pages premium" — ça couvre du petit et du gros projet.

1.2 Profils en ligne

MALT (MARCHÉ FRANÇAIS)

Titre : Développeur Fullstack Senior – MVP SaaS & Apps Web sur mesure
Sous-titre : Next.js, React, Node.js – Livraison rapide, qualité production

Bio :

Je conçois et développe des applications web et SaaS de A à Z.
Mon approche : comprendre votre besoin métier, proposer une architecture simple et efficace, livrer rapidement un produit fini et maintenable.

Spécialisations :

- MVP SaaS (auth, billing, dashboard, API)
- Landing pages haute conversion
- Dashboards et back-offices
- Intégrations API et automations
- Applications avec intelligence artificielle

Stack : Next.js, React, TypeScript, Tailwind CSS, Supabase, Stripe, Vercel

TJM : 600-800 € (ajuster selon marché local)

LINKEDIN

Headline : Founder @ [Ton Studio] – Je transforme vos idées en produits web

À propos :

J'aide les startups et PME à lancer leurs produits web rapidement.

- Vous avez une idée de SaaS ? Je la transforme en MVP fonctionnel en 1-3 semaines.
- Vous voulez une landing page qui convertit ? C'est livré en 3-5 jours.
- Vous avez besoin d'un dashboard custom ? Pareil.

Mon secret : un process de développement ultra-optimisé qui me permet de livrer en fraction du temps habituel, sans sacrifier la qualité.

Parlons de votre projet → [lien calendly]

1.3 Portfolio (Claude Code, 2jours)

Construire 5 projets vitrine avec Claude Code :

Projet 1 : "NexBoard" – Dashboard SaaS analytics

- └── Auth + multi-tenant
- └── Charts temps réel
- └── Export CSV
- └── Screenshots + démo Loom

Projet 2 : "LaunchKit" – Landing page startup

- └── Hero animé
- └── Pricing table
- └── Testimonials
- └── CTA optimisé
- └── Score Lighthouse 95+

Projet 3 : "InvoiceFlow" – App de facturation

- └── Création/envoi de factures
- └── Suivi paiements
- └── Dashboard résumé
- └── Intégration Stripe

Projet 4 : "ChatWidget" – Widget chatbot intégrable

- └── Interface conversationnelle
- └── Configuration sans code
- └── Analytics des conversations
- └── Démo interactive

Projet 5 : "DataSync" – Intégration API

- └── Connecteurs multiples (Stripe, Slack, etc.)
- └── Logs temps réel
- └── Retry + error handling
- └── Documentation API auto-générée

Chaque projet : déployé sur Vercel, code sur GitHub (public ou partiel), Loom de 2 min, screenshots dans le portfolio.

// Phase 2 – Prospection (continu, 1-2h/jour)

2.1 Canaux de prospection

Canal 1 : Malt (passif – les clients viennent à toi)

- Optimiser le profil chaque semaine
- Répondre aux missions dans l'heure
- Accepter les petits projets au début (pour les avis)
- Objectif : 4-5 étoiles, 10+ avis → flux entrant constant

Canal 2 : LinkedIn (semi-actif)

- Poster 3-5x/semaine
 - Lundi : insight technique ("Comment j'ai réduit le temps de chargement de X de 80%")
 - Mercredi : case study rapide ("Ce SaaS, je l'ai livré en 5 jours")
 - Vendredi : behind the scenes ("Mon setup dev en 2026")
 - Bonus : répondre aux posts de startups qui cherchent des devs
- Contacter 5-10 fondateurs de startups/semaine
 - Message : "J'ai vu que vous lancez [X]. Votre site/app est super.
Si vous avez besoin de dev ponctuellement, je serais ravi d'échanger."
 - PAS de pitch agressif. Juste ouvrir la porte.
- Objectif : 2-3 calls/semaine à partir du mois 2

Canal 3 : Réseau direct (le plus rentable long terme)

- Participer à 1-2 meetups/mois (tech, startup)
- Rejoindre des communautés (Discord, Slack de startups)
- Proposer de l'aide gratuite ponctuellement (code review, conseil archi)
- Le bouche à oreille est le canal #1 au bout de 6 mois

Canal 4 : Cold outreach ciblé (actif)

- Identifier des startups early-stage (ProductHunt, AngelList, Twitter)
- Trouver le fondateur sur LinkedIn
- Message personnalisé basé sur LEUR produit
 - "J'ai testé [votre produit]. L'UX est top. J'ai remarqué que [point technique]. Si vous cherchez un dev fullstack pour accélérer, je peux aider."
- Objectif : 1-2% de taux de conversion → 5-10 contacts/semaine = 1 projet/mois

2.2 Qualification des leads

Grille de qualification rapide (5 min) :

Budget :

- └─ < 1 000 € → Refuser poliment (sauf si stratégique)
- └─ 1 000-3 000 € → Petit projet, livrable en 2-5 jours
- └─ 3 000-10 000 € → Projet moyen, 1-3 semaines
- └─ 10 000-30 000 € → Gros projet, 1-2 mois
- └─ > 30 000 € → Engagement long terme

Complexité technique :

- └─ Landing page / site vitrine → Facile, Claude Code gère à 90%
- └─ CRUD app / dashboard → Moyen, Claude Code gère à 80%
- └─ SaaS avec billing → Moyen+, Claude Code gère à 75%
- └─ App temps réel / WebSocket → Complexe, Claude Code gère à 60%
- └─ Système distribué / infra lourde → Très complexe, évaluer si rentable

Red flags (fuir) :

- └─ "On a pas de budget mais on peut faire du revenue share"
- └─ "On a besoin de ça pour hier"
- └─ "Notre précédent dev est parti et on comprend pas le code"
 - └─ (sauf si tu es payé grassement pour ça)
- └─ Pas de décisionnaire identifié
- └─ Changements de spec constants sans process

2.3 Process de vente

Étape 1 : Premier contact (message/email)

- └── Répondre en < 4h
- └── Poser 3 questions :
 - └── "Quel est le problème que vous cherchez à résoudre ?"
 - └── "Avez-vous un budget et un délai en tête ?"
 - └── "Avez-vous des maquettes/specs ou c'est à définir ensemble ?"
- └── Proposer un call de 20 min

Étape 2 : Call découverte (20-30 min)

- └── 80% écoute, 20% questions
- └── Comprendre le VRAI besoin (pas ce qu'ils PENSENT vouloir)
- └── Identifier les contraintes (budget, délai, technique)
- └── Montrer 1-2 projets similaires du portfolio
- └── "Je vous envoie une proposition sous 24h"

Étape 3 : Proposition (Claude Code aide à la rédiger)

- └── Résumé du besoin (1 paragraphe – prouve que tu as compris)
- └── Solution proposée (architecture simplifiée, stack)
- └── Périmètre précis
 - └── Inclus : feature 1, feature 2, feature 3
 - └── Exclus : feature X, feature Y (pour éviter le scope creep)
- └── Planning (jalons clairs)
 - └── Semaine 1 : setup + pages principales
 - └── Semaine 2 : features core
 - └── Semaine 3 : polish + livraison
- └── Tarif
 - └── Option A : forfait fixe (recommandé pour les petits projets)
 - └── Option B : TJM (pour les projets longs/évolutifs)
- └── Conditions
 - └── 30-50% à la signature
 - └── 50% à mi-parcours ou livraison
 - └── Révisions incluses (1-2 tours)
- └── CTA : "Répondez OK pour démarrer"

Étape 4 : Signature + démarrage

- └── Facture d'acompte (Stripe)
- └── Créer un canal Slack/Discord dédié
- └── Partager un board Notion (suivi du projet)
- └── Commencer immédiatement (l'effet "wow" du démarrage rapide)

// Phase 3 – Exécution d'un projet type (le workflow quotidien)

3.1 Jour 0 — Setup (2-4h)

```
# 1. Créer le projet depuis un starter  
cp -r starters/starter-saas client-project-name  
cd client-project-name  
git init && git add . && git commit -m "init: from saas starter"  
  
# 2. Configurer l'environnement  
# Claude Code : setup env vars, DB, auth provider  
  
# 3. Déployer le squelette  
# Vercel / Railway → le client a un URL de preview dès jour 0  
  
# 4. Partager l'URL au client  
# "Voici le projet initialisé. Je commence le développement demain matin."  
# → Effet WOW : le client voit un truc déployé en quelques heures
```

3.2 Jour 1-2 — Core development (sessions de 3-4h)

Session de travail type avec Claude Code :

- 08:00 - Revue des specs du jour
 - └ Ouvrir la spec, identifier les 3-5 tasks
 - └ Prioriser : ce qui a le plus d'impact visuel d'abord
- 08:15 - Claude Code : batch 1 (features principales)
 - └ "Crée le composant X avec les specs suivantes : ..."
 - └ "Ajoute la route API pour Y avec validation Zod"
 - └ "Implémente le formulaire Z avec ces champs : ..."
 - └ Paralléliser : lancer 2-3 tâches Claude Code en même temps
- 09:30 - Review + ajustements
 - └ Tester manuellement chaque feature
 - └ Identifier les edge cases
 - └ Claude Code : fix les bugs trouvés
- 10:00 - Claude Code : batch 2 (intégrations, polish)
 - └ Intégration DB
 - └ Emails transactionnels
 - └ Loading states, error handling
 - └ Responsive
- 11:00 - Déployer la preview
 - └ git push → Vercel déploie auto
 - └ Envoyer l'URL au client
- 11:15 - Passer au projet suivant ou pause

3.3 Jour 3 — Polish & QA (2-3h)

Checklist de qualité (à suivre pour CHAQUE projet) :

Performance :

- └─ Lighthouse score > 90
- └─ Core Web Vitals dans le vert
- └─ Images optimisées (WebP, lazy loading)
- └─ Bundle < 200kb first load

Responsive :

- └─ Mobile (375px)
- └─ Tablet (768px)
- └─ Desktop (1280px+)
- └─ Tester sur un vrai device

Accessibilité :

- └─ Navigation clavier
- └─ Contraste suffisant
- └─ Alt text sur les images
- └─ aria-labels sur les éléments interactifs

SEO (si pertinent) :

- └─ Meta tags (title, description, OG)
- └─ Sitemap
- └─ Schema.org markup
- └─ robots.txt

Sécurité :

- └─ Validation côté serveur (Zod)
- └─ Rate limiting sur les API
- └─ CORS configuré
- └─ Pas de secrets exposés
- └─ Headers de sécurité (CSP, HSTS)

3.4 Jour 4 — Livraison

1. Déploiement final
 - └ Vérifier les env vars de production
 - └ DNS configuré (si domaine custom)
 - └ SSL actif
 - └ Monitoring basique (Vercel analytics ou Sentry)
2. Documentation (Claude Code la génère)
 - └ README technique
 - └ Guide d'utilisation (si admin panel)
 - └ Variables d'environnement documentées
 - └ Process de déploiement
3. Handoff au client
 - └ Loom de 5-10 min : tour du produit
 - └ Accès au repo Git
 - └ Accès au dashboard Vercel/hébergement
 - └ Facture finale
4. Follow-up (J+3)
 - └ "Comment se passe la prise en main ?"
 - └ Résoudre les derniers points
 - └ Demander un avis/témoignage

// Phase 4 – Scale (mois 3+)

4.1 Augmenter les prix

Stratégie de pricing progressif :

Mois 1-2 : Prix d'entrée (construire le portfolio + avis)

- └ Landing page : 1 500-2 500 €
- └ App web : 3 000-5 000 €
- └ MVP SaaS : 5 000-8 000 €

Mois 3-6 : Prix marché (tu as des avis + portfolio solide)

- └ Landing page : 2 500-4 000 €
- └ App web : 5 000-10 000 €
- └ MVP SaaS : 8 000-15 000 €

Mois 6-12 : Prix premium (tu es reconnu, flux entrant)

- └ Landing page : 4 000-6 000 €
- └ App web : 8 000-15 000 €
- └ MVP SaaS : 15 000-30 000 €

Règle : augmenter de 20% quand tu refuses des projets par manque de temps

4.2 Optimiser ton temps

Temps par projet avec Claude Code vs sans :

	Sans AI	Avec Claude Code Max
Landing page	3-5 jours	0.5-1 jour
App CRUD	2-3 semaines	3-5 jours
MVP SaaS	1-2 mois	1-3 semaines
Dashboard complexe	2-4 semaines	5-10 jours

Ça veut dire :

- Tu peux gérer 3-5 projets en parallèle
- Ou tu livres plus vite et tu augmentes les prix
- Ou les deux

4.3 Crée des assets réutilisables

À chaque projet, extraire et généraliser :

1. Composants UI
 - └ Formulaires, modals, tables, charts
 - └ → Bibliothèque interne qui s'enrichit
2. Patterns backend
 - └ Auth flows, billing, email, webhooks
 - └ → Starters de plus en plus complets
3. Automations
 - └ Scripts de setup projet
 - └ Templates de proposition commerciale
 - └ Templates d'email client
 - └ → Process de plus en plus rapide
4. Documentation
 - └ Checklists de qualité
 - └ Guides de déploiement
 - └ → Déléguéable à terme

4.4 Transition vers les autres stratégies

L'agence finance les projets à long terme :

Mois 1-3 : 100% agence (cash flow)
Mois 3-6 : 80% agence + 20% templates/SaaS (revenus passifs)
Mois 6-9 : 60% agence + 20% SaaS + 20% contenu
Mois 9-12 : 40% agence + 30% SaaS + 20% contenu + 10% produit AI

L'objectif final :

- L'agence génère du cash prévisible
- Les templates génèrent des revenus passifs
- Le SaaS génère du MRR croissant
- Le contenu génère de l'audience (qui alimente tout le reste)
- Le produit AI est le moonshot à long terme

// Résumé : Plan d'action semaine par semaine

SEMAINE	ACTION	OBJECTIF
1	Setup outils + starters + profils en ligne	Être prêt
2	Portfolio (5 projets) + premières candidatures Malt	Premier contact client
3	Prospection active + premiers calls	Signer le 1er projet
4	Livrer le 1er projet	Premier revenu + premier avis
5-8	2-3 projets en parallèle + optimiser le process	Cash flow régulier
9-12	Augmenter les prix + commencer les templates	Diversification
13-16	Lancer le 1er micro-SaaS	Début de MRR
17-24	Scale agence + SaaS + contenu	Multiples sources de revenus

// KPIs à suivre

MÉTRIQUE	CIBLE MOIS 1	CIBLE MOIS 6	CIBLE MOIS 12
Projets livrés	2	5/mois	6/mois
CA mensuel	5 000 €	25 000 €	40 000+ €
Taux de marge	90%+	85%+	80%+
Avis clients	2	15+	30+
Taux de closing	20%	40%	50%+
Projets récurrents	0	2-3	5+
Templates vendus	0	50/mois	200/mois
MRR SaaS	0	2 000 €	10 000 €

DEEP DIVE – AI fine-tunée pour professionnels (avocats, médecins, experts-comptables)

Cette section détaille comment construire et vendre un service d'intelligence artificielle spécialisée à des professionnels réglementés. C'est le "moonshot" à plus fort potentiel de revenus récurrents du document.

// Pourquoi ce marché est une mine d'or

Le problème des professionnels

Les avocats, médecins et experts-comptables partagent un problème commun : **80% de leur temps est consacré à des tâches répétitives à faible valeur ajoutée.**

PROFESSION	TÂCHES CHRONOPHAGES	TEMPS PERDU/SEMAINE	VALEUR HORAIRE
Avocat	Rédaction de conclusions, recherche jurisp., contrats types, courriers	15–25h	150–400 €/h
Médecin généraliste	Comptes-rendus, courriers confrères, codage CCAM/NGAP, certificats	10–15h	60–120 €/h (consult.)
Médecin spécialiste	Comptes-rendus opératoires, dictées médicales, protocoles	10–20h	100–300 €/h
Expert-comptable	Saisie/catégorisation, liasses fiscales, lettres de mission, rapports	15–25h	80–200 €/h
Notaire	Actes types, recherches cadastrales, formalités, courriers	15–20h	150–350 €/h

Pourquoi ils paieraient TOI plutôt que ChatGPT

CHATGPT / GPT-4 GÉNÉRALISTE	TON SERVICE FINE-TUNÉ
Réponses génériques, souvent fausses en droit FR	Entraînés sur du droit français réel, jurisprudence à jour
Données envoyées aux USA (violation secret pro)	Inference locale, données en France, RGPD compliant
Pas de templates métier	Templates pré-remplis par spécialité
L'utilisateur doit savoir prompter	Interface métier, pas besoin de prompt
Pas de traçabilité	Audit trail complet (obligatoire pour les pros)
Hallucinations non détectées	RAG sur sources vérifiées + citations systématiques

Chiffres du marché français

PROFESSION	NOMBRE EN FRANCE	BUDGET IT MOYEN/AN	MARCHÉ ADRESSABLE (1% PÉNÉTRATION)
Avocats	73 000	2 000-8 000 €	73M-584M €
Médecins	230 000	1 000-4 000 €	230M-920M €
Experts-comptables	21 000	3 000-10 000 €	63M-210M €
Notaires	17 000	5 000-15 000 €	85M-255M €
Total	341 000		451M-1,97Md €

Même avec 0.1% de pénétration, c'est un business à 7 chiffres.

// Les 8 niches les plus pertinentes (classées par potentiel)

Niche 1 : Assistant juridique pour avocats — ★ TOP PICK

Douleur : Un avocat passe 60% de son temps à rédiger (conclusions, contrats, mises en demeure). C'est répétitif, chronophage, et les juniors coûtent cher.

Use cases :

- Rédaction de conclusions (à partir d'un résumé des faits)
- Recherche de jurisprudence (avec citations exactes)
- Génération de contrats types personnalisés
- Mises en demeure et courriers types
- Résumé de pièces (documents adverses volumineux)
- Veille juridique automatisée par spécialité
- Analyse de risques sur un dossier

- TJM avocat = 150–400 €/h → chaque heure gagnée a une énorme valeur
- Le secret professionnel IMPOSE une solution privée → avantage concurrentiel massif
- Le droit FR est une niche mal servie par les outils US
- Les cabinets de 2–20 avocats sont les plus réceptifs (pas assez gros pour un outil interne, assez gros pour payer)

MÉTRIQUE	VALEUR
Prix cible	149–499 €/mois par utilisateur
Coût d'acquisition client	200–500 € (cycle court, douleur forte)
LTV estimée	3 600–12 000 € (24 mois rétention moyenne)
TAM France	73 000 avocats × 299 €/mois = 262M €/an

Niche 2 : Codage et facturation médicale automatisée

Douleur: Les médecins perdent 1–2h/jour sur la facturation, le codage CCAM/NGAP, et les refus CPAM. Beaucoup sous-codent (= perte de revenus).

- Use cases :
- Codage CCAM automatique à partir du compte-rendu
 - Vérification de cohérence (acte vs diagnostic)
 - Détection de sous-codage (revenus perdus)
 - Génération automatique des ententes préalables
 - Suivi des rejets CPAM + correction automatique
 - Reporting mensuel revenus par acte

MÉTRIQUE	VALEUR
Prix cible	79–199 €/mois
Gain moyen pour le médecin	500–2 000 €/mois (récupération sous-codage)
ROI client	3–10x → vente facile
TAM France	120 000 médecins libéraux × 149 €/mois = 214M €/an

Niche 3 : Rédaction médicale assistée (comptes-rendus)

Douleur: Les médecins dictent ou tapent des comptes-rendus pendant des heures. Les secrétaires médicales sont débordées ou absentes.

- Use cases :
- Transcription audio → compte-rendu structuré (Whisper + LLM)
 - Génération de courriers aux confrères
 - Comptes-rendus opératoires standardisés
 - Certificats médicaux types
 - Lettres d'adressage spécialiste
 - Résumé de dossier patient pour transmission

- Whisper Large v3 (RTX 5090) pour la transcription
- LLM fine-tuné sur la terminologie médicale FR pour la structuration
- RAG sur les référentiels HAS pour la conformité

MÉTRIQUE	VALEUR
Prix cible	99–249 €/mois
Temps gagné	5–10h/semaine
TAM France	230 000 médecins × 149 €/mois = 411M €/an

Niche 4 : Saisie comptable intelligente

Douleur: Les experts-comptables passent un temps fou sur la saisie, la catégorisation, et les rapprochements bancaires. Les outils actuels (type Dext) sont limités.

- Use cases :
- OCR factures + catégorisation automatique (plan comptable)
 - Rapprochement bancaire intelligent
 - Détection d'anomalies (doublons, montants incohérents)
 - Génération automatique des écritures comptables
 - Pré-remplissage liasses fiscales
 - Lettres de mission personnalisées
 - Rapports de gestion mensuels automatiques

MÉTRIQUE	VALEUR
Prix cible	99–399 €/mois par cabinet
Gain productivité	30–50% sur la saisie
TAM France	21 000 cabinets × 249 €/mois = 63M €/an

Niche 5 : Assistant notarial

Douleur: Rédaction d'actes répétitifs, recherches cadastrales, formalités administratives.

- Use cases :
- Génération d'actes types (vente, donation, succession)
 - Pré-remplissage à partir des pièces du dossier
 - Vérification de conformité légale
 - Calcul automatique des droits et frais
 - Courriers de formalités
 - Résumé de dossier client

MÉTRIQUE	VALEUR
Prix cible	199–599 €/mois
TAM France	17 000 notaires × 399 €/mois = 81M €/an

Niche 6 : Veille réglementaire automatisée

Douleur: Toutes les professions réglementées doivent suivre les évolutions légales. C'est un cauchemar permanent.

Use cases :

- Extraction quotidienne JO / Légifrance / EUR-Lex
- Résumé des textes pertinents par spécialité
- Alerte impact sur les dossiers en cours
- Historique des modifications législatives
- Comparaison ancien texte / nouveau texte
- Newsletter automatique personnalisée

MÉTRIQUE	VALEUR
Prix cible	49–149 €/mois
TAM France	Cross-niche, 50 000+ pros × 99 €/mois = 59M €/an

Niche 7 : Due diligence et analyse de contrats

Douleur: Lors de fusions, acquisitions ou audits, analyser des centaines de contrats manuellement.

Use cases :

- Extraction des clauses clés (pénalités, résiliation, exclusivité)
- Détection de clauses à risque
- Comparaison de contrats (matrice de synthèse)
- Résumé exécutif de data room
- Identification des manquements réglementaires
- Timeline des obligations contractuelles

MÉTRIQUE	VALEUR
Prix cible	499–1999 €/mois (ou au projet : 2 000–10 000 €)
Clients	Cabinets d'avocats M&A, Big 4, fonds d'investissement

Niche 8 : Gestion RH et paie automatisée (cabinets comptables)

Douleur: La paie est un enfer de complexité réglementaire FR. Les conventions collectives changent tout le temps.

Use cases :

- Vérification bulletins de paie vs convention collective
- Alerte changement réglementaire paie
- Rédaction contrats de travail conformes
- Calcul automatique des provisions (CP, RTT, primes)
- Courriers RH types (licenciement, avertissement, attestations)
- Simulation coût employeur

MÉTRIQUE	VALEUR
Prix cible	149–399 €/mois par cabinet
TAM France	15 000 cabinets avec pôle social x 249 €/mois = 45M €/an

// PRD & Features par niche – fiches détaillées

Chaque niche est détaillée avec : persona cible, features par priorité (P0 = MVP jour 1, P1 = mois 2–3, P2 = mois 4–6), métriques de succès, et stratégie de conversion spécifique.

PRD Niche 1 : Assistant juridique pour avocats — JurisAI

Persona principal : Maître Sophie D., 38 ans, avocate en droit du travail dans un cabinet de 5 personnes à Lyon. Elle facture 250 €/h mais passe 15h/semaine à rédiger des conclusions et chercher de la jurisprudence. Elle utilise ChatGPT en cachette mais n'ose pas y mettre de données confidentielles.

Problème : La rédaction juridique est le goulot d'étranglement. Un avocat qui rédige 2x plus vite prend 2x plus de dossiers.

Proposition de valeur : "Rédigez vos conclusions en 20 min au lieu de 4h. Vos données restent en France."

FEATURES PAR PRIORITÉ

- 1. Générateur de conclusions
 - └ Input : résumé des faits (texte libre ou formulaire)
 - └ Output : conclusions structurées (visa, discussion, dispositif) avec références légales
 - └ Export Word (.docx) formaté selon les usages
 - └ CRITIQUE : c'est la feature qui vend le produit
- 2. Recherche de jurisprudence RAG
 - └ Question en langage naturel
 - └ Résultats avec : juridiction, date, n° d'arrêt, extraits pertinents, lien Légifrance
 - └ Base : Légifrance open data + DILA
 - └ CRITIQUE : sans citations vérifiables = pas crédible
- 3. Templates de courriers juridiques
 - └ 10-15 templates de base :
 - ├ Mise en demeure
 - ├ Assignation
 - ├ Requête
 - ├ Courrier adverse
 - ├ Courrier client
 - └ Lettre de saisine
 - └ Personnalisation par champs (parties, dates, faits)
- 4. Auth + Sécurité de base
 - └ Login email/password + MFA
 - └ Chiffrement des données au repos
 - └ Isolation par cabinet (multi-tenant)
 - └ Page /securite avec certifications
- 5. Dashboard personnel
 - └ Documents récents
 - └ Crédits/usage restant
 - └ Accès rapide aux templates

- 6. Résumé de pièces adverses
 - └ Upload PDF (pièces, jugements) → résumé structuré
 - └ Extraction des arguments clés de l'adversaire
 - └ Suggestion de contre-arguments
- 7. Analyse de risques sur un dossier
 - └ Input : faits + prétentions
 - └ Output : probabilité de succès, points forts/faibles
 - └ Basé sur la jurisprudence similaire (RAG)
- 8. Historique et versioning
 - └ Toutes les générations sauvegardées
 - └ Comparaison de versions
 - └ Favoris et modèles personnels
- 9. Collaboration cabinet
 - └ Partage de documents entre associés
 - └ Commentaires internes
 - └ Rôles (associé, collaborateur, stagiaire)
- 10. Intégration email
 - └ Envoi direct de courriers depuis l'app
 - └ Suivi des envois

- 11. Veille juridique personnalisée

- └ Alertes par spécialité (nouveaux textes, jurisprudence)
- └ Résumé hebdomadaire automatique
- 12. Dictée vocale → document structuré
 - └ Whisper (RTX 5090) → transcription → mise en forme
 - └ Le must pour les avocats en déplacement
- 13. Intégration RPVA / e-barreau
 - └ Dépôt direct des conclusions
 - └ Suivi des calendriers de procédure
- 14. API pour intégration logiciels métier
 - └ Connecteurs Jarvis, Secib, Kleos
 - └ Documentation développeur
- 15. Analytics cabinet
 - └ Temps gagné par avocat
 - └ Types de documents les plus générés
 - └ ROI mesurable (pour justifier le renouvellement)
- 16. Mode on-premise
 - └ Installation sur serveurs du cabinet
 - └ Pour les gros cabinets (50+ avocats)
 - └ Premium : 2 499 €/mois

MÉTRIQUES DE SUCCÈS MVP

MÉTRIQUE	CIBLE J30	CIBLE J90
Conclusions générées/user/semaine	3+	5+
Temps moyen de rédaction (avant/après)	-60%	-75%
NPS	> 30	> 50
Rétention M1	> 70%	> 80%
Taux de conversion trial → payant	> 15%	> 25%

CONVERSION – FUNNEL SPÉCIFIQUE AVOCATS

ÉTAPE 1 : ACQUISITION (comment ils découvrent le produit)

- LinkedIn : contenu ciblé avocats (posts, articles)
 - "Comment j'ai réduit mes conclusions de 4h à 20 min"
 - "Les 5 erreurs des avocats qui utilisent ChatGPT"
 - Ciblage : titre "avocat" + France + cabinet < 50
- SEO : articles optimisés
 - "logiciel rédaction juridique"
 - "IA pour avocats"
 - "assistant juridique intelligence artificielle"
 - "alternative ChatGPT avocat RGPD"
- Partenariats barreaux
 - Présentation lors des formations continues
 - Article dans la gazette du barreau local
- Recommandation (le canal #1 à terme)
 - Les avocats parlent entre eux
 - Programme de parrainage : 1 mois offert

ÉTAPE 2 : LANDING PAGE (ce qui convertit en inscription)

- Hero : "Rédigez vos conclusions en 20 min. Vos données restent en France."
- Démo interactive live (le visiteur tape un résumé → voit le résultat)
 - CRITIQUE : la démo vaut 1000 mots
 - Limiter à 1 génération gratuite (teaser)
- Preuves :
 - Témoignages de 3-5 avocats beta testers (avec photo + nom + barreau)
 - "Données hébergées en France – chiffrement AES-256"
 - Badge RGPD + logo hébergeur FR
 - Chiffre clé : "4h → 20 min" / "75% de temps gagné"
- Pricing transparent
 - Solo : 149 €/mois (1 utilisateur)
 - Cabinet : 399 €/mois (5 utilisateurs)
 - Enterprise : sur devis
 - "14 jours d'essai gratuit, sans CB"
- CTA : "Essayer gratuitement" (pas "Demander une démo")
 - Le self-serve convertit 3-5x mieux que le formulaire de démo

ÉTAPE 3 : ONBOARDING (les 10 premières minutes = tout se joue)

- Minute 0-1 : Incription (email + mot de passe, rien d'autre)
- Minute 1-3 : Choix de la spécialité
 - "Quelle est votre spécialité principale ?"
 - Droit du travail / Droit pénal / Droit des affaires / etc.
 - Personnalise immédiatement les templates
- Minute 3-5 : Premier document généré (LE MOMENT AHA)
 - Template pré-rempli avec un cas exemple
 - L'utilisateur clique "Générer" → voit le résultat en 30 sec
 - Effet WOW : "ça marche vraiment, c'est pertinent"
- Minute 5-10 : Son propre cas
 - Prompt : "Essayez avec un de vos dossiers en cours"
 - Zone de texte pour coller son résumé des faits
 - Génération personnalisée
- Follow-up J+1 : Email
 - "Voici votre premier document. Voici 3 choses à essayer ensuite :"
 - Lien vers les templates les plus populaires
 - Lien vers la recherche de jurisprudence

ÉTAPE 4 : ACTIVATION (trial → usage régulier)

- Emails de nurturing (J+1, J+3, J+7, J+14)
 - J+1 : "3 astuces pour des conclusions parfaites"
 - J+3 : "Avez-vous essayé la recherche de jurisprudence ?"
 - J+7 : "Maître X a gagné 12h cette semaine avec JurisAI"
 - J+14 : "Votre essai se termine dans 2 jours"
- In-app prompts

- └ Badge "Nouveau" sur les features non essayées
 - └ Checklist d'onboarding (3/5 tâches complétées)
 - └ Compteur de temps gagné (gamification légère)
 - └ Support proactif
 - └ Chat live pendant le trial
 - └ Call de 10 min offert à J+3 si pas d'usage
- ÉTAPE 5 : CONVERSION (trial → payant)**
- └ Soft paywall à J+14
 - └ "Votre essai gratuit est terminé. Vous avez généré 12 documents et économisé environ 8h. Continuez pour 149 €/mois."
 - └ Résumé personnalisé de l'usage pendant le trial
 - └ Montrer le ROI : "8h × 250 €/h = 2 000 € de valeur, pour 149 €"
 - └ Réductions early adopter
 - └ -20% si paiement annuel
 - └ -10% pour les 100 premiers inscrits
 - └ Call de closing pour les cabinets (> 3 users)
 - └ Démo personnalisée 20 min
 - └ Offre Cabinet sur mesure
 - └ Garantie satisfait ou remboursé 30 jours
 - └ Réduit le risque perçu à zéro

PRD Niche 2 : Codage médical — CodAI Santé

Persona : Dr. Marc L., 52 ans, chirurgien orthopédiste libéral. Il opère 8-10 patients/jour mais sous-code systématiquement ses actes par manque de temps. Il perd ~1500 €/mois en revenus non-facturés.

Proposition de valeur : "Ne perdez plus un euro. CodAI code vos actes automatiquement et détecte le sous-codage."

FEATURES PAR PRIORITÉ

P0 – MVP

- └ 1. Codage CCAM automatique
 - └ Input : compte-rendu opératoire ou description acte (texte)
 - └ Output : codes CCAM correspondants + libellés + tarifs
 - └ Suggestions multiples si ambiguïté (le médecin valide)
 - └ Base : nomenclature CCAM officielle (Ameli)
- └ 2. Détection de sous-codage
 - └ Analyse des actes sur 1-3 mois (import fichier ou saisie)
 - └ Déetecte les actes qui auraient pu être codés différemment
 - └ Calcul du manque à gagner estimé
 - └ C'EST LA FEATURE QUI VEND : "Vous perdez 1 500 €/mois"
- └ 3. Vérification cohérence acte/diagnostic
 - └ Alerte si le code ne correspond pas au diagnostic CIM-10
 - └ Prévient les rejets CPAM avant envoi
- └ 4. Dashboard revenus
 - └ Vue mensuelle des actes codés
 - └ Comparaison vs mois précédent
 - └ Alertes sur les anomalies
- └ 5. Auth + Sécurité HDS
 - └ Hébergement certifié HDS (obligatoire données de santé)
 - └ MFA, chiffrement, audit trail

P1 – Growth (mois 2-3)

- └ 6. Ententes préalables automatiques
 - └ Génération du formulaire pré-rempli
 - └ Suivi des réponses CPAM
- └ 7. Suivi des rejets CPAM + correction
 - └ Import des retours CPAM
 - └ Analyse du motif de rejet
 - └ Proposition de correction
- └ 8. Intégration logiciels de cabinet
 - └ Connecteurs : Doctolib, Weda, Hellodoc, Maiia
 - └ Import automatique des consultations
- └ 9. Reporting avancé
 - └ Revenus par type d'acte
 - └ Taux de rejet par motif
 - └ Benchmark vs moyenne spécialité

P2 – Scale (mois 4-6)

- └ 10. Mode multi-praticien (clinique/centre de santé)
- └ 11. Prédiction de revenus mensuels
- └ 12. Formation en ligne codage CCAM (upsell)
- └ 13. API pour éditeurs de logiciels médicaux

CONVERSION SPÉCIFIQUE MÉDECINS

HOOK PRINCIPAL : l'audit gratuit de sous-codage

- └── Le médecin upload 3 mois de relevés CPAM (ou saisit ses actes)
- └── L'outil analyse et affiche : "Vous perdez ~X €/mois en sous-codage"
- └── C'est GRATUIT et IMMÉDIAT → acquisition virale
- └── Pour voir les corrections détaillées → inscription trial
- └── ROI évident : "79 €/mois pour récupérer 1 500 €/mois"

CANAUX :

- └── Congrès médicaux (stand + démo live de l'audit)
- └── Groupes Facebook médecins libéraux (très actifs)
- └── Partenariats avec les URPS (unions régionales)
- └── Bouche à oreille entre confrères (le canal #1)
- └── YouTube : "Médecins : combien perdez-vous en sous-codage ?"

PRD Niche 3 : Rédaction médicale — DictAI Med

Persona: Dr. Isabelle R., 45 ans, médecin généraliste, 30 patients/jour. Elle dicte ses comptes-rendus sur un dictaphone, sa secrétaire les tape le lendemain. Retard permanent de 2-3 jours sur les courriers.

Proposition de valeur: "Dictez, c'est rédigé. Comptes-rendus structurés en temps réel."

FEATURES PAR PRIORITÉ

P0 – MVP

- 1. Dictée vocale → compte-rendu structuré
 - Enregistrement audio dans l'app (web ou mobile)
 - Whisper Large v3 (RTX 5090) → transcription
 - LLM fine-tuné → structuration en sections médicales
 - (Motif, Examen, Diagnostic, Traitement, Suivi)
 - Terminologie médicale correcte automatiquement
- 2. Templates par spécialité
 - Consultation généraliste
 - Compte-rendu opératoire
 - Courrier au spécialiste / au généraliste
 - Certificat médical
 - Lettre d'adressage
- 3. Export et envoi
 - Export PDF/Word avec en-tête du cabinet
 - Envoi par email (intégration Resend)
 - Copie auto dans le dossier patient
- 4. Sécurité HDS
 - Audio et texte chiffrés
 - Suppression audio après transcription (option)
 - Hébergement HDS certifié

P1 – Growth

- 5. Résumé de dossier patient
 - Upload de documents passés → résumé synthétique
- 6. Traduction médicale (patient étranger)
- 7. App mobile (enregistrement hors connexion)
- 8. Intégration DMP/logiciels de cabinet
- 9. Modèles personnalisables par praticien

P2 – Scale

- 10. Transcription de consultations en temps réel
 - Le médecin parle au patient, l'app structure le CR en live
- 11. Aide au diagnostic différentiel
- 12. Multi-langues (arabe, turc, chinois → français)
- 13. Mode hôpital (multi-service)

CONVERSION SPÉCIFIQUE

- HOOK : test gratuit en 30 secondes
- Page d'accueil : bouton micro "Dictez un compte-rendu"
 - L'utilisateur dicte 30 sec → voit le CR structuré instantanément
 - Pas besoin d'inscription pour tester
 - Effet WOW immédiat → inscription pour sauvegarder
 - Cible : "Essayez maintenant, pas de CB requise"

PRD Niche 4 : Saisie comptable — ComptAI

Persona : Laurent M., 42 ans, expert-comptable, cabinet de 8 collaborateurs, 200 dossiers clients. Ses collaborateurs passent 60% de leur temps sur la saisie. Il perd des clients à cause des délais.

Proposition de valeur : "Divisez par 3 le temps de saisie. Vos collaborateurs se concentrent sur le conseil."

FEATURES PAR PRIORITÉ

P0 – MVP

- 1. OCR factures + catégorisation automatique
 - Upload photo/PDF de facture
 - Extraction : fournisseur, montant HT/TTC, TVA, date
 - Catégorisation sur le plan comptable du client
 - Apprentissage par client (s'améliore avec l'usage)
- 2. Génération d'écritures comptables
 - Écriture pré-remplie à partir de la facture OCR
 - Validation en 1 clic
 - Export FEC compatible
- 3. Rapprochement bancaire intelligent
 - Import relevés (CSV/OFX)
 - Matching automatique factures ↔ mouvements
 - Suggestions pour les non-matchés
- 4. Dashboard dossiers
 - Vue par client
 - Avancement saisie
 - Alertes retard
- 5. Multi-dossiers / multi-collaborateurs

P1 – Growth

- 6. Détection d'anomalies
 - Doublons, montants incohérents, TVA erronée
- 7. Lettrage automatique
- 8. Rapports de gestion automatiques
 - Tableaux de bord mensuels par client
 - Export PDF personnalisé
- 9. Intégration logiciels comptables
 - Connecteurs : Sage, Cegid, ACD, Pennylane
- 10. Import bancaire automatique (open banking)

P2 – Scale

- 11. Pré-remplissage liasses fiscales
- 12. Lettres de mission personnalisées
- 13. Prévisionnel de trésorerie AI
- 14. Portail client (dépôt de pièces)

CONVERSION SPÉCIFIQUE

HOOK : audit de productivité gratuit

- Formulaire : nombre de dossiers, collaborateurs, logiciel actuel
- Résultat instantané : "Vous pourriez gagner X heures/mois"
- Comparaison avant/après avec chiffres personnalisés
- CTA : "Testez sur 5 dossiers gratuitement pendant 30 jours"

PRD Niche 5 : Assistant notarial — NotarIA

Persona : Me. Claire V., 50 ans, notaire associée, étude de 4 notaires + 10 collaborateurs. La rédaction d'actes types prend un temps fou malgré les modèles Word existants.

FEATURES PAR PRIORITÉ

P0 – MVP

- └ 1. Génération d'actes types
 - └ Vente immobilière, donation, succession, PACS, SCI
 - └ Input : formulaire structuré (parties, biens, conditions)
 - └ Output : acte complet avec clauses standard + personnalisées
 - └ Références légales automatiques
- └ 2. Pré-remplissage intelligent
 - └ Upload des pièces (titre propriété, cadastre, diagnostics)
 - └ Extraction automatique des infos → injection dans l'acte
- └ 3. Calcul droits et frais
 - └ Calcul automatique des frais de notaire
 - └ Droits d'enregistrement, taxes, émoluments
 - └ Simulation pour le client
- └ 4. Export + signature
 - └ Export Word/PDF conforme
 - └ Compatible avec les logiciels notariaux

P1 – Growth

- └ 5. Vérification conformité
 - └ Check automatique des clauses obligatoires
 - └ Alerte si clause manquante ou obsolète
- └ 6. Résumé de dossier
- └ 7. Courriers de formalités automatiques
- └ 8. Gestion des délais légaux (alertes)
- └ 9. Base de clauses personnalisées par étude

P2 – Scale

- └ 10. Intégration MICEN / Télé@ctes
- └ 11. Recherches cadastrales automatisées
- └ 12. Espace client (suivi du dossier en ligne)
- └ 13. Mode multi-étude

CONVERSION SPÉCIFIQUE

- HOOK : démo sur un acte réel
- └ Webinaire mensuel : "Rédigez un acte de vente en 5 min avec l'IA"
 - └ Invitation via Chambre des Notaires + LinkedIn
 - └ Démo live avec un cas concret
 - └ Offre exclusive participants : 2 mois offerts
 - └ Follow-up individuel post-webinaire

PRD Niche 6 : Veille réglementaire — VeillAI

Persona : Tout professionnel réglementé qui doit suivre les évolutions légales. Cross-niche.

FEATURES PAR PRIORITÉ

P0 – MVP

- 1. Extraction automatique JO + Légifrance
 - Crawl quotidien des nouveaux textes
 - Filtrage par mots-clés et domaine
- 2. Résumé AI des textes pertinents
 - Chaque texte résumé en 3-5 lignes
 - Impact potentiel identifié
 - Lien vers le texte original
- 3. Newsletter personnalisée
 - Email quotidien ou hebdomadaire
 - Personnalisé par spécialité
- 4. Dashboard de veille
 - Timeline des modifications
 - Recherche dans l'historique
 - Favoris / alertes personnalisées

P1 – Growth

- 5. Comparaison ancien/nouveau texte (diff visuel)
- 6. Alerte impact sur les dossiers en cours
- 7. Veille européenne (EUR-Lex)
- 8. Intégration Slack/Teams (notifications)
- 9. Rapport mensuel PDF automatique

P2 – Scale

- 10. Prédiction de tendances réglementaires
- 11. Base de connaissances collaborative
- 12. API pour intégration dans d'autres outils
- 13. Version anglaise (expansion internationale)

CONVERSION SPÉCIFIQUE

HOOK : freemium fort

- Plan gratuit : 1 domaine de veille, newsletter hebdo
- L'utilisateur s'habitue → veut plus de domaines + quotidien
- Upgrade naturel → 49 €/mois pour le plan complet
- Viralité : "Partager cette veille avec un collègue" → acquisition

PRD Niche 7 : Due diligence — DiligenceAI

Persona : Alexandre P., 35 ans, avocat M&A dans un cabinet mid-size. Il passe 3 semaines à analyser une data room de 500+ documents lors d'une acquisition.

FEATURES PAR PRIORITÉ

P0 – MVP

- 1. Upload et ingestion de data room
 - Upload batch (ZIP, dossier, drag & drop)
 - OCR automatique si PDF scannés
 - Indexation et classification par type de document
- 2. Extraction de clauses clés
 - Pour chaque contrat : parties, durée, montant, clause de résiliation, pénalités, exclusivité, changement de contrôle, non-concurrence
 - Matrice de synthèse exportable Excel
- 3. Détection de risques
 - Clauses inhabituelles ou défavorables
 - Incohérences entre documents
 - Documents manquants par rapport à la checklist type
- 4. Résumé exécutif automatique
 - Synthèse globale de la data room
 - Points clés pour le client
 - Alertes hiérarchisées (critique/majeur/mineur)
- 5. Export rapport de due diligence
 - PDF structuré professionnel
 - Annexes avec les extraits sources
 - Prêt à envoyer au client

P1 – Growth

- 6. Comparaison de contrats (versions, fournisseurs)
- 7. Timeline des obligations contractuelles
- 8. Q&A sur la data room (chatbot RAG)
- 9. Collaboration (annotations, commentaires)
- 10. Templates de rapports personnalisables

P2 – Scale

- 11. Intégration plateformes VDR (Intralink, Datasite)
- 12. Scoring de qualité de la data room
- 13. Mode multi-deal simultané
- 14. Benchmark clauses vs marché

CONVERSION SPÉCIFIQUE

HOOK : pricing au projet (pas d'abonnement)

- "Analysez votre data room pour 2 000-5 000 €"
- Beaucoup plus facile à vendre qu'un abonnement pour du M&A ponctuel
- Upsell vers abonnement pour les repeat buyers
- Démo gratuite : upload 10 documents → voir l'extraction
- Référence : "3 semaines de travail junior → 2 heures avec DiligenceAI"

PRD Niche 8 : Gestion RH et paie — PaieAI

Persona : Nathalie T., 48 ans, responsable pôle social dans un cabinet comptable. Elle gère 300 bulletins de paie/mois avec 2 collaboratrices. Les conventions collectives changent sans arrêt et les erreurs coûtent cher.

FEATURES PAR PRIORITÉ

P0 – MVP

- 1. Vérification bulletins de paie
 - Upload bulletin PDF → analyse automatique
 - Check vs convention collective applicable
 - Détection d'erreurs : taux, primes, ancienneté, congés
 - Rapport d'anomalies avec corrections suggérées
- 2. Veille conventions collectives
 - Alerte automatique changement de convention
 - Résumé des impacts sur les bulletins
 - Checklist des modifications à appliquer
- 3. Génération contrats de travail
 - Input : poste, convention, durée, rémunération
 - Output : contrat conforme complet
 - Clauses adaptées à la convention collective
- 4. Simulation coût employeur
 - Charges patronales détaillées
 - Coût total par salarié
 - Simulation augmentation / prime
- 5. Courriers RH types
 - Avertissement, licenciement, attestation, avenant
 - Conformes au Code du travail + convention

P1 – Growth

- 6. Calcul provisions automatique (CP, RTT, 13e mois)
- 7. Tableau de bord social par client
- 8. Intégration Silae, PayFit, Sage Paie
- 9. Gestion des DSN (anomalies, corrections)
- 10. Base de connaissances paie (FAQ par convention)

P2 – Scale

- 11. Prédiction de la masse salariale
- 12. Audit social automatisé
- 13. Portail salarié (bulletins, attestations)
- 14. Mode multi-cabinet

CONVERSION SPÉCIFIQUE

HOOK : audit de conformité gratuit

- Upload de 5 bulletins + convention applicable
- Analyse automatique → "3 erreurs détectées sur 5 bulletins"
- Impact financier : "Risque de redressement URSSAF : X €"
- Pour voir le détail → inscription trial
- Peur du redressement = levier de conversion puissant

// PLAYBOOK CONVERSION UNIVERSEL – Applicable à toutes les niches

Ce playbook détaille les mécanismes psychologiques et les tactiques opérationnelles pour maximiser la conversion à chaque étape du funnel. Il s'applique à n'importe quelle niche pro.

Principe #1 : Démontrer le ROI avant de demander la CB

LA RÈGLE D'OR :

Le prospect doit VOIR la valeur AVANT de payer.

MAUVAIS funnel (taux de conversion : 1-3%)

Pub → Landing → "Demander une démo" → Call sales → Démo → Trial → Payant

✗ Trop de friction, cycle trop long

BON funnel (taux de conversion : 8-15%)

Pub → Landing avec démo live intégrée → Inscription gratuite → Moment Aha en < 5 min → Usage trial 14j → Paywall avec ROI personnalisé → Payant

✓ Le prospect se convainc lui-même

Principe #2 : Le "Moment Aha" doit arriver en < 5 minutes

Pour chaque niche, le Moment Aha c'est :

Niche 1 (Avocats) → Voir ses conclusions rédigées à partir de son résumé

Niche 2 (Codage) → Voir le montant qu'il perd en sous-codage

Niche 3 (Rédaction) → Dicter 30 sec et voir le CR structuré

Niche 4 (Comptable) → Photographier une facture et voir l'écriture comptable

Niche 5 (Notaire) → Remplir un formulaire et voir l'acte généré

Niche 6 (Veille) → Recevoir son premier résumé personnalisé

Niche 7 (Due dilig.) → Uploader 10 docs et voir la matrice d'extraction

Niche 8 (Paie) → Voir les erreurs dans ses propres bulletins

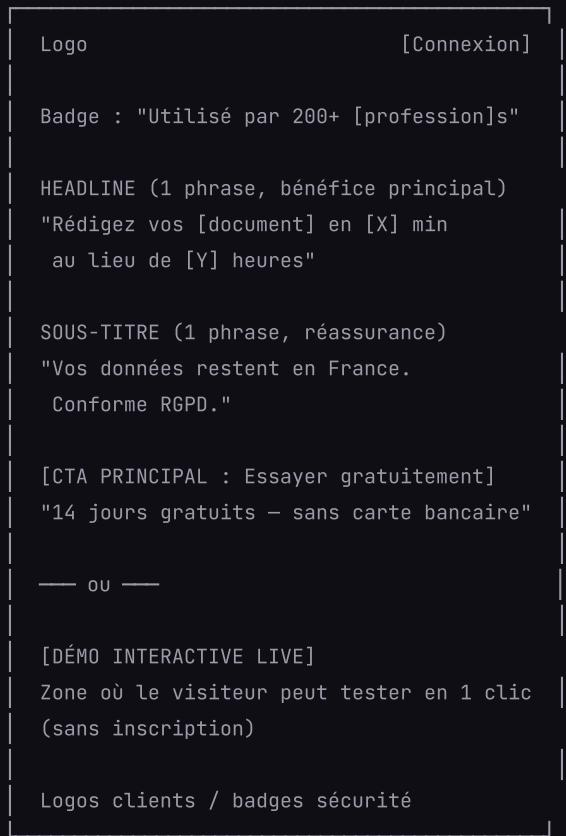
RÈGLE : Si le Moment Aha prend plus de 5 min → tu perds 80% des prospects.

Optimiser ce moment est la chose la PLUS IMPORTANTE du produit.

Principe #3 : La landing page qui convertit

STRUCTURE OPTIMALE (testée, prouvée) :

ABOVE THE FOLD (0 scroll)



SECTION 2 : Le problème (empathie)

- "Vous passez [X]h par semaine sur [tâche]"
- "Pendant ce temps, vous ne faites pas [activité à haute valeur]"
- "Et si vous pouviez récupérer ce temps ?"

SECTION 3 : La solution (3 features clés)

- Feature 1 : [la plus impressionnante] + screenshot/gif
- Feature 2 : [la plus utile au quotidien] + screenshot/gif
- Feature 3 : [le différentiateur] + screenshot/gif
- Chaque feature = 1 titre + 2 lignes + 1 visuel

SECTION 4 : Preuve sociale

- 3-5 témoignages avec photo + nom + titre + entreprise
 - Format : "Avant [problème]. Maintenant [résultat]. Chiffre clé."
- Logos des cabinets/entreprises clients
- Chiffres : "X documents générés" / "Y heures gagnées"
- Note moyenne (étoiles) si disponible

SECTION 5 : Sécurité et conformité

- "Vos données ne quittent jamais la France"
- Logos : RGPD, chiffrement AES-256, hébergeur FR
- Lien vers la page /securite détaillée
- Pour les médecins : badge HDS

SECTION 6 : Pricing

- 2-3 plans clairement différenciés
- Plan recommandé mis en avant visuellement
- Toggle mensuel / annuel (montrer l'économie)
- Feature comparison table
- FAQ sous le pricing (objections courantes)
 - "Mes données sont-elles en sécurité ?"

- └ "Puis-je annuler à tout moment ?"
- └ "L'IA fait-elle des erreurs ?"
- └ "Faut-il une formation ?"
- └ "Est-ce compatible avec mon logiciel actuel ?"
- └ CTA répété : [Essayer gratuitement]

SECTION 7 : Footer CTA

- └ Rappel du bénéfice principal
- └ CTA final
- └ "Des questions ? Contactez-nous : email / tel"

Principe #4 : Le pricing qui convertit

PSYCHOLOGIE DU PRICING POUR LES PROS

1. ANCHORING (ancrage)

- └ Montrer le coût SANS l'outil d'abord
- └ "Un collaborateur junior coûte 2 500 €/mois"
- └ "Notre outil fait le même travail pour 149 €/mois"
- └ Le prix paraît dérisoire en comparaison

2. ROI EXPLICITE

- └ Calculateur de ROI sur la landing page
- └ Input : nombre d'heures sur [tâche], taux horaire
- └ Output : "Vous économisez X €/mois, soit Yx le prix de l'outil"
- └ Le prospect se convainc avec SES propres chiffres

3. PLANS CROISSANTS (bon/meilleur/premium)

- └ Solo : pour tester et adopter (individuel)
- └ Cabinet : le plan que tu VEUX vendre (5-10 users, meilleure marge)
- └ Enterprise : pour les gros (signal de crédibilité)
- └ Mettre "POPULAIRE" sur le plan Cabinet

4. ANNUEL vs MENSUEL

- └ Toujours proposer les deux
- └ Annuel = -20% → afficher "Économisez X €/an"
- └ Le mensuel rassure les indécis
- └ Le switch annuel augmente la LTV de 30-40%

5. TRIAL SANS CB

- └ "14 jours gratuits, sans carte bancaire"
- └ Convertit 2-3x mieux qu'un trial avec CB
- └ Tu perds des tire-au-flanc mais tu gagnes plus de vrais clients
- └ Le soft paywall à J14 convertit 15-25% des trialists actifs

6. GARANTIE SATISFAIT OU REMBOURSÉ

- └ 30 jours de garantie après le paiement
- └ Réduit le risque perçu à zéro
- └ Taux de remboursement réel : < 5% (si le produit est bon)
- └ Augmente le taux de conversion de 10-20%

Principe #5 : Les métriques qui comptent (par étape)

FUNNEL METRICS – CE QUE TU DOIS TRACKER

Acquisition :

- └── Visiteurs uniques / mois
- └── Source de trafic (LinkedIn, SEO, referral, ads)
- └── Coût par visiteur (par canal)
- └── Taux de conversion visiteur → inscription : cible 5-10%

Activation :

- └── Time to Aha (temps entre inscription et Moment Aha) : cible < 5 min
- └── Taux de compléTION onboarding : cible > 60%
- └── Actions J1 (a-t-il créé un document ?) : cible > 50%
- └── Actions J7 (est-il revenu ?) : cible > 40%

Conversion :

- └── Taux trial → payant : cible 15-25%
- └── ARPU (average revenue per user) : dépend de la niche
- └── Revenue par canal d'acquisition
- └── Cycle de vente moyen : cible < 14 jours (solo), < 30 jours (cabinet)

Rétention :

- └── Churn mensuel : cible < 3-5%
- └── NPS : cible > 40
- └── Documents générés/user/mois (usage = rétention)
- └── DAU/MAU ratio : cible > 30%
- └── Expansion revenue (upsell) : cible 10-20% du MRR

Referral :

- └── Taux de parrainage : cible 10-15% des users réfèrent
- └── K-factor : cible > 0.3 (chaque user ramène 0.3 nouveau user)
- └── Source "bouche à oreille" dans les inscriptions

Principe #6 : L'email séquence qui convertit les trials

SÉQUENCE D'EMAILS PENDANT LE TRIAL (14 jours)

J+0 (immédiat) : Bienvenue

- └── Sujet : "Bienvenue sur [Produit] – voici votre 1er pas"
- └── Contenu : lien direct vers le Moment Aha
- └── CTA : "Créer votre premier [document]"
- └── Ton : chaleureux, pas vendeur

J+1 : Astuce #1

- └── Sujet : "Astuce : comment obtenir de meilleurs résultats"
- └── Contenu : 1 tip concret pour améliorer les outputs
- └── Exemple : "Ajoutez les articles de loi dans votre résumé pour des citations plus précises"
- └── CTA : "Essayer maintenant"

J+3 : Feature discovery

- └── Sujet : "Avez-vous essayé [feature #2] ?"
- └── Contenu : présentation de la 2e feature clé
- └── Inclure un GIF/vidéo de 15 sec
- └── CTA : "Découvrir [feature]"

J+5 : Preuve sociale

- └── Sujet : "Comment Me. [Nom] gagne 10h/semaine avec [Produit]"
- └── Contenu : témoignage client détaillé
- └── Chiffres concrets : avant/après
- └── CTA : "Obtenir les mêmes résultats"

J+7 : Check-in

- └── Sujet : "Comment se passe votre essai ?"
- └── Contenu : court, personnel
- └── Proposer un call de 10 min si besoin d'aide
- └── Répondre directement à cet email (pas de no-reply)

J+10 : Urgence douce

- └── Sujet : "Il vous reste 4 jours d'essai gratuit"
- └── Contenu : résumé de l'usage pendant le trial
 - └── "Vous avez créé X documents et gagné ~Y heures"
- └── ROI personnalisé : "Y heures × votre taux horaire = Z €"
- └── CTA : "Continuer pour seulement [prix] €/mois"

J+13 : Dernière chance

- └── Sujet : "Votre essai se termine demain"
- └── Contenu : offre spéciale -20% si souscription annuelle
- └── Rappel de la valeur : "Ne perdez pas accès à vos documents"
- └── CTA : "Souscrire maintenant (-20%)"

J+14 : Expiration

- └── Sujet : "Votre essai est terminé – mais vos données sont conservées"
- └── Contenu : rassurer (rien n'est perdu)
- └── Offre de dernière minute : -10% supplémentaire
- └── "Répondez à cet email si vous avez des questions"

J+21 (winback) : Relance

- └── Sujet : "Nous avons ajouté [nouvelle feature] depuis votre essai"
- └── Contenu : ce qui a changé / amélioré
- └── Offre de réessay : 7 jours supplémentaires gratuits
- └── CTA : "Réactiver mon essai"

// Réalisation de A à Z – Guide complet

Phase 1 : Validation (semaine 1-4)

SEMAINE 1-2 : Recherche terrain

- Choisir 1 niche (recommandation : Niche 1 – avocats)
- Identifier 30-50 professionnels cibles
 - LinkedIn (filtrer par profession + ville + taille cabinet)
 - Annuaires professionnels (CNB pour avocats, CNOM pour médecins)
 - Réseau personnel
- Contacter pour des interviews (15-20 min)
 - Script : "Je développe un outil AI pour [profession]. J'aimerais comprendre vos défis quotidiens. 15 min de votre temps ?"
 - Objectif : 10-15 interviews minimum
- Questions clés :
 - "Quelle tâche vous prend le plus de temps au quotidien ?"
 - "Combien de temps passez-vous sur [tâche spécifique] par semaine ?"
 - "Avez-vous déjà essayé des outils AI ? Pourquoi ça n'a pas marché ?"
 - "Si un outil pouvait faire X en 2 min au lieu de 2h, combien paieriez-vous ?"
 - "Quelles sont vos contraintes de confidentialité ?"
- Documenter les patterns (quels problèmes reviennent le plus)

SEMAINE 3-4 : Validation quantitative

- Créer une landing page (Claude Code, 1 jour)
 - Headline : "[Profession] : gagnez [X]h par semaine grâce à l'IA"
 - Explication du produit (3-5 features clés)
 - Vidéo démo (mockup, 2 min)
 - Formulaire inscription beta (email + infos cabinet)
 - Pricing indicatif (pour tester la willingness to pay)
- Acquisition de trafic
 - LinkedIn ads ciblées (budget : 500-1 000 €)
 - Posts LinkedIn organiques (3-5 posts/semaine)
 - Groupes Facebook/LinkedIn de la profession
 - Cold emails personnalisés (50-100)
- Gate de validation
 - ✅ 50+ inscriptions beta → GO
 - ⚠️ 20-50 inscriptions → ajuster le positionnement
 - ❌ < 20 inscriptions → pivoter de niche
- Livrable : liste de 10-20 beta testers engagés

Phase 2 : Dataset et fine-tuning (semaine 5-8)

SEMAINE 5-6 : Collecte de données

- Sources de données (selon la niche)
 - AVOCATS :
 - Légifrance API (textes de loi, jurisprudence – gratuit, public)
 - DILA open data (Journal Officiel – gratuit)
 - Doctrine.fr (jurisprudence – partenariat ou scraping légal)
 - Conclusions et actes anonymisés (fournis par beta testers)
 - Manuels et traités juridiques (domaine public ou licence)
 - MÉDECINS :
 - HAS (recommandations, référentiels – gratuit, public)
 - CCAM / NGAP (nomenclature – Ameli open data)
 - Vidal / BDPM (médicaments – API publique)
 - Comptes-rendus anonymisés (fournis par beta testers)
 - Terminologie SNOMED CT / CIM-10
 - COMPTABLES :
 - Plan comptable général (public)
 - BOFiP (doctrine fiscale – gratuit)
 - Conventions collectives (Légifrance)
 - Écritures comptables anonymisées (beta testers)
- Préparation du dataset
 - Format : instruction / input / output (alpaca-style)
 - Exemples pour avocats :
 - instruction: "Rédige des conclusions en réponse"
input: "[résumé des faits + arguments adverses]"
output: "[conclusions structurées avec visa des textes]"
 - instruction: "Trouve la jurisprudence applicable"
input: "[question juridique]"
output: "[arrêts pertinents avec références exactes]"
 - instruction: "Rédige une mise en demeure"
input: "[contexte + demande]"
output: "[lettre formatée]"
 - Volume cible : 2 000-5 000 exemples de qualité
 - Qualité > quantité : chaque exemple vérifié par un pro
- Nettoyage et anonymisation
 - Supprimer toute donnée personnelle
 - Remplacer noms/adresses par des placeholders
 - Vérifier la conformité RGPD du dataset

SEMAINE 7-8 : Fine-tuning sur RTX 5090

- Modèle de base recommandé
 - Llama 3.1 8B Instruct → meilleur rapport qualité/VRAM
 - Alternative : Mistral 7B v0.3 (bon en français)
 - Si besoin de plus de capacité : Llama 3.1 70B en QLoRA 4-bit
- Configuration fine-tuning (QLoRA)
 - Framework : Unslloth (2x plus rapide, économise VRAM)
 - Quantization : 4-bit (QLoRA)
 - LoRA rank : 64-128
 - Learning rate : 1e-4 à 2e-4
 - Epochs : 3-5
 - Batch size : 4-8 (RTX 5090 32 Go le permet)
 - Temps estimé : 2-6h pour 3000 exemples sur 8B
- Setup technique
 - Python 3.11 + CUDA 12.x
 - pip install unslloth transformers datasets peft
 - Script de training (voir ci-dessous)
- Évaluation
 - Métriques automatiques (perplexity, BLEU, ROUGE)
 - Évaluation humaine : 50 exemples jugés par des pros
 - Comparaison : modèle fine-tuné vs GPT-4 vs modèle de base

Phase 3 : Construction du produit (semaine 9-14)

SEMAINE 9-10 : Architecture technique

- Stack recommandée
 - Frontend : Next.js 16 + Tailwind (Claude Code)
 - Backend : Next.js API routes + tRPC
 - Base de données : Supabase (PostgreSQL + Auth + Storage)
 - Inference AI : vLLM ou Ollama sur RTX 5090
 - RAG : LangChain + pgvector (embeddings dans Supabase)
 - File processing : LlamaParse ou Docling (extraction PDF)
 - Recherche : Typesense ou Meilisearch (recherche full-text)
 - Auth : Supabase Auth (SSO pour cabinets)
 - Billing : Stripe (abonnements mensuels)
- Architecture globale
 - Client (Next.js) → API (Next.js) → Queue (BullMQ/Redis)
 - Worker (vLLM sur RTX 5090) → Résultat → Client
 - Le RAG enrichit chaque requête avec du contexte vérifié
 - Audit trail : chaque requête/réponse est loguée
- Sécurité (CRITIQUE pour les pros)
 - Chiffrement end-to-end (TLS 1.3)
 - Données stockées chiffrées au repos (AES-256)
 - Authentification forte (MFA obligatoire)
 - Isolation des données par cabinet (multi-tenancy stricte)
 - Hébergement France (ou on-premise)
 - Logs d'audit immutables
 - Conformité RGPD documentée
 - PIA (Privacy Impact Assessment) réalisé
 - Si santé : hébergement HDS certifié (OVH, Scaleway)

SEMAINE 11-12 : Développement des features core

- Claude Code : développement rapide de toute l'app
- Features MVP (OBLIGATOIRES pour le lancement)
 - 1. Authentification + onboarding
 - Inscription, login, MFA, profil cabinet
 - Choix de la spécialité (pour personnaliser les templates)
 - 2. Interface de rédaction assistée
 - Éditeur de texte riche (TipTap ou Plate)
 - Panel AI latéral (suggestions, reformulations)
 - Templates par type de document
 - Historique des générations
 - 3. Moteur RAG (recherche dans les sources)
 - Upload de documents de référence (lois, jurisprudence, interne)
 - Recherche sémantique dans la base
 - Citations automatiques avec sources
 - Détection d'hallucinations (cross-check avec les sources)
 - 4. Génération de documents
 - Input structuré (formulaire adapté au type de doc)
 - Génération en streaming (effet machine à écrire)
 - Export Word/PDF
 - Historique + versioning
 - 5. Dashboard utilisateur
 - Documents récents
 - Statistiques d'utilisation
 - Crédits restants / plan actuel
 - 6. Admin cabinet
 - Gestion des utilisateurs
 - Paramétrage des templates
 - Logs d'activité
- Features V2 (post-lancement, mois 2-4)
 - 7. Transcription audio → document (Whisper + LLM)
 - 8. Analyse de documents uploadés (extraction, résumé)

- └─ 9. Intégration outils métier (RPVA pour avocats, logiciels de cabinet)
- └─ 10. Mode collaboration (plusieurs utilisateurs sur un doc)
- └─ 11. API pour intégration dans des outils tiers
- └─ 12. App mobile (consultation hors bureau)

SEMAINE 13-14 : Beta privée

- └─ Inviter les 10-20 beta testers de la phase 1
- └─ Onboarding individuel (call de 15 min + démo)
- └─ Slack/Discord privé pour le feedback
- └─ Itérer quotidiennement (Claude Code)
 - └─ Fix bugs en priorité
 - └─ Top 3 feature requests
 - └─ Ajuster les prompts/templates selon les retours
- └─ Métriques à suivre
 - └─ NPS (Net Promoter Score) → viser > 40
 - └─ Rétention J7 et J30
 - └─ Documents générés/utilisateur/semaine
 - └─ Taux de satisfaction par type de document
- └─ Gate : 5+ utilisateurs qui reviennent SANS relance = GO pour le lancement

Phase 4 : Go-to-market (mois 4-6)

- MOIS 4 : Lancement officiel
 - Site web professionnel
 - Claude Code : site complet en 2-3 jours
 - Positionnement : "L'assistant AI privé des [avocats/médecins]"
 - Pages clés :
 - Home : problème → solution → preuve sociale → CTA
 - Features : tour des fonctionnalités avec démos
 - Sécurité : page dédiée (RGPD, chiffrement, HDS)
 - Pricing : plans clairs
 - Blog : articles SEO
 - Démo : formulaire ou vidéo Loom
 - Lancement beta publique
 - Accès sur inscription (créer de la rareté)
 - Annonce LinkedIn (post personnel + campagne ads)
 - Email aux contacts de la phase validation
 - PR : contacter LegalTech Magazine, Le Monde du Droit, etc.
 - Premiers clients payants
 - Convertir les beta testers en payants
 - Objectif : 10-20 clients payants en fin de mois 4

MOIS 5-6 : Acquisition

- Canal 1 : SEO / Content marketing
 - Articles optimisés : "Comment l'AI transforme [profession]"
 - Guides pratiques : "Automatiser [tâche] avec l'IA"
 - Comparatifs : "ChatGPT vs [ton produit] pour les avocats"
 - Claude Code : générer 2-4 articles/semaine
- Canal 2 : LinkedIn (le canal #1 pour les pros)
 - Posts 3-5x/semaine
 - Cas clients / témoignages
 - Webinaires mensuels (démo live, 30 min)
 - LinkedIn Ads ciblées (budget : 1 000-3 000 €/mois)
- Canal 3 : Partenariats
 - Barreaux et ordres professionnels
 - Éditeurs de logiciels métier (intégration)
 - Formations professionnelles (présentation dans les CFC, etc.)
 - Influenceurs du secteur (LegalTech, MedTech)
- Canal 4 : Sales directes
 - Liste de 200-500 cabinets cibles
 - Séquence email personnalisée (3-5 emails)
 - Démo live 20 min + follow-up
 - Objectif : 5-10% de conversion
- Canal 5 : Événements
 - Salons LegalTech / HealthTech / Salon des Experts-Comptables
 - Meetups professionnels locaux
 - Webinaires co-organisés avec des partenaires
- Objectif mois 6 : 50-100 clients payants

Phase 5 : Scale et moat (mois 7-12+)

INFRASTRUCTURE

- Migrer progressivement vers le cloud
 - RTX 5090 locale pour le dev + fine-tuning
 - Cloud (Scaleway GPU, OVH AI) pour la production
 - Garder une option on-premise pour les gros clients
- Haute disponibilité
 - Load balancer + auto-scaling
 - Backup automatisé
 - SLA 99.9% pour le plan Enterprise
- Monitoring
 - Uptime, latence, qualité des réponses
 - Alertes automatiques

MOAT (avantage concurrentiel durable)

- 1. Données propriétaires
 - Chaque utilisation enrichit ta base de connaissances
 - Plus de clients = meilleur modèle = plus de clients (flywheel)
 - Personne ne peut reproduire ton dataset
- 2. Fine-tuning spécialisé
 - Modèle entraîné sur des milliers d'exemples réels du domaine
 - Itération continue : re-fine-tuning mensuel avec les nouvelles données
 - RTX 5090 : coût de fine-tuning quasi nul
- 3. Intégrations métier
 - Connecté aux outils de la profession (difficilement remplaçable)
 - Plus le client intègre, plus le switching cost augmente
- 4. Conformité réglementaire
 - Certifications (HDS pour la santé, etc.)
 - Coûteux et long à obtenir → barrière à l'entrée
- 5. Communauté et réseau
 - Base d'utilisateurs fidèles
 - Partenariats avec les institutions
 - Réputation dans le secteur

EXPANSION

- Nouvelles spécialités au sein de la même profession
 - Avocat droit du travail → droit des affaires → droit de la famille
- Nouvelles professions adjacentes
 - Avocats → Notaires → Huissiers → Experts-comptables
- Nouveaux marchés géographiques
 - Belgique, Suisse, Luxembourg (droit civil similaire)
 - Canada (Québec – droit civil francophone)
- Upsell
 - Plan basique → Pro → Enterprise
 - Add-ons : transcription, API, on-premise
 - Services : formation, consulting, personnalisation
- Embauche
 - Mois 6-9 : 1 dev fullstack + 1 sales/customer success
 - Mois 9-12 : 1 ML engineer + 1 account manager
 - Mois 12+ : équipe complète selon les revenus

// Modèle économique détaillé

Plans tarifaires (exemple pour la niche avocats)

PLAN	PRIX/MOIS	CIBLE	FEATURES
Solo	149 €/mois	Avocat indépendant	1 utilisateur, 100 docs/mois, templates de base, support email
Cabinet	399 €/mois	Cabinet 2-10 avocats	5 utilisateurs, 500 docs/mois, templates avancés, RAG personnalisé, support prioritaire
Enterprise	999 €/mois	Cabinet 10-50+ avocats	Utilisateurs illimités, docs illimités, SSO, API, SLA 99.9%, account manager dédié
On-premise	2 499 €/mois	Grands cabinets, institutions	Installation sur serveurs client, personnalisation complète, formation sur site

Structure de coûts

Coûts fixes mensuels :

- Claude Code Max 20x : 200 €
- Hébergement cloud (Scaleway/OVH) : 200-800 € (selon volume)
- Supabase Pro : 25 €
- Domaine + Email : 15 €
- Outils (Stripe, Resend, monitoring) : 50-100 €
- Électricité RTX 5090 : 30-50 €
- Total fixe : ~520-1 190 €/mois

Coûts variables :

- Inference GPU cloud : ~0.02-0.05 €/requête (si cloud)
- Inference locale RTX 5090 : ~0 € (déjà payé)
- Stockage documents : ~0.01 €/Go/mois
- Support client : 0 € (toi au début) → 2 500 €/mois (1 CSM à mi-temps)

Coûts d'acquisition :

- LinkedIn Ads : 1 000-3 000 €/mois
- Contenu/SEO : 0 € (Claude Code)
- Événements : 500-1 000 €/trimestre
- Total acquisition : 1 000-4 000 €/mois

Métriques unitaires

MÉTRIQUE	VALEUR CIBLE
ARPU (Average Revenue Per User)	250–350 €/mois
CAC (Customer Acquisition Cost)	300–800 €
LTV (Lifetime Value)	6 000–12 000 € (24 mois × ARPU)
LTV/CAC ratio	8–15× (excellent, cible > 3×)
Churn mensuel	< 3% (les pros sont fidèles si l'outil marche)
Marge brute	85–92% (SaaS pur, inference locale)
Payback period	1–3 mois (CAC récupéré rapidement)

// Projections de rentabilité – 24 mois

Scénario conservateur

MOIS	CLIENTS	ARPU	MRR	COÛTS	RÉSULTAT NET	MRR CUMULÉ
1–2	0 (beta)	—	0 €	1500 €	-1500 €	0 €
3	10	200 €	2 000 €	2 000 €	0 €	2 000 €
4	20	220 €	4 400 €	2 500 €	1 900 €	4 400 €
5	35	240 €	8 400 €	3 000 €	5 400 €	8 400 €
6	55	260 €	14 300 €	4 000 €	10 300 €	14 300 €
9	120	280 €	33 600 €	8 000 €	25 600 €	33 600 €
12	200	300 €	60 000 €	15 000 €	45 000 €	60 000 €
18	400	320 €	128 000 €	30 000 €	98 000 €	128 000 €
24	700	350 €	245 000 €	50 000 €	195 000 €	245 000 €

Scénario optimiste (product-market fit fort)

MOIS	CLIENTS	MRR	RÉSULTAT NET
6	100	30 000 €	22 000 €
12	400	120 000 €	90 000 €
18	1000	320 000 €	240 000 €
24	2000	700 000 €	500 000 €

Scénario pessimiste (pivot nécessaire)

MOIS	CLIENTS	MRR	RÉSULTAT NET
6	20	4 000 €	-500 €
12	60	15 000 €	5 000 €
18	100	28 000 €	12 000 €
24	150	45 000 €	20 000 €

Seuil de rentabilité

Point mort mensuel (breakeven) :

- └── Coûts fixes : ~1 000 €/mois (début)
- └── ARPU moyen : 250 €/mois
- └── Marge brute : 90%
- └── Breakeven = $1\ 000 / (250 \times 0.90) = 5$ clients payants
- └── → Atteignable dès le mois 3

Point mort avec 1 salarié :

- └── Coûts fixes + salaire : ~5 000 €/mois
- └── Breakeven = $5\ 000 / 225 = 23$ clients
- └── → Atteignable vers le mois 5-6

Point mort avec équipe complète (4 personnes) :

- └── Coûts totaux : ~20 000 €/mois
- └── Breakeven = $20\ 000 / 225 = 89$ clients
- └── → Atteignable vers le mois 8-10

// Risques et mitigations

RISQUE	PROBABILITÉ	IMPACT	MITIGATION
Réglementation IA restrictive (EU AI Act)	Moyenne	Élevé	Se positionner comme "conforme" dès le départ, veille active
Concurrence des GAFAM (GPT-5 pour avocats)	Élevée	Moyen	Moat = données FR + intégrations métier + RGPD + relation client
Difficulté à obtenir des datasets de qualité	Élevée	Élevé	Partenariats avec des cabinets, commencer avec du RAG sur données publiques
Cycle de vente B2B trop long	Moyenne	Moyen	Freemium + self-serve pour les solos, sales dédié pour les cabinets
Hallucinations AI (erreurs juridiques/médicales)	Élevée	Critique	RAG obligatoire, citations systématiques, disclaimer, human-in-the-loop
Hébergement HDS (santé) coûteux	Faible	Moyen	Commencer par les avocats (pas de HDS requis), santé en V2
RTX 5090 insuffisante pour le scale	Moyenne	Faible	Migration cloud progressive, la RTX reste pour le fine-tuning

// Checklist de lancement

PRÉ-REQUIS TECHNIQUES :

- RTX 5090 installée et fonctionnelle
- CUDA + Python + Unsloth configurés
- Modèle de base téléchargé (Llama 3.1 8B)
- vLLM ou Ollama fonctionnel
- Dataset de 2000+ exemples collecté et nettoyé
- Fine-tuning réalisé et évalué
- RAG pipeline opérationnel (embeddings + pgvector)
- App Next.js déployée (Vercel ou VPS FR)
- Supabase configuré (auth, DB, storage)
- Stripe billing configuré
- Chiffrement E2E implémenté
- Audit trail fonctionnel

PRÉ-REQUIS BUSINESS :

- 10+ interviews professionnels réalisées
- Landing page live avec inscriptions beta
- 50+ inscriptions beta collectées
- 10-20 beta testers actifs
- NPS beta > 40
- Pricing validé par les beta testers
- Site web professionnel en ligne
- CGV / mentions légales / politique de confidentialité
- Statut juridique créé (SAS recommandé)
- Compte Stripe opérationnel
- LinkedIn professionnel optimisé
- 5+ articles SEO publiés

PRÉ-REQUIS RÉGLEMENTAIRES :

- RGPD : registre des traitements
- RGPD : PIA réalisé
- RGPD : DPO désigné (ou référent)
- Si santé : hébergement HDS
- Mentions "outil d'aide, pas de conseil juridique/médical"
- Conditions d'utilisation claires sur les limitations de l'IA

ARCHITECTURE TECHNIQUE – RAG-first, fine-tuning later

Le fine-tuning N'EST PAS un prérequis pour lancer la Niche 1 (assistant juridique). Cette section explique pourquoi, et quelle architecture adopter pour un MVP rapide et efficace.

// Pourquoi le RAG bat le fine-tuning pour le juridique

Le problème du fine-tuning pour les avocats

ASPECT	FINE-TUNING	RAG
Citations jurisprudence	Hallucine des n° d'arrêt inventés	Cite UNIQUEMENT ce qui est dans la base → vérifiable
Mise à jour des lois	Snapshot figé = obsolète en semaines	Mise à jour de la base = toujours à jour
Données d'entraînement	Besoin de 2000–5000 exemples confidentiels	Besoin de sources publiques (Légifrance)
Temps de mise en place	4–6 semaines (collecte + training + éval)	1–2 semaines (scraping + embeddings)
Coût	Temps + compute GPU	Quasi gratuit (données publiques)
Risque juridique	Le modèle peut reproduire des données client	Le modèle ne stocke rien, il consulte

Le point critique : un avocat qui cite un arrêt de la Cour de cassation qui N'EXISTE PAS devant un juge, c'est une faute professionnelle. Le RAG élimine ce risque. Le fine-tuning ne peut pas le garantir.

Ce que chaque couche apporte

COUCHE 1 : LLM de base (Llama 3.1 70B sur RTX 5090)

└ Ce qu'il sait déjà :

- ├ Écrire en français juridique correct
- ├ Structurer un raisonnement juridique
- ├ Comprendre les concepts de droit
- └ Reformuler, résumer, synthétiser

└ Ce qu'il ne sait PAS :

- ├ Les arrêts récents (coupure de connaissances)
- ├ Les références exactes (il hallucine)
- └ Les spécificités de votre client

COUCHE 2 : System prompt + few-shot

└ Ce que ça ajoute :

- ├ Le format exact des conclusions (visa, discussion, dispositif)
- ├ Les conventions de rédaction par juridiction
- ├ Le ton et le style attendus
- ├ La règle "cite uniquement les sources fournies"
- └ 2-3 exemples de documents parfaitement rédigés

└ Coût : 0 € – juste du prompt engineering

COUCHE 3 : RAG (Retrieval Augmented Generation)

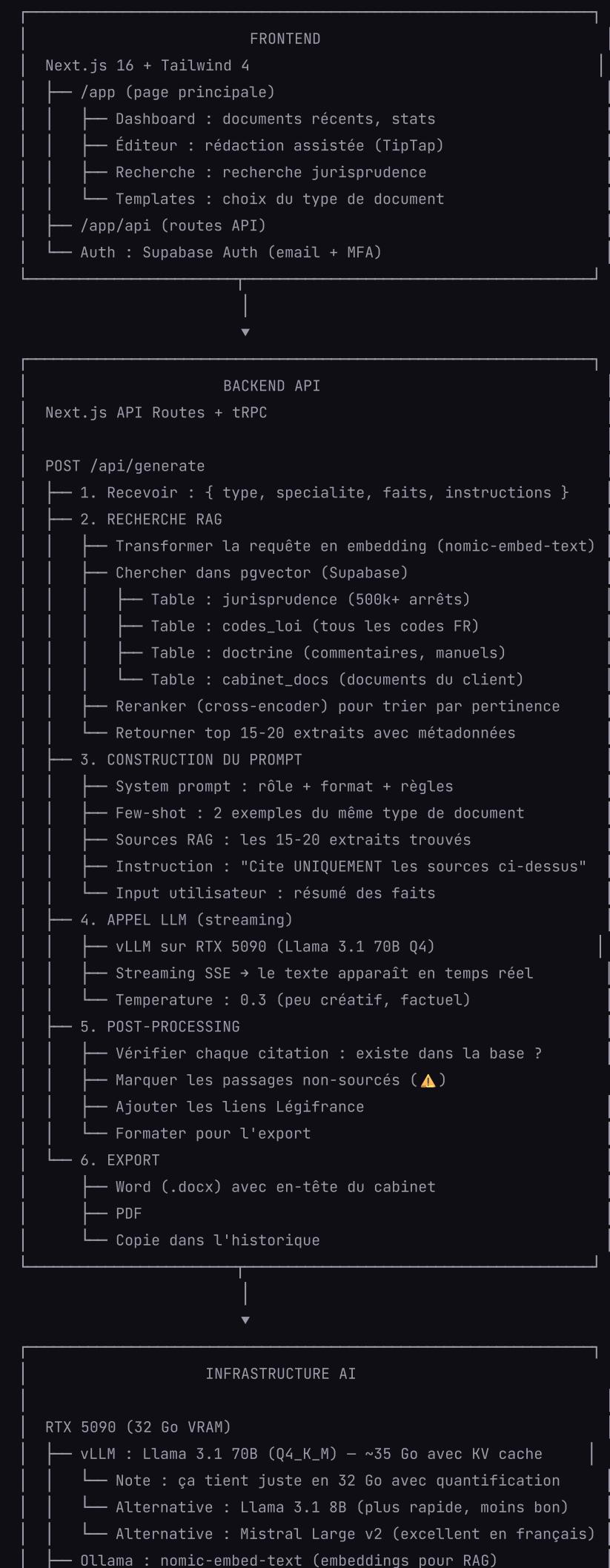
└ Ce que ça ajoute :

- ├ Jurisprudence exacte avec n° d'arrêt, date, juridiction
- ├ Articles de loi à jour (codes civil, travail, pénal, etc.)
- ├ Doctrine et commentaires
- ├ Documents internes du cabinet (si uploadés)
- └ TOUT est vérifiable et traçable

└ Coût : stockage embeddings (~10 €/mois Supabase)

RÉSULTAT : 90-95% de la qualité d'un fine-tuning, en 1/4 du temps.

L'architecture technique détaillée



Pipeline de données RAG — Légifrance

ÉTAPE 1 : Récupérer les données (1 fois + mises à jour hebdo)

- API Légifrance (PISTE) – gratuit, clé API requise
 - └ https://api.piste.gouv.fr/
 - └ Endpoints : /consult/code, /consult/juri, /search
 - └ Limites : 2000 requêtes/jour (suffisant)
- DILA Open Data
 - └ Dump complet des codes en XML/JSON
 - └ Jurisprudence : Cour de cassation, Conseil d'État, CA
 - └ ~2M de décisions disponibles
- EUR-Lex (droit européen, optionnel)
- Script Node.js/Python pour l'extraction

ÉTAPE 2 : Découper en chunks

- Chaque décision → découper en sections
 - └ Faits, Moyens, Motivation, Dispositif
 - └ Taille cible : 500-1000 tokens par chunk
 - └ Overlap de 100 tokens entre chunks
- Chaque article de loi → 1 chunk
 - └ Avec métadonnées : code, livre, titre, chapitre, article
- Enrichir avec les métadonnées
 - └ Juridiction, date, n° de pourvoi/décision
 - └ Mots-clés, domaine de droit
 - └ Textes cités (articles de loi référencés)
- Volume estimé : 2-5M de chunks

ÉTAPE 3 : Générer les embeddings

- Modèle : nomic-embed-text (Ollama, RTX 5090)
 - └ 768 dimensions, bon en français
 - └ ~3000 chunks/minute sur RTX 5090
 - └ 2M chunks = ~11h (une nuit)
- Alternative : text-embedding-3-small (OpenAI)
 - └ Meilleure qualité, mais coût API
 - └ 2M chunks × 0.02 \$/1M tokens ≈ 5-10 \$
- Stocker dans Supabase pgvector

ÉTAPE 4 : Index et recherche

- Index HNSW sur pgvector (recherche rapide)
- Requête : embedding de la question → top 20 plus proches
- Reranking : cross-encoder pour affiner (optionnel mais recommandé)
 - └ BAAI/bge-reranker-base (tourne sur RTX 5090)
 - └ Prend les top 20 → retrie les top 5-10
- Résultat : les 5-10 extraits les plus pertinents avec sources

ÉTAPE 5 : Mise à jour continue

- Cron job hebdomadaire : nouveaux arrêts + modifications législatives
- Embeddings incrémentaux (seulement les nouveaux documents)
- Notification aux utilisateurs : "Nouveau texte dans votre domaine"
- Coût marginal : quasi nul

Quand introduire le fine-tuning (roadmap)

MOIS 1-3 : RAG ONLY ← ON EST ICI

- └── Llama 3.1 70B (ou 8B) + RAG + system prompts
- └── Qualité : 85-90% du potentiel max
- └── Avantage : rapide à builder, facile à itérer
- └── Focus : valider le product-market fit

MOIS 4-6 : RAG + FINE-TUNING LÉGER

- └── Collecter 500-1000 exemples réels (avec accord clients)
 - └── Documents générés + corrections manuelles de l'avocat
 - └── = dataset "gold standard" de ce que le modèle DEVRAIT produire
- └── Fine-tuner Llama 3.1 8B en QLoRA
 - └── Le 8B fine-tuné remplace le 70B pour les tâches simples
 - └── = 4-8x plus rapide, même qualité sur le domaine
 - └── Le 70B reste en backup pour les cas complexes
- └── Qualité : 92-95% du potentiel max
- └── Avantage : meilleure latence, coûts réduits

MOIS 7-12 : RAG + FINE-TUNING SPÉCIALISÉ

- └── Un adapter LoRA par spécialité juridique
 - └── Adapter "droit-travail" (entraîné sur conclusions prud'hommes)
 - └── Adapter "droit-penal" (entraîné sur conclusions correctionnelles)
 - └── Adapter "droit-affaires" (entraîné sur assignations commerciales)
 - └── etc.
- └── Hot-swap : le bon adapter est chargé selon la spécialité de l'avocat
- └── Qualité : 95-98% du potentiel max
- └── Avantage : qualité spécialisée imbattable = MOAT technique

JAMAIS : Fine-tuning SANS RAG

- └── Le fine-tuning seul ne résout pas les hallucinations
- └── Le RAG reste TOUJOURS nécessaire pour les citations
- └── Les deux sont complémentaires, pas exclusifs

GUIDE DE DÉMARRAGE — Coder JurisAI avec Claude Code Opus 4.6

Ce guide est un plan d'exécution pas à pas pour construire le MVP de JurisAI (Niche 1: assistant juridique pour avocats) en utilisant Claude Code Max 20x avec le modèle Opus 4.6. Durée cible : 4 semaines.

// Pourquoi Opus 4.6 est le copilote idéal pour ce projet

CAPACITÉ OPUS 4.6	APPLICATION DIRECTE
Raisonnement complexe	Architecure RAG, prompt engineering, edge cases juridiques
Code fullstack	Next.js + tRPC + Supabase + vLLM – tout le stack
Contexte massif	Peut ingérer des specs longues, des schémas DB, des prompts système entiers
Itération rapide	Claude Code Max 20x = pas de rate limit, itérations continues
Qualité production	Code typé, testé, documenté dès la première génération

// Semaine 1: Fondations (jours 1-5)

Jour 1: Setup du projet

COMMANDES CLAUDE CODE – Jour 1

1. Créer le projet

Prompt : "Crée un projet Next.js 16 avec App Router, Tailwind CSS 4, TypeScript strict, Supabase (auth + DB + storage), tRPC, et Resend pour les emails. Structure en features modulaires."

2. Configurer Supabase

Prompt : "Configure Supabase avec :

- Auth email + MFA
- Tables : users, cabinets, documents, generations, templates
- pgvector extension activée
- RLS policies : chaque cabinet ne voit que ses données
- Types TypeScript auto-générés"

3. Configurer le thème

Prompt : "Crée un design system dark premium pour une app juridique. Couleurs : bleu nuit profond (#0a1628), or (#c9a84c) pour les accents, texte clair. Typographie : serif pour les headings (Playfair Display), sans-serif pour le body (DM Sans). Composants de base : Button, Input, Card, Modal, Sidebar, TopBar."

Ce que Claude Code Opus 4.6 produit en ~2h :

- Projet Next.js complet bootstrappé
- Schéma DB avec pgvector
- Auth flow (login, register, MFA)
- Design system avec composants de base
- Layout app (sidebar + topbar + main)

Jour 2 : Pipeline RAG

COMMANDES CLAUDE CODE – Jour 2

```
# 1. Script d'ingestion Légifrance
Prompt : "Crée un script Node.js qui :
1. Se connecte à l'API PISTE (Légifrance)
2. Récupère les articles des codes principaux (civil, travail, pénal,
   commerce, procédure civile)
3. Récupère les arrêts de la Cour de cassation des 5 dernières années
4. Découpe chaque document en chunks de 500-800 tokens avec overlap 100
5. Génère les embeddings via Ollama (nomic-embed-text)
6. Stocke dans Supabase pgvector
7. Ajoute les métadonnées : juridiction, date, n° pourvoi, mots-clés
Inclus un mode 'update' qui ne traite que les nouveaux documents."
```

```
# 2. API de recherche RAG
Prompt : "Crée une route tRPC 'search.query' qui :
1. Prend une question en langage naturel
2. Génère l'embedding de la question
3. Cherche les 20 plus proches dans pgvector (cosine similarity)
4. Retourne les résultats avec : texte, source, date, lien Légifrance
Ajoute un reranker optionnel (cross-encoder) pour affiner."
```

```
# 3. Test du pipeline
Prompt : "Écris un test qui :
1. Ingère 100 arrêts de la Cour de cassation (droit du travail)
2. Pose 5 questions juridiques types
3. Vérifie que les résultats sont pertinents
4. Affiche les sources trouvées"
```

Ce que Claude Code produit :

- Script d'ingestion complet (Légifrance → chunks → embeddings → pgvector)
- Route de recherche RAG avec reranking
- Tests fonctionnels
- Le pipeline tourne sur ta RTX 5090 via Ollama

Jour 3 : Moteur de génération

COMMANDES CLAUDE CODE – Jour 3

1. Connecteur vLLM

Prompt : "Crée un module qui se connecte à vLLM (local, port 8000)
pour envoyer des requêtes à Llama 3.1 70B. Support du streaming SSE.
Gestion des erreurs et retry. Timeout configurable."

2. System prompts par type de document

Prompt : "Crée un fichier de configuration avec les system prompts

pour les types de documents juridiques suivants :

- Conclusions au fond (TGI/TJ)
- Conclusions en référé
- Mise en demeure
- Contrat type
- Courrier à l'adversaire
- Requête

Chaque prompt doit :

- Définir le rôle (assistant juridique expert en droit [spécialité])
- Imposer la structure du document (visa, discussion, dispositif...)
- Imposer de ne citer QUE les sources fournies en contexte
- Inclure 1-2 exemples few-shot de documents bien rédigés
- Spécifier le ton (professionnel, précis, sans familiarité)"

3. Route de génération

Prompt : "Crée la route tRPC 'generate.document' qui :

1. Reçoit : { type, specialite, faits, instructions_specifiques }
2. Fait une recherche RAG avec les faits comme query
3. Construit le prompt complet (system + few-shot + RAG + user input)
4. Appelle vLLM en streaming
5. Post-process : vérifie les citations, ajoute les liens
6. Sauvegarde le document dans la DB
7. Retourne le stream au client"

Jour 4-5 : Interface utilisateur core

COMMANDES CLAUDE CODE – Jour 4-5

1. Page Dashboard

Prompt : "Crée la page dashboard avec :

- Liste des documents récents (type, date, aperçu)
- Stats : documents ce mois, temps estimé gagné
- Accès rapide : 'Nouveau document' avec choix du type
- Sidebar : navigation (Dashboard, Documents, Recherche, Templates)

Design dark premium, juridique, professionnel."

2. Page de génération de document

Prompt : "Crée la page de génération avec :

- Étape 1 : choix du type de document (cards visuelles)
- Étape 2 : formulaire adapté au type
 - └ Conclusions : résumé des faits, prétentions, arguments
 - └ Mise en demeure : destinataire, objet, demande, délai
 - └ Contrat : type, parties, objet, durée, conditions
- Étape 3 : génération en streaming (le texte apparaît en temps réel)
- Étape 4 : édition (TipTap editor riche)
- Actions : copier, exporter Word, exporter PDF, sauvegarder

Animations smooth, loading states, error handling."

3. Page de recherche jurisprudence

Prompt : "Crée la page de recherche avec :

- Barre de recherche en langage naturel
- Filtres : juridiction, date, domaine de droit
- Résultats : carte par arrêt avec juridiction, date, n° pourvoi, extrait pertinent surligné, lien Légifrance
- Pagination + infinite scroll
- Action : 'Utiliser dans un document' → injecte dans l'éditeur"

4. Page Sécurité (CRITIQUE pour la conversion)

Prompt : "Crée une page /securite qui détaille :

- Où sont hébergées les données (France, datacenter spécifique)
- Chiffrement (AES-256 au repos, TLS 1.3 en transit)
- Multi-tenancy stricte (isolation par cabinet)
- Audit trail (chaque action loguée)
- MFA disponible
- Politique de rétention des données
- Contact DPO
- Conformité RGPD détaillée

Design sobre, rassurant, professionnel. Badges visuels."

// Semaine 2 : Features essentielles (jours 6-10)

Jour 6-7 : Templates et export

COMMANDES CLAUDE CODE

1. Système de templates

Prompt : "Crée un système de templates de documents juridiques.

Chaque template a :

- Un nom, une description, un type (conclusions, contrat, courrier...)
- Des champs dynamiques ({{ partie_1 }}, {{ date }}, {{ objet }})
- Un system prompt spécialisé
- 1-2 exemples few-shot intégrés

Crée 10 templates de base pour le droit du travail :

conclusions prud'homales, mise en demeure employeur, contrat de travail CDI, lettre de licenciement, saisine CPH, attestation, etc.

Interface : galerie de templates avec preview et 'Utiliser'."

2. Export Word (.docx)

Prompt : "Crée un module d'export Word qui :

- Prend le contenu généré (markdown/HTML)
- Le convertit en .docx avec mise en forme professionnelle
- Ajoute un en-tête personnalisable (logo + adresse du cabinet)
- Numérote les pages
- Utilise des styles juridiques (police : Garamond 12, interligne 1.5)

Utilise docx.js (npm docx)."

3. Export PDF

Prompt : "Ajoute l'export PDF en utilisant le même template que Word.

Utilise puppeteer ou react-pdf."

Jour 8-9 : Auth, billing, onboarding

COMMANDES CLAUDE CODE

```
# 1. Flow d'inscription optimisé
Prompt : "Crée le flow d'inscription :
1. Email + mot de passe (pas de champs inutiles)
2. Vérification email
3. Choix de la spécialité (droit du travail, pénal, etc.)
4. Nom du cabinet (optionnel)
5. Onboarding : tutorial interactif en 3 étapes
- Étape 1 : 'Voici comment générer vos conclusions' (guidé)
- Étape 2 : 'Essayez avec un cas exemple' (pré-rempli)
- Étape 3 : 'Essayez avec votre propre dossier' (zone libre)
Le Moment Aha doit arriver avant la fin de l'étape 2 (< 3 min)."
```

```
# 2. Stripe billing
Prompt : "Intègre Stripe avec 3 plans :
- Solo : 149 €/mois (1 user, 100 docs/mois)
- Cabinet : 399 €/mois (5 users, 500 docs/mois)
- Enterprise : 999 €/mois (illimité)
Toggle mensuel/annuel (-20%).
Trial gratuit 14 jours sans CB.
Webhooks Stripe pour gérer les événements.
Page /pricing avec comparaison des plans."
```

```
# 3. Emails transactionnels
Prompt : "Configure Resend avec les templates React Email :
- Welcome email (après inscription)
- Document généré (résumé + lien)
- Séquence trial : J+1, J+3, J+7, J+10, J+13, J+14
- Facture (après paiement)
- Winback J+21 (après expiration trial)
Chaque email : design sobre, professionnel, responsive."
```

Jour 10 : Résumé de pièces + upload

COMMANDES CLAUDE CODE

```
Prompt : "Crée une feature 'Analyser des pièces' qui :
1. Upload de PDF (drag & drop, multi-fichiers)
2. Extraction du texte (pdf-parse ou LlamaParse)
3. Résumé automatique de chaque pièce (LLM)
4. Extraction des arguments clés
5. Suggestion de contre-arguments
6. Interface : liste des pièces avec résumé + détails expandable
7. Action : 'Utiliser dans les conclusions' → injecte le contexte
Limiter à 50 pages par pièce et 10 pièces par dossier."
```

// Semaine 3 : Polish et landing page (jours 11-15)

Jour 11-12 : QA et polish

COMMANDES CLAUDE CODE

1. Tests E2E

Prompt : "Écris des tests Playwright pour les parcours critiques :

- Inscription → onboarding → première génération
- Génération de conclusions (les 3 types principaux)
- Recherche de jurisprudence
- Export Word et PDF
- Upload et analyse de pièces
- Paiement Stripe (mode test)"

2. Responsive + accessibilité

Prompt : "Vérifie et corrige le responsive sur mobile et tablette.

L'app doit être utilisable sur iPad (beaucoup d'avocats en audience).

Vérifie l'accessibilité : navigation clavier, contraste, aria-labels."

3. Performance

Prompt : "Optimise les performances :

- Lazy loading des pages
- Streaming SSE optimisé (pas de buffering)
- Images optimisées
- Bundle splitting

Cible : Lighthouse > 90 sur toutes les métriques."

Jour 13-14 : Landing page conversion

COMMANDES CLAUDE CODE

Prompt : "Crée la landing page de JurisAI selon cette structure :

HERO :

- Badge : 'Utilisé par les cabinets les plus exigeants'
- Titre : 'Rédigez vos conclusions en 20 minutes au lieu de 4 heures'
- Sous-titre : 'L'assistant juridique AI qui respecte le secret professionnel. Vos données restent en France.'
- CTA : 'Essayer gratuitement – 14 jours, sans CB'
- Démo interactive : zone de texte où le visiteur peut tester UNE génération sans inscription

SECTION PROBLÈME :

'Vous passez 60% de votre temps à rédiger. Pendant ce temps, vos dossiers s'accumulent.'

SECTION SOLUTION (3 features) :

1. Conclusions automatiques – 'De 4h à 20 min' + GIF démo
2. Jurisprudence vérifiée – 'Citations exactes, jamais d'hallucination'
3. Secret professionnel – 'Chiffrement AES-256, hébergement France'

SECTION PREUVE SOCIALE :

3-5 témoignages (on mettra les vrais plus tard, utiliser des placeholders réalistes pour le moment)

SECTION SÉCURITÉ :

Badges RGPD, chiffrement, hébergement FR, audit trail

SECTION PRICING :

3 plans (Solo 149€, Cabinet 399€, Enterprise sur devis)

Toggle mensuel/annuel

FAQ :

6 questions/réponses sur la sécurité, la qualité, la compatibilité

FOOTER CTA :

Rappel du bénéfice + CTA final

Design : dark premium, tons bleu nuit + or, professionnel.

Animations d'entrée subtiles (Framer Motion).

Score Lighthouse > 95."

Jour 15 : Déploiement

COMMANDES CLAUDE CODE

```
# 1. Déploiement
Prompt : "Configure le déploiement :
- Frontend : Vercel (ou VPS FR si on veut tout en France)
- vLLM : tourne en local sur la RTX 5090
- Tunnel : Cloudflare Tunnel pour exposer vLLM
- Supabase : déjà en cloud
- Variables d'environnement : documenter toutes les env vars
- Domaine : jurisai.fr (ou similaire)
- SSL : automatique via Vercel/Cloudflare"

# 2. Monitoring
Prompt : "Configure le monitoring :
- Sentry pour les erreurs
- Vercel Analytics pour le trafic
- Custom metrics : documents générés, latence, satisfaction
- Alertes : si vLLM down, si latence > 10s, si erreur rate > 5%"
```

// Semaine 4 : Beta et itération (jours 16-20)

Jour 16-17 : Ingestion des données

EXÉCUTION (pas Claude Code, exécution directe)

```
# 1. Lancer vLLM sur RTX 5090
vllm serve meta-llama/Llama-3.1-70B-Instruct \
--quantization awq \
--max-model-len 8192 \
--port 8000

# 2. Lancer Ollama pour les embeddings
ollama serve
ollama pull nomic-embed-text

# 3. Lancer le script d'ingestion
node scripts/ingest-legifrance.mjs --codes civil,travail,penal \
--jurisprudence cass,ce --years 5

# 4. Vérifier
# → X arrêts ingérés
# → Y articles de loi
# → Z chunks dans pgvector
# → Tester 5 requêtes de recherche
```

Jour 18-19 : Tests avec vrais avocats

PROCESS BETA

- Inviter 5-10 beta testers (avocats contactés en phase validation)
- Call individuel de 15 min : onboarding + démo
- Leur demander de tester avec un VRAI dossier
- Collecter le feedback :
 - La qualité des conclusions est-elle suffisante ?
 - Les citations sont-elles correctes ?
 - Le format est-il conforme à leurs habitudes ?
 - Qu'est-ce qui manque ?
 - Paieraient-ils 149 €/mois pour ça ?
- Itérer avec Claude Code (fix les problèmes en temps réel)

Jour 20 : Ajustements et lancement soft

COMMANDES CLAUDE CODE – ajustements post-feedback

Les prompts ici dépendent du feedback réel.

Exemples typiques :

Prompt : "Le format des conclusions ne correspond pas aux usages du Tribunal judiciaire de Paris. Modifie le template 'conclusions_fond' pour utiliser la structure : I. Rappel des faits, II. Discussion, III. Par ces motifs..."

Prompt : "Les avocats veulent pouvoir ajouter leurs propres modèles de phrases récurrentes. Ajoute un système de 'snippets' personnalisés que l'utilisateur peut créer et insérer dans l'éditeur."

Prompt : "La latence de génération est trop longue (15 sec). Passe le modèle en Llama 3.1 8B pour les mises en demeure et courriers (documents simples), garde le 70B pour les conclusions."

// Récapitulatif — Les 20 jours

JOUR	LIVRABLE	EFFORT CLAUDE CODE
1	Projet bootstrappé, DB, auth, design system	~3h de prompting
2	Pipeline RAG complet (Légifrance → pgvector)	~3h
3	Moteur de génération (vLLM + prompts + post-processing)	~3h
4–5	UI core (dashboard, génération, recherche, sécurité)	~5h
6–7	Templates juridiques + export Word/PDF	~4h
8–9	Auth, Stripe billing, onboarding, emails	~4h
10	Upload et analyse de pièces	~2h
11–12	Tests, responsive, accessibilité, performance	~4h
13–14	Landing page conversion	~4h
15	Déploiement + monitoring	~2h
16–17	Ingestion données + configuration vLLM	~3h (exécution)
18–19	Beta tests avec vrais avocats	~2h (itérations)
20	Ajustements + lancement	~2h
Total	MVP complet prêt pour les premiers clients	~41h de travail

41 heures de travail réel sur 20 jours ouvrés = ~2h/jour en moyenne. Le reste du temps : prospection, interviews clients, contenu LinkedIn.

// Commandes essentielles pour le dev quotidien

```
# Lancer l'environnement de dev
cd /c/Users/amarm/SaaS/JurisAI

# Terminal 1 : Next.js
npm run dev

# Terminal 2 : vLLM (RTX 5090)
vllm serve meta-llama/Llama-3.1-70B-Instruct --quantization awq --port 8000

# Terminal 3 : Ollama (embeddings)
ollama serve

# Terminal 4 : Claude Code
# → C'est ici que tout se passe
# → Opus 4.6, Max 20x, pas de rate limit
# → Itérer en continu

# Workflow type avec Claude Code :
# 1. Décrire la feature en détail (contexte + specs + contraintes)
# 2. Claude Code génère le code
# 3. Tester manuellement
# 4. Si bug → décrire le bug → Claude Code fix
# 5. Commit + push → Vercel déploie
# 6. Répéter
```

Tips pour maximiser Opus 4.6

1. CONTEXTE RICHE

- └ Plus tu donnes de contexte, meilleur est le résultat
- └ Inclure : specs, schéma DB, exemples, contraintes, design tokens
- └ Opus 4.6 gère des prompts de 10 000+ tokens sans problème

2. ITÉRER VITE

- └ Claude Code Max 20x = pas de rate limit
- └ Ne pas chercher la perfection au premier prompt
- └ Générer → tester → affiner → générer → tester
- └ 5 itérations rapides > 1 prompt parfait

3. PARALLÉLISER

- └ Lancer plusieurs agents Claude Code en parallèle
- └ Agent 1 : feature frontend
- └ Agent 2 : feature backend
- └ Agent 3 : tests
- └ = 3x plus rapide

4. UTILISER LES SKILLS

- └ /commit pour les commits propres
- └ /epct pour les features complexes (Explore-Plan-Code-Test)
- └ frontend-design pour l'UI (obligatoire par convention)

5. SAUVEGARDER LES BONS PROMPTS

- └ Quand un prompt donne un excellent résultat
- └ Le sauvegarder dans un fichier prompts/
- └ Le réutiliser comme template pour les features similaires