AI Support

Every AI-assisted surface the platform offers, the providers that power them, what counts against the daily quota, what does not, and how the Premium gate shapes access.

Up to date as of v1.32.0

CONTENTS

- What AI Support is
- Two providers: Claude for Service Agreements, Gemini for everything else
- Platform-managed API keys
- The shared daily AI quota
- What counts against the daily quota
- What does not count against the quota
- AI Writing Assistant
- AI document analysis at booking
- Service Agreement — AI Populate (extract + apply)
- Service Agreement — AI Review (paid via Store)
- Service Agreement — AI gov-fees shortlist
- Active File Review — AI triage extraction
- AI ID document extraction (AFR + Transfer Room)
- Modules that reuse the SA extractor
- Matter Work Ledger — AI code suggester
- Multilingual booking-page translation
- Client-side Service Agreement translation
- The help-center chat (Help ’ AI chat)
- Subscription tier and AI access posture
- Privacy posture
- The AI-translation disclaimer discipline
- Interactions with other modules
- Troubleshooting

What AI Support is

AI Support is not one screen — it is the constellation of AI-assisted features that show up inside almost every module on the platform. The AI Writing Assistant draft-pad lives at /dashboard/ai-assistant; the booking flow has an AI document analyzer; the Service Agreement Builder has AI Populate (extract a draft from uploaded

documents) AND AI Review (audit the draft for compliance issues); Active File Review has AI triage on intake submissions AND AI auto-fill from a photographed ID; Co-Counselling Agreements and Service Proposals reuse the same Service Agreement extractor; Transfer Room's Information Card has AI auto-fill from an ID document; the Matter Work Ledger has an AI code suggester; the booking page itself can be translated at runtime into ten languages; signed Service Agreements can be translated clause-by-clause for the client. The Help center chat at /dashboard/help is also AI-backed. This chapter walks through each surface and explains the rules.

Two providers: Claude for Service Agreements, Gemini for everything else

The platform runs two large-language-model providers in parallel because each one is better at a specific class of task. Anthropic's Claude (currently the Sonnet 4.6 model) powers every Service Agreement AI surface — AI Populate extraction, AI Review compliance audit. The choice is deliberate: Claude's tool-use enforcement guarantees the model returns a strict-schema JSON object on every call, which matters for legal-document workflows where a malformed AI response cannot be salvaged at the parser. Claude also reads PDFs natively as document blocks and reads images as image blocks, so the multimodal extraction path on SA AI Populate handles passport scans + DOCX consultation notes + signed PDFs without preprocessing.

Google's Gemini (currently the 2.5 Flash model) powers everything else — the AI Writing Assistant draft-pad, the booking AI document analyzer, the AFR intake triage extractor, the AFR ID document extractor, the Matter Work Ledger AI code suggester, the help-center chat, the admin-side error formatter, and the runtime translation pipeline that powers multilingual booking pages and client-side SA translation. The Gemini surfaces benefit from Gemini's much higher rate limits and lower per-call cost, which lets the platform keep those flows in the day-to-day quota rather than rationing them as paid actions.

Platform-managed API keys

All AI provider keys are managed by Investatech at the platform level, not by individual tenants. You never have to sign up for an Anthropic account, a Google AI Studio account, or any other AI provider account; you do not have to manage API keys, billing, or quota with those providers; you do not have to worry about a credit-card expiry on the AI side breaking AI features. The platform abstracts all of that away. Every AI request goes through the platform's server-side keys, the platform absorbs the per-call cost into the subscription model (or charges your wallet for the few explicitly-paid actions), and the platform handles provider outages, key rotations, model deprecations, and rate-limit retries without involving you.

The shared daily AI quota

Most AI surfaces share one tenant-wide daily quota. Basic tenants get 10 quota-consuming AI calls per UTC day. Premium tenants (paid or in their Premium trial) get 50 calls per UTC day. Tenants whose subscription has lapsed, been cancelled, or trial-expired get 0 — they cannot fire any quota-consuming AI surface until they re-subscribe (the help-center chat stays usable; see the dedicated section on it). The quota resets at midnight UTC; the AI Writing Assistant page surfaces the remaining count for the day in the header so you know how many calls you have left.

The quota is tenant-wide, not per-user. A four-person team on Premium shares the 50 calls across the four of them; if your assistant burns 30 calls in the morning doing AI Populate runs on a complicated SA, you have 20 left for the rest of the day. Failed calls (provider down, model returned a malformed response, network glitch) still consume one quota slot — the platform logs the failed attempt against the quota so a misbehaving integration cannot accidentally burn the same slot a hundred times. The one exception is the explicitly-paid SA AI Review, which is metered separately against your Store wallet (see the dedicated section).

What counts against the daily quota

- AI Writing Assistant — every Generate click at /dashboard/ai-assistant consumes one quota slot.

- Booking AI document analyzer — when a client uploads documents to the booking flow and asks for an AI summary, each summary generation consumes one slot on the licensee's quota.
- Service Agreement AI Populate (extract) — running the extractor against uploaded documents on the SA Builder consumes one slot. The Apply step does not consume a slot (it is purely deterministic).
- Service Agreement gov-fees AI shortlist — when the SA Builder asks AI to shortlist matching government fees against your matter description, each shortlist consumes one slot.
- Active File Review intake AI extraction — when the public intake form submits with documents, AI triage extracts authority + issue type + deadlines + observations. Each intake submission consumes one slot.
- Active File Review ID extraction — when the licensee uploads a photographed ID at the intake step, AI extracts the personal-details fields. Each ID extract consumes one slot.
- Co-Counselling Agreement AI Populate — reuses the SA extractor; each extract run consumes one slot.
- Service Proposals AI Populate — reuses the SA extractor; each extract run consumes one slot.
- Booking Notes AI Populate — when the licensee asks AI to extract a structured note set from the booking's uploaded files, each extract run consumes one slot.
- Transfer Room Information Card AI ID extraction — when the RCIC uploads an ID document to populate the IC, each extraction consumes one slot.
- Matter Work Ledger AI code suggester — Premium-only; each suggestion call consumes one slot.

What does not count against the quota

- Help-center chat — the conversational AI at /dashboard/help is uncapped per tenant. Every team member can use it as much as they want, every day, regardless of subscription tier (even trial-expired tenants can

keep asking the chat because answering platform questions does not require a paid relationship). The chat is grounded to the platform's documentation; it is not a general-purpose AI you can use for unrelated work.

- Service Agreement AI Review — this is the one paid surface. AI Review is always priced at 1 Store point per run, charged against your Store wallet rather than the daily quota. The reasoning: the audit is genuinely expensive (it walks the full clause spine + runs the compliance checks against every populated field), the quota would force prioritisation between Review and the other quota surfaces, and the per-action billing is more honest. The wallet check happens before the run; insufficient wallet returns 402 with a top-up CTA.
- Client-side runtime translation (booking page + SA translation) — the translation pipeline calls Google Cloud Translation API directly, not Gemini. It does not consume the daily AI quota; access is gated by the Premium subscription itself (only Premium tenants have additional languages enabled on the booking page + the client-side SA translation modal). There is no per-call charge or quota on translation — the constraint is the Premium subscription, not the per-call quota.
- Admin-side AI error formatter — internal-only tool platform operators use to clean up error-report formatting. Does not consume the tenant's quota (tenants do not see it).

AI Writing Assistant

Sidebar ’ AI Assistant opens /dashboard/ai-assistant. The page is Premium-only — Basic tenants land on a friendly upsell card explaining that the Writing Assistant is part of the Premium tier. The Assistant itself is a draft-pad: you type a prompt describing what you want (e.g. 'draft a follow-up email to a client whose Open Work Permit application was approved, congratulating them and reminding them to upload their letter of employment'), pick an optional tone (professional, friendly, formal), and click Generate. The platform calls Gemini server-side and renders the response inline. You then copy-and-paste the draft into wherever you actually need it (an email client,

the booking notes, a custom letter); the Assistant does NOT send anything itself — it is a draft-pad, not a delivery channel.

The Assistant has no memory between runs — each Generate is a fresh call with only the prompt you typed; the model does not remember your previous prompts, your tenant identity, or your client list. That isolation is deliberate: the Assistant is for one-off drafting help, not for a conversational session that builds context over time. If you need a different output, edit the prompt and click Generate again — each click consumes one quota slot.

AI document analysis at booking

When a client attaches documents to a booking, the booking flow surfaces an AI document summary feature. The client (or the licensee on the dashboard) can ask the AI to read the uploaded documents and produce a structured summary highlighting key facts, dates, and potential issues. This is the licensee's first-look-at-the-matter aid — instead of opening every PDF before the consultation, you skim the AI-generated summary on the booking detail page and arrive at the call with the context already in hand.

The summary is informational and explicitly does NOT constitute legal advice — the AI is not making representations about the matter, just summarising the documents the client uploaded. The summary is generated against the licensee's daily quota (one call per summary generation); the feature is Premium-only at the platform level. The summary is regenerable — if the client uploads a new document after the original summary, you can re-run the summary to incorporate it; each rerun consumes one quota slot.

Service Agreement — AI Populate (extract + apply)

Inside the SA Builder, the AI Populate workflow is the headline AI feature on the platform. You drag in supporting documents (a consultation memo, a passport scan, a refusal letter, a prior application), optionally add a tenant prompt with extra context ('the client wants to add their spouse as a dependent'), click Extract, and Claude returns a structured set of populated-field proposals — client name, client email, matter type, client business

details if applicable, family members, scope items, recommended template, recommended service components. Each proposal is rendered as a pre-ticked checkbox row you can review; ticking Apply All threads the accepted proposals through the SA draft, applies the recommended template at a chosen fee, and the Builder reloads with everything populated.

Two design choices keep AI Populate trustworthy. First, the field-path allowlist: Claude can only propose values for fields the platform explicitly exposes on the allowlist (around 30 paths covering client, matter, scope, parties, designated persons). A proposal targeting a path outside the allowlist is silently dropped server-side; the AI cannot accidentally rewrite a clause body or change the fee structure. Second, the Apply step is deterministic — it walks each accepted proposal, applies it via a switch-case on the field path, and respects the Fill-empty vs Replace mode toggle (Fill-empty skips proposals targeting non-empty fields). Together the two mean the AI Populate output is bounded and reviewable; you decide what to keep before anything writes to the draft.

Service Agreement — AI Review (paid via Store)

AI Review is the one paid AI surface on the platform. From the SA Builder's detail page (status=draft only), clicking the AI Review button runs Claude against the full populated draft and returns a structured audit report: which clauses look complete, which have wording that may need a second look, which CICC Code of Professional Conduct obligations the draft addresses vs leaves silent on, which fields look like placeholder values that probably should not ship to the client. The report is read-only — AI Review does not write back to the draft; you use the findings to manually edit. Each run costs 1 Store point against your wallet (not against the daily AI quota); the wallet is checked before the run fires, so a known-insufficient wallet returns 402 with a top-up CTA before any AI cost is incurred.

AI Review is always priced at 1 point per run — there is no Premium discount or bundle option. Failed runs (the model returned malformed output, Claude was briefly down, the draft has no content to audit) do not consume the wallet — the platform only charges after the audit completes successfully. The audit run also writes a marker

row to ai_assistant_usage with a sentinel flag indicating it should NOT count against the daily quota; this keeps the Reports surface clean about who used what.

Service Agreement — AI gov-fees shortlist

The Government Fees section of the SA Builder offers a 300+ item catalogue (IRCC + CBSA + IRB + provincial / territorial fees). Picking from a list of 300+ is tedious; the AI shortlist accelerates the pick. With a one-sentence matter description ('this is a spousal sponsorship from inside Canada with one dependent child'), AI returns a ranked shortlist of 8-12 catalogue items that probably apply — application fees, biometric fees, RPRF for the principal applicant + dependents, etc. The Builder renders the shortlist as a pre-ticked checklist; you uncheck anything that does not apply and Add the rest in one click. Each shortlist call consumes one daily quota slot.

Active File Review — AI triage extraction

Every Active File Review intake submission fires an AI extraction step against the documents the client uploaded. The extractor (Gemini, with a strict response schema) returns a structured triage payload: authority (IRCC / CBSA / IRB / other), issue type (RFE / PFL / refusal / interview invitation / I don't know / other), application type if identifiable, application number if visible on the letter, response deadline if present, deadline confidence (high / medium / low), and a one-paragraph triage summary plus observations. The dashboard surfaces the extraction alongside the visitor's own answers, so when an Owner or Admin opens the new AFR row they see both 'what the client said' and 'what the AI saw' side by side. The extraction also drives the pathway recommendation card on the triage view.

AFR intake extraction runs against the tenant's daily quota — failed extractions consume a slot but do not block the AFR submission itself (the row is still created with the visitor's typed answers; only the AI-derived fields are missing). If you have hit the daily cap, new AFR submissions still record but skip the AI triage step; you can manually triage the matter from the dashboard detail page.

AI ID document extraction (AFR + Transfer Room)

Two surfaces share the same AI ID-document extractor. On Active File Review, the licensee can upload a photographed ID at the intake step; AI extracts the personal-details fields (full name, date of birth, sex, document number, country of issue, expiry date). On Transfer Room's Information Card, the RCIC can upload an ID document to populate the IC; the same extractor runs against the photo. Both surfaces are RCIC-only — the platform takes the position that AI-derived identity data is a regulated-professional input, not something staff should be filling in autonomously.

The ID extractor never stores the image. The photo is sent to the AI provider, the structured response is returned, the response is shown to the licensee for review-and-accept, the photo bytes are dropped from server memory at the end of the request. The accepted fields write to the AFR row or the Information Card; the image itself never persists to disk or to your tenant's Drive. Each ID extract consumes one quota slot. Failed extractions (image too blurry, ID type the model has not seen before, network glitch) do not block the licensee from typing the fields in by hand instead.

Modules that reuse the SA extractor

Three other modules reuse the SA AI Populate extractor verbatim rather than maintaining their own AI integration. Co-Counselling Agreements (CCA) reuses the extractor because CCA documents share the same identity + matter + family + scope shape as SAs; SA-shaped paths map onto CCA fields at the apply layer (SA-only paths like Sponsor and Third-Party Payer are silently dropped because CCA has no equivalent). Service Proposals reuses it because a proposal is essentially a pre-SA — same client identity, same matter context, same service components. Booking Notes reuses it because the structured-note shape is a subset of an SA's identity + scope shape. All three reuses consume one quota slot per extract run; the underlying extractor and the field-path allowlist are identical to SA's, just mapped to different destinations at the apply step.

Matter Work Ledger — AI code suggester

When you log a Matter Work Ledger entry inside a Transfer Room, the entry needs one or more work codes (MWL-100 client communication, MWL-200 intake, MWL-500 forms preparation, etc. — a 62-code catalogue). Picking the right codes by hand is doable but tedious; the AI suggester accelerates it. Type a one-paragraph description of what you did ('reviewed the client's spousal sponsorship application, prepared the questionnaire for the sponsor, sent the questionnaire to both parties by email'), click Suggest codes, and Gemini returns a ranked shortlist of 1-5 codes it thinks apply, each with a one-sentence rationale. Click each suggestion to add it to your entry. The suggester is Premium-only (the underlying Gemini call counts toward the daily quota); Basic tenants still have full access to the MWL — they just pick codes from the dropdown manually.

Multilingual booking-page translation

Premium tenants can enable additional languages on their public booking page beyond the platform's two native languages (English and Quebec French). The choices are the eight other Phase 10C languages: Spanish, Arabic, Mandarin Chinese, Portuguese, Russian, Persian (Farsi), Hindi, and Urdu. When enabled (Settings ’ Branding ’ Client page languages), the public booking page shows a language picker; visitors pick a language and the page's strings translate at runtime via Google Cloud Translation API. The translation is cached per (source string, target language) so the second visitor in Mandarin does not re-pay the translation cost. Service titles, descriptions, your firm's bio, the agreement preface — all of it flows through the same pipeline.

Two important caveats. First, the translation is machine-assisted and best-effort; legal-precision strings (the actual CICC-aligned Initial Consultation Agreement clauses) ship in English / Quebec French for legal force, with the translated version provided as a courtesy preview only. Second, the translation pipeline does NOT consume the daily AI quota — it calls a different Google API (Cloud Translation v2) with its own per-character cost the platform absorbs into the Premium subscription. The constraint on access is the Premium subscription itself, not a per-call quota.

Client-side Service Agreement translation

Once a Service Agreement is sent to a client, the public portal at /agreement/<token> renders the agreement in either English or Quebec French (whichever you set as the agreement's render locale). For Premium tenants who have additional client-page languages enabled, the portal offers a small globe icon next to every clause that opens a translation modal — the client can pick a target language from the enabled set, the platform calls Google Cloud Translation against the clause text, and the modal shows the translated version side-by-side with the original. A standing disclaimer on the modal explicitly notes the translation is machine-generated and the legal-force version remains the original-language clause.

When the agreement itself is rendered in Quebec French (because the licensee set serviceOfficialLanguage to French at the SA Builder stage), the rendered document carries an AI-translation disclaimer block at the end of the body — the platform treats the FR rendering as machine-assisted from the canonical EN clause text, so the disclaimer is built into the document itself rather than being a portal-only courtesy. The disclaimer wording is locked in code so every tenant's FR agreements carry the same legal framing.

The help-center chat (Help ’ AI chat)

Sidebar ’ Help opens /dashboard/help. Alongside the FAQ panel, the page carries an AI-backed chat box: type a platform question ('how do I add a co-counsel to a Transfer Room', 'what does the trial deadline cron actually do', 'why is my AI Populate proposal greyed out') and the chat replies with a grounded answer derived from the platform's documentation. The chat is conversational — it remembers the current session's question-and-answer history so you can ask follow-ups in plain language ('and what about deleting that co-counsel later'). The chat is uncapped per tenant: every team member can use it as much as they want, every day, regardless of subscription tier (Basic, Premium, comp-Premium, trial, trial-expired tenants all have unlimited access).

The chat is grounded — it is not a general-purpose AI you can use for unrelated work. The system prompt is locked to the platform's documentation and the chat will politely decline (or admit ignorance) on off-topic questions. The trade-off is that the chat is reliable as a platform reference but cannot help you write a letter to a client or do legal

research on an immigration question. For those needs, the AI Writing Assistant (Premium-only, draft-pad) is the right tool.

Subscription tier and AI access posture

AI access scales with subscription tier. Basic tenants get the help-center chat (uncapped), the AI document analyzer at booking (10 calls/day shared with other quota surfaces), the AFR intake extraction (same shared 10/day pool), the ID document extractor (same pool), and the SA-extractor-reuse modules' AI Populate (same pool). Basic tenants do NOT get the AI Writing Assistant page or the Matter Work Ledger code suggester (Premium-gated). Premium tenants get all of the above with the daily pool raised to 50 calls/day, plus the Writing Assistant + the MWL suggester + the multilingual booking translation + client-side SA translation. Trial-expired / lapsed / cancelled tenants get the help-center chat only (everything else returns 0 quota and a 'subscribe to keep using AI' upsell on the relevant route).

AI Review is the one explicitly-paid surface — every tenant on every tier can run it as long as they have at least 1 Store point in their wallet. The choice is deliberate: the audit is genuinely expensive, the value is clear, and the per-action billing is more honest than rationing it inside the daily pool. There is no Premium discount on AI Review; it costs 1 point per run for everyone.

Privacy posture

Every AI call sends content to the provider (Anthropic or Google) — there is no way around that for a cloud-LLM workflow. The platform treats provider-side privacy as table stakes: both Anthropic and Google have published commitments that API content is not used to train their models, both provide enterprise-grade data-handling contracts that govern what they can and cannot do with content sent over their APIs, and the platform's commercial terms with both providers track those commitments. None of this should be confused with end-to-end encryption:

the content is decrypted on the platform server before being sent to the provider, and decrypted on the provider server to be processed by the model. The model itself sees the content in plaintext for the duration of the API call.

What the platform does NOT do: train any model on your data, share content with other tenants, retain AI request bodies beyond the time needed to log the call against the daily quota, or send AI content to any provider beyond the two named (Anthropic for Claude, Google for Gemini + Cloud Translation). What the platform DOES retain: a usage record per AI call (the date, the tenant, the cost in either quota slots or Store points, the success/failure outcome) — this is the data the quota counter and the Reports surfaces read from. The /security marketing page carries the canonical disclosure of what gets sent where; refer to it when you need the wording to share with a client who asks about AI use on their matter.

The AI-translation disclaimer discipline

Three different surfaces carry an explicit AI-translation disclaimer because three different surfaces let AI translation reach a legally-meaningful output. The booking page's multilingual translation carries an inline notice on every page reminding the visitor that translated strings are machine-generated. The client-side SA translation modal carries the same notice on every clause translation. The most prominent one is the Quebec French rendering of the full SA: when an SA is rendered with serviceOfficialLanguage='French', the document itself carries a standing disclaimer block at the end of the body explicitly noting the FR text is machine-assisted from the canonical EN clauses and that in any conflict between the two, the canonical EN text governs. The disclaimer wording is locked in code (no per-tenant override) so the legal framing is uniform across the platform.

Interactions with other modules

AI Support is by definition cross-cutting — every module that surfaces an AI feature is touching the same daily-quota counter (or the Store wallet for AI Review). One practical implication: if you run a heavy AI Populate session against three complicated Service Agreements in the morning, you may not have many quota slots left for

AFR intake triage in the afternoon. The Reports surface and the AI Writing Assistant page both show the current usage count so you can see when you are approaching the cap. The other cross-cutting interaction is with Store: the Store wallet is the source of truth for AI Review's per-run cost, so a low Store balance can block AI Review even when the daily quota is full.

Troubleshooting

- AI Writing Assistant page shows the Premium upsell — Basic tenants do not have access to the dedicated Writing Assistant page. The help-center chat is available on Basic; use that for platform questions. For client-letter drafting on Basic, the per-module AI surfaces (AFR intake summary, booking AI analyzer) cover most of what you would draft from scratch.
- AI call returns 429 'Daily quota exceeded' — you have hit the 10/day (Basic) or 50/day (Premium) cap. The cap resets at midnight UTC; wait until then. Premium tenants close to the cap should be aware that the AI Writing Assistant, AI Populate, and AI gov-fee shortlist all share the same pool. Trial-expired tenants get 0/day — re-subscribe to restore AI access.
- AI Review returns 402 'Insufficient Store points' — AI Review is not in the daily-quota pool; it is paid per run from your Store wallet. Top up the wallet (Sidebar ’ Store ’ pick a top-up tier) and retry. Each AI Review costs 1 point; the wallet check happens before the run fires.
- AI Populate proposals look fragmentary or wrong — the model returned what it returned. Two likely causes: the documents you uploaded did not actually contain the information the AI was looking for, or the documents were lower-quality scans the model could not read cleanly. Try uploading higher-resolution scans, or add a clearer tenant prompt at the top of the modal ('the client is a temporary resident applying for a study permit extension; the spousal information in these documents is for context only and should not be added as a family member').

- ID extractor returned wrong personal-details fields — image quality matters. The model performs best on a flat, well-lit, in-focus scan of the ID document. Photos taken at an angle, with glare, or with a blurred section frequently produce field-by-field errors. Re-take the photo and re-upload; the previously-extracted values can be edited manually before accepting the new extraction.
- Booking page is supposed to be translated into Spanish but visitors see English — confirm three things: Premium subscription is active (translation is Premium-only), Spanish is enabled in Settings ’ Branding ’ Client page languages, and the visitor's URL or browser locale resolves to 'es'. The language picker on the booking page lets the visitor manually override; if the picker works but the auto-detect does not, the browser is sending an unexpected locale string.
- Help-center chat refuses to answer — the chat is grounded to platform documentation and politely declines off-topic questions ('how do I file a Spousal Sponsorship application' is an immigration question, not a platform question; the chat will steer you to use the AI Writing Assistant or talk to a regulated consultant). Stay on platform-mechanics questions and the chat will engage.