

Data Protection Impact Assessment — spelling.live

Status: Completed by the controller, AK Shah Ltd, and approved by its director on 2026-10-13. This is Spelling.Live's own assessment of its own processing. A school using the service remains a controller for its pupils' data and should carry out its own DPIA; this document is published to support that process, not to replace it. Summary for schools at /for-schools/trust ; the full text is in the schools data protection pack. This is an internal accountability document (UK GDPR Art. 35 + ICO Children's Code). It is **not** legal advice — a one-off professional review is recommended.

Service	spelling.live — a spelling-practice app for children (ages ~5–11)
Controller	AK Shah Ltd (company number 17491910), trading as Spelling.Live, contact: support@spelling.live . Was Mr Manoj Shah as a sole trader until the company took over the service.
DPO	Not appointed — below the mandatory threshold (no large-scale systematic monitoring; no large-scale Art. 9 data). The controller is the responsible person.
Date of this DPIA	2026-06-30
Next review	On any material change to processing, or annually

1. Why a DPIA is needed

The service is likely to be accessed by children, so under the ICO **Age Appropriate Design Code (Children's Code)** a DPIA is expected. It also involves profiling of learning performance and processing of children's data, which are DPIA screening triggers.

2. Describe the processing

Nature & purpose. Parents create an account and set up child profiles. The app delivers spelling practice and tracks progress to personalise learning and report to the parent. AI features generate practice content (sentences, hints, clues, meanings), read handwriting/photographed worksheets, transcribe spoken spelling, and produce British-English audio.

Data subjects. (a) Parents/guardians (account holders); (b) Children (learners).

Personal data processed.

- *Parent:* email, account identifiers, sign-in provider metadata, IP/device/server logs (technical).
- *Child:* name or label, learning level (Early/Fluent), year group, practice activity (words attempted, correct/incorrect, hint usage, misspellings), per-word stats, **handwriting image thumbnails**, transient voice input, report events.
- *Child (imported from Google Classroom, teachers only):* the pupil's **Google Classroom user ID** (`child_external_identities`) - an opaque dedupe key so a repeat import never creates the same child twice; never an email or photo (the two read-only scopes we request do not grant

them). Google Classroom here is a **data source** the teacher authorises under their own Google account, not a processor of our data: the teacher's access token is used server-side for one import and never stored.

- *Child (imported from a school MIS via Wonde, teachers only)*: the pupil's **Wonde student ID** (`child_external_identities`, provider `wonde`) - the same opaque dedupe key; only the pupil's preferred forename and surname are read, never date of birth, UPN, SEN, attendance, contacts or photos. Wonde is a **data source** the SCHOOL authorises in the Wonde portal (the school is the controller of its MIS data); our app-wide Wonde token is server-side only and never logged.
- *Child (derived/inferred)*: spelling-error **pattern classifications** produced deterministically from the child's own misspellings, **Skills Check feature profiles** (`skills_checks.feature_profile`), and **AI coach tips personalised to that child's recurring patterns** (`child_word_tips`). This is inference about a child's *learning difficulty*, a step beyond recording what they did, so it is treated as its own data category: RLS-scoped to the owning parent (or the child's own login), covered by account/child delete and the data export, and never shared between children. Tips that are a function only of the word and the misspelling (`word_tips`, `mistake_tips`) hold **no child identifier** and are deliberately shared across learners — the boundary between the shared and child-scoped stores is enforced in code and documented in `docs/coach-tips-personalisation.md`.

Special category data (Art. 9). None intended. Static handwriting images are used for letter-formation feedback, **not** for biometric identification, so they are not Art. 9 biometric data. Voice input is transient (transcribed, not stored as a voiceprint). Verified: no free-text fields are presented to children; the only textareas in the app are parent-side word-list inputs (paste/type words). No field invites disclosure of health, religion, ethnicity, or other Art. 9 data.

Volume/scope. Small-scale, parent-mediated. Children do not sign up independently; an adult sets up and supervises.

Processors / recipients. Vercel (hosting), Supabase (database, storage, auth), AI providers OpenAI, Google and Anthropic, and PostHog (product analytics, EU Cloud). See the AI sub-processor register (`data/ai-subprocessors.json`, rendered at `/ai-transparency.html`). Under our API configuration, **inputs are not used to train the providers' models**. PostHog receives anonymous profile UUIDs and numeric/boolean event properties — no child names, practice words, or free-text input; autocapture is disabled. **One exception, on the public pages only:** the site-search box on the help, blog, patterns and resources pages (not the app) records the search term (`site_search_performed`, `site_search_result_clicked`), lowercased, cut to 60 characters, and withheld entirely when it contains an email address, a phone number or a long run of digits, a web address, or runs past 100 characters (`searchTermForAnalytics` in `js/site-search.js`). Anonymous visitors create no person profile there, and the term is stripped from the `/search` URL on every event so it never travels unscrubbed (`redactSearchQuery`). Residual risk: a reader could type a child's first name into the box; the scrub cannot catch a name, which is why the box is not offered inside the app a child uses. Session replay is enabled to analyse where **parents** drop off (churn): it is privacy-by-default — every displayed text node and every typed input value is masked, and only parent-facing static UI (the sign-in modal and "add a child" copy, which contain no child name or practice word) is selectively unmasked via a `.ph-no-mask` marker. Child names, practice words, handwriting canvas and all typed values are never captured. Separately, for genuine parent/tutor accounts (never child logins) the parent's own account email is set as a person property (`parent_email`) plus an `is_internal` flag, so

real-parent churn can be distinguished from the team's own test accounts — adult account data, lawful basis legitimate interests (product improvement), balanced against a low-privacy-impact, EU-hosted, masked replay.

International transfers. Policy: the EU option is chosen wherever a provider offers one —

Supabase, PostHog and Vercel's function region are all set to EU. Where a service is inherently global (a CDN / edge network) or US-only (the AI vendors), that is recorded plainly below rather than implied away.

Data at rest is in the EEA — both stores that hold personal data are EU-hosted (Supabase in AWS `eu-west-1`, PostHog on EU Cloud), and transfers to the EEA are covered by the UK's adequacy regulations.

Compute is EU-configured but not EU-only, and the distinction matters: Vercel's function region is Frankfurt (`fra1`), which governs Serverless (Node) functions but does **not** pin Edge Functions — those run at the point of presence nearest the request, and most of the API is `runtime: "edge"`. So request *handling* can occur outside the EEA even though nothing is *stored* outside it. Together with AI request content in flight (processed and returned, not stored) and vendor support access, that is why the SCC-backed DPAs below are retained rather than treated as redundant. Mechanisms by provider:

Provider	Location	Transfer mechanism
Vercel	Serverless (Node) functions: EU — Frankfurt (<code>fra1</code>). Edge Functions + static assets: global edge network — an Edge Function executes at the point of presence nearest the request, which the project's function-region setting does not pin. Today 53 of 64 API routes are <code>runtime: "edge"</code> , including every route that handles child practice data (<code>chat</code> , <code>coach-tips</code> , <code>whisper</code> , <code>handwriting</code> , <code>word-lists</code> , <code>tts</code>), so this is the normal path, not an edge case	DPA with SCCs (vercel.com/legal/dpa)
Supabase	EU — Ireland (AWS <code>eu-west-1</code>)	Data at rest stays in the EEA, which is covered by the UK's adequacy regulations, so no SCC reliance is needed for the storage location itself. Supabase is a US company, so its own support/sub-processor access can still involve a transfer — covered by its DPA with SCCs (supabase.com/legal/dpa)

Provider	Location	Transfer mechanism
OpenAI	US	DPA with SCCs (openai.com/policies/data-processing-addendum)
Google Cloud (incl. Cloud Vision for Write it handwriting reading - image of the drawn word only, no name/ID/expected word; behind the <code>handwriting_google_reader</code> flag from 2026-10-04)	US/global	DPA with SCCs (cloud.google.com/terms/data-processing-addendum)
Anthropic	US	DPA with SCCs (anthropic.com/legal/commercial-terms)
PostHog	EU (<code>eu.i.posthog.com</code>)	EU Cloud instance; DPA (posthog.com/privacy/dpa)

All provider DPAs accepted by the controller (confirmed 2026-06-30).

Retention. Automated minimisation via `purge_expired_data()` (daily cron): usage events 14 months; email log 12 months; feedback + screenshots 12 months; challenge data 90 days; **handwriting image thumbnails 90 days** (image cleared, numeric scores kept); never-confirmed dormant accounts 12 months. Account/child deletion erases data on demand. **Unused accounts** are removed by two sunset crons, each warning by email first with a one-tap keep link: an account with no child is warned at day 16 and deleted at day 30 (`api/admin/sunset-childless.js`); a family with no sign-in or practice for 6 months is warned, reminded at day 21 and deleted at day 30 after the warning, children and progress included (`api/admin/sunset-inactive.js` ; teachers, admins and holders of bought credits are never swept). Both ship in dry run and are armed deliberately after the queue is reviewed at `/admin/sunset` .

3. Consultation

Solo operator; no internal stakeholders. ICO guidance (Children's Code, DPIA guidance) relied upon. Product design informed by ongoing parent usage patterns.

4. Necessity & proportionality

- **Lawful bases:** contract (delivering the practice the parent signed up for); legitimate interests (security, service improvement that does not override rights, progress report emails enabled by default with opt-out); consent (optional features the user turns on — voice, AI features). Handwriting/formation analysis defaults on as part of the learning service (contract) with a per-child parent toggle to turn it off.
- **Practice reminders (native app, 2026-10-12):** a parent may switch on a daily reminder per child on the device the child practises on. It is a **local** notification: the child's first name, time and days are stored in that device's `localStorage` and scheduled by the device itself, with no

push token, server or Apple/Google push service involved, and nothing is sent to us. Off by default, offered to the parent only, limited to 06:00–21:59, names no word or score, and skips a day the child already practised on that device. Not a nudge towards more use or less privacy: the parent sets it to support a practice routine they chose.

- **Data minimisation:** only data needed for practice/progress is collected; the model-training retention idea was **dropped**; handwriting images auto-expire.
- **Accuracy:** correctness computed server-side; parents can edit/review.
- **Transparency:** privacy policy + AI transparency page; provider list reviewed on a schedule.
- **Rights:** self-service **export and delete** (per child and whole account); only the owning parent can delete a child.

5. Risks to individuals

#	Risk	Likelihood	Severity	Mitigation	Residual
1	Children's handwriting images retained too long / over-collected	Low	Medium	Auto-purge at 90 days (image only); per-child toggle to turn analysis off; deletable. Formation analysis defaults on (migration 20260516) with parent toggle — see Standard 7 note in Children's Code self-assessment	Low
2	Child data exposed via weak access control	Low	High	Supabase RLS (owner-scoped), private storage buckets, JWT auth, service-role keys server-only	Low
3	Secondary use of children's data to train AI models	Low	High	API config with no-training stance; documented in AI register; reviewed on schedule	Low
4	A non-owner (guardian/tutor) deletes/exports a child	Low	Medium	Endpoints re-verify <code>owner_id</code> ; guardians can only unlink	Low
5	Accidental account deletion wiping all children	Low	High	Named-children confirmation + type-DELETE step	Low
6	International transfer without safeguards	Low	Medium	Provider DPAs/SCCs accepted (2026-06-30)	Low
7	Parent can't exercise rights	Low	Medium	In-app export/delete + support email; respond within 1 month	Low
8	Breach not reported in time	Low	High	Breach plan (see ROPA doc); 72-hour ICO reporting	Low

#	Risk	Likelihood	Severity	Mitigation	Residual
9	A personalised coach tip leaks one child's profile to another family	Low	High	Tips are tiered by what they are a <i>function of</i> : only cross-word-profile tips are personalised, and those write exclusively to <code>child_word_tips</code> (RLS-scoped, unique per <code>child_profile_id</code>). The shared caches are keyed on word/misspelling and never receive child history. Enforced in <code>api/coach-tips.js</code> , documented in <code>docs/coach-tips-personalisation.md</code>	Low
11	Google Classroom import over-collects, leaks the teacher's Google token, or imports a roster the teacher is not entitled to	Low	Medium	Exactly two read-only scopes, no email/photo/identity scopes (asserted by test); the access token never reaches the browser (sealed grant, in-memory only, no refresh token, nothing persisted, never logged - asserted by test); Google itself only lists courses the account teaches, the app only imports into a class the caller owns, and the school's Google admin controls whether the app may be used at all	Low
12	Wonde import lets a teacher import pupils from a class they do not teach (the app token can read every class at an approved school)	Low	High	The server re-derives permission on every roster request: the caller's CONFIRMED Spelling.Live email must be a Wonde employee's school email (home addresses never match; an ambiguous match is refused) and that employee must teach the class; a classId from the browser is never trusted (asserted by test). Only <code>spelling_app_admins</code> may browse other classes, for sandbox/support. The token never reaches the browser; the school can revoke access in Wonde at any time	Low
10	An inference about a child's learning difficulty is wrong, stale, or over-confident on thin data	Medium	Medium	Personalisation is gated on the <i>same</i> low-data thresholds as the parent Patterns report (<code>REPORT_DEFAULTS</code>), so the coach cannot claim a trend the report considers unrankable; tips are stamped with <code>classifier_version</code> + <code>profile_signature</code> and regenerated when either moves; every tip shows a parent-visible "Why this tip?" basis so the inference is contestable rather than opaque	Low

6. Measures & sign-off

Measures above are implemented in the product (RLS, retention cron, rights tooling, transparency register, owner-only erasure). ICO registration: AK Shah Ltd holds its own registration, ref **ZC266285** (registered name: Ak Shah Ltd). The earlier sole-trader registration (ref ZC173462, Mr Manoj Shah, paid 2026-06-16) is historical: an ICO registration does not transfer to a company, so the company registered in its own name before becoming the controller. All processor DPAs accepted. Keep the ROPA current.

Role	Name	Date	Outcome
Controller / decision-maker	Spelling Live (sole trader)	2026-06-30	Proceed — residual risks low and justified
Controller / decision-maker	AK Shah Ltd (Manoj Shah, Director)	2026-10-13	Re-confirmed on change of controller - processing unchanged. Published to schools in summary at /for-schools/trust ; full text sent on request

Spelling.Live is operated by AK Shah Ltd (company number 17491910, England and Wales; ICO registration ZC266285). Rendered 2026-10-14 from the maintained source; the current version is always at <https://spelling.live/for-schools/trust>.