

Safe AI for Children

Bookbot Pty Ltd – Policy on Child-Centric Artificial Intelligence

Version 1.0 – July 2026

Approver: Adrian DeWitts, Founder and CEO

Owner: Bookbot Leadership Team

1. Why this policy exists

Bookbot builds an AI-powered oral reading tutor used by children in the early primary years. Children read aloud to the app; the app listens, follows along, and tells them when a word needs another try. That interaction is the product. It is also, unavoidably, an AI system operating on a child's voice, in a child's most vulnerable learning moment – the moment they get something wrong.

Most AI safety work treats children as a filtering problem: keep the bad content out, add an age gate, move on. That is mitigation, and mitigation is not enough. Child-centric AI is a design philosophy that asks a harder question: **how do we design AI systems that actively support children's rights, development, dignity and wellbeing by default?**

This policy is our answer. It states what our AI systems do, what they will never do, where our training data came from, who controls a child's data, how we handle uncertainty, how we monitor for bias, and who is accountable when something goes wrong.

It applies to every Bookbot product, model and deployment.

1.1 Where we place ourselves

There are broadly three stages an organisation can occupy. **Awareness:** recognising that children might use the system. **Mitigation:** adding safeguards after the fact – filters, age gates, warnings. **Child centricity:** children considered from design inception.

Bookbot has no adult product. There is no general-purpose mode that children wandered into. Every model we train, every book we produce, and every interface decision is made for a child learning to decode. We consider ourselves to be operating at the third stage, and we are publishing this policy so that claim can be checked rather than merely asserted.

2. Principles

2.1 Developmental appropriateness. Children's cognitive and emotional capacities change quickly. A four-year-old at the letter-sound stage and a ten-year-old reading chapter books need different feedback, different pacing, and different tone. Our systems respond to a child's demonstrated reading level, not to a stated age or a generic average.

2.2 Privacy by default. Children cannot meaningfully consent to data collection on their own behalf. We therefore collect the minimum needed to teach reading, we process the most sensitive data on the device wherever possible, and we put control of what is collected and what is kept in the hands of the adults responsible for that child.

2.3 Genuine transparency. Not a terms-of-service page. Children, parents and teachers should be able to understand, in the moment and in plain language, what the app is listening to, what it can tell them, and what it cannot.

2.4 Inclusive design. An AI reading tutor that only works for children with fast connectivity, standard accents, quiet rooms and current-generation devices is not an inclusive product – it is a product for advantaged children with an accessibility statement attached. We design for the constrained case first.

2.5 Active empowerment. The purpose of the system is to make itself unnecessary. Bookbot exists to build a child's independent reading ability, not to maximise the time they spend with Bookbot. We do not build for engagement as an end in itself, and we do not build companionship.

3. The AI systems we operate

System	What it does	Who it faces	Where it runs
Automatic speech recognition	Listens to a child read aloud and identifies mispronounced or skipped words	Children	On device
Text-to-speech	Reads text aloud, models pronunciation, voices books and prompts	Children	On device
AI-assisted content creation and review	Drafts, illustrates and reviews decodable texts under editorial control	Children (output only)	Internal
Support assistant	Answers technical support questions	Adults only	Cloud

Book matching and levelling – the system that decides what a child reads next – is deliberately **not** an AI system. See §3.4.

3.1 Speech recognition – the child's voice

Speech recognition is the system with the greatest potential for harm in our product, because it processes a child's voice and because it renders a judgement on their performance.

- **Recognition happens on the device.** The engine runs locally. A child's speech is turned into a decision about a word in real time, on the device in the child's hands, without requiring a network connection. This is what makes the product work in a classroom with no internet, and it is also what keeps the child's voice out of transit by default.
- **Recording is opt-in and separate.** By default, audio is processed and discarded – nothing is retained. Where we have explicit permission, and only where the responsible adult has opted in, audio may be recorded and retained so that we can improve recognition for children who are currently served badly by it. Opt-in is a deliberate, separate choice. It is never bundled into general terms of use, never a condition of using the app, and never the default state. A child whose parent or school has not opted in has no audio retained, ever.
- **What opt-in means in practice.** Recorded audio is used to improve speech recognition and pronunciation feedback. It is not used for advertising, profiling, resale, or any purpose outside improving the product. Consent can be withdrawn at any time, and withdrawal removes the recordings (see §4.4).

- **No inference beyond reading.** The recogniser tells us whether a word was read as expected. It does not, and is not permitted to, infer emotional state, attention, mental health, disability, home language proficiency, or family background. We do not build these features and will not accept requirements to build them.
- **The microphone is visible.** The app displays a clear on-screen indicator whenever the microphone is active. In shared settings we recommend headsets, both for accuracy and so that children can see and control when they are being listened to.
- **A non-audio path exists.** Where a school, parent or jurisdiction does not permit microphone use, alternative non-audio and teacher-led activities are available. Declining the microphone must never lock a child out of reading.

3.2 Text-to-speech – the voice the child hears

Bookbot's speech synthesis reads books aloud, models correct pronunciation and voices in-app prompts.

- **Our voices are adult voices, recorded with consent, for this purpose.** They are built from purpose-recorded sessions with adult voice talent and storytellers working from a Bookbot recording script, under agreements that cover synthesis and distribution.
- **We do not synthesise children's voices.** No Bookbot voice is derived from a child's speech. No child's voice is synthesised, imitated, or made available for synthesis, under any circumstances.
- **We do not replicate anyone's voice without agreement.** No voice model is trained without a written agreement from the speaker or rights holder covering synthetic reproduction.
- **The voice is a narrator, not a friend.** Our voices read books and give reading feedback. They do not hold open conversations, express affection, claim to remember the child, or present themselves as a person. See §5.3.

3.3 AI-assisted content creation and review

We use AI in the production of decodable books – drafting candidate texts against a phonics scope and sequence, generating illustration candidates, and performing much of the review.

We want to be straightforward about this rather than claim a level of human scrutiny we do not apply. AI does a substantial share of the reviewing: decodability checking against the target scope and sequence, phonogram analysis, image-text fidelity checks, and flagging of content that breaches our style guide. Machine review is more consistent and more exhaustive than a human reading a thousand short books, and for these mechanical properties it is genuinely better. Human editorial review is applied on a risk-weighted basis – new content types, new phonics sequences, new cultural contexts, anything the automated checks flag, and ongoing sampling across the library.

Our commitments:

- **The review pipeline is designed on the assumption that models fail.** Automated review is layered – independent checks that would have to fail together to let a defect through – and calibrated against human-reviewed samples so we can measure how often it is wrong rather than assume it is right.
- **Cultural review involves people from the context.** Content contextualised for a particular country or community is reviewed for cultural relevance of settings, names and contexts by people from or working in that context, not by a model alone and not solely from our own office.
- **We check for stereotype and erasure.** Generative models reproduce the biases of their training corpora. Review – automated and human – explicitly examines representation across gender, family

structure, ability, skin tone and setting, and rejects or rewrites material that narrows a child's sense of who reading is for.

- **Accuracy matters even in a six-page book.** Informational texts are fact-checked. Decodable constraints never override truthfulness.
- **We can find and fix anything we have shipped.** Content is versioned and traceable, so that when a defect is found in one book we can identify every book produced the same way and correct them together.

3.4 Book matching and levelling is deterministic – deliberately

The system that decides which book a child reads next is not a model. It is a set of explicit rules based on a published scope and sequence and on the child's demonstrated mastery of specific target words and phonics patterns.

This is a deliberate choice, and we consider it a safety decision rather than a technical shortcut:

- **It is inspectable.** A teacher can be told exactly why a book was suggested, in terms of the phonics patterns the child has and has not mastered. There is no confidence score to interpret and no black box to trust.
- **It is reproducible.** The same evidence produces the same recommendation, every time, for every child. It cannot drift, and it cannot quietly develop a preference for one group of children over another.
- **It cannot encode bias it was never given.** A learned recommender trained on historical progression data would inherit whatever inequities are in that data. A rule cannot.
- **Control sits with the teacher and the child.** Recommendations are suggestions. Teachers can move a child's level in either direction, assign specific books, and disregard the suggestion entirely. Children can choose freely from what is available to them.

Two things follow that we hold to regardless of implementation. **We do not diagnose** – Bookbot does not identify dyslexia, learning disability, or any clinical condition, and does not produce output a teacher could reasonably mistake for a diagnosis. Where reading patterns suggest a child is struggling, the system's job is to route that observation to a human, not to name a condition. And **there are no permanent labels** – a child's level is a current state that moves in both directions, not a rank. We do not expose ability rankings between children, and we do not build features that let children compare their level to their classmates'.

3.5 Support assistant

Our automated technical support assistant answers questions about installation, devices, sync and account issues. It is **adult-facing only** – available to teachers, school coordinators and parents. It is not reachable from any learner-facing surface, and children have no route to a conversational AI inside Bookbot.

3.6 What we do not build

These are red lines, not current-roadmap statements.

- Any feature designed to create emotional attachment to the app or to a synthetic character
- Synthesis or cloning of a child's voice
- Emotion, attention, or mental-health inference from a child's voice or face
- Advertising, commercial profiling, or sale of child data

- Open social features, unmoderated communication, free-text input from children, or links out to the open web
 - Automated decisions that determine a child's educational placement without a teacher in the loop
-

4. Data: what we collect, and who controls it

4.1 We collect what teaching reading requires, and not more. Learning data covers reading accuracy, fluency indicators, books completed and levels gained, phonics patterns mastered, time on task, and app interaction events. Learners are identified pseudonymously – by an identifier linked to their class or school, not by a full name. We do not collect home addresses, photographs, biometric data, health records, or family information.

4.2 Parents and teachers control the data – not us.

- A parent or guardian can see what has been collected about their child.
- A parent or guardian decides whether their child's progress data is shared with a teacher or school. Sharing is a choice, and it can be withdrawn.
- A parent, guardian or school can delete a child's data. Deletion means deletion, not archival.
- Where a school holds the relationship, the school's designated lead exercises these controls under the school's own consent arrangements, and parents retain the right to withdraw their child.

The principle is that a child's reading history belongs to the child and the adults responsible for them. Our role is custodial.

4.3 What we do with data. Learning data is used to teach reading, to show progress to the child, their teacher and their parent, and – in aggregated and pseudonymised form – to improve the product and to evaluate whether it works. It is not sold, traded, shared for commercial purposes, or used for advertising, profiling, or any form of targeting. There is no third-party advertising or analytics SDK in the learner experience.

4.4 Training data: provenance, consent, de-identification. Our speech recognition and synthesis models are trained on speech data collected through dedicated data-collection and research projects, on opt-in recordings where permission has been given, and on openly licensed public datasets.

- Lawful basis and consent. All speech data collected by Bookbot for model development was collected with permission. For children's speech, informed consent was obtained from a parent or guardian, with age-appropriate assent from the child where applicable. Consent materials were provided in the relevant local language and explained what was being collected, what it would be used for, who would have access, and how to withdraw. Adult voice contributors recorded under written agreements that expressly cover synthetic reproduction.
- Purpose limitation. Speech data is used to improve reading-tutor speech recognition and synthesis. It is not used for advertising, profiling, resale, or any purpose outside the improvement of the product it was collected to serve.
- De-identification. Personally identifying information is removed from our speech corpora. Recordings are not linked to a child's name, contact details, address, date of birth, or individual identity in the training pipeline. Content that would identify a speaker indirectly – a child saying their own name, or the name of their school – is removed or redacted during dataset preparation.

- **Opt-in, never ambient.** Recording for model improvement is a separate, explicit choice, never bundled into general terms and never the default. Children whose responsible adults have not opted in contribute no audio to any model.
- **Withdrawal.** Consent may be withdrawn at any time. On withdrawal, recordings are removed from the working corpus and excluded from subsequent training runs. We are candid about the limit: a model already trained and shipped cannot have one speaker surgically removed from its weights. What we commit to is removal from the stored corpus, exclusion from all future training, and – given the de-identification above – that no shipped model reproduces or exposes an individual child's voice.
- **Third-party data.** Where we use public or licensed datasets, we check the licence permits commercial use and model training, and we prefer datasets with documented collection ethics. We do not use scraped speech data of unknown provenance.

4.5 Security. Data is encrypted in transit and at rest. Access is controlled by role: a child sees their own progress; a teacher sees the children in their class whose parents have permitted it; our own staff access is restricted by function and logged. Access to production data is audited.

4.6 If something goes wrong. A suspected breach is reported to the CEO, as data protection lead, within 24 hours of discovery. We assess the scope and the categories of data and individuals affected, notify the relevant supervisory authority within the timeframe its law requires, and notify affected individuals without undue delay where the breach is likely to present a high risk to their rights. Containment is immediate; a post-incident review follows, and its findings change the system rather than the documentation.

4.7 Open contribution. Where we are able to, we release models and tooling for low-resource languages as open source, so that the children who speak them are not permanently dependent on a single commercial vendor's commercial priorities. Any such release applies the same de-identification standard described above.

5. How the system behaves toward a child

Design decisions carry more weight than principles. These are the behavioural rules our learner-facing AI follows.

5.1 Uncertainty defaults to silence, not correction. When the recogniser has low confidence in its assessment, the system gives a neutral response rather than a potentially wrong correction. The asymmetry is deliberate: a missed error costs a child one unreinforced word; a false accusation – being told they are wrong when they were right – costs confidence, and confidence is the scarce resource in early reading. Where the two conflict, we protect the child's confidence.

5.2 Errors are handled without judgement. Feedback identifies the word, models the correct pronunciation, and invites another attempt. It does not evaluate the child. There is no scolding, no disappointment, no "you got four wrong."

5.3 The system does not simulate a relationship. Bookbot does not tell a child it missed them, does not express feelings about their absence, and does not use social pressure to bring them back. Rewards are earned by reading and are available when the child chooses to read. AI for children should support independence, not create dependency; a product that makes a child feel guilty for not opening it has crossed that line.

5.4 The child can always stop. No countdowns, no forced sequences, no penalty for leaving mid-book. A child who closes the app has lost nothing.

5.5 Age-appropriate transparency. Children are told, in simple language they can act on, that the app listens while they read to help them get better, and – where sharing is enabled – that their reading practice is tracked so their teacher can help. This is delivered at the moment it is relevant, at profile setup and when the microphone first activates, not buried in a document written for adults.

6. Bias, evaluation, and the accent problem

Speech recognition trained predominantly on adult, high-resource-language, standard-accent speech performs measurably worse on children, on second-language speakers, and on regional and non-standard accents. This is the central equity risk in our product. A system that under-recognises one child's English is not a neutral tool – it will tell that child they are wrong more often than it tells another child they are wrong, and it will do so for reasons that have nothing to do with their reading.

We treat this as a measurable engineering problem with an ongoing obligation, not a disclaimer.

- **We evaluate by subgroup, not in aggregate.** Recognition accuracy is measured separately across accent and language-variety groups, age bands, and device classes. Aggregate error rates hide exactly the disparities we are looking for.
 - **We calibrate before deploying into a new population.** Piloting exists in part to collect the evidence needed to calibrate recognition thresholds for a new population before the system makes judgements at scale.
 - **Content review is checked against human judgement.** Automated review of books is itself evaluated – sampled against human review – so that we know its error rate rather than assuming it has none.
 - **Disparities are documented, not buried.** Where a subgroup performs worse and we have not yet closed the gap, we say so, and we adjust confidence thresholds in that context so the system errs toward silence rather than toward false correction.
 - **Field evidence outranks lab metrics.** Classroom observation, teacher feedback, and real usage data are treated as primary evidence about whether a model works for real children in real rooms.
-

7. Human oversight

7.1 The teacher is the decision-maker. Every automated output in a teacher-facing surface is a suggestion. Teachers can override recommendations, adjust levels, assign books directly, and disregard flags. No Bookbot output determines a child's placement, grouping, or assessment outcome on its own.

7.2 Escalation runs to humans. Patterns that may indicate a child needs support are surfaced to a teacher, who decides what to do. The system does not act on them.

7.3 Safeguarding concerns follow the safeguarding pathway. Any child-safety concern arising in connection with the product is escalated to the designated safeguarding lead for that setting and handled under the applicable child protection procedures. It is not routed through a product support channel.

8. Governance and accountability

8.1 Named accountability. The CEO is accountable for AI safety and data protection at Bookbot. This is a named role, not a committee.

8.2 Child AI impact assessment. Before a new AI feature, model, or deployment context reaches children, we complete a child AI impact assessment covering:

1. Which children will encounter this, at what ages, in what contexts?
2. What can it get wrong, and what does a child experience when it does?
3. What data does it process, where does that processing happen, and what leaves the device?
4. Could it be experienced as judging, ranking, or labelling a child?
5. Does it create dependency, or build independence?
6. Does it work in the low-connectivity, shared-device, non-standard-accent case?
7. Who is accountable, and how would we know if it were causing harm?

A feature that cannot answer question 7 does not ship.

8.3 External expertise. Reading science and child development expertise is brought into product decisions rather than consulted after the fact. We work with academic and clinical advisors in reading science, cognitive neuroscience and reading difficulty so that people who study how children actually learn to read have standing to challenge product decisions.

8.4 Testing in real conditions. Models and features are validated in the environments they will be used in – shared devices, noisy rooms, low-specification tablets, no connectivity – before scale deployment, not only on developer hardware.

8.5 Review cycle. This policy is reviewed at least annually, and additionally: before deployment into a new country or language, before the release of any new learner-facing AI capability, and following any data or safeguarding incident. Reviews are documented.

8.6 Raising a concern. Concerns about Bookbot's AI systems – from parents, teachers, researchers, partners or staff – can be raised at team@bookbotkids.com. We commit to acknowledging within five working days and to responding substantively, including where the answer is that we got something wrong.

Policy governance

Item	Detail
Approver	Adrian DeWitts, Founder and CEO – Bookbot Pty Ltd
Owner	Bookbot Leadership Team
Version	1.0 – July 2026
Applies to	All Bookbot products, models, and deployments
Review cycle	Annually; on new country or language deployment; on new learner-facing AI capability; following any incident
Contact	team@bookbotkids.com