Specialist Language Courses

AI-Driven Communication Training for Overseas Social Carers: Design, Development, and Initial Evaluation of an LLM-Based Learning Tool

Chris Moore, Specialist Language Courses

SMLC 2026

Specialist Language Courses

SLC – 850+ hours of digital English for Healthcare

Fully developed teaching platform.

All devices.

Institutional licences from €6 per month

THE PROBLEM – WHERE WE STARTED

390,000

overseas-born workers in the UK direct care workforce

and rising.

- Skilled workers in demanding care settings — but little or no structured language support
- Existing solutions are incompatible with long, unpredictable care shifts
- Learners rely on simple sentences and habitual errors from their first language
- When conversations go off-script, confidence collapses — and care quality suffers
- The care training market is not structured to solve this problem

THE PROBLEM – MEDICAL ENGLISH (ONE EXAMPLE)

5-10m

undergraduate nurses worldwide

and rising.

- English needed to read research, access professional development, deliver care in multilingual clinical settings
- English language training programmes are squeezed around clinical training and work placements
- Programmes don’t provide the time for learners to build and practice practical, care-specific clinical communication
- In practice, when conversations go off-script, confidence collapses — and care quality suffers
- Nursing training programmes are not structured to solve this problem

THE SOLUTION

AI Service Users / Patients

AI characters play a range of care service users — responding in character to the learner in multiple realistic scenarios. Unlimited practice. Marginal cost.

AI Tutor

Gives structured, scored feedback grounded in SLC's CPD-accredited courses and the Care Certificate after every roleplay. Two levels – one for students, one for teachers/admins.

Combines seamlessly with face-to-face teaching.

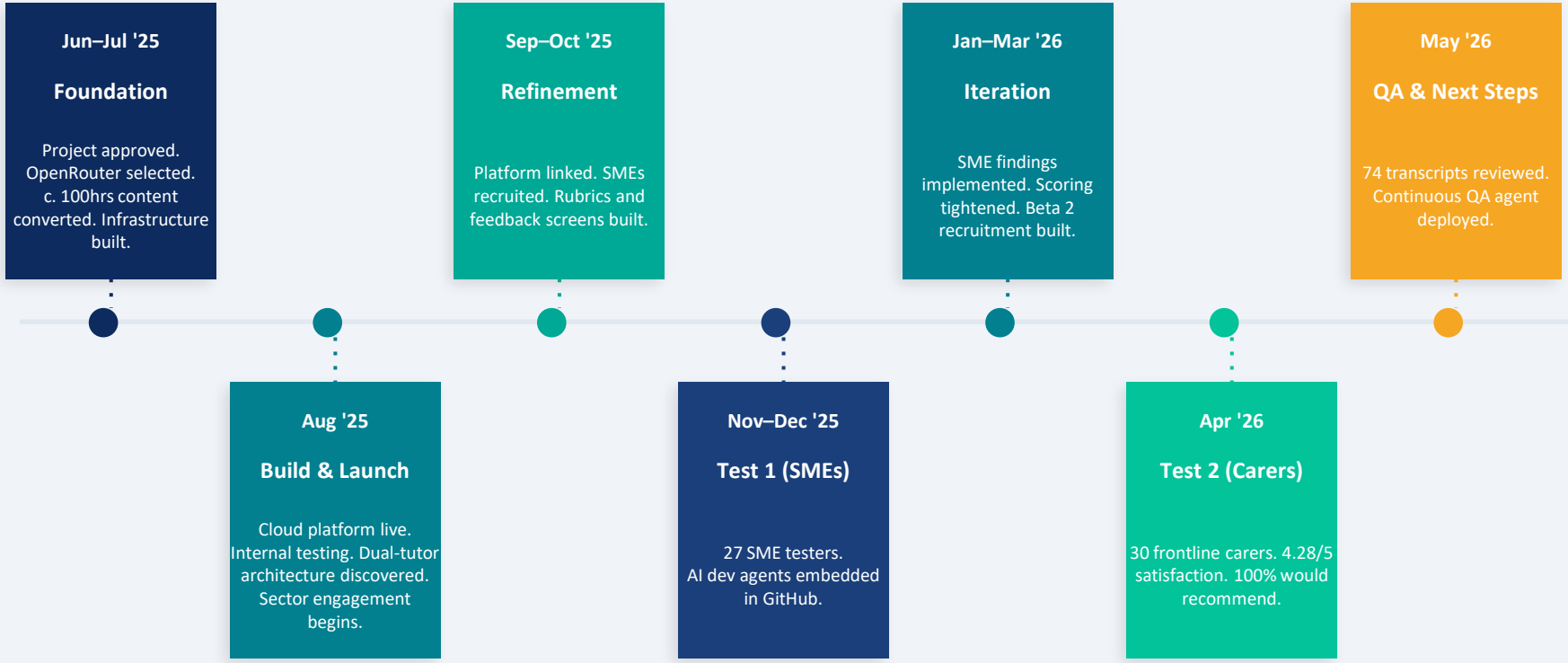
Mobile-First

Runs on any smartphone, tablet and PC. Learners can practise difficult conversations on a break, in their own time, as often as they need.

RAG-Grounded Safety

All AI responses are anchored via Retrieval-Augmented Generation in validated SLC care content, Care Certificate and Care Quality Commission — **not free-form AI generation.**

DEVELOPMENT – 12 MONTHS AT A GLANCE



THE TECHNICAL BUILD

LLM Gateway

OpenRouter — single API to GPT, Claude, Mistral, Command R and more. Transparent pricing, model-agnostic, no vendor lock-in.

AI Service User

Character-based AI playing care residents. In-character responses with hard guardrails: no responsibility-shifting, forced escalation for risk.

RAG Layer

c. 100hrs SLC content (English for Care, Mastering Communication, Medical Terminology) restructured and mapped to Care Certificate units and Care Quality Commission guidelines.

Dual-Tutor Model

Student Tutor (level-appropriate feedback) + Teacher Tutor (detailed feedback and analytics for teachers and admins). Discovered during internal testing.

Observability

Langfuse for LLM monitoring and transcript capture across every test session.

Voice Integration

ElevenLabs text-to-speech (4.5/5 quality rating). Speech-to-text input in final development.

EXTERNAL TEST 1 — SUBJECT MATTER EXPERTS

27	40	82%	0
Structured responses	Additional transcripts	Clinical educators	Unsafe outputs reported

What Testers Praised

- Empathetic, calm and supportive AI character
- Realistic care language and scenarios
- RAG layer consistently cited SLC content and Care Certificate units
- No explicitly unsafe output across all testers

Risks Identified (all actioned)

- Responsibility-shifting language (AI inviting learner decisions)
- Delayed escalation of pain/distress/mobility risk
- Response verbosity and repetition increasing cognitive load
- Scoring leniency — high scores despite weak interactions

RISK-SEVERITY FRAMEWORK

Introduced after Test 1. Now the working language for all AI quality decisions.

GREEN	**Definition:**	Safe, appropriate and effective
	Response:	Preserve and reuse as training exemplars
AMBER	**Definition:**	Suboptimal but not unsafe
	Response:	Prompt tuning and memory control — fix before wider release
RED	**Definition:**	Boundary or safeguarding risk
	Response:	Hard guardrails, forced escalation, immediate fix required

EXTERNAL TEST 2 — FRONTLINE CARERS

30	4.28/5	4.50/5	100%
Frontline carers tested	Overall satisfaction	Useful for EAL learners	Would recommend

Question	Mean / 5	Agree+SA %
Scenarios felt realistic	4.33	93%
Enjoyable / fun	4.17	93%
Easy to use	4.25	73%
Improved confidence	4.25	93%
AI feedback was helpful	4.25	73%
Helps with real conversations	4.25	83%
Will improve safety	4.25	93%
[illegible] English learners in care	4.50 ★	93%

TRANSCRIPT REVIEW — 74 BETA CONVERSATIONS

Developer-led review surfaced findings surveys alone would never have found.

[CRITICAL] Role reversal

AI resident acted as the carer in several transcripts — delivering a structured care plan with markdown headings.

Fix →

Hard prompt constraint added: 'You are the resident. Never act as carer or staff.'

[CRITICAL] Stage directions & emoji in chat

Asterisk-wrapped stage directions and emoji hearts appeared — inappropriate in a realistic care simulation.

Fix →

Explicit ban on asterisks, markdown formatting, bullets, and emoji in AI chat output.

[HIGH] Scoring leniency (quantified)

12 words across 3 turns scored 8/10 because action buttons handled tasks. Flawless 6-turn transcript: 10/10. Only 2 points apart.

Fix →

Action buttons no longer carry weight in communication score. Turn-count multiplier added.

[HIGH] 100–130 word AI responses

AI residents disclosed all clinical information in the first turn — leaving nothing for learners to discover.

Fix →

Response capped at 2–3 sentences. Information released progressively.

CONTINUOUS QUALITY ASSURANCE AGENT

100% of new transcripts are now automatically reviewed by the QA agent.

✓	In-character consistency of the AI service user
✓	Appropriate response length and pacing
✓	Absence of formatting artefacts (stage directions, markdown, emoji)
✓	Appropriateness of tutor scoring against the actual conversation
✓	Presence of unprofessional language
✓	Adherence to escalation expectations

Early Warning

Regressions in AI behaviour caught within hours — not waiting for the next test cycle.

Evidence Base

Every platform behaviour claim backed by a population of reviewed transcripts.

Continuous Improvement

The same review logic drives prompt updates — improvements compound over time.

KEY PEDAGOGICAL LEARNINGS

This is a language-learning project before it is an AI project.

Non-determinism is a pedagogical advantage

The LLM responds differently every time. Unlike fixed e-learning or linear chatbots, each attempt feels like a real care interaction rather than a drill. Carers told us this realism was what they valued most.

Repetition with variation builds real skill

Learners can attempt the same scenario many times, try different approaches, and work towards a higher score. That cycle is closer to how communication skills are actually built — and fits task-based language teaching well.

Curriculum-grounded feedback is safer and more useful

The RAG layer grounds tutor feedback in SLC course content. Testers noticed: one said the tool 'gave correction to wrong use of English and suggested what should be said instead.' Generic AI feedback cannot do this.

Feedback must be pitched at the learner's level

The dual-tutor split taught us this directly. The more capable model was too much for a learner still building confidence — better suited to a teacher wanting a fuller analysis of their staff's communication competence.

IMPLICATIONS FOR LANGUAGE TEACHERS

This is a language-learning project before it is an AI project.

Teachers are complemented not substituted

Learners get levels of practice that are impossible to achieve within classroom time. Teachers integrate this into tutored delivery, guiding learners towards desired outcomes.

Teachers have access to rich language data sets

Teachers access transcripts, a rich set of language information for each learner in a professional context – grammar, syntax, communication skills. Perfect for planning lessons and creating plans focusing on what learners need most.

Teachers becomes coaches

Teachers are freed up to do what matters most – guide, mentor, explain, provide tailored support and feedback. They provide levels of personalization, empathy and engagement that AI cannot.

Teachers extend and focus learning

The teacher creates learning inputs that consolidate and extend learning – sessions on communication skills that cross different scenarios, tailored assignments, additional skills work.

NEXT STEPS

We want you!

Try the app with your learners

Give us your feedback

Be part of the future of Medical English course design

- chris@specialistlanguagecourses.com
- +44 7788 294853