

---

# SentenceBot: An Adaptive Social Robot for Language Therapy in Children with Developmental Language Disorder

---

**Group Portfolio**  
HRI for Social Robots  
University of Groningen, 2026

| Name             | Student Number | Contribution       |
|------------------|----------------|--------------------|
| Efe Gorkem Sirin |                | Equal contribution |
| Baran Akkanat    |                | Equal contribution |
| Nihat Aksu       |                | Equal contribution |

**Group 3**  
Deadline: 17 June 2026

---

# Contents

|          |                                             |          |
|----------|---------------------------------------------|----------|
| <b>1</b> | <b>Project Report</b>                       | <b>2</b> |
| 1.1      | The Challenge . . . . .                     | 2        |
| 1.2      | Investigation Phase: Key Findings . . . . . | 2        |
| 1.3      | Solution Description . . . . .              | 3        |
| 1.4      | Solution Summary . . . . .                  | 4        |
| <b>2</b> | <b>Group Reflection</b>                     | <b>4</b> |
| 2.1      | Process and Progress . . . . .              | 4        |
| 2.2      | Knowledge and Skills Acquired . . . . .     | 5        |
| 2.3      | Division of Work and Satisfaction . . . . . | 5        |
| 2.4      | Honest Assessment . . . . .                 | 5        |
| <b>3</b> | <b>Evidence</b>                             | <b>6</b> |
| 3.1      | Source Code . . . . .                       | 6        |
| 3.2      | QR Topic Cards . . . . .                    | 6        |
| 3.3      | System Architecture Diagram . . . . .       | 6        |
| 3.4      | Papers Reviewed . . . . .                   | 6        |
| 3.5      | Video Demo . . . . .                        | 7        |

**Word counts:** Project Report  $\approx$ 1,490 words (excluding tables and figures); Group Reflection  $\approx$ 480 words.

## Project Report

---

### The Challenge

Developmental Language Disorder (DLD) is a lifelong neurodevelopmental condition affecting approximately 7.6% of school-age children [1]. Children with DLD experience persistent difficulty acquiring, understanding, and producing language despite adequate hearing and non-verbal intelligence [2]. The consequences are far-reaching: reduced academic achievement, impaired social integration, and a substantially elevated risk of reading difficulties [4]. Effective intervention requires frequent, intensive, one-on-one sentence-production practice with an experienced speech-language pathologist (SLP). In practice, however, access to SLT is severely constrained by workforce shortages and long waiting lists [6].

Our challenge was to design a social robot that can serve as a supplementary speech and language therapist for children with DLD, delivering structured, engaging sentence-production practice adapted to individual performance.

### Investigation Phase: Key Findings

A thorough literature review informed our design, covering both the clinical DLD literature and the HRI/SAR literature. We summarise the most relevant findings below.

**Understanding DLD as a clinical condition.** DLD is defined as a persistent language disorder not attributable to a known biomedical condition such as hearing loss or brain injury [2]. Bishop et al. [2] established the CATALISE consensus criteria, clarifying that DLD encompasses difficulties across expressive and receptive language domains and is not outgrown without intervention. Leonard [3] provides a comprehensive account of the linguistic profile: children with DLD consistently produce shorter, less syntactically complex utterances (reduced mean length of utterance), show word-finding difficulties, and struggle to use function words and grammatical morphemes correctly. Conti-Ramsden and Durkin [4] demonstrated in a longitudinal study that these deficits persist into adolescence, with significant impacts on literacy and social outcomes. Rice and Wexler [5] identified the Extended Optional Infinitive stage as a key marker: children with DLD omit tense morphology far beyond the age at which typically developing children resolve this, a finding that directly informed our decision to accept morphologically incomplete but communicatively valid sentences as correct in the validation prompt. McGregor et al. [3] showed that children with DLD have significantly smaller vocabularies and less robust word representations than age-matched peers, supporting our use of a structured, topic-constrained vocabulary set to scaffold sentence production.

**Effectiveness of speech and language therapy for DLD.** Law, Garrett, and Nye [6] conducted a systematic review demonstrating that targeted SLT produces significant improvements in expressive language outcomes for children with DLD, with effect sizes largest for vocabulary and sentence production tasks. Ebbels et al. [7] synthesised evidence-based pathways to intervention, emphasising that therapy must be intensive and individualised to be effective — a requirement that motivates technology-mediated supplements. Reilly et al. [7] showed in a longitudinal cohort study that early identification and intervention are associated with substantially better long-term outcomes, underscoring the importance of accessible, frequent practice opportunities.

**Socially assistive robotics for language therapy.** Spitale et al. [9] demonstrated that SAR can produce significant linguistic improvements in children with language disorders over eight weeks, with therapists acknowledging the robot’s role in sustaining motivation. Azizi et al. [10] identified three core challenges in clinical SLT that a robot can address: maintaining child engagement, providing varied activities, and tracking progress. Their co-design work with SLPs revealed that physical, tangible materials are especially effective with children who have communication difficulties, directly motivating our printed QR card design. Yu et al. [11] demonstrated the viability of LLM-generated personalised feedback in rehabilitation and highlighted that human oversight remains necessary for emotional support, reinforcing our positioning of the robot as a therapeutic supplement rather than a replacement.

From this review, we derived three core design requirements: the interface must minimise non-therapeutic linguistic burden; the system must produce targeted, sentence-level feedback rather than binary judgements; and task difficulty must adapt to individual performance within the session.

## Solution Description

We developed *SentenceBot*, a sentence-production game implemented on the Alpha Mini robot. The system presents the child with a target word and asks them to produce a grammatically complete, semantically coherent sentence using that word. A large language model validates the child’s utterance in real time and generates child-appropriate feedback. The session comprises six questions across three adaptive rounds.

**Topic selection via QR cards.** To eliminate the linguistic overhead of menu navigation, session topic selection is handled entirely through physical QR cards. Four cards were printed and laminated, each encoding a topic name (*animals*, *food*, *school*, *weather*) as plain text in a standard QR format. The robot reads the card using its front-facing camera via the `rom.sensor.sight.read` API and decodes it using the OpenCV `QRCodeDetector`. This design choice was directly motivated by Azizi et al.’s finding that tangible materials reduce cognitive load for children with communication difficulties [10]. In practice, the QR detection worked reliably under normal indoor lighting at distances of 20–60 cm, and the laminated cards proved durable across repeated testing sessions.

**LLM-based sentence validation.** Rather than a rule-based keyword matcher — which would reject any valid sentence not matching a predefined template — we use a large language model as the validator. The system supports two backends: a locally-running Gemma 3 model served via Ollama (default, no API key required), and Google Gemini 1.5 Flash (cloud, enabled with `--gemini`). Both receive a structured system prompt that defines the validator’s clinical role, specifies three evaluation criteria (grammatical validity, logical coherence, and correct use of the target word), and constrains the output to a fixed JSON schema containing a validity flag, a reason string, and an encouragement message. This approach can handle the open-ended, linguistically diverse productions characteristic of natural child speech, a capability impossible to achieve with pattern matching [11].

**Adaptive three-stage difficulty.** Vocabulary is divided into two tiers per topic: easy words (high-frequency, phonologically simple: *dog*, *sun*, *cake*) and hard words (lower-frequency, phonologically complex: *elephant*, *thunder*, *homework*). A third stage requires the child to incorporate two randomly selected words from the full pool into a single coherent sentence. The game consists of three rounds of two questions each. If a child

answers both questions in a round correctly, the difficulty advances for the following round; otherwise, the same difficulty level is maintained. This mechanism is grounded in Vygotsky’s zone of proximal development [12]: tasks should be calibrated to be slightly beyond current ability while remaining achievable.

**Celebration gesture and streak feedback.** A correct response triggers both verbal encouragement from the LLM and a physical celebration gesture: the robot raises both arms simultaneously via direct motor control (`rom.actuator.motor.write`), then returns to the neutral standing pose. After every third consecutive correct answer, a special streak message is delivered alongside the gesture. This multimodal positive reinforcement is consistent with evidence that gesture-accompanied feedback enhances engagement in child-robot interaction [8].

**Key design parameters.**

Table 1: System configuration parameters

| Parameter                 | Value                   | Rationale                       |
|---------------------------|-------------------------|---------------------------------|
| Questions per session     | 6 (3 rounds $\times$ 2) | Manageable session length       |
| Max attempts per question | 4                       | Prevents frustration            |
| STT timeout               | 15 s                    | Accommodates word-finding delay |
| Streak threshold          | 3 correct               | Variable reward schedule        |
| Stage advance criterion   | 2/2 correct             | Mastery-based progression       |

**Known limitation: STT accuracy.** The Alpha Mini’s built-in microphone and speech recognition module exhibited non-trivial misrecognition errors during testing. For example, the target word *rice* was transcribed as *price*, causing a valid sentence to be incorrectly flagged as not containing the target word. This is a known limitation of general-purpose STT systems applied to child speech and to phonologically similar word pairs. The 15-second timeout further revealed that children (or testers) occasionally paused longer than expected before speaking. Future work should explore STT models specifically trained on child speech corpora to mitigate these issues [9].

## Solution Summary

SentenceBot delivers a structured, engaging, and adaptive sentence-production therapy experience on the Alpha Mini robot. The integration of LLM-based open-ended validation, tangible QR topic selection, adaptive difficulty, and physical celebration gestures represents a substantive proof-of-concept contribution. The system is fully implemented, locally testable via a terminal mode, and operationally demonstrated with the physical robot. While formal clinical evaluation with DLD-diagnosed children remains future work, the proof-of-concept demonstrates the viability of LLM-augmented SAR for language therapy and establishes a concrete foundation for future clinical trials.

## Group Reflection

### Process and Progress

Our group made steady and ultimately satisfying progress toward the challenge. We began by dividing the investigation phase evenly: each group member took responsibility

for a distinct strand of the literature — clinical DLD research, SAR therapy systems, and LLM-based feedback mechanisms respectively — and synthesised findings in shared sessions. This division allowed us to cover the relevant literature broadly while maintaining shared ownership of the conclusions.

The implementation phase revealed a steeper learning curve than anticipated. None of us had prior experience with the WAMP messaging protocol or the RIE robot API, and the first week of coding was largely consumed by understanding the asynchronous architecture and debugging connection issues. Once the core event loop was working, however, development accelerated significantly. We adopted a practice of testing each new feature in terminal mode (`--local-test`) before deploying to the physical robot, which substantially reduced iteration time.

A significant practical challenge was the robot’s speech recognition quality. During testing, we observed that phonologically similar words were frequently confused — the most notable example being *rice* transcribed as *price* — causing valid child responses to be incorrectly rejected. This limitation led us to add explicit acknowledgment of STT constraints to our system documentation and shaped our future work recommendations. It also reinforced our understanding of why research groups working with real clinical populations invest heavily in specialised ASR pipelines.

## Knowledge and Skills Acquired

As a group, we deepened our understanding of the clinical landscape of DLD, the design principles of socially assistive robotics, and the practical challenges of deploying LLMs in constrained, real-time interaction contexts. Technically, we gained hands-on experience with asynchronous Python programming, the WAMP/Twisted framework, prompt engineering for structured LLM outputs, computer vision-based QR decoding, and direct robot motor control. The dual-backend LLM architecture — supporting both a locally running Gemma 3 model via Ollama and a cloud-based Gemini deployment — was a technically demanding addition that significantly improved the system’s flexibility and offline capability.

## Division of Work and Satisfaction

Efe led the LLM validation pipeline and prompt engineering; Baran led the adaptive difficulty engine, QR integration, and celebration gesture; Nihat led the clinical literature review, systematic testing, and documentation. All three members collaborated closely on integration throughout, understand the full codebase, and contributed equal effort. The group dynamic was constructive; disagreements were resolved through quick experiments rather than prolonged debate.

## Honest Assessment

Our principal unfulfilled goal is clinical validation. The system has been tested only by the group members themselves, not by children with DLD. This limits the evidence base for our therapeutic claims and is the most important direction for future work.

## Evidence

### Source Code

The complete source code, README, and `.env.example` are submitted as a separate `.zip` file alongside this portfolio. The repository contains:

- `main.py` — main application (robot + LLM + adaptive game loop)
- `README.md` — setup instructions and usage guide
- `.env.example` — environment variable template
- `1st_assignment/` — Coding Assignment submission (PwA getting-to-know conversation)

Key flags:

```
python main.py          # local Ollama/Gemma3 backend
python main.py --genini  # Gemini 1.5 Flash cloud backend
python main.py --local-test # terminal mode, no robot required
```

### QR Topic Cards

Four A4-format QR cards were printed and laminated, each encoding one topic name as plain text (**animals**, **food**, **school**, **weather**). The card design was produced in LaTeX using the `qrcode` package and printed on standard office paper. Detection was reliable at 20–60 cm under normal indoor lighting.

### System Architecture Diagram

Table 2: System component overview

| Layer               | Component               | Technology                             |
|---------------------|-------------------------|----------------------------------------|
| Physical interface  | Alpha Mini robot        | RIE WAMP API                           |
| Application logic   | Python 3.9+ application | Twisted / autobahn                     |
| Speech input        | Robot STT module        | <code>rie.dialogue.stt.stream</code>   |
| Speech output       | Robot TTS (animated)    | <code>rie.dialogue.say_animated</code> |
| Vision / QR         | Robot camera + OpenCV   | <code>rom.sensor.sight.read</code>     |
| Motor control       | Arm gesture             | <code>rom.actuator.motor.write</code>  |
| LLM backend (local) | Gemma 3 via Ollama      | HTTP REST ( <code>/api/chat</code> )   |
| LLM backend (cloud) | Gemini 1.5 Flash        | Google GenAI SDK                       |

### Papers Reviewed

The following papers were reviewed during the Investigation phase and directly informed the design:

- Azizi et al. (2024/2025) — Co-design of SAR for children’s SLT
- Yu et al. (2026) — LLM personalisation for aphasia rehabilitation
- Spitale et al. (2023) — SAR for children with language impairments
- Tomblin et al. (1997) — DLD prevalence in school-age children
- Bishop et al. (2016) — CATALISE consensus definition of DLD

- Leonard (2014) — Linguistic profile of DLD: MLU, morphology, word-finding
- Conti-Ramsden & Durkin (2012) — Longitudinal outcomes of DLD
- Rice & Wexler (1996) — Tense morphology as clinical DLD marker
- Law, Garrett & Nye (2004) — Efficacy of SLT for DLD
- Ebbels et al. (2019) — Evidence-based intervention pathways for DLD
- Vygotsky (1978) — Zone of proximal development
- Belpaeme et al. (2018) — Social robots for education: a review

## Video Demo

A video demonstration of the complete system is submitted alongside this portfolio. The recording captures the full interaction flow: QR card topic selection, all three adaptive rounds of sentence production, real-time LLM validation feedback, and the robot's arm-raise celebration gesture on correct responses. One group member acted as the child user, responding verbally to the robot's prompts, which allowed us to demonstrate realistic turn-taking, the STT pipeline under natural speech conditions, and the difficulty progression across stages. The demo also illustrates the known STT limitation: occasional misrecognitions of phonologically similar words under the Alpha Mini's built-in microphone.

## Video Demo

A video demonstration of the complete system is available at: [Click here to view the video demo \(Google Drive\)](#)

[https://drive.google.com/drive/folders/1gsnvTPFLvUGdPTj2J8rDVmCv6il0gpD8?usp=drive\\_link](https://drive.google.com/drive/folders/1gsnvTPFLvUGdPTj2J8rDVmCv6il0gpD8?usp=drive_link)

## Source Code

The complete source code is available on GitHub: [Click here to view the repository \(GitHub\)](#)

[https://github.com/verynewusername/human\\_robot\\_interaction\\_course](https://github.com/verynewusername/human_robot_interaction_course)

## References

- [1] J. B. Tomblin *et al.*, “Prevalence of specific language impairment in kindergarten children,” *Journal of Speech, Language, and Hearing Research*, vol. 40, no. 6, pp. 1245–1260, 1997.
- [2] D. V. M. Bishop *et al.*, “Phase 1 of CATALISE: A multinational Delphi consensus study identifying language impairments in children,” *PLOS ONE*, vol. 11, no. 7, p. e0158753, 2016.
- [3] L. B. Leonard, *Children with Specific Language Impairment*, 2nd ed. Cambridge, MA: MIT Press, 2014.
- [4] G. Conti-Ramsden and K. Durkin, “Language development and assessment in the preschool period,” *Neuropsychology Review*, vol. 22, no. 4, pp. 384–401, 2012.
- [5] M. L. Rice and K. Wexler, “Toward tense as a clinical marker of specific language impairment,” *Journal of Speech, Language, and Hearing Research*, vol. 39, no. 6, pp. 1239–1257, 1996.
- [6] J. Law, Z. Garrett, and C. Nye, “The efficacy of treatment for children with developmental speech and language delay/disorder,” *Journal of Speech, Language, and Hearing Research*, vol. 47, no. 4, pp. 924–943, 2004.
- [7] S. H. Ebbels *et al.*, “Evidence-based pathways to intervention for children with language disorders,” *International Journal of Language & Communication Disorders*, vol. 54, no. 1, pp. 3–19, 2019.
- [8] T. Belpaeme *et al.*, “Social robots for education: A review,” *Science Robotics*, vol. 3, no. 21, p. eaat5954, 2018.
- [9] M. Spitale *et al.*, “Using socially assistive robots in speech-language therapy for children with language impairments,” *International Journal of Social Robotics*, vol. 15, no. 9, pp. 1525–1542, 2023.
- [10] N. Azizi *et al.*, “Development of robot-assisted speech-language therapy: Co-design with speech-language pathologists,” in *Proc. ICSR*, Springer, pp. 366–380, 2025.
- [11] M. Yu *et al.*, “Embodied digital therapists with LLM personalization for aphasia rehabilitation,” in *Proc. IUI*, pp. 941–954, 2026.
- [12] L. S. Vygotsky, *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press, 1978.