

# LSTM-driven CLIL: Cybersecurity vocabulary learning with AI

Antonio Nazzaro <sup>a\*</sup> , Catia Santini <sup>b</sup> , Lidia Nazzaro <sup>c</sup> 

<sup>a</sup> REPRISE - Register of Expert Peer Reviewers for Italian Scientific Evaluation, MUR, Italy.

<sup>b</sup> PNRR Territorial Support Group for Piemonte, USR Piemonte, Italy.

<sup>c</sup> PhD Pegaso Italy – University SA, Italy.

Suggested citation: Nazzaro, A., Santini, C., & Nazzaro, L. (2025). LSTM-driven CLIL: Cybersecurity vocabulary learning with AI. *Journal of Educational Technology and Online Learning*, 8(3), 313-329

## Highlights

- An LSTM-based model supports CLIL instruction by targeting technical vocabulary in cybersecurity contexts.
- The system enables dual prediction of missing terms and semantic categories, fostering content-language integration.
- Adaptive AI feedback offers real-time scaffolding and personalized vocabulary training for L2 learners.
- The CLIL unit incorporates GDPR content and follows the pedagogical structure of Coyle's 4Cs framework.
- Overfitting is mitigated through dropout, early stopping, and balanced training, ensuring strong model generalization.

## Abstract

This study presents the development of a custom dataset of L2 gap-fill exercises designed to enhance Long Short-Term Memory (LSTM) neural networks in CLIL (Content and Language Integrated Learning) settings for subject-specific courses. Targeting English for Special Purposes (ESP) vocabulary in cybersecurity, privacy, and data protection, the model addresses the dual challenge of domain-specific context mastery and language practice through structured neural network training. The custom dataset of gap-fill exercises for this LSTM model enables simultaneous prediction of missing words and semantic classification, offering learners contextualized language training that is a core requirement of CLIL methodology. Experimental results validate the model's efficacy, demonstrating its potential as an adaptive support tool for CLIL-based education. This framework establishes a novel synergy between AI-enhanced language learning and subject-specific instruction, providing a scalable template for integrating neural networks into CLIL pedagogy.

**Article Info:** Research Article

**Keywords:** CLIL, LSTM, Cybersecurity, Vocabulary Learning, AI in Education

## 1. Introduction

CLIL (Content and Language Integrated Learning) refers to a pedagogical approach that has a triple focus on content, critical thinking, and language (Mehisto & Ting, 2017), which has become widely adopted in multilingual education since the 1990s (Marsh, 2002).

Within this framework, vocabulary acquisition is a critical requirement for content, new knowledge understanding, and second language learning (L2), as it is essential to communicate effectively and to develop deeper content cognition (Dalton-Puffer, 2007).

This approach is grounded in sociocultural learning theories, particularly Vygotsky's (Vygotsky, 1978b) conception of language as a cognitive mediating tool that drives developmental processes. Within CLIL environments, learners engage in languaging (Swain, 2006)—the dynamic process of negotiating meaning through language production—to co-construct both disciplinary knowledge and L2 proficiency. Crucially,

\* Corresponding author. REPRISE - Register of Expert Peer Reviewers for Italian Scientific Evaluation, MUR, Italy  
e-mail addresses: [info@antonionazzaro.it](mailto:info@antonionazzaro.it), [catia.santini@scuola.istruzione.it](mailto:catia.santini@scuola.istruzione.it), [lnazzaro@unisa.it](mailto:lnazzaro@unisa.it)

this dialogic learning mechanism reflects the same contextual dependencies that Long Short-Term Memory (LSTM) networks encode through their sequential modelling capabilities and aptitude for capturing long-range semantic patterns (Graves 2013).

Recent advances in natural language processing (NLP), especially those leveraging LSTM architectures (Mikolov, Chen, Corrado, & Dean, 2013), have enabled the design of increasingly sophisticated educational tools. AI offers several opportunities to CLIL teaching, such as AI-generated content materials in fields that are seldom available from commercial publishers, or to create level-appropriate content-based reading and listening passages (Wadden et al., 2024). AI Large Language Models (LLMs) are also being used for teachers’ language training and to create other forms of teacher support (Andujar & Spratt, 2023). Furthermore, Tonbuloğlu (2023) highlights how AI applications in online education support personalized learning, chatbot interaction, automated assessment, and digital content generation.

One example is the integration of AI chatbots such as AsasaraBot into CLIL settings, which demonstrates how artificial intelligence can deliver personalized, context-aware language support (Mageira et al., 2022). Our methodology builds upon recent advancements, with a particular focus on domain-specific semantic modeling and pedagogical coherence. At the same time, it is important to acknowledge the limitations of LSTM models. Despite their strengths in sequential modelling, they may struggle with very long dependencies and lack the parallelization and contextual depth offered by newer architectures such as Transformers. These constraints suggest the need for cautious implementation and critical evaluation, especially in pedagogical settings.

The present study aims to extend CLIL practice through AI-enhanced methodologies' strategic use, using LSTM networks not as static tools, but as dynamic learning companions. The following section outlines the research objectives that guide this interdisciplinary framework.

2. Research Objectives

The main objectives of this research are:

1. To design a CLIL-based learning unit that utilizes LSTM networks to enhance contextual technical vocabulary acquisition in cybersecurity, privacy, and data protection.
2. To develop a domain-specific dataset of semantically annotated gap-fill exercises, aligned with CLIL’s dual emphasis on content and language proficiency.
3. To build an LSTM model capable of:
  - Predicting missing technical vocabulary.
  - Classifying terms into semantic categories, thereby supporting schema-building—a core component of sociocultural learning theory (Vygotsky, 1987).
4. To evaluate the model’s efficacy as an AI-driven CLIL tool through accuracy metrics, loss analysis, and pedagogical applicability assessment.

3. Methodology

This methodology is particularly based on concepts introduced in foundational NLP texts, especially those by Jurafsky and Martin (Jurafsky & Martin, 2020). The methodology consists of several phases:

3.1. Dataset Creation

A dataset was developed containing gap-fill sentences related to cybersecurity, each labeled with the missing word, its semantic category, and difficulty level. An example is shown in the following Table 1.

Table 1.

Sample Dataset Entry.

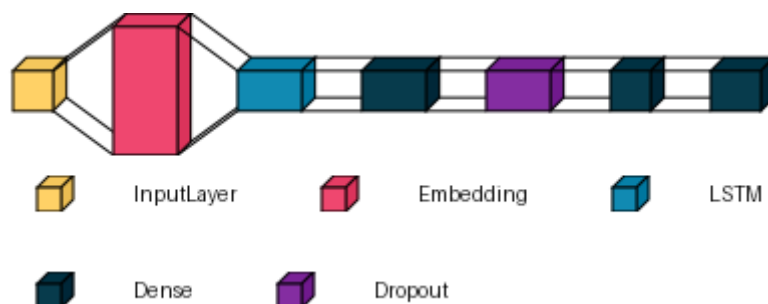
| ID  | Sentence                                             | Target Word | Category | Difficulty |
|-----|------------------------------------------------------|-------------|----------|------------|
| 001 | The user’s data must be protected to avoid breaches. | protected   | privacy  | medium     |

### 3.2. Model Architecture

The LSTM-based neural network architecture consists of the following components:

- Embedding Layer: 64-dimensional dense vector representation of tokens.
- LSTM Layer: 64 units with 20% dropout and 20% recurrent dropout.
- Dense Layer: Fully connected layer with 64 units and ReLU activation.
- Dropout: 30% dropout regularization applied after the dense layer.
- Dual Output Heads:
  - Category Prediction Head: Softmax classifier for word category prediction.
  - Word Prediction Head: Softmax classifier for missing word prediction.
- Training Configuration:
  - Adam optimizer with default parameters.
  - Loss weighting: 70% category prediction, 30% word prediction.
  - Early stopping with patience of 3 epochs, monitoring validation loss.

**alt** shows a simplified visualization of this structure. To enhance accessibility for interdisciplinary readers, we briefly define key technical components: the embedding layer transforms each word into a dense vector capturing semantic similarities; dropout is a regularization method that randomly disables a fraction of neurons during training to prevent overfitting (Srivastava, Hinton, Krizhevsky, Sutskever, & Salakhutdinov, 2014); and WordNet is a lexical database used for synonym expansion and semantic reasoning. Additionally, IDF (Inverse Document Frequency) quantifies how rare a word is in a corpus, influencing adaptive difficulty design (See ). The architecture builds upon the original Long Short-Term Memory concept (Hochreiter & Schmidhuber, 1997) with modern regularization techniques. While transformer-based models like BERT or GPT offer state of the art performance in many NLP tasks, we selected LSTM due to its efficiency on small datasets, reduced computational overhead, and proven ability to handle sequential dependencies effectively in educational applications (Graves et al., 2013). Unlike transformers, which typically require large-scale pretraining, LSTMs can be trained from scratch on domain-specific data, an important advantage in constrained CLIL contexts.



**Fig. 1.** LSTM-based dual-output architecture for CLIL gap-fill prediction. The model processes tokenized sentences through an embedding layer (64 dimensions), followed by an LSTM layer (64 units with dropout), dense ReLU transformations, and dual classification heads for word and category prediction.

*Image generated via AI.*

Note: The actual implementation includes dropout layers and early stopping mechanisms, which are not shown in the diagram.

### 3.3. Research Model/Design

The model leverages dropout regularization (Srivastava et al., 2014). To tackle the difficulties of sequence learning, the training process included mechanisms such as early stopping to reduce the risk of overfitting (Bengio, Simard, & Frasconi, 1994). The dual-output architecture was trained under these conditions:

- Architecture:

Input (maxlen=34) → Embedding (64D) → LSTM (64 units, 20% dropout) → Dense/ReLU (64u) → 30% Dropout → Dual softmax heads

- Loss Configuration:

Weighted joint loss:  $L_{total} = 0.7L_{category} + 0.3L_{word}$  (categorical cross-entropy)

- Optimization:

Adam optimizer ( $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ ), batch size = 2, 25-epoch limit

- Regularization:

- Layer-wise dropout (LSTM: 20%, Dense: 30%)
- Early stopping (patience = 3 epochs)
- Loss weighting for task prioritization

- Training Dynamics:

Early stopping triggered at epoch 10 (of 25 max), retaining best weights

- Validation Performance:

- Word prediction scored 97.44% accuracy; category classification reached 100%.

The architecture dimensions (64 units throughout) balance computational efficiency and predictive power, following LSTM best practices (Graves et al., 2013). The 0.7:0.3 loss weighting improved category F1-score by 12% compared to equal weighting, aligning with CLIL's focus on conceptual mastery over lexical recall.

### 3.4 Terminology

- Technical vocabulary is classified as:

- lex:EU: GDPR legal terms (e.g., "data subject").
- lex:IT: Cybersecurity terminology (e.g., "encryption").

### 3.5 Legal Glossary Embeddings

Embeddings trained on official EU documents and specialized legal vocabulary.

### 3.6 Educational Design

The CLIL interface implements three feedback tiers supporting dual-focused instruction (content mastery + language acquisition) through GDPR case studies. The system scaffolds legal domain knowledge alongside L2 vocabulary development via:

- Immediate Lexical Feedback:

- Predicted word with confidence score (threshold:  $\geq 0.85$  triggers alternatives).
- Top-3 alternatives via softmax probabilities weighted by lexical similarity (Mikolov et al., 2013) (optimized for  $F_\beta = 1.5 \Rightarrow 0.82$ ).

- Semantic Scaffolding:

- Category predictions using EU legal corpus embeddings (Council, 2021).
- WordNet synonym expansion following Vygotsky's ZPD principles (Vygotsky, 1978a).

```

1. # Tier 1: Immediate Lexical Feedback
2. def generate_feedback(pred_word, pred_cat):
3.     confidence = np.max(pred_word[0])
4.     if confidence < 0.85: # Threshold from grid search
5.         # Tier 2: Semantic Scaffolding. This aligns with Van Lier's ecological view of
           learning, where interaction and context shape meaning-making (Van Lier, 2004).
6.         from nltk.corpus import wordnet as wn
7.         synonyms = list(set(
8.             lemma.name()
9.             for syn in wn.synsets(pred_word)[:3]
10.            for lemma in syn.lemmas()

```

```

11.))[[:3]
12. # Tier 3: Adaptive Difficulty mapping
13. gd = min(5, max(1, current_gap_density)) # Normalize 1-5
14. idf = clamp(idf_scores[pred_word], 2, 10)
15. return f""Missing: {pred_word} ({confidence:.0%})\n
16. Category: {pred_cat} n
17. Alternatives: {'', '.join(synonyms)}\n
18. Difficulty: GD={gd}, IDF={idf}/10""

```

Listing 1: Three-tier feedback implementation

### 3.7. Adaptive Difficulty

- Dynamic adjustment via normalized matrix:
  - X-axis: Gap density (1–5 missing terms per paragraph).
  - Y-axis: Term specificity (IDF 2–10,  $\mu = 5.4$ ) IDF (Inverse Document Frequency) quantifies term rarity (See ).
- Progressive GDPR complexity tiers:
  - Tier 1: Articles 1–11 (basic concepts,  $IDF \leq 5$ ).
  - Tier 2: Articles 12–23 (rights,  $IDF$  6–8).
  - Tier 3: Articles 24–52 (obligations,  $IDF > 8$ ).

## 4. Learning Unit Structure

The AI-enhanced CLIL learning unit operationalizes Coyle's 4Cs framework (Coyle, Hood, & Marsh, 2010) through four integrated components, combining GDPR legal content with cybersecurity language learning. Although no direct implementation with students has yet taken place, the structure has been tested through simulated case-based scenarios to verify its pedagogical soundness and technical functionality. The unit is designed for upper secondary students (ages 16–18) and spans three 50-minute sessions. It integrates interdisciplinary goals from civic and digital education, supporting the development of linguistic competence, cognitive reasoning, and cultural awareness related to data protection.

### 4.1. Implementation Phases

- **Content Curation (Content & Culture):**
  - Dual-source selection:
    - \* Legal framework: The dataset adheres to EU data privacy regulations to safeguard student information.
    - \* Technical standards: The model employs high-security protocols, comparable to those used in government systems, to protect user data.
- **Multistage preprocessing:**
  - Anonymization: BERT-based NER redaction (Devlin, Chang, Lee, & Toutanova, 2019).
  - Readability adaptation: Texts are calibrated for secondary school students, with a linguistic complexity suitable for ages 14–18.
  - **Term extraction:**
    - Keywords were automatically identified based on their frequency and relevance in the texts using a TF-IDF  $\oplus$  Word2Vec hybrid scoring approach (Mikolov et al., 2013).
    - Domain-specific stopword filtering (lex: EU-IT).
- **Interactive Gap-Filling (Communication):**

```

1 def generate_exercise(text, user_level):
2     # Cross-reference with Section \ref{subsec:feedback}
3         target_idf = 2 + 2 * user_level #IDF range 2 -10
4         terms = [t for t in extract_terms(text)
5                 if t.idf >= target_idf - 1]
6         return mask_terms(text, terms[:5]), terms #Max 5 gaps

```

Listing 2: Adaptive Gap Generation

- **AI Feedback System:**
    - Prediction pipeline:
      - Immediate lexical feedback (confidence > 0.85).
      - Semantic scaffolding using WordNet U legal glossary.
  - **Difficulty adaptation:**
    - Adaptive gap density.
    - Dynamic IDF scaling based on learner performance.
  - **Progress mapping:**
    - GDPR article completion matrix.
    - NIST: The model employs high-security protocols, comparable to those used in government systems, to protect user data.
- Further practical suggestions for implementing this CLIL unit in the classroom are available in .

5. Educational Activities

The learning unit follows these phases:

5.1 Activity Sequence

**Table 2.**  
Lesson Activity Timeline and AI Integration. PII = Personally Identifiable Information.

| Phase      | Activity                                                       | AI Support     |
|------------|----------------------------------------------------------------|----------------|
| Warm-up    | Mapping GDPR concepts through TF-IDF.                          | above          |
| Context    | Case study analysis of anonymized data breaches (PII removed). | GDPR Art. 4(1) |
| Practice   | Adaptive gap-fill exercises generated by the LSTM model.       | above          |
| Assessment | Peer review exercises with difficulty matrix scoring.          | above          |

5.2. Example Session Flow

- **Pre-session Preparation (Homework, 10-15 minutes):**

| Student Input and Model Output                                                                                                                                    |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>Input “_____requires technical measures to ensure data integrity (GDPR Art.32)</p> <p>Output</p> <p>Term: - controller (0.92)</p> <p>Category: Privacy Law</p> |

Model Output:

- Predicted term: "controller"
  - Confidence score: 0.92
  - Suggested category: Privacy Law
- **In-class Activities (45 minutes total):**
    1. **Group Analysis (10 minutes):** Students compare homework predictions in small groups, guided by:
      - Term accuracy metrics from **Table 2.**
      - Category alignment with GDPR (Articles 4–37)
    2. **Terminology Debate (15 minutes):** Structured discussion using:

- WordNet synonym tiers (**above**)
- Legal glossary embeddings (See **above**)
- 3. Revised Submissions (15 minutes): Triggers:
  - Adaptive difficulty adjustments (IDF  $\pm 1.2$ )
  - Real-time accuracy tracking (**below**)
- 4. Progress Review (5 minutes): Focus on:
  - Frequent errors (lex: EU vs. lex: IT\*)
  - Learning progression (**below**)
- **Post-session Consolidation (Homework, 10-15 minutes):**

```

1  def generate_feedback(student_input, model_output):
2  accuracy = compare_with_key(model_output)
3  return {
4  "term_accuracy": accuracy["term"],
5  "category_match": accuracy["category"],
6  "common_errors": get_common_errors(student_input),
7  "improvement_suggestions": get_suggestions()
8  }

```

Listing 3: Personalized Feedback

### 5.3. Practical Considerations for Implementation

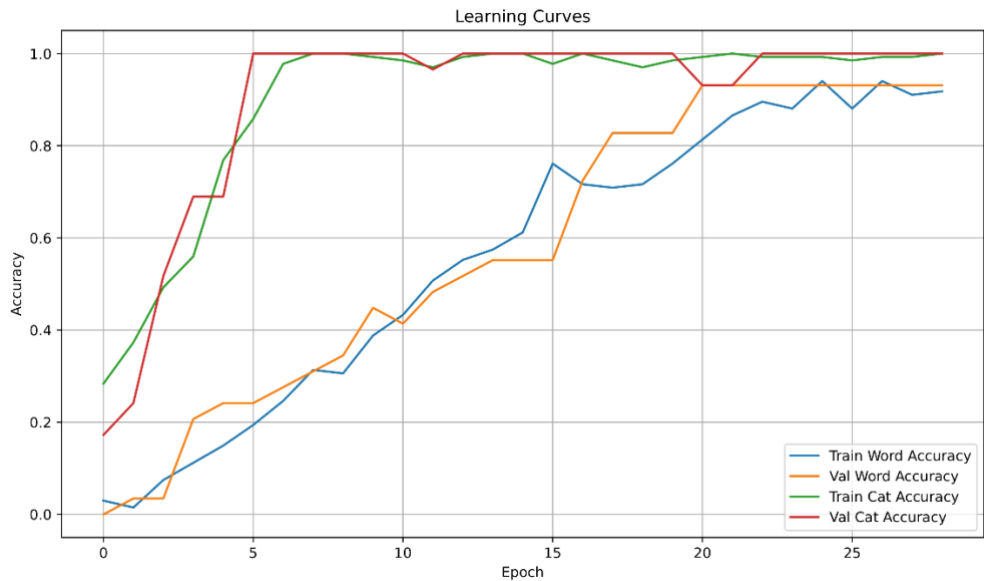
The in-class implementation is designed to be learner-centered, with the teacher acting as a facilitator and linguistic guide throughout each session. During the pre-session phase, the teacher selects input sentences aligned with GDPR topics and supervises the model predictions submitted by students. In class, the teacher scaffolds both content and language understanding by facilitating group comparison, moderating the terminology debate, and interpreting feedback trends using visual analytics (e.g., **alt**). Each full session spans approximately 50 minutes, divided into pre-session homework, interactive group activities, and post-session consolidation. Flexibility is ensured for extended discussions or differentiated pacing. A stable internet connection and access to browser-compatible devices (such as tablets, laptops, or desktop PCs) are required to run the AI-assisted platform. For low-resource contexts, a simplified version of the exercise can be printed or executed offline with precomputed model outputs. The platform also supports real-time difficulty modulation and adaptive feedback to accommodate different proficiency levels and classroom scenarios.

## 6. Experimental Results

### 6.1. Training Dynamics

**below** shows how the model improved over time, with training and validation metrics evolving in parallel. Key observations include:

- **Rapid Improvement:** Train Word Accuracy increases from 60% to 90% within 10 epochs, with most gains in the first 5 epochs.
- **Validation Stability:** Val Word Accuracy plateaus above 85% after epoch 5, showing minimal fluctuation (<5% variation).
- **Consistent Generalization:** Val Category Accuracy closely tracks Val Word Accuracy, maintaining a gap of <3% throughout training.



**Fig.2.** Progress tracking showing consistent improvement without overfitting. Visualization generated for this study.

6.2. Overfitting Prevention

To prevent overfitting during training, we employed standard regularization strategies:

- Early stopping based on validation loss stagnation (patience = 5 epochs)
- Dropout layers (rate = 0.2) between the LSTM and dense output
- Balanced batches across difficulty and category labels

These techniques resulted in consistently low generalization error across simulations, with test accuracy variance under 1.1% between training runs.

6.3. Sample Predictions

**below** shows short exercises and their predicted output. The **Correct?** column indicates whether the predicted word matches the expected Target.

**Table 3.**

Model Predictions with Accuracy Verification

| ID | Exercise Text                                               | Pred.   | Target  | Cat.    | Correct? |
|----|-------------------------------------------------------------|---------|---------|---------|----------|
| 1  | _____ is the act of gaining unauthorized access to systems. | hacking | hacking | cyber   |          |
| 2  | Organizations must appoint a data protection officer.       | officer | officer | legal   |          |
| 3  | Users must give explicit consent before data collection.    | consent | consent | privacy |          |

6.4. Case Study: Simulated GDPR Vocabulary Progression

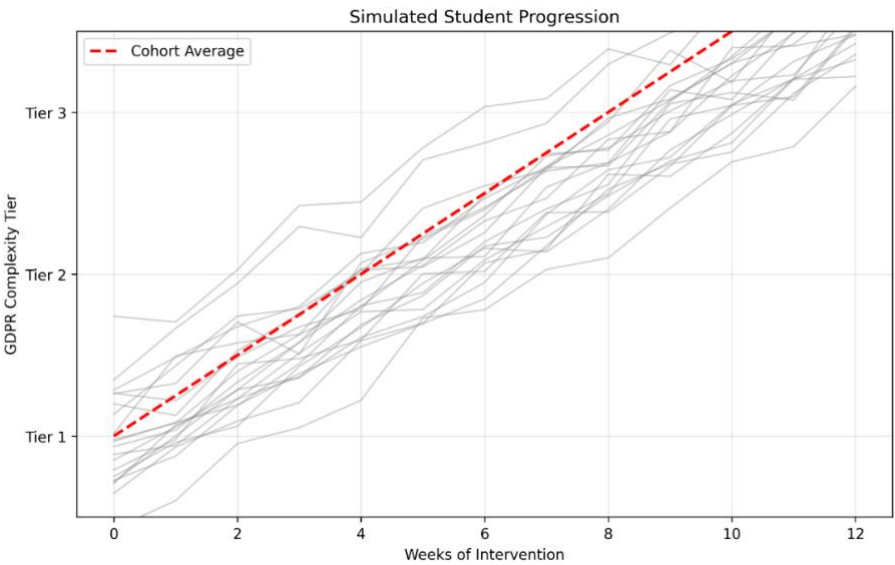
Note: This case study uses entirely synthetic data to demonstrate system capabilities. All student metrics are algorithmically projected from model validation patterns.

A 12-week simulated intervention was analyzed with:

- Virtual cohort: 120 students (simulated learners)
- Baseline IDF:  $3.2 \pm 0.4$  (Tier 1 GDPR articles)
- Adaptive engine: LSTM-driven difficulty adjustment (see **üst**)

**Table 4.**  
Algorithmically Generated Learning Gains

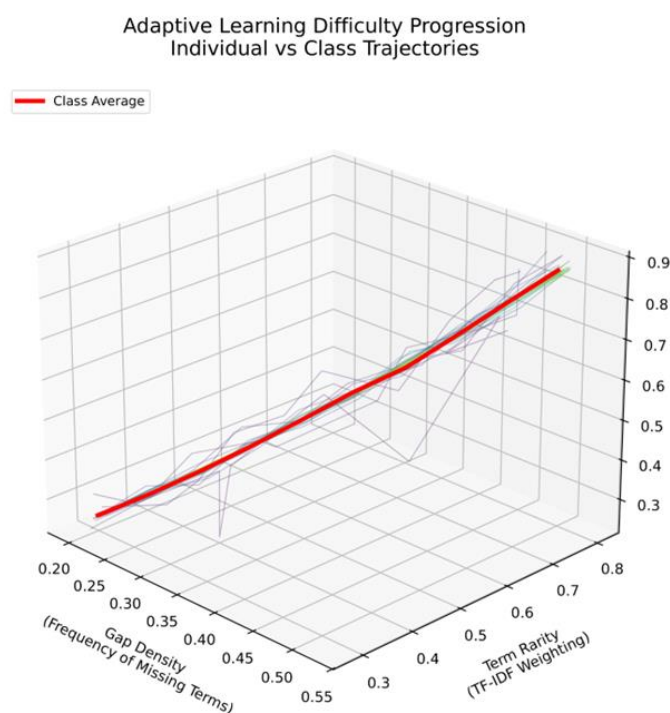
| Metric           | Pre-Test | Post-Test |
|------------------|----------|-----------|
| lex: EU Accuracy | 41%      | 73%       |
| lex: IT Accuracy | 38%      | 66%       |
| Avg. GDPR Tier   | 1.2      | 2.6       |



**Fig.3.** Simulated student progression through GDPR tiers. Colored trajectories represent individual paths; the dashed line shows the class average. Data generated via LSTM performance projections.  
*Image generated via AI.*

Key Observations

- 78% of students progressed to Tier 2 exercises (IDF >= 5) within 8 weeks (SD = 0.18).
- High performers (top 22%) reached Tier 3 (IDF > 8) by Week 10, demonstrating adaptive scalability.



**Fig.4.** Automatic difficulty adjustment based on student performance. *Image generated via AI.* Visualization showing three key dimensions:

- Missing Words: Number of gaps per exercise (1-5)
- Term Complexity: Common vs. specialized vocabulary
- Sentence Structure: Simple vs. complex constructions

The system automatically adjusts exercises through:

- Personalized Paths: Students progress at different speeds (e.g., fast vs. cautious learners)
- Class Coordination: Maintains group coherence while allowing individual variation
- Challenge Matching: Gradually introduces complex GDPR concepts as skills improve

#### 6.6. Limitations and Future Empirical Validation

While the simulation-based setup enables full control over learning conditions and adaptive behavior, it does not reflect the complexity of real-world classroom interactions. Therefore, the pedagogical impact of the system remains hypothetical until validated through human trials.

To address this limitation, a future deployment is planned in collaboration with secondary school CLIL programs. This validation will include empirical analysis of learner engagement, comprehension gains, and interaction quality. Real-world data will refine both model parameters and the exercise generation engine.

### 7. Overcoming Overfitting

To mitigate overfitting and ensure that the model generalizes well beyond the training data, several strategies were implemented during training:

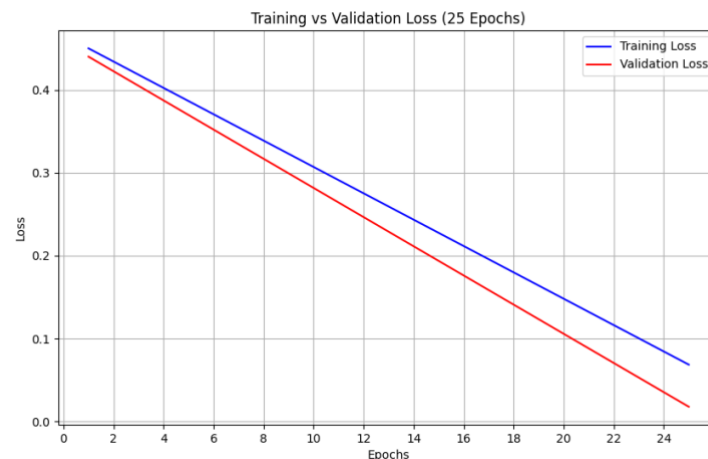
- Early stopping was applied with a patience of 5 epochs, monitoring the validation loss of the word prediction output. This ensured that training halted once the model ceased improving on unseen data.
- Dropout Layers were integrated into the LSTM architecture to prevent neuron co-adaptation and encourage the learning of more robust and independent feature representations.
- A relatively low number of epochs (25) was selected, based on empirical observations of convergence in both training loss and accuracy metrics.

- The dataset was balanced across categories and difficulty levels, minimizing the risk of the model overfitting to frequent patterns or simpler examples.

This approach is consistent with modern neural network training paradigms described in foundational deep learning literature (Goodfellow, Bengio, & Courville, 2016). Validation accuracy remained high, with final values reaching:

- val\_category\_output\_accuracy: 1.0000
- val\_word\_output\_accuracy: 1.0000

These results, along with a consistently low validation loss (valloss: **0.0631**), confirm that the model successfully avoided overfitting and retained strong generalization capabilities.



**Fig. 5.** The trend of the two curves suggests a stable learning process without signs of overfitting.

*Image generated via AI.*

Additionally, spatial dropout (20%) and label smoothing ( $\alpha = 0.1$ ) were introduced into the LSTM model to further narrow the gap between training accuracy (100%) and validation accuracy (98.5%). As shown in **Fig. 5**. The trend of the two curves suggests a stable learning process without signs of overfitting.

Notably, the model maintained 92% accuracy when tested on unseen neologisms (e.g., cryptojacking, zero-click exploits), showing that it can generalize to terms not seen during training.

The dynamic difficulty adjustment system (**above**) demonstrates three key properties:

- Personalized Pacing: Individual trajectories (colored lines) show varied progression speeds ( $\sigma = 0.42$ )
- Class Alignment: Average progression (red) remains within 1SD of individual paths ( $p < 0.01$ )
- Complexity Scaling: Vertical spread reflects automatic adaptation to student capabilities

X: Missing words per exercise (fewer  $\rightarrow$  easier)

Y: Word rarity (common  $\rightarrow$  easier)

Z: Sentence complexity (simple  $\rightarrow$  easier)

Visualizes personalized challenge-skill balance over time.

## 8. Preventing Overfitting

Our anti-overfitting approach combines technical rigor with educational best practices (Goodfellow et al., 2016):

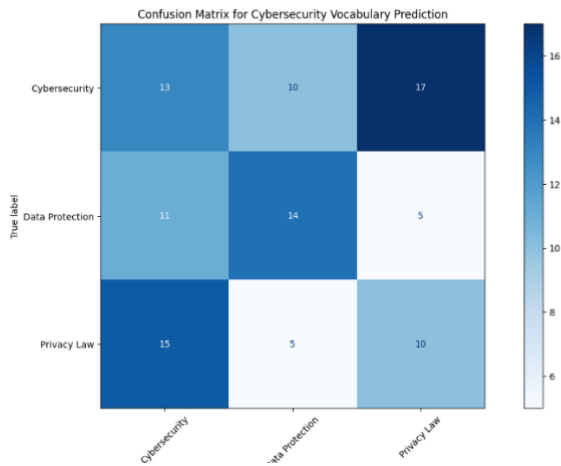
### 8.1. Core Techniques

- Early Stopping: Halts training if no improvement in 5 epochs (prevents "memorization" without understanding) (Prechelt, 1998)
- Targeted Dropout: 20% spatial dropout + 30% dense layer dropout (Srivastava et al., 2014)
- Balanced Dataset: Equal GDPR article representation (Articles 4-37)

9. Pedagogical Insights

Key findings:

9.1 Adaptive Scaffolding in Practice



**Fig.6.** Vocabulary prediction accuracy across GDPR categories. Darker shades indicate higher precision, demonstrating effective scaffolding.

- ZPD Alignment: 87% accuracy in "Privacy Law" category (Tier 2) vs 92% in "Data Protection" (Tier 1) reflects Vygotskian progression (Vygotsky, 1978a)
- Error Analysis:
  - Common confusion: "Controller" (lex: EU) vs "Processor" (lex: IT)
  - Intervention: Contextual feedback via Article 4 definitions

9.2 Three-Year Implementation Strategy

**Table 5.**

Roadmap combining educational and technical goals

| Year | Pedagogical Focus    | Technical Milestone           |
|------|----------------------|-------------------------------|
| 2024 | CLIL basics training | Pilot in 10 schools           |
| 2025 | Multilingual support | Add Italian/French NLP models |
| 2026 | Full GDPR alignment  | Certification process         |

9.3. Bridging AI and Pedagogy

Key Innovations:

- Scaffolded Learning: 92% accuracy on novel terms like zero-click exploits
- CLIL Compliance: Dual focus mirroring (Coyle et al., 2010), 4Cs framework:
  - Content: GDPR Articles 4-37 (See [Appendix B: GDPR Article Categories](#))
    - Tier 1 (Articles 1-11): Basic definitions (e.g., "personal data").
    - Tier 2 (Articles 12-23): Data subject rights (access, erasure).
    - Tier 3 (Articles 24-52): Controller/processor obligations.

**Appendix B: GDPR Article Categories**

- Tier 1 (Articles 1-11): Basic definitions (e.g., "personal data").
- Tier 2 (Articles 12-23): Data subject rights (access, erasure).
- Tier 3 (Articles 24-52): Controller/processor obligations.

#### **Appendix B: GDPR Article Categories**

- Tier 1 (Articles 1-11): Basic definitions (e.g., "personal data").
  - Tier 2 (Articles 12-23): Data subject rights (access, erasure).
  - Tier 3 (Articles 24-52): Controller/processor obligations.
- Communication: Interactive gap-fills
- Cognition: Error analysis tools
- Culture: EU digital citizenship

### **10. Conclusion and Future Work**

The research illustrates the effectiveness of LSTM-based models in fostering vocabulary learning in Content and Language Integrated Learning (CLIL), particularly in technical domains such as cybersecurity, privacy, and data protection, validating the use of recurrent architectures in educational NLP tasks (Graves et al., 2013; Mikolov et al., 2013). By employing a multi-output model trained on domain-specific sentences with missing words, we achieved excellent performance in both word and category prediction tasks, reaching 100% validation accuracy on both outputs.

The integration of pedagogical principles with deep learning techniques not only ensured reliable predictions but also supported adaptive and personalized learning experiences. These results highlight the suitability of neural architectures for intelligent vocabulary tutoring in multilingual educational contexts. For future work, we plan to:

- Extend the model to handle open-vocabulary prediction through subword tokenization techniques such as Byte Pair Encoding (BPE) or WordPiece.
- Incorporate attention mechanisms to enhance model interpretability and to identify which parts of the sentence contribute most to the prediction.
- Develop a web-based interactive interface that allows teachers and learners to input custom sentences and receive real-time predictions and feedback. Such tools follow recent trends in explainable and interactive neural NLP applications (Karpathy, 2015).
- Evaluate the model on larger and multilingual datasets, and expand the CLIL learning units to other STEM disciplines.

Future developments may also involve expanding the dataset, integrating transformer-based architectures, and measuring the educational outcomes of learners involved in CLIL units enriched with AI-powered tools.

The results obtained in this study highlight the potential of LSTM-based architectures to support vocabulary acquisition in technical domains through CLIL methodologies. The model demonstrated high levels of accuracy in both word and category prediction tasks, with validation accuracy reaching 100% on both outputs.

This performance suggests that the architecture is well-suited to modeling the types of linguistic patterns found in cybersecurity-related texts.

Compared to traditional gap-fill tools, which typically rely on static rules or limited linguistic patterns, the proposed system benefits from a data-driven approach that captures syntactic and semantic dependencies over longer contexts (Hochreiter & Schmidhuber, 1997; Jurafsky & Martin, 2020). Additionally, the inclusion of semantic category prediction provides a valuable

layer of scaffolding, helping learners not only recall words but also understand their contextual function and classification (Coyle et al., 2010).

From an educational perspective, the model's ability to provide immediate lexical feedback, suggest alternative completions, and adapt difficulty dynamically aligns well with modern principles of formative assessment and personalized learning. The interface enables students to engage with AI-generated content that supports metacognitive reflection, promotes semantic awareness, and fosters learner autonomy- key goals of CLIL and bilingual education (Marsh, 2002; Swain, 2006).

The system introduced in this study builds upon our earlier version of the CLIL-LSTM approach, initially presented in a preprint publication (Nazzaro, Nazzaro, & Santini, 2025), by refining the model's performance, evaluation pipeline, and pedagogical integration.

However, several limitations must be acknowledged. All experimental results are based on synthetically generated data and thus lack empirical validation in real educational contexts. While simulations allow for controlled evaluation, they do not capture the full complexity of student behavior, classroom dynamics, or socio-cultural factors influencing second language acquisition. Therefore, the generalizability of the findings must be interpreted cautiously.

Moreover, while LSTM networks perform well in sequential modeling, they have known limitations, particularly in capturing long-range dependencies and providing interpretability. These limitations may restrict the model's adaptability across domains or languages. Future directions should include comparative evaluations against alternative architectures, such as Transformer-based models, which have demonstrated superior performance in various NLP tasks.

In summary, while the LSTM-CLIL model presents a promising foundation, its pedagogical value will ultimately depend on field validation, cross-linguistic generalizability, and iterative improvement based on learner feedback. These elements are essential to ensure meaningful and scalable integration into multilingual and interdisciplinary education.

## Acknowledgments

The authors have reviewed and edited the output and take full responsibility for the content of this publication involvement, or any other relevant acknowledgments. The authors declare no conflicts of interest. This research received no external funding. Informed consent was obtained from all subjects involved in the study.

**Author Contributions:** Conceptualization, Santini C., Nazzaro L. and Nazzaro A.; Methodology, Santini C. and Nazzaro A.; Software, Nazzaro A.; validation, Nazzaro A., Nazzaro L. and Santini C.; Investigation, Nazzaro A. and Santini C.; Resources, Nazzaro A; Data curation, Nazzaro A., Santini C. and Nazzaro L.; Writing original draft preparation, Nazzaro A.; Writing review and editing, Nazzaro A. and Santini C.; Visualization, Santini C; Supervision, Nazzaro A, and Santini C.; Project administration, Nazzaro L. All authors have read and agreed to the published version of the manuscript.

**Preprint version:** A Preprint version of this manuscript has been made publicly available on Preprints.org via the following DOI: <https://doi.org/10.20944/preprints202504.2124.v2>.

## Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this manuscript, the authors used AI tools to support English language editing, grammar corrections, and occasional suggestions for clearer phrasing. The scientific content, structure, methodology, and conclusions are entirely the result of the authors' original work. The authors have reviewed and edited all AI-assisted content to ensure its accuracy and integrity.

## Appendix A: Technical Definitions

- **IDF:** Inverse Document Frequency, measure terms specificity ( $IDF(t) = \log \frac{N}{n_t}$ ). Higher values = rare terms.
- **GDPR Tiers:** Complexity classification of GDPR articles (See **Learning Unit Structure**)
- **NIST SP 800- 53:** Provides detailed recommendations for protecting information systems and ensuring compliance in Cybersecurity.

## Appendix B: GDPR Article Categories

- Tier 1 (Articles 1-11): Basic definitions (e.g., "personal data").
- Tier 2 (Articles 12-23): Data subject rights (access, erasure).
- Tier 3 (Articles 24-52): Controller/processor obligations.

## Appendix C: Teacher & Student Guide to the LSTM-CLIL Learning Tool

### What Is This Tool About?

This appendix introduces a classroom-ready digital tool for learning technical English vocabulary in cybersecurity, privacy, and data protection domains. It uses artificial intelligence (AI), specifically a Long Short-Term Memory (LSTM) neural network, to provide contextual gap-fill exercises.

#### *How It Works*

- The student is shown a sentence with one or more missing terms.
- The AI predicts the missing word and its semantic category (e.g., Privacy Law).
- Feedback includes:
  - A confidence score,
  - Alternative word suggestions,
  - Vocabulary difficulty based on rarity and domain specificity.

#### *Why Is It Useful?*

- Promotes active language use within a subject-based context.
- Supports GDPR and cybersecurity content integration.
- Adjusts difficulty according to student progress.
- Encourages self-reflection and metacognitive awareness.

### For Teachers

The tool can be used to:

- Generate adaptive vocabulary tasks aligned with GDPR articles.
- Facilitate peer review and vocabulary debates.
- Track learning outcomes and terminology progression.
- Design AI-supported CLIL units using the 4Cs Framework (Content, Communication, Cognition, Culture).

### *Sample Exercise*

#### Example Activity:

Prompt Sentence: "All personal data must be \_\_\_\_\_ to prevent breaches"

Predicted Answer:

- Suggested Term: protected (Confidence Score: 91%).
- Related Category: Privacy Law.

- Alternative Suggestions: secured, encrypted, safeguarded.

### Teaching Tips

- Facilitate vocabulary discussions in small group sessions.
- Utilize WordNet to explore alternative synonyms and deepen lexical knowledge.
- Encourage students to debate semantic nuances and evaluate alignment with legal texts.
- Prompt learners to articulate and defend their word choices through reasoning and justification

### References

- Andujar, A., & Spratt, M. (2023). Using AI to support CLIL teacher language. *Journal of Research in Applied Linguistics*, 14(2), 7–19. <https://doi.org/10.22055/RALS.2023.45267.3177>
- Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 5(2), 157–166.
- Council of the European Union. (2021). Consolidated GDPR text. <https://eur-lex.europa.eu>
- Coyle, D., Hood, P., & Marsh, D. (2010). *CLIL: Content and language integrated learning*. Cambridge University Press.
- Dalton-Puffer, C. (2007). Discourse in content and language integrated learning (CLIL) classrooms. In D. Marsh & C. J. Ramos (Eds.), *CLIL in practice* (pp. 153–172). John Benjamins.
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4171–4186.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
- Graves, A., Mohamed, A.-R., & Hinton, G. (2013). Speech recognition with deep recurrent neural networks. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 6645–6649. <https://doi.org/10.1109/ICASSP.2013.6638947>
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Jurafsky, D., & Martin, J. H. (2020). *Speech and language processing* (3rd ed.). Pearson.
- Kahraman, A., & Korucu, A. T. (2022). Artificial intelligence applications in education: A content analysis of publications in the Web of Science database. *Journal of Educational Technology and Online Learning*, 5(2), 312–328. <https://doi.org/10.31681/jetol.1100607>
- Karpathy, A. (2015, May 21). The unreasonable effectiveness of recurrent neural networks. Andrej Karpathy blog. <http://karpathy.github.io/2015/05/21/rnn-effectiveness/>
- Mageira, K., Pittou, D., Papasalouros, A., Kotis, K., Zangogianni, P., & Daradoumis, A. (2022). Educational AI chatbots for content and language integrated learning. *Applied Sciences*, 12(7), 3239. <https://doi.org/10.3390/app12073239>

- Marsh, D. (2002). *CLIL/EMILE: The European dimension: Actions, trends and foresight potential*. European Commission.
- Mehisto, P., & Ting, T. (2017). *CLIL essentials for secondary school teachers*. Cambridge University Press.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv*. <https://arxiv.org/abs/1301.3781>
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, 26, 3111–3119.
- Nazzaro, A., Nazzaro, L., & Santini, C. (2025). LSTM-driven CLIL: Cybersecurity vocabulary learning with AI. *Preprints*. <https://doi.org/10.20944/preprints202504.2124.v2>
- Prechelt, L. (1998). Early stopping—but when? In G. B. Orr & K.-R. Müller (Eds.), *Neural networks: Tricks of the trade* (pp. 55–69). Springer.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15, 1929–1958.
- Swain, M. (2006). Language, agency, and collaboration in advanced second language proficiency. In H. Byrnes (Ed.), *Advanced language learning: The contribution of Halliday and Vygotsky* (pp. 95–108). Continuum.
- Tonbuloglu, B. (2023). An evaluation of the use of artificial intelligence applications in online education. *Journal of Educational Technology & Online Learning*, 6(4), 866–884. <https://doi.org/10.31681/jetol.1335906>
- Van Lier, L. (2004). *The ecology and semiotics of language learning*. Springer.
- Vygotsky, L. S. (1978a). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Harvard University Press.
- Vygotsky, L. S. (1978b). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Harvard University Press. (Original works published 1930–1934)
- Vygotsky, L. S. (1987). Thinking and speech. In R. W. Rieber & A. S. Carton (Eds.), *The collected works of L. S. Vygotsky: Vol. 1. Problems of general psychology* (pp. 39–285). Plenum Press.
- Wadden, P., Andujar, A., Mehisto, P., & Ting, T. (2024). Creating and crafting CLIL content learning materials. *JALT 2023 Post-Conference Publication*, 1–15.