
AI4LANG – Integrating Artificial Intelligence for the Digital Transformation of Language Teaching

Benedikt Perak¹[\[0000-0003-4177-5307\]](#), Dragana Špica²[\[0000-0003-1008-2700\]](#)

¹ University of Rijeka, Faculty of Humanities and Social Sciences / Syntagent d.o.o., Rijeka, Croatia, bperak@uniri.hr

² University of Jurja Dobrile in Pula, Faculty of Humanities, Croatia, dragana.spica@unipu.hr

Scientific paper

DOI: 10.18485/judig.2026.2.ch16

Abstract

The rapid advancement of generative AI and natural language processing has created transformative opportunities for language education, yet the integration of these technologies into pedagogically coherent, CEFR-aligned learning environments remains a significant challenge. This paper presents AI4LANG, a multinational research and development initiative that expands a production-ready AI Language Tutor platform across three typologically distinct languages – English, Japanese, and Croatian – addressing the global need for adaptive, AI-mediated language instruction that serves both high-resource and underrepresented languages. The platform’s architecture is built on a multi-agent orchestration paradigm where specialized AI agents handle distinct pedagogical functions – exercise generation, pronunciation evaluation, grammar feedback, and learning path adaptation – coordinated through a central orchestrator operating over a dual-database knowledge layer combining vector embeddings with structured lexical and grammatical graphs. We present the technical architecture, the CEFR-aligned pedagogical framework underpinning adaptive exercise generation, and the methodology for creating language-specific knowledge resources through LLM-assisted inference validated by expert review. The paper contributes a replicable model for developing multilingual AI tutoring systems that balance technological sophistication with pedagogical rigor across diverse linguistic contexts.

Keywords: AI language tutoring, multi-agent architecture, CEFR, multilingual NLP, adaptive learning, knowledge graphs, speech recognition

1. Introduction

Language learning is undergoing a fundamental transformation driven by advances in artificial intelligence. Large language models (LLMs) can now generate contextually appropriate texts, evaluate grammatical correctness, explain linguistic structures, and engage in open-ended dialogue at levels approaching human tutors. Simultaneously, automatic speech recognition (ASR) and text-to-speech (TTS) systems have reached accuracy thresholds that make real-time pronunciation feedback and spoken interaction feasible for educational applications. Yet the majority of AI language tools available to learners remain disconnected from established pedagogical frameworks, offering generic chatbot interactions without systematic alignment to proficiency standards, adaptive sequencing, or structured knowledge representation (Kasneci et al., 2023; Godwin-Jones, 2022).

The gap between AI capability and pedagogical deployment is particularly acute for languages beyond the small set of high-resource languages that dominate commercial platform development. While English, Spanish, and Mandarin benefit from extensive digital resources and LLM training data, the majority of the world’s languages – including many European languages with rich cultural and educational traditions – lack AI tutoring infrastructure that meets the standards of formal language education. This asymmetry not only limits learning opportunities for speakers of these languages but also perpetuates a digital divide in which AI educational benefits accrue disproportionately to already well-resourced linguistic communities (Bender et al., 2021; Jespersen, 1922).

The AI4LANG project – Integrating Artificial Intelligence for the Digital Transformation of Language Teaching – addresses these challenges through a coordinated effort to expand an existing AI Language Tutor platform to three languages spanning different linguistic families, writing systems, and resource levels. The platform, currently in production use for Japanese language instruction, combines generative AI with structured knowledge representation, speech processing, and adaptive exercise generation within a unified multi-agent architecture. The project extends this system to English (a high-resource Germanic language with global reach) and Croatian (a South Slavic language with approximately 5.5 million speakers and limited AI tutoring resources), creating a testbed for multilingual AI tutoring system design that spans the resource spectrum.

The initiative brings together a multidisciplinary consortium spanning computational linguistics, language pedagogy, software engineering, and digital humanities across five European countries. University partners in Croatia (Rijeka, Pula, Zagreb), Ireland (Dublin), Poland (Krakow), and Serbia (Belgrade) contribute language-specific expertise and pilot testing infrastructure, while industry partner Syntagent provides the technical architecture and platform development capabilities. This combination of academic depth and engineering capacity creates a development environment in which pedagogical requirements directly inform technical design decisions, and empirical classroom evidence drives iterative platform refinement.

This paper contributes: (1) a multi-agent orchestration architecture for language tutoring that separates pedagogical functions into specialized, independently optimizable AI agents; (2) a methodology for constructing language-specific knowledge graphs – lexical, grammatical, and proficiency-aligned – through LLM-assisted inference with expert validation; (3) a CEF¹-anchored framework for adaptive exercise generation that maintains pedagogical coherence across languages while accommodating typological diversity; and (4) a replicable model for extending AI language tutoring infrastructure to under-resourced languages.

2. AI Language Tutor Platform Architecture

The AI Language Tutor is a full-stack web application that provides an integrated environment for language learning, combining generative AI with structured pedagogical content and multimodal interaction. The platform has been developed through an interdisciplinary collaboration involving computational linguists, language educators, and software engineers, and has been in active production use for Japanese language instruction at the university level since 2025. The overall architecture of the platform is shown in Figure 1.

2.1. System Design

The platform follows a modular, service-oriented architecture. The backend, built with Python (FastAPI), interfaces with two complementary database systems: PostgreSQL with the pgvector exten-

¹ Common European Framework of Reference for Languages (CEFR) <https://www.coe.int/en/web/common-european-framework-reference-languages/level-descriptions>

sion for dense vector embeddings enabling semantic similarity search, and Neo4j for graph-based representation of lexical relations, grammatical patterns, and CEFR-aligned learning objectives. This dual-database design reflects a key architectural insight: vector embeddings excel at fuzzy semantic matching and similarity-based retrieval, while graph structures enable precise traversal of explicitly defined relationships – synonymy, hypernymy, grammatical governance, proficiency prerequisites – that are essential for pedagogically grounded exercise generation. The frontend is developed with Next.js (React), communicating with the backend via a RESTful API. Figure 1 summarises this layered design and shows how the orchestration layer, the dual knowledge layer and the external AI services are separated by explicit interfaces, so that each can be revised independently of the others. External AI services – LLMs, ASR, TTS, and image generation – are accessed through provider abstraction layers supporting multiple API backends with automatic fallback.

The knowledge representation layer merits particular attention as it distinguishes the platform from general-purpose LLM chatbots applied to language learning. Lexical knowledge is modeled as a graph where word nodes are connected by typed semantic relations – synonymy, hypernymy, hyponymy, antonymy, meronymy, and semantic field membership – inferred through a combination of LLM-based analysis, existing lexical resources, and expert linguistic validation. For Japanese, this graph currently encompasses over 17,000 word nodes with more than 67,000 typed relations. Grammar patterns are similarly structured as graph entities, linked to the lexical nodes they govern and to the CEFR descriptors they support. Can-Do descriptors from the CEFR Companion Volume ([Council of Europe, 2020](#)) are modeled as nodes connected through prerequisite and extension relations, forming a navigable proficiency network that the learning path agent traverses to construct individualized exercise sequences.

2.2. Exercise Modalities and AI Integration

The platform supports multiple exercise modalities, each leveraging different AI capabilities within a coherent pedagogical design. Vocabulary exercises use retrieval augmented generation (RAG) over the lexical graph to produce contextually appropriate example sentences, multiple-choice questions, and fill-in-the-blank tasks cali-

AI Language Tutor — platform architecture

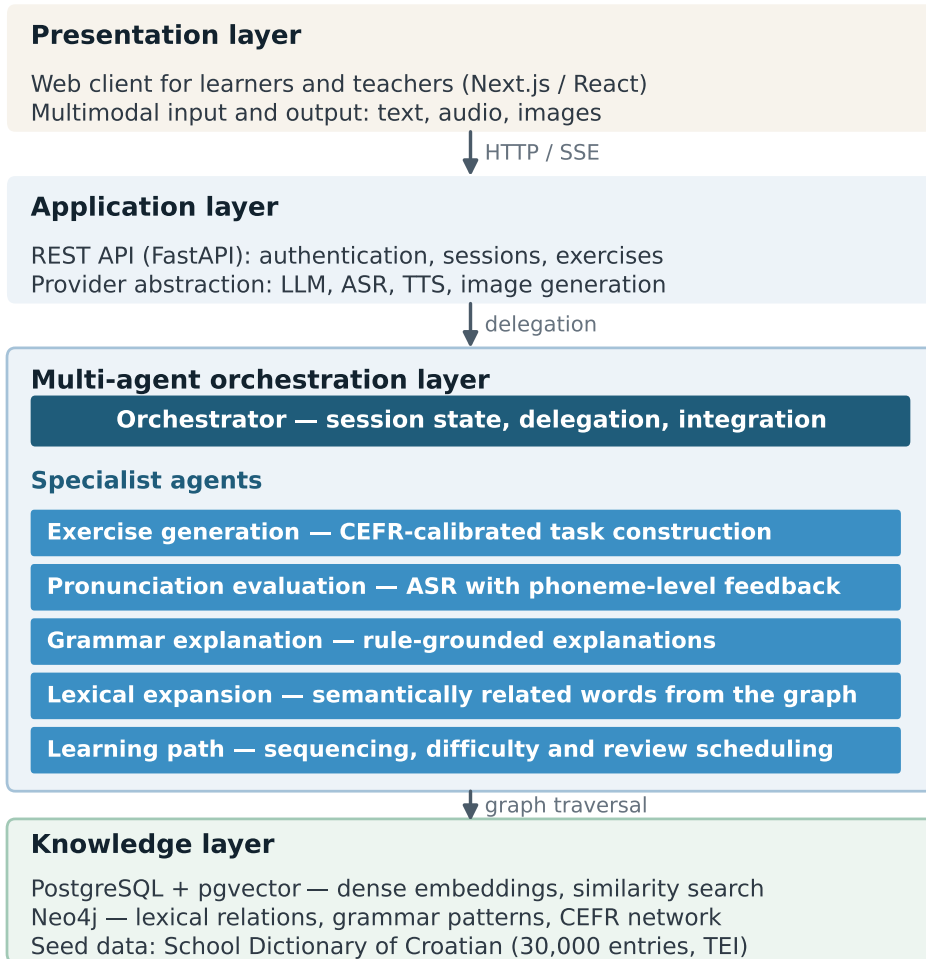


Figure 1: AI Language Tutor – platform architecture. The presentation and application layers are separated from a multi-agent orchestration layer whose specialist agents are grounded in a dual knowledge layer (dense vector embeddings plus a lexical, grammatical and CEFR knowledge graph) and in external AI services accessed through a provider abstraction layer with automatic fallback.

brated to the learner’s proficiency level. Pronunciation exercises employ ASR with forced alignment to evaluate learner speech against target utterances, providing phoneme-level feedback rather than simple word-level accuracy scores. Grammar exercises draw on the gram-

mar pattern graph to generate transformation and completion tasks targeted to specific grammatical structures. Reading and listening comprehension exercises retrieve or generate texts at specified CEFR levels, followed by comprehension questions with automated evaluation of open-ended responses. A conversation mode enables free-form dialogue practice with an AI partner that dynamically adapts its language complexity to the learner’s demonstrated level.

A defining characteristic of the platform is that exercises are generated dynamically rather than retrieved from a static bank, enabling virtually unlimited practice opportunities while maintaining pedagogical coherence. The generation process is governed by structured prompts parameterized by the target vocabulary scope, grammatical focus, proficiency constraints, and exercise type. Learner performance data – response accuracy, time on task, error patterns, and help requests – is continuously collected and feeds into the learning path agent’s recommendation logic (Perak and Špica, 2026).

2.3. Current Deployment

As of mid-2026, the platform is in active production use for Japanese language instruction, supporting university-level learners with exercises spanning CEFR levels A1 through C1. The Japanese module includes 993 Can-Do descriptors across all proficiency levels, a lexical graph of 17,000+ entries with semantic relations, and a growing inventory of grammar patterns. The Croatian module is under active development with foundational vocabulary, lexical relations covering 53,000+ word nodes and 67,000+ semantic edges, and initial A1–A2 Can-Do descriptors in place. The English module is in the design phase, with content development scheduled as part of the multilingual expansion described below.

The transition from a primarily Japanese-focused platform to a multilingual system raises architectural questions that extend beyond content translation. Japanese and English share no genetic relationship and employ fundamentally different writing systems, yet both have benefited from extensive LLM training – albeit with differing levels of tokenization efficiency. Croatian, by contrast, represents a Slavic language with rich inflectional morphology whose digital footprint is orders of magnitude smaller. Extending the platform across this typological and resource spectrum serves as a stress test for the

architecture’s language-agnostic design principles and will produce insights applicable to other language pairs facing similar asymmetries.

3. Multi-Agent Orchestration for Language Tutoring

At the core of the platform’s AI capabilities is a multi-agent orchestration framework that represents a departure from the monolithic LLM-application pattern common in current educational AI tools. Rather than relying on a single large language model to handle all aspects of language instruction – from content generation to feedback to curriculum sequencing – the platform decomposes the tutoring task into specialized functions, each implemented by a dedicated AI agent operating within a structured coordination framework.

3.1. Agent Specialization

The agent ensemble includes: an exercise generation agent that creates language tasks calibrated to specified CEFR levels, target vocabulary, and grammatical focus areas; a pronunciation evaluation agent that processes learner audio through ASR, aligns it with target utterances using forced alignment, and produces phoneme-level diagnostic feedback; a grammar explanation agent that accesses the grammar pattern graph to provide contextualized, rule-based explanations of grammatical structures; a lexical expansion agent that retrieves semantically related words from the knowledge graph to support vocabulary acquisition through meaningful associations; and a learning path agent that analyzes cumulative performance history to recommend optimal exercise sequences, difficulty progression, and review scheduling.

These specialist agents are coordinated by an orchestrator agent that manages the overall learning session: interpreting learner inputs, determining the appropriate pedagogical response, delegating tasks to specialist agents, and integrating their outputs into a coherent interaction. The orchestrator maintains session state and ensures that the sequence of activities follows a pedagogically sound progression. Figure 2 traces a complete learner turn through this ensemble, from the initial request through graph-grounded task generation to the adapted exercise sequence returned by the learning-path agent.

Multi-agent orchestration — learner session flow

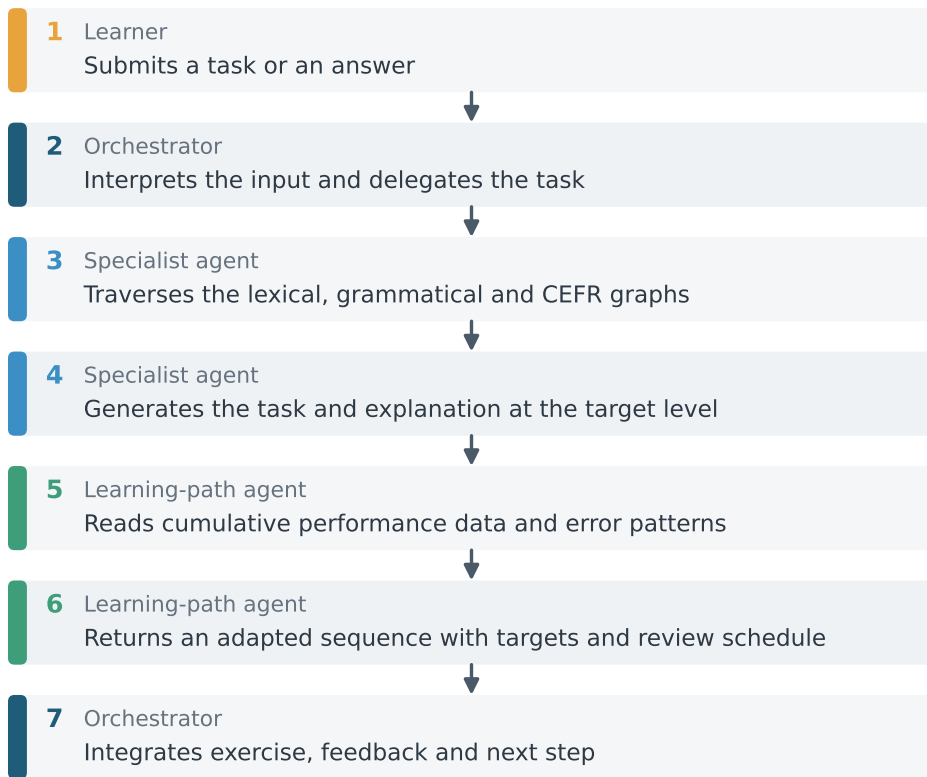


Figure 2: Learner session flow across the agent ensemble. The orchestrator delegates the task to a specialist agent, which retrieves task parameters by traversing the knowledge graph; the learning-path agent then reads the learner’s cumulative performance data and returns an adapted sequence before the orchestrator integrates the final response.

3.2. Architectural Advantages

The multi-agent design provides several advantages over monolithic approaches. First, each agent can be optimized independently: different language models, prompting strategies, and evaluation criteria can be assigned per function, allowing the pronunciation agent to use a model optimized for phonetic analysis while the grammar agent uses one with superior metalinguistic reasoning. Second, the architecture exhibits graceful degradation: if one agent produces suboptimal output or encounters an error, others continue functioning, and the

orchestrator can implement fallback strategies. Third, the modular design facilitates incremental improvement: individual agents can be upgraded or replaced without affecting the overall system, allowing the platform to benefit from rapid advances in AI capabilities without architectural overhaul.

This design draws on recent advances in agentic AI systems (Wu et al., 2023) while addressing a specific limitation of general-purpose agent frameworks: the need for deep integration with structured, domain-specific knowledge representations. The agents are grounded in lexical and grammatical knowledge graphs that encode linguistically validated relationships, ensuring that generated content respects genuine language structure rather than reproducing plausible but linguistically inaccurate patterns that LLMs may hallucinate when operating purely from parametric knowledge.

3.3. Explainability and Pedagogical Transparency

An important design consideration for AI systems in education is explainability: learners and educators need to understand not just what the system recommends, but why (Khosravi et al., 2022; Conati et al., 2018). In the multi-agent architecture, explainability is implemented at the agent level: the grammar explanation agent references specific grammatical rules and provides examples; the pronunciation agent identifies specific phonemes and articulatory features; the learning path agent articulates the rationale behind exercise recommendations by citing performance history and the pedagogical sequence connecting current and target competencies. This transparency is essential for building learner trust and for enabling educators to understand and potentially override automated recommendations, maintaining the teacher’s role as the ultimate pedagogical authority.

4. Multilingual Knowledge Graph Construction

A central challenge in extending AI tutoring to new languages is the creation of structured linguistic knowledge resources that can drive pedagogically meaningful exercise generation. Unlike general-purpose LLMs that rely on patterns learned during pre-training, an AI tutor requires explicit representations of lexical relationships, grammatical structures, and proficiency progressions that can be queried, traversed, and reasoned over deterministically. This section describes the methodology developed for constructing these re-

Knowledge-graph construction pipeline

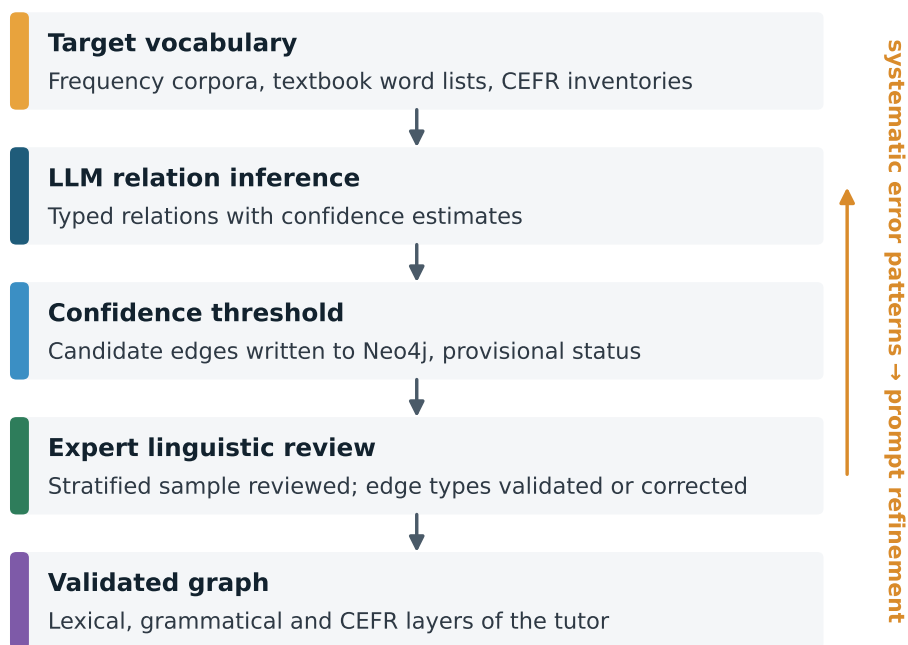


Figure 3: Knowledge-graph construction pipeline. LLM-based relation inference produces typed candidate edges with confidence estimates; edges above a configurable threshold are written into Neo4j in provisional status, reviewed by expert linguists on a stratified sample, and only then promoted to the validated graph. Systematic error patterns identified during review feed back into prompt refinement.

sources across languages with varying levels of existing digital infrastructure. Figure 3 shows the construction pipeline, including the expert-validation step that separates machine-proposed candidate relations from relations admitted into the validated graph.

4.1. Lexical Relation Inference Pipeline

The lexical knowledge graph construction employs a semi-automated pipeline that combines LLM-based inference with expert linguistic validation. For a given target vocabulary — drawn from frequency-ordered corpora, textbook word lists, and CEFR-aligned vocabulary inventories — the pipeline proposes semantic relations between word pairs using structured LLM prompts that specify the rela-

tion taxonomy (synonymy, hypernymy, hyponymy, antonymy, near-synonymy, semantic field membership, meronymy, holonymy) and require confidence estimates. Proposed relations above a configurable confidence threshold are inserted into the Neo4j graph as candidate edges. Expert linguists then review a stratified sample of proposed relations, validating or correcting edge types and identifying systematic error patterns that inform prompt refinement.

This methodology has been validated on Croatian, a language for which comprehensive digital lexical resources comparable to WordNet (Fellbaum, 1998) were not available at project outset. The initial pipeline run processed over 53,000 word nodes, generating 67,698 typed semantic relations across eight relation types. Expert review of the output identified specific areas requiring prompt adjustment — particularly the distinction between near-synonyms and synonyms, and the appropriate granularity of semantic field classification — leading to iterative improvements in relation quality. The same methodology will be applied to English, where existing resources such as WordNet and FrameNet (Fillmore et al., 2003) provide seed data that can be enriched and extended through LLM-assisted inference, and to Japanese, where the existing graph of 17,000+ nodes with 67,000+ relations will be expanded with additional vocabulary domains.

4.2. Grammar Pattern Modeling

Grammar patterns are modeled as graph structures connecting grammatical constructions to the lexical items that instantiate them and the CEFR descriptors that require them. For example, the Croatian genitive case following prepositions is represented as a grammar node connected to all prepositions that govern it, to the CEFR descriptors at which case government is introduced (typically A2 for basic prepositions), and to related grammar nodes such as accusative case for prepositions of motion. This graph representation enables the exercise generation agent to produce tasks that target specific grammatical structures while respecting the learner’s proficiency level — generating a genitive-case exercise at A2 using high-frequency prepositions, or a more complex case-alternation exercise at B1 requiring discrimination between genitive and accusative.

The grammar pattern modeling approach is language-specific by design, acknowledging that grammatical categories and their acquisition sequences differ across languages. For English, the grammar

graph encodes tense-aspect systems, article usage, and phrasal verb constructions; for Japanese, it represents particle usage, honorific registers, and the topic-comment structure; for Croatian, it encodes the seven-case declension system, verbal aspect pairs, and clitic ordering constraints. This language-specific modeling, grounded in the knowledge graph rather than embedded in agent prompts, allows the exercise generation agent to operate in a largely language-agnostic manner, parameterized by the graph structures it traverses.

4.3. CEFR Descriptor Networks

The descriptor mapping is adapted per language, acknowledging that linguistic typology affects the feasibility and ordering of communicative tasks. The Japanese writing system introduces orthographic complexity not present in English or Croatian, requiring CEFR descriptor adaptation for literacy skills. Croatian grammatical gender and Japanese honorific registers introduce language-specific pedagogical sequencing considerations. These adaptations are performed by domain experts for each language, with the resulting descriptor networks serving as the pedagogical backbone for exercise generation across all three languages.

5. CEFR-Aligned Adaptive Exercise Framework

The pedagogical framework governing exercise generation and sequencing is anchored in the action-oriented approach of the CEFR, which defines language proficiency in terms of what learners can do with the language in real-world communicative contexts. This section describes how the CEFR taxonomy is operationalized within the platform’s AI architecture to produce pedagogically coherent, adaptive learning experiences.

5.1. Parameterized Exercise Generation

Exercise generation is governed by a structured prompting framework that decomposes each Can-Do descriptor into task parameters: language activity type (reception, production, interaction, mediation), topic domain, vocabulary scope, grammatical constraints, and expected output complexity. The exercise generation agent uses these parameters to produce tasks that are both linguistically targeted and pedagogically varied. For receptive skills, the agent generates or se-

lects texts at the specified CEFR level and produces comprehension questions targeting the descriptor’s competence. For productive skills, the agent generates prompts that elicit language at the target level and evaluates learner output against CEFR-aligned criteria including range, accuracy, fluency, coherence, and phonological control.

Critically, the parameterization is grounded in the knowledge graph rather than encoded directly in prompts. When the agent receives a generation request for a specific descriptor, it traverses the graph to retrieve the associated vocabulary, grammar patterns, and proficiency constraints, then constructs a prompt that incorporates these retrieved elements. This architecture ensures that exercise content remains consistent with the curated knowledge base even as the underlying LLM is updated or replaced — the pedagogical parameters are stored in the graph, not in the prompt template.

To illustrate this parameterization concretely: when generating a B1-level writing exercise for Croatian aligned to the descriptor “Can write accounts of experiences, describing feelings and reactions in simple connected text,” the system retrieves from the graph the vocabulary domain for personal experiences and emotions, grammar patterns for past tense (perfect) and adjectival agreement, and discourse connectors appropriate for B1 narrative cohesion. The exercise generation agent constructs a prompt incorporating these retrieved elements — for instance, asking the learner to describe a memorable trip using specified vocabulary and grammatical structures — while the evaluation agent checks the response against the same parameter set. This graph-grounded generation stands in contrast to purely prompt-based approaches where the LLM alone must infer appropriate CEFR constraints from its parametric knowledge, which may be unreliable for less-resourced languages and proficiency levels.

5.2. Multi-Scale Adaptivity

Adaptivity operates at multiple timescales within the platform. Within a single session, exercise difficulty and topic are adjusted based on immediate performance: successful completion leads to increased complexity or introduction of new vocabulary, while errors trigger review exercises targeting the specific linguistic feature that caused difficulty. Across sessions, the learning path agent analyzes cumulative performance patterns — error clusters, response times, help request frequency — to identify persistent difficulties and

under-practiced skill areas. This analysis informs both exercise selection and, at longer timescales, recommendations for learning strategy adjustments that are communicated to the learner and, where applicable, to the instructor.

The multi-scale adaptivity design is informed by research on spaced repetition and personalized difficulty calibration in language learning (Settles and Meeder, 2016; Van der Linden and Glas, 2010). Unlike fixed-interval spaced repetition systems, the platform’s approach adapts review intervals based on error patterns: items repeatedly confused are presented more frequently and in varied contexts, while items consistently mastered are reviewed at expanding intervals. The learning path agent also detects potential L1 interference patterns — for example, a Croatian learner of English systematically omitting articles (a grammatical category absent in Croatian) — and generates targeted contrastive exercises addressing the specific cross-linguistic challenge (Selinker, 1972; Ellis, 2002).

5.3. Pilot Testing and Evaluation Methodology

The platform’s pedagogical effectiveness will be evaluated through a structured pilot testing programme involving learners across all three target languages at multiple proficiency levels. The evaluation design employs a mixed-methods approach combining standardized CEFR-aligned pre- and post-tests to measure proficiency gains, the System Usability Scale adapted for educational technology (Brooke, 1996), structured learner and educator interviews to capture qualitative experience, and automated learning analytics extracted from platform interaction logs. The pilot programme is structured in three sequential cycles of approximately four months each, allowing iterative platform refinement based on empirical evidence from each cycle.

Particular attention in the evaluation will be paid to differential effects across languages and proficiency levels. The three target languages — English, Japanese, and Croatian — represent distinct typological profiles (Germanic, Japonic, Slavic), writing systems (Latin alphabet, mixed kanji-kana, Latin alphabet), and resource levels (high, medium, low), creating a natural experiment for assessing how AI tutoring effectiveness varies with linguistic and resource characteristics. This comparative dimension of the evaluation design aims

to produce insights that generalize beyond the specific languages studied to inform AI tutoring system design for diverse linguistic contexts.

6. Toward Multilingual AI Tutoring Infrastructure

Beyond the specific technical and pedagogical contributions described above, AI4LANG aims to establish a replicable model for developing AI language tutoring infrastructure that can serve languages across the resource spectrum. This section outlines the broader implications of the project’s approach for the field of AI-enhanced language education.

6.1. Addressing the Language Resource Gap

The disparity in digital language resources between high-resource and under-resourced languages represents one of the most significant barriers to equitable AI-enhanced education. While LLMs trained on web-scale data exhibit impressive capabilities in English and a handful of other languages, their performance degrades substantially for languages with limited digital presence — the very languages that could benefit most from AI tutoring to compensate for scarce human instructional resources (Joshi et al., 2020). The AI4LANG methodology — combining LLM-assisted lexical and grammatical resource creation with expert validation — offers a pathway for systematically addressing this gap without requiring the massive digital corpora that underpin LLM performance in high-resource languages.

The approach is fundamentally different from fine-tuning LLMs for specific languages, which requires substantial computational resources and curated training data. Instead, it leverages LLMs as inference engines operating over structured knowledge representations that have been curated and validated by human experts. This decoupling of linguistic knowledge from model parameters means that improvements to the knowledge graph immediately benefit the tutoring system without requiring model retraining, and that the system can be extended to new languages by creating new knowledge graphs using the established pipeline methodology.

6.2. Open Science and Reproducibility

In line with European Open Science priorities and FAIR data principles (Wilkinson et al., 2016), the project commits to making

its language resources, evaluation datasets, and core platform components publicly available under open licenses. Lexical relation datasets for Croatian and English, anonymized learner performance corpora, and the knowledge graph construction methodology will be deposited in established language resource repositories, including CLARIN-HR for South Slavic language resources. Open-source platform components will be released to facilitate adoption and adaptation by other research groups and educational institutions.

The project also contributes to the broader research infrastructure for digital language resources in the South Slavic region. Croatian, as a less-resourced EU language, currently lacks many of the digital language tools and datasets available for larger European languages. The lexical relations, grammar patterns, and CEFR descriptor mappings created through this project will enrich the CLARIN infrastructure and provide a foundation for future research and development in Croatian language technology (Tadić et al., 2021).

7. Conclusion

This paper has presented the AI4LANG initiative as a systematic effort to bridge the gap between the rapid advancement of AI language technologies and the structured requirements of formal, proficiency-oriented language education. By expanding a production-tested AI Language Tutor platform across three typologically diverse languages, the project addresses both the technical challenges of multilingual AI system design and the pedagogical imperatives of CEFR-aligned, adaptive instruction that serves languages across the resource spectrum.

Three architectural principles distinguish the platform’s approach. First, the multi-agent orchestration paradigm enables specialized optimization per pedagogical function while maintaining a coherent learner experience – a pattern that we believe will become increasingly important as AI tutoring systems grow in complexity and capability. Second, the dual-database knowledge representation layer – combining vector embeddings for semantic similarity with graph structures for structured knowledge navigation – provides a scalable foundation for language-aware exercise generation that remains grounded in linguistically validated content. Third, the decoupling of linguistic knowledge (stored in curated graphs) from model parameters (in LLMs accessed via API) creates a clean separation of

concerns that facilitates both incremental improvement and language extensibility.

The project’s emphasis on under-resourced languages – demonstrated through the Croatian language module and the lexical resource construction methodology developed for it – addresses a significant and persistent gap in the AI language learning landscape. The approach of combining LLM-assisted inference with expert validation provides a practical, scalable methodology for creating structured language resources that can be replicated for other languages facing similar resource constraints.

Looking forward, the modular, language-agnostic architecture and the open-data commitments position AI4LANG as a foundation for continued development of multilingual AI tutoring infrastructure. The methodology established for extending the platform to new languages, the evaluation framework for assessing pedagogical effectiveness across diverse linguistic contexts, and the open resources produced through the project collectively contribute to the broader goal of making high-quality, AI-enhanced language education accessible to learners regardless of the size or digital presence of their target language.

Acknowledgements

This work was carried out within the **AI4LANG** project — *Integrating Artificial Intelligence for the Digital Transformation of Language Teaching* — an Erasmus+ KA220-HED Cooperation Partnership (project number 2026-1-HR01-KA220-HED-000468605), co-funded by the European Union. Views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

The work was further supported by the **STUDIA** project (uniri-iz-25-248), co-financed by the National Recovery and Resilience Plan (NPOO) through the University of Rijeka, and by the Juraj Dobrila University of Pula institutional research project *From Dynamic Vocabulary to Virtual Interaction: Development of a Foreign Language Learning Model in Tourism Using Corpora, Artificial Intelligence, and Extended Reality* (VIRAI — Language for Real and

Virtual Worlds, project number IIP_UNIPU_010138, leader: Prof. Irena Srdanović), funded under EU NextGeneration.

Research reported here was conducted within the **FORMALS2** project, supported by the Croatian Science Foundation under grant HRZZ-IP-2024-05-3882 (PI: Tin Perković), and contributes to the project *Epistemic Democracy in the Digital Age*, supported by the Croatian Science Foundation under grant HRZZ-IP-2024-05-4113 (PI: Ivan Cerovac).

We thank the Institute for the Croatian Language and Linguistics (IHJJ) for making the *School Dictionary of the Croatian Language* available as a TEI-encoded lexical resource, the Language Resources and Technologies Society (JeRTeh) for the regional research context, and the students and teaching staff at the University of Rijeka and the Juraj Dobrila University of Pula who took part in the platform pilots.

Language Resources

The following resources are used by the platform directly or as seed data for the knowledge-graph construction pipeline described in Section 4.

- *Školski rječnik hrvatskoga jezika* (School Dictionary of the Croatian Language). Institute for the Croatian Language and Linguistics (IHJJ) and Školska knjiga, Zagreb, 2012; online edition 2020. Approximately 30,000 entries in TEI XML, with morphological, syntactic, semantic and phraseological information and explicit synonym and antonym links. Seed lexical resource for the Croatian module. <https://rjecnik.hr/>
- *Hrvatska jezična riznica* (Croatian Language Repository). IHJJ, Zagreb. Reference corpus used for lexical validation and example checking. <http://riznica.ihjj.hr>
- *Common European Framework of Reference for Languages: Learning, Teaching, Assessment — Companion Volume*. Council of Europe, 2020. Source of the Can-Do descriptor network represented in the knowledge graph. <https://www.coe.int/en/web/common-european-framework-reference-languages>
- *WordNet: An Electronic Lexical Database*. Fellbaum, C. (ed.), MIT Press, 1998. Seed resource for the English lexical layer. <https://wordnet.princeton.edu>

- *FrameNet*. Fillmore, C. J., Johnson, C. R., Petruck, M. R. L., International Computer Science Institute, Berkeley. Seed resource for the English semantic layer. <https://framenet.icsi.berkeley.edu>

References

- Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. [On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?](#) In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pages 610–623. ACM.
- John Brooke. 1996. SUS: A Quick and Dirty Usability Scale. In Patrick W. Jordan, Bruce Thomas, Bernard A. Weerdmeester, and Ian L. McClelland, editors, *Usability Evaluation in Industry*, pages 189–194. Taylor & Francis.
- Cristina Conati, Kaśka Porayska-Pomsta, and Manolis Mavrikis. 2018. [AI in Education Needs Interpretable Machine Learning: Lessons from Open Learner Modelling](#). In *Proceedings of the 2018 ICML Workshop on Human Interpretability in Machine Learning*.
- Council of Europe. 2020. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment – Companion Volume*. Council of Europe Publishing, Strasbourg.
- Nick C. Ellis. 2002. [Frequency Effects in Language Processing: A Review with Implications for Theories of Implicit and Explicit Language Acquisition](#). *Studies in Second Language Acquisition*, 24(2):143–188.
- Christiane Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, MA.
- Charles J. Fillmore, Christopher R. Johnson, and Miriam R. L. Petruck. 2003. [Background to FrameNet](#). *International Journal of Lexicography*, 16(3):235–250.
- Robert Godwin-Jones. 2022. [Emerging Technologies: Partnering with AI for Language Learning](#). *Language Learning & Technology*, 26(2):5–13.
- Otto Jespersen. 1922. *Language: Its Nature, Development, and Origin*. Allen and Unwin.

- Pratik Joshi, Sebastin Santy, Amar Budhiraja, Kalika Bali, and Monojit Choudhury. 2020. [The State and Fate of Linguistic Diversity and Inclusion in the NLP World](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6282–6293. Association for Computational Linguistics.
- Enkelejda Kasneci, Kathrin Sessler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günnemann, Eyke Hüllermeier, Stephan Krusche, Gitta Kutyniok, Tilman Michaeli, Claudia Nerdel, Jürgen Pfeffer, Oleksandra Poquet, Michael Sailer, Albrecht Schmidt, Tina Seidel, Matthias Stadler, Jochen Weller, Jochen Kuhn, and Gjergji Kasneci. 2023. [ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education](#). *Learning and Individual Differences*, 103:102274.
- Hassan Khosravi, Simon Buckingham Shum, Guanliang Chen, Cristina Conati, Yi-Shan Tsai, Judy Kay, Simon Knight, Roberto Martinez-Maldonado, Shazia Sadiq, and Dragan Gašević. 2022. [Explainable Artificial Intelligence in Education](#). *Computers and Education: Artificial Intelligence*, 3:100074.
- Benedikt Perak and Dragana Špica. 2026. An Agentic Architecture for Personalized Language Tutoring: Multi-Agent Orchestration, Teacher Amplification, and Learning Analytics. In *Proceedings of AIET 2026*, Lecture Notes on Data Engineering and Communications Technologies. Springer.
- Larry Selinker. 1972. [Interlanguage](#). *International Review of Applied Linguistics in Language Teaching*, 10(1–4):209–232.
- Burr Settles and Brendan Meeder. 2016. [A Trainable Spaced Repetition Model for Language Learning](#). In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pages 1848–1858. Association for Computational Linguistics.
- Marko Tadić, Dunja Brozović Rončević, and Amir Kapetanović. 2021. Digitalna infrastruktura za hrvatski jezik: CLARIN-HR. In Marko Tadić, editor, *Jezične tehnologije i digitalna humanistika*, pages 15–32. FF Press.
- Wim J. Van der Linden and Cees A. W. Glas, editors. 2010. [Elements of Adaptive Testing](#). Springer, New York.

- Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, et al. 2016. [The FAIR Guiding Principles for Scientific Data Management and Stewardship](#). *Scientific Data*, 3:160018.
- Qingyun Wu, Gagan Bansal, Jieyu Zhang, Yiran Wu, Beibin Li, Erkang Zhu, Li Jiang, Xiaoyun Zhang, Shaokun Zhang, Jiale Liu, Ahmed Hassan Awadallah, Ryen W. White, Doug Burger, and Chi Wang. 2023. [AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation](#).

AI4LANG – Integracija umjetne inteligencije za digitalnu transformaciju nastave jezika

Benedikt Perak, Dragana Špica

Sažetak

Brz napredak generativne umjetne inteligencije i obrade prirodnog jezika otvorio je mogućnosti za korjenito unaprjeđenje nastave jezika, no integracija tih tehnologija u pedagoški usklađena okruženja za učenje, utemeljena na Zajedničkom europskom referentnom okviru za jezike (CEFR), i dalje predstavlja značajan izazov. U ovom se radu predstavlja AI4LANG, međunarodna istraživačko-razvojna inicijativa koja proširuje platformu AI Language Tutor, spremnu za produkcijsku uporabu, na tri tipološki različita jezika – engleski, japanski i hrvatski – odgovarajući na globalnu potrebu za prilagodljivom nastavom jezika posredovanom umjetnom inteligencijom, namijenjenom i jezicima s bogatim resursima i nedovoljno zastupljenim jezicima.

Arhitektura platforme temelji se na paradigmi višeagentne orkestracije, u kojoj specijalizirani agenti umjetne inteligencije obavljaju različite pedagoške funkcije – generiranje vježbi, procjenu izgovora, pružanje povratnih informacija o gramatici i prilagodbu putanje učenja. Njihov rad koordinira središnji orkestrator koji djeluje nad slojem znanja utemeljenim na dvjema bazama podataka, kombinirajući vektorske reprezentacije sa strukturiranim leksičkim i gramatičkim grafovima.

Predstavljamo tehničku arhitekturu, pedagoški okvir usklađen sa CEFR-om na kojem se temelji prilagodljivo generiranje vježbi te metodologiju izrade jezično specifičnih resursa znanja zaključivanjem uz pomoć velikih jezičnih modela (LLM), čiji se rezultati potvrđuju stručnom provjerom. Doprinos rada jest model za razvoj višejezičnih sustava za poučavanje utemeljenih na umjetnoj inteligenciji koji se može primijeniti i u drugim sredinama, a koji u različitim jezičnim kontekstima uravnotežuje tehnološku naprednost i pedagošku utemeljenost.

Ključne riječi: Poučavanje jezika uz pomoć umjetne inteligencije, višeagentna arhitektura, CEFR, višejezična obrada prirodnog jezika, prilagodljivo učenje, grafovi znanja, prepoznavanje govora

Works are licensed under the terms and conditions of the Creative Commons Attribution 4.0 International License.

<https://creativecommons.org/licenses/by/4.0/>

