

Back Issues Online

CONTENTS

[Special Theme \(/en137/special\)](#)
[Research and Innovation \(/en137/r-i\)](#)
[Joint ERCIM Actions \(/en137/jea\)](#)
[Announcements / In Brief \(/en137/ib\)](#)

Personalised, Context-Aware XR Training Applications Driven by Large Language Models (/en137/special/personalised-context-aware-xr-training-applications-driven-by-large-language-models)

Special Theme (/En137/Special) 📅 26 April 2024 ⌚ Last Updated: 26 April 2024 👁 Hits: 4068

by George Fatouros and John Soldatos (INNOV-ACTS Limited)

The use of Extended Reality (XR) technologies for industrial workers' training provides immersive and cost-effective learning experiences that improve safety, efficiency, and productivity in the workplace. Nevertheless, most XR-based training applications are based on static and inflexible content, which limits their customisation and personalisation potential. This article introduces a novel approach to developing and deploying XR applications for training industrial workers. The approach leverages Large Language Models (LLMs) and Generative Artificial Intelligence (GenAI) to produce customisable, context-aware, and personalised content that significantly enhances the workers' training experience.

LLM and XR Integration

Industrial training applications of the Industry 5.0 era must be highly personalised and worker centric. This asks for a shift from inflexible “one-size-fits-all” applications to person-centric applications based on content tailored to the workers' preferences and context. The development of such person-centric XR requires novel approaches to the world-building, content-production and flow-control elements of XR systems. In this direction, Artificial Intelligence (AI) can be used to develop context-aware and personalised synthetic data [1], which are integrated with XR-based training applications. The latter leverage AI models like GAN (Generative Adversarial Networks) to generate multimedia content that drives the development of customised cyber representations of industrial training processes. Such approaches increase the flexibility of training applications yet provide limited room for generating content in an interactive way that considers the context, preferences, and experience of the workers.

During the past year, the advent of Large Language Models (LLMs) (e.g. OpenAI's GPT variants, Google's Gemini, Meta's open source LLaMA2) and Generative AI tools (e.g. ChatGPT, MidJourney, Copy.ai, Perplexity.AI) have revolutionised the generation of realistic context-aware content, along with the ways humans interact with AI systems towards retrieving information [2]. Furthermore, the development of domain-specific LLMs based on proprietary knowledge have demonstrated the power of LLMs in carrying out specialised tasks. Specifically, the domain-specific enhancement, training, and customisation of LLMs enables the production of relevant and accurate content for industrial tasks based on proper prompts. Upon integration with XR devices, custom LLMs could facilitate the interaction with the workers based on user-friendly chat interfaces. Nevertheless, the development of domain-specific LLMs towards interactive development of customised content is associated with several challenges: (i) The

enhancement of LLMs with knowledge missing from their training datasets through Retrieval Augmented Generation (RAG); (ii) Their fine-tuning towards domain-specific LLMs; (iii) The development of proper prompts for effective interactions via prompt engineering; and (iv) The development of efficient agents and functions for LLMs to effectively respond to such prompts.

The XR2IND project, developed in the scope of the Open Calls of the EU funded XR2Learn project (Horizon Europe Grant Agreement No: 101092851), designs and implements a naturalistic, chat-based, human-centric system for XR-based industrial training applications, notably applications that train workers in the maintenance and repair of industrial equipment. The framework interconnects XR-based industrial training applications to a customised domain-specific GenAI/LLM system, which leverages existing mainstream models (e.g., GPT-4 and LLaMA2). This LLM system is enhanced with knowledge for the industrial equipment and maintenance task at hand based on available information, including instruction manuals and Frequently Asked Questions (FAQ) about the equipment operation and troubleshooting (see Figure 1). The XR2IND system comprise the following main components:

- **Domain-Specific Knowledge Models:** Domain-specific knowledge in XR2IND is modelled based on embeddings [L1]. Specifically, the ADA model is used to generate text embeddings of documents that describe the industrial task (e.g. texts within training manuals).
- **Vector Database:** Documents with text embeddings will be persisted in a Vector Database (such as PineCone, Marqo, and Facebook faiss). The Vector Database will comprise data that are not in the base training set of the mainstream LLMs used (e.g. GPT-4). It will enable searching for instructions, how-to answers, and other user queries.
- **LLM Pipelines Engine:** A LLM Pipelining engine based on the LangChain framework [L2] will be implemented to support entire workflow from the user's question to the generation of the proper answer by the LLM and its retrieval from the Vector Database. As part of the engine, the project implements a prompt manager that handle the users' prompts, as well as various LLM agents that provide context-aware content generation strategies.
- **XR Application for Industrial Training:** The training application can be implemented in two setups:
 - VR application, based on Unity that will allow the importing, and holographic 3D visualisation of the use case environment, while allowing the trainee to interact with 3D content.
 - AR application ingesting AI generated feedback to the user based on its interactions with actual environment performing the task at hand.

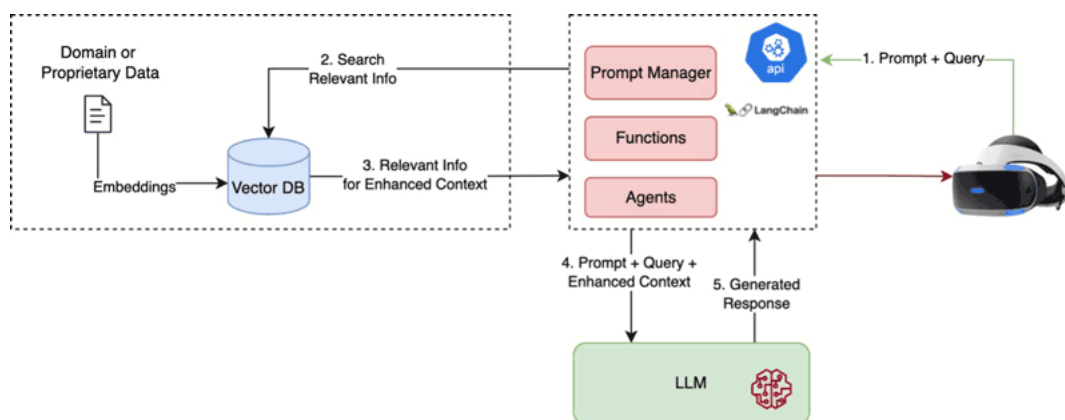


Figure 1: Generative AI enhancing XR applications.

Overall, the XR2IND system will empower the development of immersive industrial 5.0 training applications leveraging three state-of-the-art domains: (i) GenAI libraries, (ii) XR tools; and (iii) Educational knowledge and content. The project will therefore demonstrate the future personalised XR-based training based on a first of a kind integration between LLMs and XR training applications. The latter will be ergonomic, interactive, user-friendly, and very well aligned with the Industry 5.0 concept.

The work presented in this article has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101092851 (XR2Learn).

Links:

[L1] <https://platform.openai.com/docs/guides/embeddings> (<https://platform.openai.com/docs/guides/embeddings>)

[L2] https://python.langchain.com/docs/get_started/introduction (https://python.langchain.com/docs/get_started/introduction)

References:

[1] T. Hirzle, et al., "When XR and AI meet – a scoping review on extended reality and artificial intelligence," in Proc. of the CHI '23, Hamburg, Germany, ACM, 2023.

<https://doi.org/10.1145/3544548.3581072>.

[2] G. Fatouros, et al., "Transforming sentiment analysis in the financial domain with ChatGPT," Machine Learning with Applications, vol. 14, 2023, 100508, ISSN 2666-8270.

<https://doi.org/10.1016/j.mlwa.2023.100508>.

Please contact:

John Soldatos, INNOV-ACTS, Limited, Cyprus

jsoldat@innov-acts.com (<mailto:jsoldat@innov-acts.com>)

Next issue: April 2026

Special theme:

Open Science Experiences and Prospects

Call for the next issue (/call)



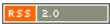
(/images/stories/EN137/EN137-web.pdf)

This issue in pdf (/images/stories/EN137/EN137-web.pdf)



This issue in ePub format (/images/stories/EN137/EN137.epub)

Get the latest issue to your desktop



(/?format=feed&type=rss)