

Glossary as a Compass: Domain Knowledge Artifacts in Human-Centered AI Development

Felipe França¹, Eduardo Oliveira¹, Leonardo Neri¹, Emilio Coutinho², Hugo Silva¹, Luciana Franci¹, and Virgínia Ribeiro¹

¹CESAR School, Recife–PE, 50030390, Brazil

²Petrobras, Rio de Janeiro–RJ, 21941915, Brazil

ABSTRACT

The development of Artificial Intelligence systems in complex environments faces a persistent challenge: translating users' natural language, rich in context, specialized terms, and cultural nuances, into formal structures that inform interface design and algorithmic logic. A promising approach to this challenge is the collaborative construction of knowledge artifacts such as technical glossaries, which serve as semantic mediation tools between multidisciplinary teams and technology. More than simple collections of definitions, these glossaries act as methodological compasses guiding human-centered AI projects and exposing gaps that drive improvement. This paper reports a case in which a collaborative glossary was developed during user immersion and adopted as a guiding artifact for AI design and engineering through participatory sessions involving domain experts, designers, software engineers, and data scientists. The resulting living document defined technical and operational terms while translating everyday practices, metaphors, and user needs; it also mapped synonyms, terminological variations, and contexts of use to align ambiguous expressions with real intentions. The glossary served as a semantic bridge linking natural language to computational representations, as a shared reference that reduced ambiguities and accelerated design decisions, and as support for integration with large language models. By centering the glossary, teams aligned expectations, improved usability, and ensured algorithms reflected business rules and field practices, demonstrating glossaries as strategic artifacts that enhance explainability, trust, and interdisciplinary collaboration in human-centered AI.

Keywords: Human-centered AI, Knowledge management, Domain ontologies

INTRODUCTION

Artificial intelligence has increasingly influenced complex technical domains, including energy, geosciences and engineering. These areas are characterised by dense bodies of knowledge, marked by detailed information and a high degree of interdependence among concepts (Kim et al., 2025). Recent advances allow AI systems not only to process and interpret technical content but also to generate new material that supports decision-making and fosters learning within specialised settings. Yet, despite these advances, working with highly specific domains introduces persistent difficulties, particularly when systems

must interpret terminology whose meaning depends heavily on context, tacit practices and professional conventions. Although the growing maturity of large language models (LLMs) has made interactions with computational systems more fluid, the challenge no longer rests solely on producing coherent text. Rather, it lies in understanding and situating terms, expressions and contextual nuances that define expertise in each field.

The complexity of applying LLMs in specialised environments is not new. Arcan (2017), for example, identified critical issues associated with semantic conversion and the removal of ambiguity in discipline-specific vocabularies. The author argues that a central difficulty in translating ontologies is interpreting each label according to the conceptual structure established by that ontology. Similar concerns are echoed by Estevez et al. (2022), who note that terminological precision is a fundamental requirement for ensuring the reliability and explainability of AI systems.

In professional and scientific settings, therefore, semantic translation extends far beyond literal word equivalence. It involves building shared understanding, harmonising terminology and establishing alignment between diverse technical viewpoints. More recently, research focusing on AI within user experience studies has pointed to an emerging intersection between design practice and the creation of meaning. As Zhang et al. (2024) observe, although AI tools are already widely adopted for tasks such as transcription and automated analysis, their contribution to collaborative interpretation and insight generation remains relatively limited. This limitation appears to stem partly from unfamiliarity with AI, and partly from the absence of mechanisms capable of effectively bridging expert vocabulary and computational reasoning. In such contexts, an AI system's interpretive capacity depends heavily on the clarity, organisation and stability of the conceptual foundations provided to it, reinforcing Arcan's (2017) observations on the need for semantic coherence when aligning expressions in multilingual ontologies.

Within this broader landscape, generative AI presents opportunities to act as an intermediary between highly specialised knowledge and the routines of professional practice. It is in this context that Chat Avaliação, a solution developed through a partnership between Petrobras and CESAR, was conceived. The system operates in a domain where precision is not simply desirable, but necessary, which made it essential to establish a shared terminological base between domain experts and the development team.

From this standpoint, the present article examines the creation and implementation of a collaborative glossary as a mechanism for alignment across multidisciplinary teams, and discusses its role in supporting the development of a generative AI system designed to analyse data and reports from formation testing environments. By doing so, the paper aims to contribute to ongoing discussions concerning linguistic resources in specialised AI design, and to highlight how design practice can serve as a bridge between technical complexity and human understanding.

THEORETICAL FRAMEWORK

Human-Centered AI

The human-centered perspective has proven to be a way to reconsider the role of AI in broader social and technical systems. Instead of limiting the discussion to the efficiency of algorithms, this view argues that technology should be guided by human values, respond to the real needs of users, and be constantly subject to oversight. Recent research, such as that by Schmager, Pappas, and Vassilakopoulou (2023), indicates that this movement arose to address the limitations of AI models that operate unclearly or without connection to everyday professional life.

Two ideas, in particular, become central: amplification and control. Amplification is using AI to augment what humans already do, not to replace them. In contrast, the concept of “Control” refers to the creation of mechanisms that allow experts to understand, verify, and adjust the functioning of systems. Reference documents, such as the *Glossary of Human-Centric Artificial Intelligence* by Estevez et al. (2022), emphasize the need for shared terminology to support trustworthy and human-centered AI discussions across stakeholders.

This continuous interaction between people and systems depends on a shared conceptual basis. According to Mark, Lyytinen, and Bergman (2007), these artifacts are defined as boundary objects, as they are capable of connecting distinct professional domains and facilitating collaborative work. In complex projects, their importance goes beyond the merely formal, since they contribute to the standardization of language, the reduction of ambiguities, and the strengthening of collective competence to understand and direct artificial intelligence systems.

Technical Glossaries as Instruments for Representing Knowledge

In highly technical areas, where concepts have complex connections and relevant practical implications, a glossary transcends the idea of a list of definitions. It represents a first step toward making knowledge visible, organized, and accessible to different people. Hilera et al. (2010) describe this as a gradual development where the glossary serves as a basis for more elaborate constructions, such as formalized conceptual structures. As new relationships between concepts are identified and refined, this resource can transform into a semantic architecture that enables logical reasoning and facilitates integration between different systems.

That is precisely why the collaborative glossary takes on the role of a strategic artifact, functioning as a semantic map that allows teams and systems to navigate the same conceptual terrain, reducing ambiguities and strengthening mutual understanding. The literature emphasizes the relevance of instruments of this nature for the construction of explainable and reliable AI systems. Hilera et al. (2010) describe glossaries as the starting point of an evolutionary process that culminates in ontologies, highlighting their role in the formalization of knowledge and interoperability between technical

domains. This approach allows knowledge to become more expressive and reusable, favoring its integration into complex computational systems.

Arcan (2017) emphasizes that the translation of technical terminology does not mean a simple linguistic equivalence, but a deep and structuring understanding of the conceptual context of each term. After all, small changes in terminology can change parameters or affect calculations, even compromising critical evaluations. In other words, the glossary plays a dual systematic role as a recording instrument and a new attempt at a game, ensuring that everyone will strive to use the same technical basis.

In this way, the technical glossary plays a dual function: it serves both as a recording tool and as a coordination mechanism. It guides dialogues, assists in decision-making, and offers a conceptual framework through which computational systems can be integrated into the domain vocabulary. In large-scale projects, their presence generally acts as a semantic stabilizer, ensuring that everyone is truly speaking the same language.

Chat Avaliação: A Support Tool for Formation Testing

Formation tests are fundamental procedures in reservoir engineering, used to evaluate how fluid and rock behave near the well (FRANCESCHINI, 2018). Based on pressure variation over time, these tests allow the estimation of essential parameters, such as permeability, transmissibility, productivity, skin, and damage factor. This information guides crucial technical decisions, such as the productive potential of the well, the need for interventions, and the adequacy of field development strategies. However, interpreting these tests is a complex task that requires combining specialized geoscientific knowledge, mastery of analytical methods, and familiarity with different sources of operational and historical data.

This complexity is intensified by the heterogeneous nature of the data involved. Data relating to formation evaluations are found in textual reports, result databases, files generated by specialized programs, graphs, and operational documents. In addition, the technical vocabulary used by specialists varies between teams, disciplines, and corporate systems. For an analyst, putting all these pieces together means navigating multiple platforms, reconciling nomenclatures, and manually consolidating indicators that influence the understanding of reservoir behavior.

In this context, the Chat Avaliação establishes a direct link with the context of formation testing by acting as an auxiliary layer for consultation and interpretation. Developed to understand the natural language and technical language specific to the sector, the system integrates information from various corporate systems, analyzes and interprets textual reports, extracting significant parameters, and allows the user to formulate questions related to wells, zones, or fields. In this way, it does not replace technical interpretation, but simplifies access to the information required for its execution, presenting answers that consolidate dispersed information and reflect the operational context of the reservoir.

Thus, the relationship between formation testing and the Chat Avaliação can be understood as an intersection between technical complexity and

computational mediation. The tests, by highlighting essential reservoir characteristics, allow the agent to perform the function of an efficient mediator, which structures, analyzes, and presents this information in a comprehensible and effective manner. For specialists and multidisciplinary teams, this mediation reduces the effort of data collection and reconciliation and strengthens the shared understanding of well and reservoir behavior.

PROJECT CONTEXT

During the immersion phase, conducted jointly by teams from CESAR and Petrobras, informational and semantic challenges were identified. A first obstacle lay in the dispersion of data sources: redundant, incomplete, or contradictory information was distributed across different systems and formats. Secondly, a semantic challenge emerged: the technical development team initially needed to acquire and internalize knowledge about the geoscientific data of high interest to the client. Additionally, it was crucial to map the different forms and variations of natural language by which experts could refer to this data. This need to align the technical jargon of the domain with the interpretation capabilities of the AI, coupled with the natural communication barrier between multidisciplinary teams (engineering, data science, and design), created the conditions for noise to occur both in the formulation of requirements and in the interpretation of the results produced by the system.

In this context, the glossary takes on the role of a strategic artifact, functioning as a semantic map that will enable the team and the system to navigate the same conceptual terrain, reducing ambiguities and strengthening mutual understanding. The literature emphasizes the relevance of instruments of this nature for the construction of explainable and reliable AI systems. Hilera et al. (2010) describe glossaries as the starting point of an evolutionary process that culminates in ontologies, highlighting their role in the formalization of knowledge and interoperability between technical domains. This approach will enable knowledge to become more expressive and reusable, favoring its integration into complex computational systems.

Based on this diagnosis, the design team recognized that the construction of a structured glossary would be an indispensable step in the project. The literature indicates that natural language is inherently ambiguous and, in the context of complex AI systems, can lead to inconsistencies in algorithmic interpretation (Kassab et al., 2014). The glossary, therefore, would not be limited to documenting terminologies, but would function as an artifact for conceptual and cognitive alignment between humans and algorithms, promoting semantic consistency throughout the development cycle. This decision finds support in the literature on knowledge representation, which understands glossaries as the first step in an evolutionary process towards formal ontologies. According to Hilera et al. (2010), “a glossary constitutes a natural starting point for the development of terminological ontologies, by offering a shared and agreed vocabulary within a specific knowledge domain.” Thus, the creation of this artifact would be, simultaneously, a participatory design strategy and a foundation for the future semantic integration of the system.

The challenge, however, was not limited to the standardization of terms. As Arcan (2017) points out in his analysis of the translation of domain-specific expressions, linguistic equivalence between experts and computer systems is not only syntactic, but contextual. The author emphasizes that, in multilingual technical environments, conceptual accuracy depends on semantic disambiguation and terminological adaptation processes capable of preserving the functional meaning of the original concepts. This same principle was observed in the Chat Avaliação, where the glossary's function went beyond the translation of words, operating as an interpretive layer that mediated differences in understanding between disciplines. In the organizational context, the project followed an agile and iterative approach, guided by the delivery of a Minimum Viable Product (MVP). This methodological choice made it possible to quickly test hypotheses with users, reduce uncertainties, and adjust the scope as new needs arise. By conducting brief validation cycles, the MVP accelerated the team's learning and prevented inappropriate investments, ensuring that the system evolves in a manner consistent with effective use and the requirements of the experts. At this stage, the glossary acted as a center of convergence: it guided discussions, supported design deliberations, and directed linguistic modeling in experiments conducted by the data science team.

The roles distributed reflected the multidisciplinary nature of the project. Reservoir engineers provided technical expertise and validated the concepts included in the glossary. The designer's role focused on understanding the reasoning modes and communication practices of users, transforming technical vocabulary into input for more intuitive conversational flows. Data scientists conducted semantic modeling and integration between heterogeneous databases. Software developers took care of the agent's infrastructure and interoperability with corporate systems. Finally, quality and testing analysts verified the consistency of the responses, both from a functional and terminological point of view.

At the end of the immersion phase, the glossary was consolidated as a common reference base, supporting decisions on data modeling, prioritization of functionalities, and validation of results. As Hilera et al. (2010) point out, the transition from glossaries to ontologies represents organizational maturation, in which the explicitness of meanings becomes a cognitive and strategic advantage. This view converges with Müller and Thoring (2011), who argue that knowledge artifacts not only document design results but also shape the cognitive and communicative processes of teams. Thus, the creation of a collaborative glossary represents a movement of externalization and formalization of collective reasoning.

Next, we describe the artifact's construction process and how it integrated into the AI system development, outlining the methodological steps that transformed a scattered set of technical terms into a unified and operational semantic base.

GLOSSARY CONSTRUCTION PROCESS

The glossary construction process followed a practical and iterative sequence, consisting of three central steps: collection, organization, and validation.

Each step sought to reduce the terminological ambiguity of the domain and generate a useful artifact for both the design team and the engineers and data scientists.

Before the initial phase, a central aspect that is often underestimated in highly specialized domains became evident: the nuances of communication between specialists and computer systems. During the interviews, it was observed that different professionals not only used distinct terminologies but also formulated questions in different ways to refer to the same indicators. This asymmetry between what is asked and what is desired as an answer reinforces the need to understand that, in real practices, the act of consulting data is also a communicative act. Research on language in human-computer interaction, such as Clark's (1996) on communicational grounding, indicates that interpretation depends on both the content and the pragmatic intention of the user. In line with this, scholars such as Furnas et al. (1987) demonstrate that different people tend to employ distinct "vocabularies of choice" when describing the same technical concept, which reinforces the importance of instruments that bring these modes of expression closer together.

The initial phase consisted of semi-structured interviews with reservoir engineers, conducted remotely via Teams. The focus was on understanding which geological data and indicators were used on a daily basis, how they were named, context of use, where they could be found, and especially possibilities of prompts to refer to and the expected responses to the relevant prompts.

After collection, the data were consolidated into a spreadsheet that became the first version of this glossary. This spreadsheet contained standardized columns for term, synonyms, definition, context of use, data type (structured or unstructured), calculation formulas when applicable, representative prompts, and source of information. This level of detail was intentional: by recording not only the definition, but also the context of use and the method of calculation, the glossary ceased to be a simple dictionary and began to operate as an operational map, guiding data extraction and validation.

The consolidated material was submitted to rounds of internal review, involving the designer, data scientist, and representatives of the reservoir engineers. These reviews aimed to validate definitions, correct inconsistencies, and map correspondences between terms and columns in the corporate databases.

RESULTS AND IMPACTS

The use of the glossary as a guiding artifact in the project demonstrated practical and strategic impacts on the management and improvement of the system. This section describes the impact of the glossary on the various areas involved and the tangible results that emerged from its collective use.

From a design perspective, the glossary served as a conceptual reference that guided terminological and content choices from the early stages of the project. Although it did not have as much of a direct effect on the visual definition of the interfaces, it provided the semantic basis that enabled the designer to understand the technical domain and ensure

consistency in the textual communication of the product. In addition, it was possible to identify patterns of prompts that frequently arose during the interviews: direct queries, queries of temporal data, correlation with other indicators, and common statistics. This segmentation contributed to a better organization of this collected data and also indicated which flows would be prioritized.

In the development of the application, the glossary was used as context in the query routing done by the large language model (LLM). For this, the glossary was used in the form of a table containing the columns Indicator, Synonyms, Definition, Database Table, Columns, Unit of Measurement. This table, in turn, was transformed into text containing the topics (1) Available Tables and (2) Indicators and related columns. The first topic focused on mapping which indicators each table and its respective columns could answer; the second topic contained the information relating to each indicator (synonyms, definition, tables, columns and unit of measurement).

The text created from the glossary was used in query routing, in which the LLM receives the glossary in the system prompt. Query routing is a decision-making mechanism in which the LLM determines the best path to obtain the best answer to the question. In the case of this application, the LLM receives the user's question and uses the content of the system prompt containing the glossary to identify which tables and columns should be used to answer the question. The LLM then returns a dictionary containing the keys 'indicator', 'dataframes', 'columns', 'unit_measure'. After routing, this dictionary is used to filter the tables with the most appropriate columns for generating the answer and to supplement the answer with the unit of measurement information.

The creation of the glossary was fundamental to structuring the quality tests. From it, it was possible to organize all relevant indicators, clarify their meanings, and correctly delimit the scope of each query. This allowed the prompts to be defined more precisely for each test scenario, ensuring that the AI was always evaluated based on the same criteria, reducing ambiguities and increasing the reliability of the results.

In addition, the glossary helped transform complex scenarios into clear, complete prompts aligned with operational needs. The prior organization of synonyms, scope variations (well, field, zone, platform), expected exceptions, applicable fluids, and formation parameters contributed to the construction of a comprehensive test matrix representative of real user interactions. This structuring enabled not only the effective validation of AI responses but also the identification of gaps and divergent behaviors, allowing for targeted interventions to improve the system. Thus, the glossary has established itself as an essential methodological tool, providing rigor, traceability, and precision to the testing cycle.

Representatives from Petrobras reported, as an observed impact, the identification of gaps and inconsistencies in the systems that fed the Chat Avaliação with corporate data. During the review of indicators in the glossary with the technical team, it was observed that some parameters requested by the engineers were not present in the available databases. This absence resulted in concrete actions, such as the implementation of new indicators in

the system that provides internal data. In this context, the glossary functioned as an instrument for eliciting requirements for internal systems.

- Among the most relevant results observed, we can highlight:
- Conceptual alignment between multidisciplinary teams, fostering a common understanding of the main indicators.
- Minimization of ambiguities in requirements documentation and technical conversations.
- More precise prioritization of critical flows and fundamental indicators for the MVP.
- Identification of data gaps, enabling improvements in the provision of data in corporate repositories.
- Optimization of verification and validation processes, serving as a common basis for data consistency tests and reviews.
- Semantic mapping and improvement of the consistency of queries and responses.

Therefore, the results of using the glossary demonstrate its role as a guiding element in the project. Beyond being a repository of terms, it also acts as a mechanism for integrating technical knowledge and product development. Its implementation exposed gaps and improvements, fostered conceptual alignment between teams, and helped build a common vision regarding the system's objectives.

CONCLUSION

The application of an artifact such as the glossary in this work demonstrates that its function goes beyond the simple standardization of terms, but is a structuring element for mediating between specialized technical knowledge and generative AI systems, operating as a cognitive infrastructure capable of supporting consistent reasoning in a highly complex domain. This role corroborates what Hilera et al. (2010) describe as the beginning of an evolutionary path towards semantic formalization, while also engaging with Arcan (2017), who emphasizes that coherence and clarity are prerequisites for any form of meaningful automatic interpretation.

The project results reveal opportunities to expand this semantic ecosystem. Transforming the glossary into formal ontologies, integrating it with automatic terminological extraction tools, creating continuous semantic validation mechanisms, and using computational representations capable of expressing causal and functional relationships are promising avenues. These developments not only promote greater accuracy for the Chat Avaliação, but also contribute to building an organizational environment in which technical knowledge becomes more explicit, structured, and interconnected.

In short, the collaborative glossary proved fundamental in connecting the complexity of formation testing with the daily operation of the generative AI agent, acting as a bridge between human expertise and the logic of systems. Its contribution goes beyond the aforementioned work, but can be seen as a possibility for a more mature semantic infrastructure, capable of

supporting future AI initiatives, integrating teams, and consolidating shared understanding patterns. Moving towards more formal structures, continuous maintenance practices, and expansion into other domains represents not only a technical opportunity, but a strategic step to strengthen the organizational capacity to handle critical knowledge at scale.

ACKNOWLEDGMENT

The authors thank Petrobras for the established partnership and the support provided throughout the project's development. The involvement of the technical teams and their willingness to contribute with information, discussions, and validations were essential to the completion of this work.

REFERENCES

- Arcan, M. (2017). Machine translation of domain-specific expressions within ontologies and terminological resources. In: Proceedings of the International Conference on Semantic Computing.
- Clark, Herbert H.; Carston, Robyn. Using language. *Journal of Linguistics*, v. 35, n. 1, pp. 167–222, Mar. 1999.
- D'Arcy, P., & Doolin, B. (2009). Knowledge artifacts, boundary spanning and the design process. *Information and Organization*, 19(2), 79–91.
- Estevez, E., Janowski, T., & Misuraca, G. (2022). Glossary of Human-Centric Artificial Intelligence. Publications Office of the European Union.
- Franceschini, Joana. *Visão Geral Do Processo De Testes De Poços De Petróleo Como Parte Da Avaliação De Formações*. 2018.
- Furnas, G. W. et al. The vocabulary problem in human-system communication. *Commun. ACM*, v. 30, n. 11, pp. 964–971, 1 Nov. 1987.
- Hilera, J. R., et al. (2010). An Evolutive Process to Convert Glossaries into Ontologies. *Journal of Software Engineering and Applications*, 3(10), 964–974.
- Kassab, M., Al-Namani, K., & Abdellatif, A. (2014). State of practice on software requirements engineering: an empirical study. *Journal of Software Engineering and Applications*, 7(10), 711–729.
- Kim, Sungil; Kim, Tea-Woo; JO, Suryeom. Artificial intelligence in geoenery: bridging petroleum engineering and future-oriented applications. *Journal of Petroleum Exploration and Production Technology*, v. 15, n. 2, p. 35, 10 Feb. 2025.
- Mark, Gloria; Lytinen, Kalle; Bergman, Mark. Boundary Objects in Design: An Ecological View of Design Artifacts. *Journal of the Association for Information Systems*, v. 8, n. 11, 1 Nov. 2007.
- Müller, R., & Thoring, K. (2011). Understanding Artifact Knowledge in Design Science. *Design Research Society Conference Proceedings*.
- Schmager, Stefan; PAPPAS, Ilias; Vassilakopoulou, Polyxeni. Defining Human-Centered AI: A Comprehensive Review of HCAI Literature. In: set. 2023.
- SLB. (2023). SLB Awarded Five Year Contract for Petrobras' Digital Transformation. Press Release, July 6.