



## Analysis of Chunk Size and Overlap in Optimizing Vector Databases for AI Agent Automation

Análisis del tamaño de fragmentos y el solapamiento en la optimización de bases de datos vectoriales para la automatización de agentes de IA

Análise do tamanho dos fragmentos e a sobreposição na otimização de bancos de dados vetoriais para automação de agentes de IA

Francisco Valdés-Souto<sup>1</sup>  

Daniela Richard-Ramírez<sup>2</sup>  

Manuel Rodríguez-Flores<sup>3</sup>  

Natalia Edith Mejía-Bautista<sup>4</sup>  

Emilio Domínguez-Valenzuela<sup>5</sup>  

**Recibido:** 19 de diciembre de 2025

**Aceptado:** 18 de febrero de 2026

**Para citar este artículo:** Valdés-Souto, F., Richard-Ramírez, D., Domínguez-Valenzuela, E., Rodríguez-Flores, M. y Mejía-Bautista, N. E. (2026). Análisis del tamaño de fragmentos y el solapamiento en la optimización de bases de datos vectoriales para la automatización de agentes de IA. *Revista Científica*, 53(1), 1-22. <https://doi.org/10.14483/23448350.24600>

### Abstract

Recent advances in language models and retrieval-augmented generation (RAG) systems have highlighted the importance of optimizing parameters such as chunk size and overlap when vectorizing unstructured data for efficient information retrieval. This study explores the impact of varying chunking configurations on the performance of a RAG system designed to answer queries regarding the Regulations of the Social Service of the Faculty of Science at UNAM. Using N8N as the integration platform and OpenAI's gpt-4o-mini as the LLM tool, different chunking strategies were evaluated based on four key metrics: correctness, semantic similarity, context relevance, and answer relevance. The results suggest that, while the overall performance differences between chunking strategies were not statistically significant, variations in chunk size and overlap can influence the consistency and quality of responses, particularly regarding answer relevance. These findings offer practical guidelines for optimizing chunking parameters in a RAG system for retrieving information from regulations documents, in order to enhance retrieval accuracy and response reliability.

**Keywords:** artificial intelligence, retrieval-augmented generation, large language models, information retrieval, document vectorization, AI agent automation

1. Universidad Nacional Autónoma de México, CDMX, México, Contacto: [fvaldes@ciencias.unam.mx](mailto:fvaldes@ciencias.unam.mx)
2. Universidad Nacional Autónoma de México, CDMX, México, Contacto: [d.richard@ciencias.unam.mx](mailto:d.richard@ciencias.unam.mx)
3. Universidad Nacional Autónoma de México, CDMX, México, Contacto: [manuel.rodr.flores@ciencias.unam.mx](mailto:manuel.rodr.flores@ciencias.unam.mx)
4. Universidad Nacional Autónoma de México, CDMX, México, Contacto: [nmejia30@ciencias.unam.mx](mailto:nmejia30@ciencias.unam.mx)
5. Universidad Nacional Autónoma de México, CDMX, México, Contacto: [emiliodominguez@ciencias.unam.mx](mailto:emiliodominguez@ciencias.unam.mx)

## Resumen

Los avances recientes en modelos de lenguaje y sistemas de generación aumentada por recuperación (RAG) han destacado la importancia de optimizar parámetros como el tamaño de los fragmentos y su solapamiento al vectorizar datos no estructurados para una recuperación de información eficiente. Este estudio explora el impacto de diferentes configuraciones de fragmentación en el rendimiento de un sistema RAG diseñado para responder consultas sobre el Reglamento del Servicio Social de la Facultad de Ciencias de la UNAM. Utilizando N8N como plataforma de integración y OPENAI gpt-4o-mini como herramienta de LLM, se evaluaron diferentes estrategias de fragmentación con base en cuatro métricas clave: corrección, similitud semántica, relevancia del contexto y relevancia de la respuesta. Los resultados sugieren que, si bien las diferencias generales de rendimiento entre las estrategias de fragmentación no fueron estadísticamente significativas, las variaciones en el tamaño de los fragmentos y su solapamiento pueden influir en la consistencia y calidad de las respuestas, particularmente en lo que respecta a la relevancia de la respuesta. Estos hallazgos ofrecen lineamientos prácticos para optimizar los parámetros de fragmentación en un sistema RAG destinado a recuperar información de documentos reglamentarios, con el fin de mejorar la precisión de la recuperación y la confiabilidad de las respuestas.

**Palabras clave:** inteligencia artificial, generación aumentada por recuperación, modelos de lenguaje a gran escala, recuperación de información, vectorización de documentos, automatización de agentes de IA

## Resumo

Avanços recentes em modelos de linguagem e sistemas de geração aumentada por recuperação (RAG) têm destacado a importância de otimizar parâmetros como o tamanho dos fragmentos e sua sobreposição ao vetorizar dados não estruturados para uma recuperação eficiente de informação. Este estudo explora o impacto de várias configurações de fragmentação no desempenho de um sistema RAG projetado para responder a consultas sobre o Regulamento do Serviço Social da Faculdade de Ciências da UNAM. Usando o N8N como plataforma de integração e o OpenAI gpt-4o-mini como ferramenta de LLM, diferentes estratégias de fragmentação foram avaliadas com base em quatro métricas-chave: correção, similaridade semântica, relevância do contexto e relevância da resposta. Os resultados sugerem que, embora as diferenças gerais de desempenho entre as estratégias de fragmentação não tenham sido estatisticamente significativas, variações no tamanho dos fragmentos e na sobreposição podem influenciar a consistência e a qualidade das respostas, particularmente no que diz respeito à relevância da resposta. Esses achados oferecem diretrizes práticas para otimizar os parâmetros de fragmentação em um sistema RAG destinado à recuperação de informações a partir de documentos regulamentares, a fim de melhorar a precisão da recuperação e a confiabilidade das respostas.

**Palavras-chaves:** inteligência artificial, geração aumentada por recuperação, modelos de linguagem de grande escala, recuperação de informação, vetorização de documentos, automação de agentes de IA

## INTRODUCTION

Recent developments in natural language processing models have led to their extensive use in daily life. These models are proficient tools for tasks such as translating texts, assisting problem-solving processes, or even as software development assistants. Such growth has led to the development of different language models, as is the case with OpenAI's GPT models ([Narasimhan et al., 2018](#)), Google's BERT models ([Chang et al., 2019](#)), and Meta's Llama models ([Touvron et al., 2023](#)), just to mention a few. While each of these are now widely used all around the world, the accuracy of their response has been shown to vary depending on different factors such as prompt quality, the task at hand, or even the lack of context to provide a satisfactory answer to the user's input ([Marcus, 2020](#); [Marcus, 2020](#)).

There are different ways of optimizing a model's response, e.g., fine-tuning it to deal with very specific tasks—as seen, for example, with Copilot, which has been fine-tuned to act as a very efficient coding assistant. However, this strategy still relies on the initial database with which the model is trained. Therefore, when the model is exposed to novel situations, its accuracy may decrease significantly ([Lewis et al., 2020](#); [Zhang et al., 2023](#)). Another relevant strategy is retrieval-augmented generation (RAG), which provides the model with relevant and correct context for the task at hand. In other words, the model is handed relevant information for the possible queries from the user. Thus, the model's response will be retrieved from the provided knowledge base, yielding more accurate and correct answers for the user.

While there are different ways of implementing a RAG system, a common approach is to store the knowledge in a vector database ([Juvekar & Purwar, 2024](#)). During the conversion process aimed at storing the information in the vector database, translating from binary into numerical data, the vectorization relies on two parameters: the vector chunk size and the vector overlap. However, variations in these parameters may influence the efficiency of answer retrieval ([Juvekar & Purwar, 2024](#)), thus leading to changes in the length, context, and relevance of the model's responses. This could entail deficiencies in the information provided to the user.

When searching the literature for information about the criteria or parametrization for vector chunk size and vector overlap, we found a lack of studies on the matter. While implementing an artificial intelligence (AI) agent with a RAG system, a case study was developed, aiming to provide some practical guidelines regarding the definition of vector chunk size and vector overlap. The main question addressed in this paper is *which chunk and overlap size are optimal for the best answer in a RAG system?* Therefore, our work aims to explore how variations in chunk size and overlap affect the efficacy of a RAG system's response, as well as how to optimize it for a vectorized database.

## Background

### Artificial intelligence

**Agents.** AI agents software components that can perform concrete tasks specified by the user. They are also capable of planning, reasoning, and demonstrating memory, and they exhibit a significant degree of autonomy in decision-making and learning processes. A relevant feature of these agents is their capability to interact with other agents and tools ([Google Cloud, 2025](#)). The tools that can be used range from specific software products such as Wolfram to crypto tools, Google Drive, and large language models (LLMs).

To perform a task, the agent needs to receive an instruction from the user. To understand said instruction, it relies on a subfield of computer science and artificial intelligence called *natural language processing*

(NLP). NLP not only helps the agent understand the user's request; it also allows it to communicate its answer to the user ([Holdsworth & Stryker, 2024](#)). The commands submitted to LLMs or agents are called *prompts*, and they are commonly defined as statements, questions, or commands ([Sheldon & Yasar, 2024](#)).

**Large Language Models.** LLMs are pre-trained language models that hold a considerable amount of knowledge in diverse topics and languages, which was learned through training on specific datasets ([Juvekar & Purwar, 2024](#); [Duarte et al., 2024](#)). These models' capacity for processing natural language has allowed for a dynamic and user-friendly interaction with users. While this is a great step forward, as already mentioned, said models are pre-trained, which leads to occasional 'hallucinations' ([Singh et al., 2024](#); [Duarte et al., 2024](#)) that exhibit low coherence or are irrelevant to the user's input. RAG is one of the solutions proposed in this regard ([Lewis et al., 2020](#); [Singh et al., 2024](#)).

When retrieving information via a pre-trained language model, computational demands can be substantial depending on the model's architecture and limitations. Models like GPT, for example, impose a maximum context window of 128 000 tokens, which roughly corresponds to 96 000 words in English ([OpenAI, 2025](#)). Processing texts that exceed this token limit remains a challenge across all major models. To address this issue, several strategies have been developed, with one of the most common being document chunking, *i.e.*, the process of breaking down strings of text into smaller ones that act as retrieval units ([Chen et al., 2024](#)). These units will be referred to as *chunks* throughout this paper.

Chunk size (measured in tokens) and the degree of overlap are hyperparameters, *i.e.*, user-defined settings that are independent of the model's internal functioning. Selecting an appropriate chunk size is critical: excessively large chunks may lead to reduced relevance to the user's query, while chunks that are too small may lack sufficient context to generate accurate and complete responses ([Juvekar a Purwar, 2024](#); [Duarte et al., 2024](#)). Furthermore, processing documents without any overlap between chunks often results in semantic discontinuities at chunk boundaries, highlighting the importance of maintaining overlap to preserve coherence ([Alfarraj et al., 2024](#)).

While chunking divides text into smaller segments to facilitate processing by LLMs, it does not inherently involve vectorized storage. Only when these chunks are encoded into embeddings and stored in a vector database does vector retrieval become part of the system's architecture.

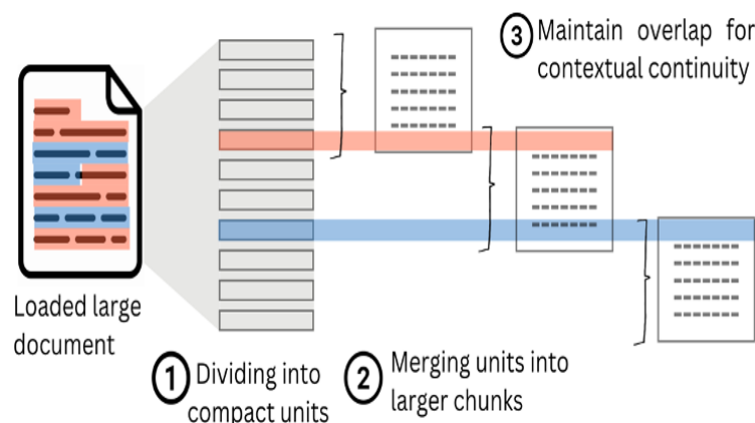


Fig 1. Overview of chunking and overlap

Source: Kimothi (2024)

**Retrieval-augmented generation.** RAG is a framework for enhancing an LLM's output by augmenting the model with access to an external knowledge base containing information beyond its original training data ([Juvekar & Purwar, 2024](#)).

This approach allows the model to reference external knowledge in order to generate more accurate and contextually relevant responses to user inputs. There are various strategies for implementing RAG systems, with one common approach involving the vectorization of relevant texts and their storage in what is typically referred to as a *vector database* ([Juvekar & Purwar, 2024](#)).

Vector databases. Vectorization can be varied out through different methods—e.g., Word2Vec ([Chen et al., 2013](#)). This results in a database that stores what is initially a string of text as a vector, not only encoding a text and storing it in a database, but also providing it with semantic meaning. These semantics allow for easier comparisons between strings, which in turn gives the model a way of traversing the database in search of relevant information to answer the user's input ([Taipalus, 2024](#)).

The methods used for comparing strings are usual vector operations that will not be described in this paper, as they fall outside its scope. Nevertheless, this short description should be insightful on the usefulness of vectorizing data to optimize an LLMs' answers through queries directed to a vector database.

**Pinecone vector databases.** These databases are used to store the vectorized data. The specific vector embedding model used herein is OpenAI's *text-embedding-3-small*, which, by vectorizing the unstructured data, generates a database with the following parameters:

- a. *Cosine metric*: this measures the cosine of the angle between two vectors in a vector space. Let  $k$  be the similarity, where  $-1 < k < 1$ .  $k = 1$  represents identical vectors,  $k = 0$  represents orthogonal vectors, and  $-1$  represents vectors that are diametrically opposed ([Chang et al., 2019](#)).
- b. *Dimensions*: 1536
- c. *Cloud*: AWS (Amazon)
- d. *BlobType*: "application/JSON" (Type of MIME)
- e. *Capacity mode*: Serverless

Pinecone databases use a table base index, listing the elements with a specific ID and indicating the values of the vector.

Table-based indices partition the vector collection into buckets, and each bucket can be retrieved by looking up a key like the rows in the hash table. They are easy to maintain, but their performance may worsen if the buckets are very large. Large buckets can improve recall but are harder to scan, and small buckets can suffer from low recall but are easier to scan. There are existing techniques aimed at improving performance at high recall ([Cui et al., 2024](#)).

## METHODOLOGY AND CASE STUDY

In this analysis, a RAG system was implemented, using N8N due to its ability to visually and flexibly automate the entire retrieval-augmented generation workflow. This platform facilitates seamless integration with language models and vector databases without the need for complex infrastructure. Additionally, N8N enabled dynamic control over the critical parameters (chunk size and overlap).

The aim was to build a RAG chatbot meant to answer questions regarding the Regulations of the Social Service of UNAM's Faculty of Science. To create the RAG system, we used OpenAI's *gpt-4o-mini*.

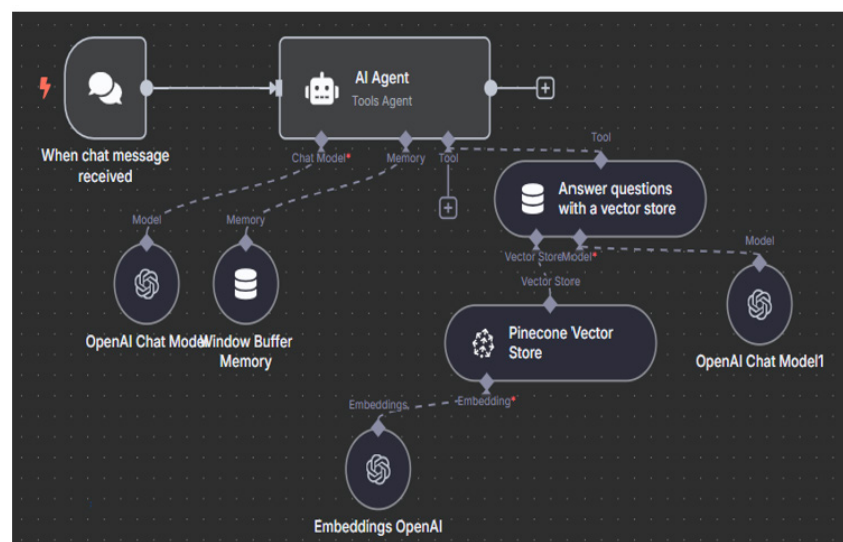
The RAG system was specifically requested to answer all questions concerning social service. If it did not have the requested information, it should redirect the user to the Faculty site, and, if the query is not related to the social service, then there should be no answer. The prompt given to the agent is presents below.

Hello, you are acting as an assistant for the Faculty of Sciences at UNAM. You are responsible for providing information about social service procedures. For this purpose, I am providing you with the Faculty of Sciences Social Service Regulations [regulations], and the frequently asked questions (FAQs) with their answers, organized by topic and subtopic. You must use a friendly tone when answering and cannot answer anything that is not in the regulations or the FAQs. If there is something you do not understand, you should answer with the response to question #51.

[regulations] = University social service, in its closest adherence to its original ideals, has a reciprocal character and focuses on community development, with a formative and professional purpose for the student.

Definition. University Social Service (SSU) is understood as the program of mandatory activities, included in the curriculum and developed temporarily, with a community and professional focus.

The Regulations of the Social Service of the Faculty of Science at UNAM correspond to a significantly large file, consisting of around 40 pages with specific information. These data were stored in the Pinecone vector store using a N8N workflow, which consists of retrieving the Social Service document, using the OpenAI vector embedding tool to convert the data from a binary into a numerical representation, and then vectorizing and uploading it into the database.



**Fig 2.** AI agent workflow created with N8N using window buffer memory, a chat model, and vector store retrieval (Pinecone) to answer questions

To study the effect of the chunk size and overlap, different configurations were used:

- *Large-size vectorization*: The data are vectorized with a 1000 chunk size and 150 overlap parameters.
- *Small-size vectorization*: The data are vectorized with a 325 chunk size and 25 overlap parameters.
- *Dual-size vectorization*: The data are vectorized with a 350 chunk size and 30 overlap parameters; then, it are re-vectorized with a 1000 chunk size and 150 overlap. Both vectorizations integrate the vector database.

The proposed methodology consists of three steps:

**Step 1.** A set of questions (prompts) are generated and selected, which consists of 36 questions in order for the sample to be statistically significant (see [Appendix A](#)). The prompts are generated based on important information that can be found on the document, and common questions are also selected.

When generating a RAG chatbot, the agent can be programmed to have certain degrees of freedom in their responses. This is done by indicating a value from 0 to 1. The former implies the lowest degree of freedom, and the latter corresponds to the highest degree of freedom. Our agent was set to 0.1, allowing it to be more cohesive and giving it the freedom to add information that might not be directly presented in the text (interpretation).

**Step 2.** The responses are evaluated with four basic metrics ([Taipalus, 2024](#)) against an ideal response or reference answer (see Annex 2) gathered from the original document. The four criteria are

- correctness
- semantic similarity
- context relevance
- answer relevance

To evaluate the four basic metrics, a number ranging from 0 to 1 is assigned, where 1 indicates that the answer is identical to the ideal response, and 0 indicates that there is no correlation between the two.

Firstly, correctness evaluates if the generated response matches the reference one, answering the following question: *Does the generated answer exactly match the answer you would get directly from the document? Evaluate literal or factual accuracy.*

Semantic similarity aims to assess whether the generated answer is similar to the reference in terms of semantic meaning. The prompt question is *Does the generated answer convey the same meaning as the answer you would get directly from the document? Even if it does not use the same words?*

Context relevance checks whether the retrieved context is related to the query. The prompt question is as follows: *Is the context retrieved by the RAG system relevant and appropriate to the answer you would get directly from the document?*

Finally, answer relevance evaluates whether the generated answer is relevant to the query, if it lacks important information, or, on the contrary, if it has irrelevant information. The prompt question reads: *Is the generated answer relevant to the question? Is it missing essential information or including irrelevant data in comparison with the ideal response?*

To evaluate the response, it is compared to a direct quote from the text, i.e., the 'controlled' variable. As the data are vectorized, we do not expect the answer to be a quote from the text, but to exhibit the previously properties listed.

**Step 3.** A repeated-measures analysis of variance (ANOVA) was applied to determine whether there were statistically significant differences between the four evaluation metrics when assessed across the same set of responses (which were based on the same question). The assumptions of the test include normality, sphericity, and independence. The given data were the p-value, the F-value, the effect size, and descriptive statistics (Algina et al., 2001).

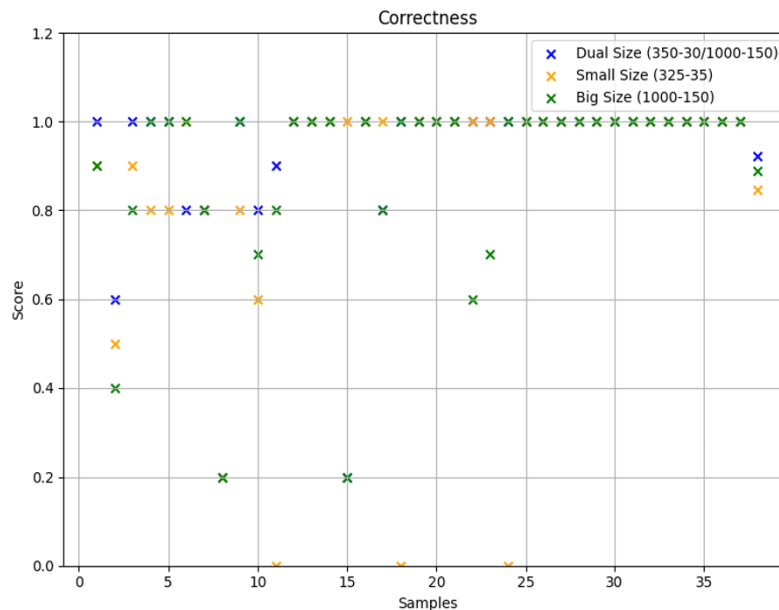
This test was conducted automatically using a Python code, expecting a p-value indicating whether the differences were statistically significant. The null hypothesis was as follows: *the median of the three conditions (differently sized chunks, small chunks, and large) is equal*. The alternative hypothesis was *at least one of the means significantly differs from the rest*. If  $p < 0.05$  the alternative hypothesis would be accepted, and, if  $p > 0.05$  the null hypothesis would be accepted.

## RESULTS AND DISCUSSION

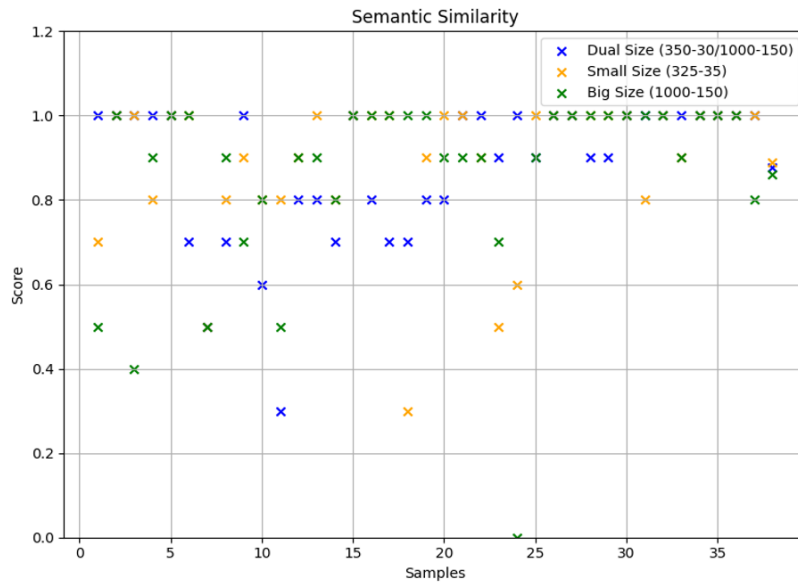
In [Appendix A, Table 1](#) summarizes the evaluation results of the different chunking configurations across the four key metrics. Each metric was rated for 36 different queries related to social service information.

To provide a clearer interpretation of the results, we present graphic representations with scatter and box plots of the metric evaluations. These visualizations aim to highlight trends, performance variations among chunking strategies, and the impact of chunk size on RAG system effectiveness (Cui et al., 2024).

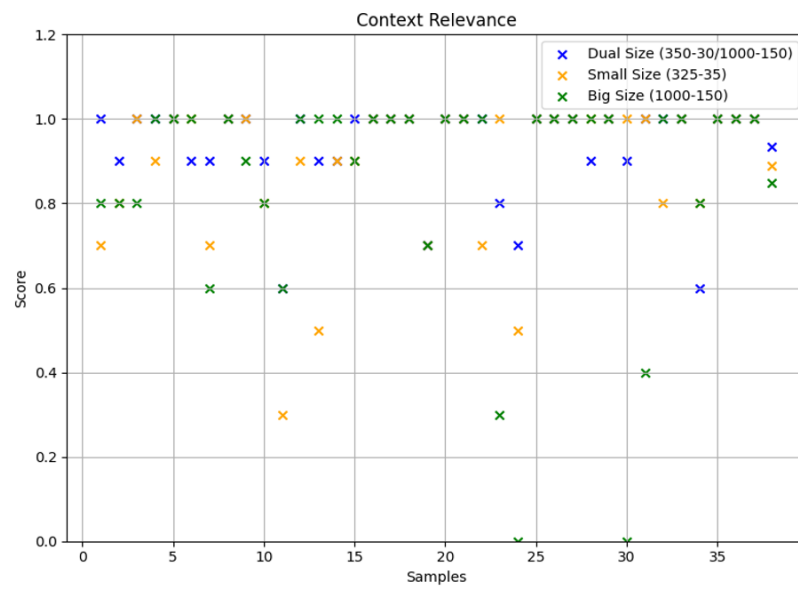
In the scatterplots, the x-axis represents the queries used, organized from the most ideal answer to the least ideal one. The y-axis corresponds to the score obtained. These plots aim to showcase the relationship between the variables.



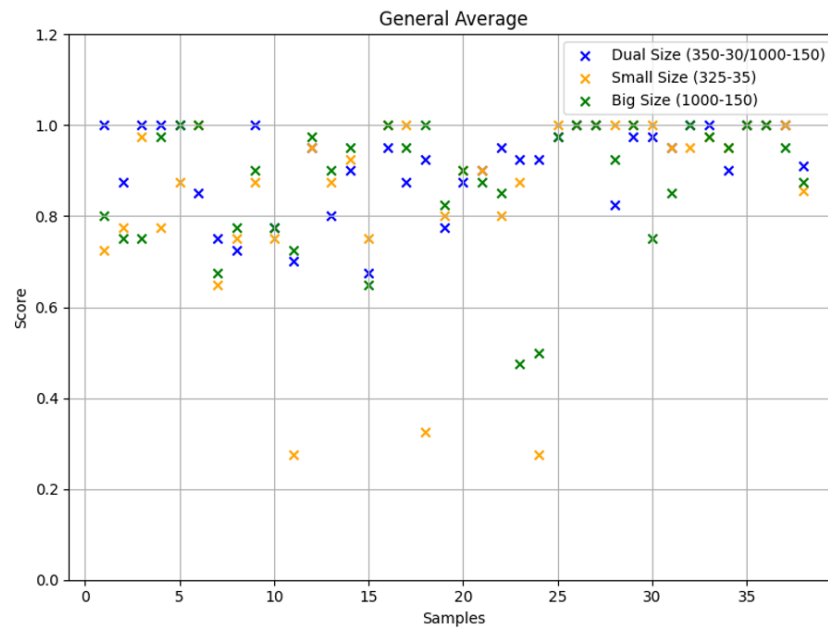
**Fig 3.** Assessment of correctness according to chunk size and overlap



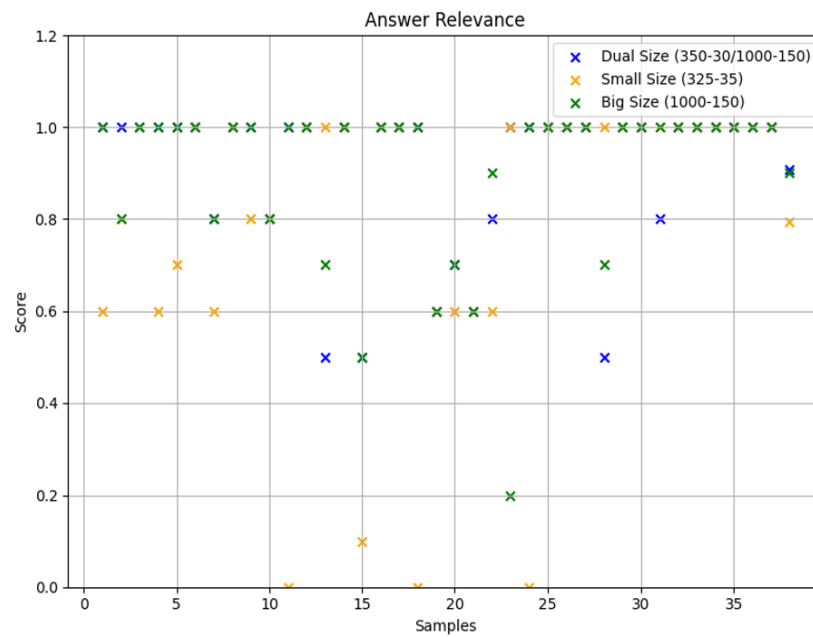
**Fig 4.** Assessment of semantic similarity according to chunk size and overlap



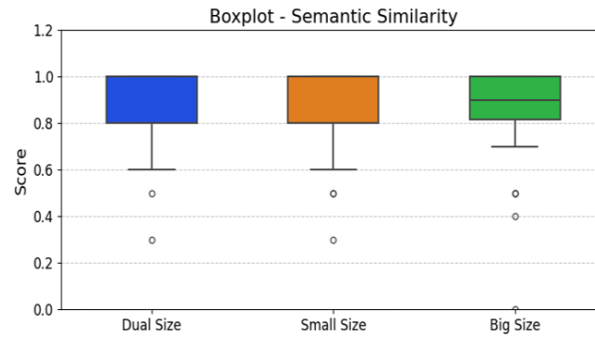
**Fig 5.** Assessment of context relevance according to chunk size and overlap



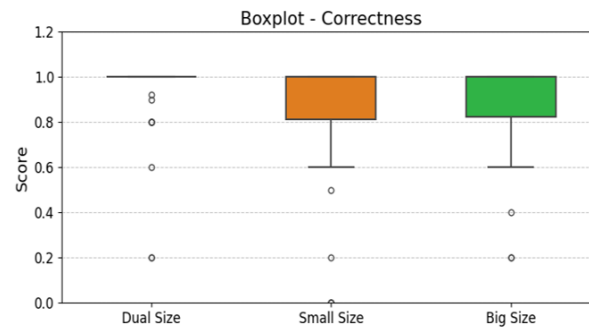
**Fig 6.** Assessment of answer relevance according to chunk size and overlap



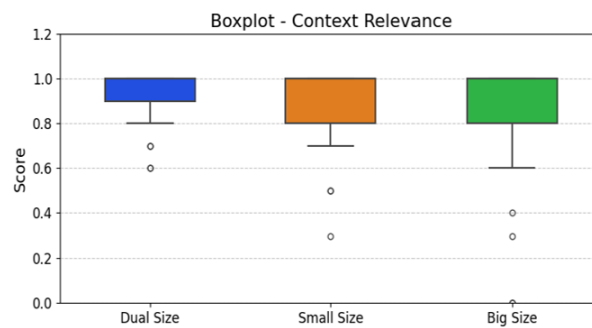
**Fig 7.** Plot comparing the average of the four evaluated metrics regarding the agent's response to chunk size and overlap



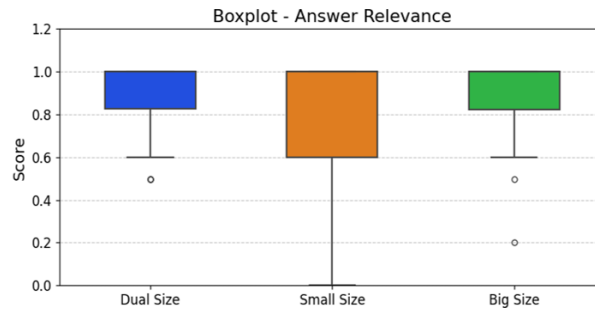
**Fig 8.** Boxplot comparing semantic across three groups with varying chunk sizes and overlaps (small, large, and mixed sizes). The boxes represent the interquartile range (IQR), the central horizontal line indicates the median and the whiskers show the distribution range



**Fig 9.** Boxplot comparing correctness across three groups with varying chunk sizes and overlaps (small, large, and mixed sizes). The boxes represent the interquartile range (IQR), the central horizontal line indicates the median and the whiskers show the distribution range



**Fig 10:** Boxplot comparing context across three groups with varying chunk sizes and overlaps (small, large, and mixed sizes). The boxes represent the interquartile range (IQR), the central horizontal line indicates the median and the whiskers show the distribution range



**Fig 11:** Boxplot comparing answer relevance across three groups with varying chunk sizes and overlaps (small, large, and mixed sizes). The boxes represent the interquartile range (IQR), the central horizontal line indicates the median and the whiskers show the distribution range

Figures 3 to 6 show that large chunk sizes consistently yield higher scores. However, some variability is observed, with some relatively low scores. Small sizes provide inconsistent and low scores, as well as high variability, especially regarding answer relevance, with some zero scores. Dual chunk sizes yield high scores with low variability—the only graph that shows low scores is semantic similarity; in the rest of the plots, the lowest values are greater than 0.5.

Figure 7 confirms the above: higher scores are obtained via dual- and large-size chunking, while small-size chunking shows low scores and high variability. Different sizes provide less variability, with most of the information being grouped.

In Figures 8 to 11, the median similarity scores are high across all conditions, reaching a value of 1. This indicates a strong alignment depending on the metric used and regardless of the chunk size. In Figure 8, the interquartile ranges (IQRs) suggest that the large chunks offer lower variability in comparison with the other configurations, representing a more consistent score similarity. The three scenarios show a few extreme cases below 0.2, which correspond to occasional semantic mismatches.

In Figure 9, small- and large-size chunking show slightly lower medians and a wider IQR in comparison with dual-size chunking. While differently sized chunks exhibit more outliers, this is due to the extremely low variability in the group's correctness score. In general, Figure 10 shows a more similar performance across groups, especially regarding the context relevance metric. Finally, in Figure 11, there is greater variability than in the previous metrics, especially for the small-size chunking. The IQR is wide, including extremely low outliers (in comparison with the other configurations), indicating an inconsistent answer quality. In contrast, dual- and large-size chunking exhibit higher and stable medians.

**Table 1:** Result of the repeated-measures ANOVA for the overall metrics

| Source    | SS       | DF | MS       | F        | p-unc    | p-GG-corr | ng2      | eps      |
|-----------|----------|----|----------|----------|----------|-----------|----------|----------|
| Condition | 0.296757 | 2  | 0.148378 | 3.654586 | 0.030784 | 0.051198  | 0.048464 | 0.654134 |
| Error     | 2.923243 | 72 | 0.040601 | NaN      | NaN      | NaN       | NaN      | NaN      |

**Table 2:** Result of the repeated-measures ANOVA for the correctness metric

| Source    | SS       | DF | MS       | F        | p-unc    | ng2      | eps     |
|-----------|----------|----|----------|----------|----------|----------|---------|
| Condition | 0.059944 | 2  | 0.029972 | 2.290502 | 0.108545 | 0.023932 | 0.88911 |
| Error     | 0.94214  | 72 | 0.013085 | NaN      | NaN      | NaN      | NaN     |

**Table 3:** Result of the repeated-measures ANOVA for the semantic similarity metric

| Source    | SS       | DF | MS       | F        | p-unc    | p-GG-corr | ng2      | eps     |
|-----------|----------|----|----------|----------|----------|-----------|----------|---------|
| Condition | 0.106667 | 2  | 0.053333 | 1.531915 | 0.223082 | 0.225772  | 0.016318 | 0.57679 |
| Error     | 2.506667 | 72 | 0.034815 | NaN      | NaN      | NaN       | NaN      | NaN     |

**Table 4:** Result of the repeated-measures ANOVA for the context relevance metric

| Source    | SS       | DF | MS       | F        | p-unc    | ng2      | eps      |
|-----------|----------|----|----------|----------|----------|----------|----------|
| Condition | 0.016757 | 2  | 0.008378 | 0.347375 | 0.707718 | 0.004292 | 0.889727 |
| Error     | 1.736577 | 72 | 0.024119 | NaN      | NaN      | NaN      | NaN      |

**Table 5:** Result of the repeated-measures ANOVA for the answer relevance metric

| Source    | SS       | DF | MS       | F        | p-unc    | p-GG-corr | ng2      | eps      |
|-----------|----------|----|----------|----------|----------|-----------|----------|----------|
| Condition | 0.138559 | 2  | 0.069279 | 2.897634 | 0.061611 | 0.082473  | 0.031992 | 0.687983 |
| Error     | 1.721441 | 72 | 0.023909 | NaN      | NaN      | NaN       | NaN      | NaN      |

A repeated-measures ANOVA was conducted to compare the effect of three vectorization parameters (varying in chunking and overlap size) across five evaluation metrics. The results are summarized in [Tables 1](#) through [5](#). The analysis revealed a statistically significant difference in the overall performance between the parameters ( $p$ ). However, no statistically significant differences were found for correctness, semantic similarity, context relevance and answer relevance, as their  $p$ -values exceeded the 0.05 significance threshold.

## CONCLUSIONS

The efficiency of RAG system's response can be compromised in vectorizing unstructured data, as it is necessary to set the chunk size and overlap. This can lead to changes in the length, context awareness, and relevance of the model's responses.

This paper presented an analysis between three vectorization configurations (based on chunking and overlap) to determine which is more efficient when working with the Regulations of the Social Service of the Faculty of Science (an UNAM document comprising around 40 pages). The results were compared using a repeated-measures ANOVA, and boxplots were used to represent how the data were grouped.

Our analysis of the vectorization parameters showed that there is no significant difference in content when the vectorization size varies, as the  $p$  parameter in the average repeated-measures ANOVA test is greater than 0.05. As clearly seen in [Figures 8 to 11](#); there is a difference in the grouping of the data, which is generally not significant, except in the case of answer relevance.

The small-size configuration shows a wide IQR, indicating higher variability within the parameter data, *i.e.*, there is a limitation regarding answer relevance. This is also shown on the scatterplots from [Figures 3 to 7](#). In general, dual- and large-size chunking yields better scores; the small-scale configuration shows lower scores and more variability in all metrics.

The ANOVA's results are shown in [Tables 1 to 5](#). They suggest that there is statistically significant deficiency in the RAG system's response which depends on the vectorization parameters. There may be a deficiency related to small chunking and overlap, arising with source documents that are similar in size to the one used herein. RAG systems are expected to perform better when the database vectorization parameters include dual chunking configurations. As for large chunking, the RAG system may also offer a good performance, but it will be exposed to more variability, which can affect the consistency of the responses.

Our case study shows that while the vectorization parameters have a statistically significant difference in the overall performance, they do not produce a statistically significant difference when metrics are evaluated individually. This suggests that the size of the data should be augmented, and that other metrics should be considered. The authors acknowledge that the data presented herein can be subjective, so, for future work, it will be necessary to look for metrics with greater objectivity.

## LIMITATIONS

While this study provides valuable insights into the effects of chunk size and overlap on RAG system performance, several limitations must be considered. Firstly, the scope was limited to a specific document comprising 40 pages, with a domain-specific focus on social service regulations, which may not generalize to broader, structured, or more heterogeneous datasets. Secondly, only three chunking configurations were analyzed, potentially limiting the exploration of other combinations that could yield different performance outcomes. Lastly, the study employed a single LLM and embedding method, meaning that the results might vary when different models or vectorization techniques are used.

## FURTHER WORK

The results obtained represent an initial step towards understanding how chunking strategies affect retrieval performance in vector databases. However, further research is necessary to deepen and broaden these findings. The next step is to generalize the conclusions drawn by using different, larger, and more diverse texts, vectorizing them as it was done in this work while testing different embedding models.

An interesting phenomenon that was observed during this case study, although not part of the initial research objectives, was the RAG system's ability to provide accurate responses to queries whose answers were not explicitly in the source document. In several cases, the model appeared to infer or synthesize information by combining semantically related segments retrieved from different parts of the text. Understanding this phenomenon could contribute to developing more explainable and cognitively aligned AI agents for new tasks.

## AUTHORSHIP CONTRIBUTIONS

**Daniela Richard Ramírez:** Conceptualization, Methodology, Formal analysis, Investigation, Writing – Original Draft, Project administration.

**Manuel Rodríguez Flores:** Formal analysis, Investigation, Writing – Original Draft.

**Natalia Edith Mejía Bautista:** Data curation, Visualization.

**Emilio Domínguez Valenzuela:** Investigation.

**Francisco Valdés Souto:** Supervision, Writing – Review & Editing.

All authors read and approved the final manuscript.

## ACKNOWLEDGEMENTS

The authors acknowledge the Facultad de Ciencias, UNAM, for its academic environment. The views and conclusions presented in this work are solely the responsibility of the authors.

## DECLARATION OF USE OF ARTIFICIAL INTELLIGENCE

The authors declare that ChatGPT was used as an Artificial Intelligence (AI) tool, exclusively during the methodology phase to assist in generating a set of question available in the Annex. No AI tool were used in the drafting, editing or data análisis of the resto f the manuscript. The authors reviewed, modified and approved the generated questions.

## REFERENCES

- Chen, T., Wang, H., Chen, S., Yu, W., Ma, K., Zhao, X., Zhang, H., & Yu, D. (2024). Dense X retrieval: What retrieval granularity should we use? *arXiv preprint*. <https://doi.org/10.48550/arXiv.2312.06648>
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint*. <https://doi.org/10.48550/arXiv.1810.04805>
- Duarte, A. V., Marques, J., Graça, M., Freire, M., Li, L., & Oliveira, A. L. (2024). LumberChunker: Long-form narrative document segmentation. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2406.17526>
- Google Cloud. (n.d.). *What are AI agents? Definition, examples, and types*. <https://cloud.google.com/discover/what-are-ai-agents>
- Juvekar, K., & Purwar, A. (2024). Introducing a new hyper-parameter for RAG: Context window utilization. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2407.19794>
- Keselman, H. J., Algina, J., & Kowalchuk, R. K. (2001). The analysis of repeated measures designs: A review. *British Journal of Mathematical and Statistical Psychology*, 54(1), 1-20. <https://doi.org/10.1348/000711001159357>
- Kimothi, A. (2024, September 23). Breaking it down: Chunking for better RAG. *Medium*. <https://medium.com/data-science/breaking-it-down-chunking-techniques-for-better-rag-3fd288bf25a0>
- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2005.11401>
- Marcus, G. (2020). The next decade in AI: The steps towards robust artificial intelligence. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2002.06177>

- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint*. <https://doi.org/10.48550/arXiv.1301.3781>
- OpenAI (n.d.). What are tokens and how to count them? *OpenAI Help Center*. <https://help.openai.com/en/articles/4936856-what-are-tokens-and-how-to-count-them>
- Pinecone (n.d.). *Vector similarity explained*. <https://www.pinecone.io/learn/vector-similarity/>
- Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training. OpenAI. <https://openai.com/research/language-unsupervised>
- Sheldon, R., & Yasar, K. (2024, November 13). What is an AI prompt? *TechTarget*. <https://www.techtarget.com/searchenterpriseai/definition/AI-prompt>
- Singh, I. S., Aggarwal, R., Allahverdiyev, I., Taha, M., Akalin, A., Zhu, K., & O'Brien, S. (2024). ChunkRAG: Novel LLM-chunk filtering method for RAG systems. *arXiv*. <https://doi.org/10.48550/arXiv.2410.19572>
- Stryker, C., & Holdsworth, J. (2024, August 11). What is NLP (natural language processing)? *IBM Think*. <https://www.ibm.com/think/topics/natural-language-processing>
- Superteams.ai (2025). *A deep dive into chunking strategy, chunking methods, and precision in RAG applications*. <https://www.superteams.ai/blog/a-deep-dive-into-chunking-strategy-chunking-methods-and-precision-in-rag-applications>
- Taipalus, T. (2024). Vector database management systems: Fundamental concepts, use-cases, and current challenges. *Cognitive Systems Research*, 85, 101216. <https://doi.org/10.1016/j.cogsys.2024.101216>
- Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E., & Lample, G. (2023). LLaMA: Open and efficient foundation language models. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2302.13971>
- Zhang, Y., Li, Y., Cui, L., Cai, D., Liu, L., Fu, T., Huang, X., Zhao, E., Zhang, Y., Chen, Y., Wang, L., Luu, A. T., Bi, W., Shi, F., & Shi, S. (2023). Siren's song in the AI ocean: A survey on hallucination in large language models. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2309.01219>
- Zhong, Z., Liu, H., Cui, X., Zhang, X., & Qin, Z. (2024). Mix-of-granularity: Optimize the chunking granularity for retrieval-augmented generation. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2406.00456>
- Zhou, J., Zhang, G., Alfarraj, O., Li, X., Tolba, A., & Zhang, H. (2024). DC-Graph: A chunk optimization model based on document classification and graph learning. *Artificial Intelligence Review*, 57, Article 143. <https://doi.org/10.1007/s10462-024-10771-w>

## APPENDIX A

### Evaluation results

The following table contains the evaluation results of the different chunking configurations across the four key metrics and their average

|    | CORRECTNESS                            |                               |                                           | SEMANTIC SIMILARITY                    |                               |                                           | CONTEXT RELEVANCE                      |                               |                                           | ANSWER RELEVANCE                       |                               |                                           | AVERAGE                                |                               |                                           |
|----|----------------------------------------|-------------------------------|-------------------------------------------|----------------------------------------|-------------------------------|-------------------------------------------|----------------------------------------|-------------------------------|-------------------------------------------|----------------------------------------|-------------------------------|-------------------------------------------|----------------------------------------|-------------------------------|-------------------------------------------|
|    | Dual Size<br>(350-<br>30/1000-<br>150) | Small<br>Size<br>(325-<br>35) | Big<br>Size<br>(1000-<br>30/1000-<br>150) | Dual Size<br>(350-<br>30/1000-<br>150) | Small<br>Size<br>(325-<br>35) | Big<br>Size<br>(1000-<br>30/1000-<br>150) | Dual Size<br>(350-<br>30/1000-<br>150) | Small<br>Size<br>(325-<br>35) | Big<br>Size<br>(1000-<br>30/1000-<br>150) | Dual Size<br>(350-<br>30/1000-<br>150) | Small<br>Size<br>(325-<br>35) | Big<br>Size<br>(1000-<br>30/1000-<br>150) | Dual Size<br>(350-<br>30/1000-<br>150) | Small<br>Size<br>(325-<br>35) | Big<br>Size<br>(1000-<br>30/1000-<br>150) |
| 1  | 1                                      | 0.9                           | 0.9                                       | 1                                      | 0.7                           | 0.5                                       | 1                                      | 0.7                           | 0.8                                       | 1                                      | 0.6                           | 1                                         | 1                                      | 0.725                         | 0.8                                       |
| 2  | 0.6                                    | 0.5                           | 0.4                                       | 1                                      | 1                             | 1                                         | 0.9                                    | 0.8                           | 0.8                                       | 1                                      | 0.8                           | 0.8                                       | 0.875                                  | 0.775                         | 0.75                                      |
| 3  | 1                                      | 0.9                           | 0.8                                       | 1                                      | 1                             | 0.4                                       | 1                                      | 1                             | 0.8                                       | 1                                      | 1                             | 1                                         | 1                                      | 0.975                         | 0.75                                      |
| 4  | 1                                      | 0.8                           | 1                                         | 1                                      | 0.8                           | 0.9                                       | 1                                      | 0.9                           | 1                                         | 1                                      | 0.6                           | 1                                         | 1                                      | 0.775                         | 0.975                                     |
| 5  | 1                                      | 0.8                           | 1                                         | 1                                      | 1                             | 1                                         | 1                                      | 1                             | 1                                         | 1                                      | 0.7                           | 1                                         | 1                                      | 0.875                         | 1                                         |
| 6  | 0.8                                    | 1                             | 1                                         | 0.7                                    | 1                             | 1                                         | 0.9                                    | 1                             | 1                                         | 1                                      | 1                             | 1                                         | 0.85                                   | 1                             | 1                                         |
| 7  | 0.8                                    | 0.8                           | 0.8                                       | 0.5                                    | 0.5                           | 0.5                                       | 0.9                                    | 0.7                           | 0.6                                       | 0.8                                    | 0.6                           | 0.8                                       | 0.75                                   | 0.65                          | 0.675                                     |
| 8  | 0.2                                    | 0.2                           | 0.2                                       | 0.7                                    | 0.8                           | 0.9                                       | 1                                      | 1                             | 1                                         | 1                                      | 1                             | 1                                         | 0.725                                  | 0.75                          | 0.775                                     |
| 9  | 1                                      | 0.8                           | 1                                         | 1                                      | 0.9                           | 0.7                                       | 1                                      | 1                             | 0.9                                       | 1                                      | 0.8                           | 1                                         | 1                                      | 0.875                         | 0.9                                       |
| 10 | 0.8                                    | 0.6                           | 0.7                                       | 0.6                                    | 0.8                           | 0.8                                       | 0.9                                    | 0.8                           | 0.8                                       | 0.8                                    | 0.8                           | 0.8                                       | 0.775                                  | 0.75                          | 0.775                                     |
| 11 | 0.9                                    | 0                             | 0.8                                       | 0.3                                    | 0.8                           | 0.5                                       | 0.6                                    | 0.3                           | 0.6                                       | 1                                      | 0                             | 1                                         | 0.7                                    | 0.275                         | 0.725                                     |
| 12 | 1                                      | 1                             | 1                                         | 0.8                                    | 0.9                           | 0.9                                       | 1                                      | 0.9                           | 1                                         | 1                                      | 1                             | 1                                         | 0.95                                   | 0.95                          | 0.975                                     |
| 13 | 1                                      | 1                             | 1                                         | 0.8                                    | 1                             | 0.9                                       | 0.9                                    | 0.5                           | 1                                         | 0.5                                    | 1                             | 0.7                                       | 0.8                                    | 0.875                         | 0.9                                       |
| 14 | 1                                      | 1                             | 1                                         | 0.7                                    | 0.8                           | 0.8                                       | 0.9                                    | 0.9                           | 1                                         | 1                                      | 1                             | 1                                         | 0.9                                    | 0.925                         | 0.95                                      |
| 15 | 0.2                                    | 1                             | 0.2                                       | 1                                      | 1                             | 1                                         | 1                                      | 0.9                           | 0.9                                       | 0.5                                    | 0.1                           | 0.5                                       | 0.675                                  | 0.75                          | 0.65                                      |
| 16 | 1                                      | 1                             | 1                                         | 0.8                                    | 1                             | 1                                         | 1                                      | 1                             | 1                                         | 1                                      | 1                             | 1                                         | 0.95                                   | 1                             | 1                                         |
| 17 | 0.8                                    | 1                             | 0.8                                       | 0.7                                    | 1                             | 1                                         | 1                                      | 1                             | 1                                         | 1                                      | 1                             | 1                                         | 0.875                                  | 1                             | 0.95                                      |
| 18 | 1                                      | 0                             | 1                                         | 0.7                                    | 0.3                           | 1                                         | 1                                      | 1                             | 1                                         | 1                                      | 0                             | 1                                         | 0.925                                  | 0.325                         | 1                                         |
| 19 | 1                                      | 1                             | 1                                         | 0.8                                    | 0.9                           | 1                                         | 0.7                                    | 0.7                           | 0.7                                       | 0.6                                    | 0.6                           | 0.6                                       | 0.775                                  | 0.8                           | 0.825                                     |
| 20 | 1                                      | 1                             | 1                                         | 0.8                                    | 1                             | 0.9                                       | 1                                      | 1                             | 1                                         | 0.7                                    | 0.6                           | 0.7                                       | 0.875                                  | 0.9                           | 0.9                                       |
| 21 | 1                                      | 1                             | 1                                         | 1                                      | 1                             | 0.9                                       | 1                                      | 1                             | 1                                         | 0.6                                    | 0.6                           | 0.6                                       | 0.9                                    | 0.9                           | 0.875                                     |
| 22 | 1                                      | 1                             | 0.6                                       | 1                                      | 0.9                           | 0.9                                       | 1                                      | 0.7                           | 1                                         | 0.8                                    | 0.6                           | 0.9                                       | 0.95                                   | 0.8                           | 0.85                                      |
| 23 | 1                                      | 1                             | 0.7                                       | 0.9                                    | 0.5                           | 0.7                                       | 0.8                                    | 1                             | 0.3                                       | 1                                      | 1                             | 0.2                                       | 0.925                                  | 0.875                         | 0.475                                     |
| 24 | 1                                      | 0                             | 1                                         | 1                                      | 0.6                           | 0                                         | 0.7                                    | 0.5                           | 0                                         | 1                                      | 0                             | 1                                         | 0.925                                  | 0.275                         | 0.5                                       |
| 25 | 1                                      | 1                             | 1                                         | 0.9                                    | 1                             | 0.9                                       | 1                                      | 1                             | 1                                         | 1                                      | 1                             | 1                                         | 0.975                                  | 1                             | 0.975                                     |
| 26 | 1                                      | 1                             | 1                                         | 1                                      | 1                             | 1                                         | 1                                      | 1                             | 1                                         | 1                                      | 1                             | 1                                         | 1                                      | 1                             | 1                                         |
| 27 | 1                                      | 1                             | 1                                         | 1                                      | 1                             | 1                                         | 1                                      | 1                             | 1                                         | 1                                      | 1                             | 1                                         | 1                                      | 1                             | 1                                         |
| 28 | 1                                      | 1                             | 1                                         | 0.9                                    | 1                             | 1                                         | 0.9                                    | 1                             | 1                                         | 0.5                                    | 1                             | 0.7                                       | 0.825                                  | 1                             | 0.925                                     |

|         |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |
|---------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 29      | 1      | 1      | 1      | 0.9    | 1      | 1      | 1      | 1      | 1      | 1      | 1      | 1      | 0.975  | 1      | 1      |
| 30      | 1      | 1      | 1      | 1      | 1      | 1      | 0.9    | 1      | 0      | 1      | 1      | 1      | 0.975  | 1      | 0.75   |
| 31      | 1      | 1      | 1      | 1      | 0.8    | 1      | 1      | 1      | 0.4    | 0.8    | 1      | 1      | 0.95   | 0.95   | 0.85   |
| 32      | 1      | 1      | 1      | 1      | 1      | 1      | 1      | 0.8    | 1      | 1      | 1      | 1      | 1      | 0.95   | 1      |
| 33      | 1      | 1      | 1      | 1      | 0.9    | 0.9    | 1      | 1      | 1      | 1      | 1      | 1      | 1      | 0.975  | 0.975  |
| 34      | 1      | 1      | 1      | 1      | 1      | 1      | 0.6    | 0.8    | 0.8    | 1      | 1      | 1      | 0.9    | 0.95   | 0.95   |
| 35      | 1      | 1      | 1      | 1      | 1      | 1      | 1      | 1      | 1      | 1      | 1      | 1      | 1      | 1      | 1      |
| 36      | 1      | 1      | 1      | 1      | 1      | 1      | 1      | 1      | 1      | 1      | 1      | 1      | 1      | 1      | 1      |
| 37      | 1      | 1      | 1      | 1      | 1      | 0.8    | 1      | 1      | 1      | 1      | 1      | 1      | 1      | 1      | 0.95   |
| AVERAGE | 0.9216 | 0.8459 | 0.8892 | 0.8784 | 0.8892 | 0.8595 | 0.9351 | 0.8892 | 0.8486 | 0.9081 | 0.7946 | 0.9000 | 0.9108 | 0.8547 | 0.8743 |

## APPENDIX B

### Queries and expected responses

The following table contains the queries used for evaluation together with their respective expected

| Queries                                       | Expected response                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|-----------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| De que formas puedo hacer el servicio social? | <p>Los alumnos de la Facultad podrán realizar su SSU colaborando con cualquiera de sus profesores, dentro de cinco grandes Programas registrados ante DGOAE:</p> <p>A.- APOYO A LA INVESTIGACIÓN: Cuando las actividades a realizar están dirigidas a apoyar proyectos que complementen el conocimiento teórico o experimental de una disciplina, que se apegan a los lineamientos de este reglamento. Ejes de acción: Ciencia, Tecnología e Investigación; Medio Ambiente y Desarrollo Sustentable; Salud; Medios- de Comunicación o Educación.</p> <p>B.- APOYO A LA DOCENCIA Y ASESORÍA ACADÉMICA: Cuando los estudiantes participen en actividades relacionadas con aspectos del proceso enseñanzaaprendizaje. Ayudantías de materias, materiales basados en TICS, educación no presencial, revisión de contenidos temáticos, etc. Eje de acción: Educación.</p> <p>C.-APOYO A LA DIVULGACIÓN: Cuando los estudiantes participen en actividades relacionadas con difusión y divulgación de la Ciencia. Arte, Cultura y Recreación o Educación.</p> <p>D.- APOYO A LOS SERVICIOS ESTUDIANTILES Y ACADÉMICOS: Actividades de carácter comunitario que apoyan los servicios de atención a los estudiantes dentro de la Facultad. Ejes de acción: Educación o Modernización y Automatización de la Gestión Administrativa.</p> <p>E.- APOYO A LOS PROGRAMAS INSTITUCIONALES Y CUERPOS COLEGIADOS: Actividades de apoyo a los programas institucionales, así como al Consejo Técnico de la Facultad. Ejes de acción: Educación; Arte, Cultura y Recreación o Ciencia, Tecnología e Innovación.</p> <p>Adicionalmente se cuenta con el siguiente programa dirigido a estudiantes de otras Facultades o escuelas que deseen realizar su servicio social en diferentes áreas de la Facultad:</p> <p>F.- EL SERVICIO SOCIAL EN LA FACULTAD DE CIENCIAS: Actividades de apoyo a la Docencia, Investigación o Extensión de estudiantes de otras Facultades o escuelas. Ejes de acción: Educación; Arte, Cultura y Recreación; Ciencia, Tecnología e Innovación, Medio Ambiente y Desarrollo Sustentable; Salud o Medios de Comunicación.El Servicio Social en la UNAM sólo puede ser realizado por nuestros alumnos que cuenten con 70% de créditos o más, en instituciones públicas y organizaciones civiles que cuenten con programas de Servicio Social debidamente registrados y aprobados en el Sistema de Información Automatizada de Servicio Social y Prácticas Profesionales (SIASSyPP) de la DGOAE. Para saber dónde podrías realizar tu Servicio Social como estudiante de la Facultad de Ciencias, puedes consultar los Programas disponibles para tu carrera ingresando al sistema SIASSyPP: <a href="https://siassypp.unam.mx/">https://siassypp.unam.mx/</a></p> <p>Puedes consultar algunas ofertas para tu carrera aquí.</p> <p>Te recomendamos revisar la información sobre el Servicio Social que está publicada en la sección Alumnos de este sitio. Tus documentos los deberás enviar a través de nuestra ventanilla virtual.</p> <p>IMPORTANTE: No es posible realizar el Servicio Social en empresas o instituciones privadas.</p> <p>A partir de que envías tus documentos a través de la Ventanilla Virtual, podrán pasar 16 días hábiles para que tengas noticias sobre tu trámite.</p> |

|                                                                                            |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|--------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Puedo colaborar con profesores?                                                            | <p>Artículo 13. Los alumnos de la Facultad podrán realizar su SSU colaborando con cualquiera de sus profesores, dentro de cinco grandes Programas registrados ante DGOAE: A.- APOYO A LA INVESTIGACIÓN: Cuando las actividades a realizar están dirigidas a apoyar proyectos que complementen el conocimiento teórico o experimental de una disciplina, que se apegan a los lineamientos de este reglamento. Ejes de acción: Ciencia, Tecnología e Investigación; Medio Ambiente y Desarrollo Sustentable; Salud; Medios de Comunicación o Educación. B.- APOYO A LA DOCENCIA Y ASESORÍA ACADÉMICA: Cuando los estudiantes participen en actividades relacionadas con aspectos del proceso enseñanza-aprendizaje. Ayudantías de materias, materiales basados en TICS, educación no presencial, revisión de contenidos temáticos, etc. Eje de acción: Educación. C.-APOYO A LA DIVULGACIÓN: Cuando los estudiantes participen en actividades relacionadas con difusión y divulgación de la Ciencia. Arte, Cultura y Recreación o Educación. D.- APOYO A LOS SERVICIOS ESTUDIANTILES Y ACADÉMICOS: Actividades de carácter comunitario que apoyan los servicios de atención a los estudiantes dentro de la Facultad. Ejes de acción: Educación o Modernización y Automatización de la Gestión Administrativa. E.- APOYO A LOS PROGRAMAS INSTITUCIONALES Y CUERPOS COLEGIADOS: Actividades de apoyo a los programas institucionales, así como al Consejo Técnico de la Facultad. Ejes de acción: Educación; Arte, Cultura y Recreación o Ciencia, Tecnología e Innovación. Adicionalmente se cuenta con el siguiente programa dirigido a estudiantes de otras Facultades o escuelas que deseen realizar su servicio social en diferentes áreas de la Facultad: F.- EL SERVICIO SOCIAL EN LA FACULTAD DE CIENCIAS: Actividades de apoyo a la Docencia, Investigación o Extensión de estudiantes de otras Facultades o escuelas. Ejes de acción: Educación; Arte, Cultura y Recreación; Ciencia, Tecnología e Innovación, Medio Ambiente y Desarrollo Sustentable; Salud o Medios de Comunicación.</p> |
| ¿En qué consiste el programa "Apoyo a la docencia y asesoría académica"?                   | <p>El programa "Apoyo a la docencia y asesoría académica" es un programa general de servicio social de la Facultad de Ciencias en el que todos sus académicos de tiempo completo pueden fungir como asesores y en el que puedes participar siendo ayudante en alguna asignatura, impartir asesorías académicas, elaborar material didáctico y textos para la docencia, colaborar en talleres y prácticas, así como apoyar en las diversas actividades de docencia que se realizan en la Facultad de Ciencias, en beneficio de estudiantes de licenciatura o de bachillerato.</p> <p>Para realizar tu Servicio Social en este programa, lo recomendable es que te acerques con algún académico de tiempo completo de la Facultad en el área de tu interés y que platiques con él para ver qué temas específicos puedes desarrollar en el programa.</p> <p>Si tu asesor es profesor de asignatura (y no cuenta con ningún otro nombramiento en alguna otra entidad de la UNAM), tu Servicio Social lo podrás realizar bajo su asesoría el semestre que el profesor esté activo en la Facultad; en este caso, tu asesor deberá contar con el Visto Bueno (Vo. Bo.) de la coordinación de la Licenciatura en donde el profesor colabore. Este Visto bueno lo solicitará la Secretaría de Asuntos Estudiantiles directamente a la coordinación que corresponda cuando realices tu trámite.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| Puedo realizar mi servicio social en un proyecto de investigación dentro de la UNAM?       | <p>Sí, el programa de servicio social "Apoyo a la investigación" es un programa general de la Facultad de Ciencias en donde, como asesores, pueden fungir académicos de tiempo completo de la Facultad, que busca apoyar en actividades de investigación que se realizan en la Facultad de Ciencias, con el propósito de complementar la formación académica de los estudiantes mediante la adquisición de nuevas habilidades y conocimientos. Para realizar tu Servicio Social en este programa, lo recomendable es que te acerques con algún académico de tiempo completo de la Facultad en el área de tu interés y que platiques con él para ver qué temas específicos puedes desarrollar en el programa. Si tu asesor es profesor de asignatura (y no cuenta con ningún otro nombramiento en alguna otra entidad de la UNAM), tu Servicio Social lo podrás realizar bajo su asesoría el semestre que el profesor esté activo en la Facultad; en este caso, tu asesor deberá contar con el Visto Bueno (Vo. Bo.) de la coordinación de la Licenciatura en donde el profesor colabore. Este Visto bueno lo solicitará la Secretaría de Asuntos Estudiantiles directamente a la coordinación que corresponda cuando realices tu trámite.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| ¿Puedo hacer mi servicio social en cualquier institución o debe estar avalada por la UNAM? | <p>Artículo 17. Los estudiantes de la Facultad de Ciencias podrán presentar su servicio social en cualquier institución u organismo público descentralizado o de servicio, fundación o asociación civil sin fines lucrativos, que contenga en sus funciones la prestación de servicio en beneficio de la comunidad, o en dependencias de servicio, investigación o académicas, en la inteligencia de que la función a desempeñar sea congruente con los ejes de acción presentados en este reglamento, y con la formación académica recibida durante sus estudios. 5</p> <p>Artículo 18. La entidad que requiera del apoyo de estudiantes de la Facultad de Ciencias para la realización del SSU en cualquiera de las modalidades, deberá contar con un Programa aprobado por la DGOAE y con un área de atención al SS; no podrá asignar a los estudiantes al área de Recursos Humanos.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| Cuales son las funciones de la coordinación?                                               | <p>Serán funciones de la Coordinación de Servicio Social.</p> <ol style="list-style-type: none"> <li>Definir los criterios generales para la realización del SSU de común acuerdo con las coordinaciones de las licenciaturas y con el Consejo Técnico.</li> <li>Atender todo asunto o trámite de los estudiantes de las diferentes licenciaturas relacionados con el SSU.</li> <li>Establecer comunicación y mantener relación de trabajo con las entidades solicitantes de prestadores de SSU.</li> <li>Difundir los programas de SSU dando a conocer las solicitudes, perfiles, incentivos y requisitos a cumplir.</li> <li>Promover en la comunidad de Ciencias, la importancia académica y social del SSU.</li> <li>Apoyar la revisión y aprobación de programas acordes con las líneas rectoras establecidas.</li> <li>Apoyar el registro y seguimiento de alumnos que realizan el SSU en los diferentes programas aprobados.</li> <li>Certificar la prestación del Servicio Social.</li> </ol>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |

|                                                                                                           |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|-----------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Hay algún tipo de seguro o cobertura médica durante la realización del servicio social?                   | Si entre las actividades a realizar se incluyen salidas al campo o estancias en localidades foráneas, el asesor es responsable de atender las indicaciones del Reglamento de salidas al Campo de la Facultad de Ciencias, y por tanto contar con los seguros médicos y de accidentes necesarios. Este reglamento puede consultarse en: <a href="http://biologia.fcencias.unam.mx/">http://biologia.fcencias.unam.mx/</a> . En este caso, se deberá requisitar el formato Aviso de Salidas en Servicio Social, el cual deberá ser entregado con la hoja de inicio de SS, sin menoscabo del trámite de los seguros correspondientes en su momento. También, por el Artículo 22 Son derechos de los estudiantes en Servicio Social: Recibir las prestaciones, apoyos acordados, seguros médicos y de accidentes, capacitación y entrenamiento estipulados en su programa de actividades. |
| Puedo obtener una constancia antes de concluir todas las horas requeridas?                                | no se menciona                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| Existe la posibilidad de exentar el servicio social en algún caso?                                        | Pregunta 26: ¿Cómo puedo liberar mi servicio social por artículo 91 del Reglamento de la Ley Reglamentaria del Artículo 5° Constitucional?<br>Para solicitar la Constancia de Liberación de Servicio Social por artículo 91 debes ser empleado federal activo con una antigüedad de 1 año y 2 meses y tener al menos el 70% de créditos totales de avance en la carrera.<br>Sobre los documentos y proceso a seguir, consulta la información aquí.<br>Pregunta 27: ¿Cómo puedo liberar mi servicio social por artículo 52 de la Ley Reglamentaria del Artículo 5° Constitucional?<br>Para solicitar la Constancia de Liberación de Servicio Social por artículo 52 de la Ley Reglamentaria del Artículo 5° Constitucional, debes ser mayor de 60 años o estar impedido por enfermedad grave.                                                                                          |
| Puedo cambiar de programa si no estoy conforme con el servicio social que elegí?                          | Pregunta 32: ¿Cómo puedo cancelar o dar de baja mi servicio social?<br>Para solicitar la cancelación de tu proceso de Servicio Social, deberás enviar los siguientes documentos en formato PDF:<br><ul style="list-style-type: none"> <li>▪ Credencial de la UNAM o INE: escaneada por ambos lados.</li> <li>▪ Carta de motivos: en formato libre, fechada y firmada por ti, con el visto bueno (Vo. Bo.) de tu asesor de Servicio Social (del programa que vas a dar de baja). En caso de no poder obtener el Vo. Bo., deberás especificar en la carta los motivos por los cuales no fue posible.</li> </ul> A través del formulario disponible en el siguiente enlace:<br><a href="https://bit.ly/SS-Cancelacion">https://bit.ly/SS-Cancelacion</a><br>a la que deberás acceder con tu cuenta de correo @ciencias.                                                                  |
| El servicio social puede realizarse en más de una institución?                                            | No mencionado explícitamente. El artículo 17 menciona lo siguiente: Los estudiantes de la Facultad de Ciencias podrán presentar su servicio social en cualquier institución u organismo público descentralizado o de servicio, fundación o asociación civil sin fines lucrativos, que contenga en sus funciones la prestación de servicio en beneficio de la comunidad, o en dependencias de servicio, investigación o académicas, en la inteligencia de que la función a desempeñar sea congruente con los ejes de acción presentados en este reglamento, y con la formación académica recibida durante sus estudios.                                                                                                                                                                                                                                                                |
| en donde puedo presentarlo?                                                                               | Los estudiantes de la Facultad de Ciencias podrán presentar su servicio social en cualquier institución u organismo público descentralizado o de servicio, fundación o asociación civil sin fines lucrativos, que contenga en sus funciones la prestación de servicio en beneficio de la comunidad, o en dependencias de servicio, investigación o académicas, en la inteligencia de que la función a desempeñar sea congruente con los ejes de acción presentados en este reglamento, y con la formación académica recibida durante sus estudios.                                                                                                                                                                                                                                                                                                                                    |
| Qué pasa si la institución donde realizo mi servicio social incumple con lo establecido en el reglamento? | La Coordinación de Servicio Social podrá suspender su relación con cualquiera de las entidades a que se refiere el Artículo 17 del presente Reglamento, cuando la entidad: a) Modifique el programa de actividades sin autorización del Departamento correspondiente y de la Coordinación de Servicio Social. b) No proporcione las condiciones adecuadas para el desarrollo del programa de actividades de Servicio Social. c) No dé a los estudiantes un trato profesional y respetuoso. d) No cumpla con lo estipulado en este Reglamento.                                                                                                                                                                                                                                                                                                                                         |
| ¿Dónde puedo consultar las opciones de programas aprobados por la UNAM?                                   | La vigencia de los programas aprobados es sólo de un año calendario. Si el programa estará vigente durante un año dado (por ejemplo 2023), la convocatoria se publica desde noviembre del año anterior para el que se registra el programa (en el ejemplo noviembre de 2023) y está abierta hasta abril del siguiente año que es el año en el que estará vigente el programa (en este ejemplo abril de 2024). ▪ Registro de nuevos programas de servicio social en el sitio web de Servicio Social de la DGOAE: consulte la información aquí.                                                                                                                                                                                                                                                                                                                                         |
| El servicio social cuenta como experiencia profesional en mi currículum?                                  | No mencionado explícitamente. El artículo 5 dice lo siguiente: El SSU tendrá un carácter esencialmente retributivo-formativo, y ofrecerá recursos para el desarrollo profesional de los alumnos y vinculación académica.<br>Clausula I. Tendrá como una de sus principales metas, contribuir a la formación profesional de los pasantes, a través de una práctica pertinente de los conocimientos y habilidades desarrolladas durante sus estudios universitarios. Esto es algo que puede ser mencionado en el currículum.                                                                                                                                                                                                                                                                                                                                                            |

|                                                                                                              |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
|--------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| ¿Cómo puedo realizar mis trámites de inicio o término de servicio social?                                    | El trámite de inicio o término de servicio social lo podrás realizar a través de la ventanilla virtual que ponemos a tu disposición en: Inicio Término a la que deberás acceder con tu cuenta de correo @ciencias. La Secretaría de Asuntos Estudiantiles canalizará tu documentación a la comisión de Servicio Social de tu carrera para que sea revisada y no pierdas tiempo en la realización de estos trámites. Te notificaremos por correo electrónico el resultado de la evaluación de tus trámites. |
| Puedo realizar el servicio social durante las vacaciones universitarias?                                     | Artículo 3. De acuerdo con el artículo 55 de la Ley Reglamentaria del Artículo 5 Constitucional no se computará el tiempo que, por vacaciones, enfermedad u otras causas justificables el estudiante se ausente de las actividades relativas a su Servicio Social. En el caso en que el programa de trabajo requiera realizar actividades durante el periodo vacacional de la UNAM, este tiempo se computará dentro de la prestación del servicio, previa justificación firmada por el tutor.              |
| ¿Debo firmar algún convenio o acuerdo antes de comenzar el servicio social?                                  | Sí, el artículo 10 establece lo siguiente: El Servicio Social de un alumno sólo podrá dar inicio, si el Plan de Trabajo ha sido revisado y autorizado por el Responsable de la licenciatura correspondiente. Por ello deberá considerarse un periodo de 10 días hábiles entre el registro del trámite y el inicio propiamente del Plan de Trabajo. Este plan de trabajo junto con el resto de la documentación deberán ser firmados por el alumno donde sea indicado.                                      |
| Cuáles son los requisitos para poder iniciar el servicio social en la UNAM?                                  | Contar con al menos el 70% de créditos totales en DGAE-SIAE, al momento de iniciar el proceso de registro del Servicio Social.<br><br>Sólo podrás realizar tu servicio social en alguna institución pública o en organismos no gubernamentales sin fines de lucro, que tengan programas registrados y aprobados para tu licenciatura. La duración mínima 6 meses de trabajo efectivo y máxima de un año calendario, debiendo cubrir un total de 480 horas.                                                 |
| Qué pasa si la institución donde realizo mi servicio social cierra o deja de operar antes de que lo termine? | El artículo 14 estipula lo siguiente: En el caso de que por alguna razón el asesor de SSU no pueda continuar asesorando a un prestador, la Coordinación de la licenciatura correspondiente asignará, en la medida de lo posible, a un nuevo asesor que tenga conocimientos sobre el tema desarrollado en el mismo. En el caso de que no haya ningún profesor que pueda sustituir al asesor del SSU, éste se cancelará.                                                                                     |
| Puedo recibir algún apoyo económico por realizar el servicio social?                                         | No mencionado explícitamente. Artículo 22 d) dice que es derecho del estudiante recibir las prestaciones, apoyos acordados, seguros médicos y de accidentes, capacitación y entrenamiento estipulados en su programa de actividades. Por lo que los apoyos acordados dependen de la institución y programa en el que se realice el servicio social.                                                                                                                                                        |
| Qué sucede si no termino el servicio social en el tiempo estipulado?                                         | En el caso de no cubrir el trámite en tiempo y forma, se deberá entregar una carta justificante firmada por el alumno con el Vo.Bo. del asesor; Cuando el Servicio Social sea invalidado, el estudiante queda obligado a reiniciar sus actividades en otro programa o en el mismo programa con otro plan de trabajo.                                                                                                                                                                                       |
| Cuánto tiempo tarda la aprobación de un programa de servicio social?                                         | El Servicio Social de un alumno sólo podrá dar inicio, si el Plan de Trabajo ha sido revisado y autorizado por el Responsable de la licenciatura correspondiente. Por ello deberá considerarse un periodo de 10 días hábiles entre el registro del trámite y el inicio propiamente del Plan de Trabajo.                                                                                                                                                                                                    |
| Puedo realizar el servicio social en modalidad presencial y a distancia?                                     | La entidad que requiera del apoyo de estudiantes de la Facultad de Ciencias para la realización del SSU en cualquiera de las modalidades, deberá contar con un Programa aprobado por la DGOAE y con un área de atención al SS; no podrá asignar a los estudiantes al área de Recursos Humanos.                                                                                                                                                                                                             |
| Puedo hacer el servicio social en una organización privada o solo en dependencias públicas?                  | No, no es posible realizar el Servicio Social en empresas o instituciones privadas. El Servicio Social en la UNAM sólo puede ser realizado por nuestros alumnos en los Programas disponibles para tu carrera listados en el sistema SIASSyPP: <a href="https://siassypp.unam.mx/">https://siassypp.unam.mx/</a>                                                                                                                                                                                            |
| Cual es la duración del servicio?                                                                            | La duración mínima del Servicio Social en la Facultad de Ciencias es de seis meses, y la duración máxima es de un año. Durante este tiempo, debes cubrir un total de 480 horas de trabajo efectivo. Esto está estipulado en el reglamento correspondiente.                                                                                                                                                                                                                                                 |
| Recibo un documento oficial al finalizar el servicio social?                                                 | Sí, el artículo 29 menciona lo siguiente: Será responsabilidad de la Coordinación de Servicio Social, certificar el cumplimiento del Servicio Social, previo Vo. Bo. del Responsable del SSU de la licenciatura correspondiente.                                                                                                                                                                                                                                                                           |
| Dónde puedo solicitar asesoría si tengo problemas o dudas con mi servicio social?                            | Revisa si tu inicio de Servicio Social ya ha sido registrado con éxito consultando en el ecosistema de tu cuenta de DGAE-SIAE (ingresando al SIAE -> Registros escolares -> Trayectoria escolar -> Ecosistema UNAM).                                                                                                                                                                                                                                                                                       |

|                                                                                                         |                                                                                                                                                                                                         |
|---------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Qué pasa si tengo una discapacidad o una condición médica que me impide cumplir con el servicio social? | Para solicitar la Constancia de Liberación de Servicio Social por artículo 52 de la Ley Reglamentaria del Artículo 5° Constitucional, debes ser mayor de 60 años o estar impedido por enfermedad grave. |
| Cuántas horas debo completar para cumplir con el servicio social?                                       | La duración mínima será de seis meses y la máxima de un año, debiendo cubrir un total de 480 horas de trabajo efectivo.                                                                                 |
| Puedo dividir las horas en diferentes periodos o deben ser continuas?                                   | La duración mínima será de seis meses y la máxima de un año, debiendo cubrir un total de 480 horas de trabajo efectivo.                                                                                 |
| Existe un tiempo límite para concluir el servicio social?                                               | La duración mínima será de seis meses y la máxima de un año, debiendo cubrir un total de 480 horas de trabajo efectivo.                                                                                 |
| El servicio social puede contar como prácticas profesionales o son cosas distintas?                     | En el reglamento no hay especificaciones al respecto ya que son cosas ajenas.                                                                                                                           |
| ¿Existen restricciones de edad o estatus académico para realizar el servicio social?                    | No se menciona                                                                                                                                                                                          |
| Hay programas de servicio social en el extranjero avalados por la UNAM?                                 | No se menciona                                                                                                                                                                                          |
| El servicio social puede ayudarme a conseguir un empleo en el futuro?                                   | No se menciona                                                                                                                                                                                          |

