

Selección de algoritmos

Se recolectaron 10 algoritmos de código abierto disponibles en Internet, dedicados a la tarea de generación de texto y respuesta a preguntas, la mayoría de estos algoritmos están subidos en repositorios de github o en la página de Hugging Face.

A cada algoritmo se le asignó un valor de relevancia entre uno (1) y cinco (5) de acuerdo a las ventajas y desventajas encontradas en cada uno. Las ventajas y desventajas en su mayoría se obtuvieron de la documentación presentada por cada autor.

Para información detallada sobre la recolección de información ver **Capítulo 3** del documento principal.

Se realizaron pruebas más detalladas a los algoritmos con mayor potencial con el fin de seleccionar el mejor algoritmo de generación de texto y el mejor algoritmo de respuesta a preguntas de acuerdo a la necesidad del proyecto.

Los algoritmos con mayor potencial son los que en el análisis de los papers se pudo notar que eran citados frecuentemente y que, en el análisis de los algoritmos son los que están actualizados relativamente hace poco (soporte), buena documentación y facilidad de uso.

Para la selección final de ambos algoritmos (generación de texto y generación de respuesta a preguntas) se tuvo en cuenta diferentes aspectos importantes como el rendimiento, los resultados de evaluación obtenidos por los autores, la documentación y las limitaciones. Finalmente para los algoritmos de generación de texto se aplicó una encuesta a un Docente de Inglés para obtener una calificación comparativa entre algoritmos.

A continuación se presentan pruebas realizadas y análisis respectivos.

Evidencia de ejecución de pruebas:  Evidencia_prueba_algoritmos.pdf

GENERACIÓN DE TEXTO

Nombre: GPT2-Pytorch with Text-Generator [1]

Enlace: <https://github.com/graykode/gpt-2-Pytorch>

Nota: En las siguientes pruebas los parámetros del algoritmo están por defecto.

Prueba 1: Generación de 3 textos con longitud de 100 caracteres

PRUEBA 1	
Entrada	“Google is a company “
Texto 1	<p>Google is a company that has a lot of things going for it. It has all sorts of cool features, and it is already getting pretty popular.</p> <p>And the one thing I am missing is a sense of satisfaction. It is hard to get my hands on a new device without feeling guilty for not buying it. I had a lot of hard drives, and I also had a lot of stuff going on for my Android device. It would be nice if I could get my hands on a device that is just</p>
Texto 2	<p>Google is a company that has been around for a long time. But it doesn't have a big social network yet.</p> <p>But now, with Facebook's new Facebook Pages , we'll see if the company can make a big impact on the way we interact with our friends, family and coworkers.</p> <p>When we first heard about Facebook Pages, we didn't expect it to be a major player, but now we know that it will certainly be a huge player.</p> <p>In the next few weeks we'll get an overview</p> <p>Las marcas más importantes del mundo están creciendo a una tasa del 20% anual. Cuando miro nuestras marcas, solo unas pocas no están creciendo a ese ritmo. Pero con este crecimiento, estamos viendo mayores</p>
Texto 3	<p>Google is a company which is a very different type of company than some other companies in the US. It is an American company which is a very different type of company from some other companies in the US. It is a company which is a very different type of company from some other companies in the US. It is a company which is a very different type of company from some other companies in the US. As I said before, it is a company which is a very different type of company</p>

Tabla 1

Prueba 2, 3 – Generación de 3 textos con longitud mediana de 100 caracteres

En esta prueba se ingresa como input una sola palabra clave con parámetros por defecto con una longitud de generación de 100 caracteres. Los resultados son bastante peores cuando es solo una palabra en vez de un inicio de oración, por esto, se ejecuta dos (2) veces con palabra clave para comprobar si en ambas pruebas genera texto carente de lógica.

PRUEBA 2	
Entrada	"Google "
Texto 1	Google was not able to find the address of my bank account. I was also unable to find the address of my partner's bank account, which was located on the back of my car. I also was unable to find my partner's bank account. After returning to my car I found that I had no money left. I called my partner's bank to see if they were able to find my deposit. I was told that my partner's account was not in
Texto 2	-----
Texto 3	We can call this file.txt in this order: 0: 0 1: 1 2: 2 3: 3 4: 4 5: 5 6: 6 7: 7

Tabla 2

PRUEBA 3	
Entrada	"Google "
Texto 1	Google "Aww, yeah, that's just... uh, I'll just take it off." "Oh, uh... no, I'm fine." "Okay, ok, but I don't want to do this any time soon. I'm not interested in anything like that. I'm just... I don't want to do anything that I don't want to do." "Well? Then why not? What are you doing here? What are you
Texto 2	Google and the new game, The Legend of Zelda: Breath of the Wild, will be available for PS4, Xbox One and iPad on March 23, 2018. "The Legend of Zelda: Breath of the Wild is an experience in and of itself that will leave you wanting more," said Nintendo of America President Satoru Iwata. "With the new games, we are bringing to Nintendo's marketplace a strong sense of adventure and adventure that will push players to play the game with more

Texto 3	<p>Google (if you think your favorite game is not on this list, please share the link on Reddit and/or Hacker News and let us know about it!)</p> <p>How to make a 3D printable version of this piece of hardware:</p> <p>Using the 3D printer, you can make the 2D printed version of this piece of hardware. You'll need a 3D printer (or a 3D printer that doesn't have any 3D printers) to print the parts and the STL</p>
----------------	--

Tabla 3

Prueba 4 – Generación de tres (3) textos con inicio de oración y longitud de 100 caracteres

PRUEBA 4	
Entrada	"Computer science is the study of algorithmic processes, "
Texto 1	<p>Computer science is the study of algorithmic processes like learning, prediction and training. The problem is to create algorithmic processes that are accurate and easy to apply in practice. One of the first algorithmic processes is to learn to predict a task, learn how it works and how to perform it well. This process is called training and is done in a set manner. The next set of processes is to learn to predict a task or learn how to do it well. This process is called training</p>
Texto 2	<p>Computer science is the study of algorithmic processes that are used to predict the outcome of a given problem. There are many different strategies that can be used as a tool to understand algorithmic processes like this. They are the following:</p> <p>Lethal algorithms are the most common. This is what I call a "preference" algorithm. An algorithm is a process that is able to learn to make decisions based on specific information. It is not simply an ability to learn to do it. It is also a tool to make decisions based</p>
Texto 3	<p>Computer science is the study of algorithmic processes that allow us to design and manipulate artificial intelligence systems, such as machine learning, machine learning, and artificial intelligence (AI).</p> <p>If you want to read more on the subject of machine learning, go here . The book is called Machine Learning from the start, and it gives a comprehensive overview of machine learning in real world usage.</p> <p>More on machine learning and AI</p> <p>Machine learning is a technique that takes the data and transforms it into an AI-powered system. This process can be</p>

Tabla 4

Prueba 5, 6 – Generación de 3 textos por prueba con longitud de 200 caracteres

PRUEBA 5	
Entrada	“Computer science is the study of algorithmic processes, ”
Texto 1	<p>Computer science is the study of algorithmic processes, and has the potential to improve human behavior. It involves understanding how the algorithms of algorithms, including the algorithms of social networks, affect human behavior.</p> <p>The current research into how algorithms affect human behavior is aimed at understanding how algorithms affect human behavior. The research is intended to be used to develop a practical, cost-effective model of human behavior. The model should take into account the natural tendency to choose the best choices to maximize social benefits in the pursuit of their own personal goals. If this is the case, then the model should provide a way for human behavior to be measured and understood as a whole. The next part of the research will explore how algorithms affect human behavior and how they affect human behavior in general.</p> <p>What to look for in the model</p> <p>As mentioned before, the model should take into account the natural tendency to choose the best choices to maximize social benefits in the pursuit of their own personal goals. The model should also explore how algorithms</p>
Texto 2	<p>Computer science is the study of algorithmic processes, using scientific data. It's an area in which I've found myself getting into quite a bit of trouble. It took me a while to realize how to write the code that led to this problem (see below). In the meantime, I've also learned that when you use a lot of code you won't end up with a problem—as you write it. This is because you don't know how to do it correctly. You've never been able to figure out how to use an algorithm well enough to make it work. You have to figure out what the right thing to do is to do to get the right result. A good example of this is the example we've seen before. The problem is that you don't know what the right algorithm is, so you're forced to guess it out loud. In fact, you don't even know what to do with it or think about it, let alone</p>
Texto 3	<p>Computer science is the study of algorithmic processes, _____. It has been suggested that the most effective mechanism in the field of Artificial Intelligence is through the use of computational algorithms. However, it is not clear whether artificial intelligence is an algorithmic process or just a way of thinking. Some researchers have argued that algorithmic processes are not just a way of thinking. For instance, cognitive scientists argue that the ability to think of something is not just a function of how well an algorithm does (e.g., an algorithm is more effective at solving a problem than a computer), but also that such reasoning can be a useful tool for learning AI.</p> <p>Some argue that human intelligence can have many properties that are different from the properties of artificial intelligence. For instance, human behavior is highly dependent on social networks and other interactions that involve different types of social interactions (e.g., social interactions with other people and with AI systems). However, it is difficult to test whether such social interactions can be used as human agents. Although human humans have a lot</p>

Tabla 5

PRUEBA 6	
Entrada	"Computer science is the study of algorithmic processes,"
Texto 1	<p>Computer science is the study of algorithmic processes vernacular terms for computer-science research. It is a discipline that has been recognized as a source of new knowledge in the field of computer science and with which many companies are seeking to develop new and exciting products.</p> <p>The research literature has been extensively reviewed by experts in computer science. The review of computer science literature has led to several key developments in research and application of computer science to computer science and related fields. As a result of these developments, the field of computer science has become a central focus of scientific research and has been the subject of a number of recent publications and articles. In particular, computer science has been cited by many researchers as the key to understanding how computers work, their interactions, and their own lives, and has been used to inform and inspire new concepts in computer science.</p> <p>In recent years, there have been more advances in the field of computer science such as the development of the first full-scale computer system (CDS) that is capable of running</p>
Texto 2	<p>Computer science is the study of algorithmic processes vernacular of scientific concepts and practices. It is a discipline that is concerned with the measurement and measurement of the performance of mathematical processes, and the application of mathematical principles. It is an area that has been studied by scientists from several disciplines, including computer science, mathematics and computer science.</p> <p>The main task of computer science is to understand the operation of mathematical processes. The fundamental principles of computer science are:</p> <p>The operation of mathematical processes is to determine the result of the operation of a mathematical process by evaluating the results of mathematical operations. The operation of mathematical processes is to understand the operation of a mathematical process by evaluating the results of mathematical operations.</p> <p>The mathematical operations of mathematical processes are to determine the result of the operation of a mathematical process by evaluating the results of mathematical operations. The mathematical operations of mathematical processes are to understand the operation of a mathematical process by evaluating the results of mathematical operations.</p> <p>The operation of mathematical processes is to understand the operation of a mathematical process by evaluating</p>
Texto 3	<p>Computer science is the study of algorithmic processes which is how computers work. This is an old story, but is still worth repeating. Computer science is a science that is designed to understand how a computer works. Computer science involves understanding how processes and systems interact, and understanding how computers work. Computer science is a field that is primarily concerned with building machines, which are the most important components of society. Computer science is taught in school, and in the workplace. Computer science requires that computers be used for tasks that are performed in a group and that are performed in a distributed manner, and that a computer be used in a machine learning environment. Computer science is also taught in the homes of government officials, and in the schools of</p>

	children. Computer science focuses on algorithms, rather than the computer's performance. Computer science teaches how to design computers using computer science's best practices and practices. Computer science is not a science
--	---

Tabla 6

Nombre: GPT2 Hugging Face [2]

Enlace: <https://huggingface.co/gpt2>

Pruebas con longitud mediana de 100 caracteres y parámetros por defecto

Prueba 1 – Generación de 3 textos con un solo input de cuatro cadenas de texto

PRUEBA 1	
Entrada	“Google is a company “
Texto 1	Google is a company who has had lots riding on the rise of search in Google Fiber to create an open, seamless, and innovative Internet that is as fast and user friendly as it can get. Their focus should be on expanding their Internet business and making Google available even to those who are not interested in how the other side of the ocean works. I've been trying to get access to Google Fiber for over 20 years and now this is my chance to share with you what's so great
Texto 2	Google is a company that makes very slim changes to people's homes and a company that creates online cards and tricks to get around an existing situation. I am aware of some things to look out for in that company and of their applications
Texto 3	Google is a company I'm still pretty comfortable with, though I do think some folks in the field have a little bit of a bit of an issue with how they're handling these new rules (the current 'legacy' rules can seem very vague and ambiguous). So the biggest takeaway is, it's not for your benefit and should only be handled in a private setting. So here's what I mean about this post: For anyone wondering if I've left off some of

Tabla 7

Prueba 2 – Generación de 3 textos con un solo input seguido por un espacio en blanco orientado a definiciones

En esta prueba se ingresa como input una sola palabra clave con parámetros por defecto con una longitud de generación de 100 caracteres. Los resultados son bastante peores cuando es solo una palabra en vez de un inicio de oración, por esto, se ejecuta tres veces con palabra clave para comprobar si en todos los test genera texto poco lógico.

PRUEBA 2	
Entrada	“Google “
Texto 1	<p>Google and https://www.reddit.com/r/The_Donald/comments/7vbtm5/is_president_donald_trump_a/ The "Trump for America" rally begins at 6:00pm at the North Dakota Historical Society at Fort McHenry in the Fort McKinley area. (Photo via Facebook) There are also at least two pro-Trump signs that line the road outside of the rally. The sign on</p>
Texto 2	<p>Google - " It's just a few little notes that explain how I'm doing this, but they really make it easy." "So that's how this was done?" *snicker* And this was the way to go. Once I had finished with my notes I started working on the design and creating a system to represent the game world and interact with the main characters. Initially it took me a few months of time trying to find a way to show</p>
Texto 3	<p>Google ----- Thanks for coming out Lenny · Joined Apr 2015 · Points: 100 Apr 15, 2015 · Mt. Sharp encoding='utf-8' ? 1 comment 2 Next All</p>

Tabla 8

Prueba 3 – Generación de 3 textos con una misma entrada en forma de oración inicial

PRUEBA 3	
Entrada	“Computer science is the study of algorithmic processes, ”
Texto 1	<p>Computer science is the study of algorithmic processes, vernacular names for any number of processes. The first step in solving any problem is to first understand the problem. It starts with some basic mathematics and then develops from there into a better understanding of the computer science and also the problem-solving process known as software engineering. One of the major things the computer science faculty in Michigan University takes on is Computer Science and Design. After that they focus on the next logical step: Programming.</p>
Texto 2	<p>Computer science is the study of algorithmic processes, the measurement and modeling of the complex worlds of human beings. I argue that, since human beings can only measure the world of us, or the things we measure, we must strive to obtain a deeper understanding of this world. As an attempt to see how we can change the world around us, I have developed a way of thinking about this and I have taken it upon myself to start thinking on the subject.</p>

	That
Texto 3	Computer science is the study of algorithmic processes, from the formation of systems to the final decision making required in every single computer science system. This analysis takes into account the fact that software is not necessarily the way it is made, rather than the way computers are made. One of the things I like about computer science is that it takes in a detailed understanding of a whole class of questions like which programming language should be used. As I say, I really like the way mathematics

Tabla 9

Pruebas con longitud de sentencia de 200

Prueba 4 – Generación de 3 textos con una misma entrada en forma de oración inicial

PRUEBA 4	
Entrada	“Computer science is the study of algorithmic processes, ”
Texto 1	<p>Computer science is the study of algorithmic processes, vernacular verbs. From ancient Roman Rome, classical mathematics was the study of the laws of physics, applied to all fields of human psychology and human behavior, since it is the research of those sciences to study the natural world. The classical philosophers and theologians were all familiar with the development of mathematical sciences over a period of centuries, as has been mentioned by Bocci, G.G.D., S.B. and others. All these economists understood that each of their mathematical sciences had an associated goal: to develop and test ideas in the natural sciences, in nature, in philosophy and psychology, or in theory. The problem of making general progress in mathematics is then solved.</p> <p>According to modern mathematical theories, natural numbers and mathematical formulas are not simply words or rules but rather represent the human and machine experience. According to mathematical models of man, they can be described as propositions which are logically true and so are not limited to an</p>
Texto 2	<p>Computer science is the study of algorithmic processes, when they turn a certain way to understand behavior. I'll mention some examples in order to show that this is true of any method.</p> <p>Here is a simple algorithm that describes the process from within a text. I know what it is, but I have trouble understanding it's nature. The problem is that it isn't understood on a computer. It isn't quite within the language of a text... But it is certainly clear-cut in terms of meaning.</p> <p>If we look beyond this to the underlying ideas I will see other examples. Here, a set of machine learning algorithms. These are called Deep Learning methods but, like some machine learning, they are not computer science-level (I've highlighted the names in the middle and bottom areas because none are fully applicable to these cases), but more for a deep learning machine learning method.</p>

Texto 3	<p>Computer science is the study of algorithmic processes, which is to say, how they were put together, and how they fit together. Some people think about it as "learning curve" (because they'll find out if they succeed). But "learning curve" is actually a general term that can encompass a variety of phenomena such as the way computers perform their computation. The general phenomenon of computer science is that computer science involves a general discussion of the fundamental mathematical principles of computing. If we want a proper definition of the term "learning curve," I will tell the story of the "learning curve," which is what most computer scientists refer to as a "learning curve," to illustrate how they relate to other fundamental physics.</p> <p>The main problem with an idea like mathematical learning curve is one major one that is often misunderstood or is misunderstood. In short, it is hard to understand mathematics. It is hard to understand the mathematical concept of what constitutes an important property of our universe. For example, it</p>
----------------	---

Tabla 10

PRUEBA 5	
Entrada	"Computer science is the study of algorithmic processes, "
Texto 1	<p>Texto 1:</p> <p>Computer science is the study of algorithmic processes, vernacular term for algorithmic processes that follow a predetermined pattern. Computational algorithms are, in their application, the process-based approach whereby one performs a task from one's own memory, which then moves further and further to the next task. In this form of learning, computers learn from experience, with specific instructions and concepts that are based on their current knowledge base and experience. (This concept is described in the book "How Systems Learn, but Never Die")</p> <p>A few examples of the different kinds of learning mechanisms are presented under the heading of "The Nature of Linguistic Learning" in the following. In this case, one learns by observing words on the Wikipedia page about the "Groups", where the group is defined in a very simple way, rather than by looking up and looking up words, like in the above analogy or the text on the Wikipedia. In all these cases, the language "Groups" actually represents</p>
Texto 2	<p>Computer science is the study of algorithmic processes, where "unlearning" is a process in which new rules are built up to make sense of existing algorithms. As soon as you learn a new algorithm, it is a good idea to put some of that knowledge inside your head. The same techniques described above are useful if you use a learning algorithm for training. Then if your learning algorithm is still too difficult or does not work well, it can be replaced by a more sophisticated learning algorithm. But, if not all of the learning algorithms you have learned in previous years are still hard to use, you are in danger of losing your own skills in this area.</p> <p>3. What's a "learn" algorithm?</p> <p>As already discussed, learning algorithms are a powerful tool for training algorithms. The difference isn't so much that the trained algorithm is working efficiently and easily, but that its algorithms are learning rather than learning.</p> <p>Learning algorithms can also help build learning algorithms for applications in a variety</p>

Texto 3	<p>Computer science is the study of algorithmic processes, including the natural history of learning and the history of information processing. The scientist has long thought that such thinking (or not, or not, not) is based on theory - which has now proved to be the case . Therefore, this new view is simply incorrect. First the incorporation of information by machines is incorrect because one could easily make use of existing ideas like computing with just machine knowledge in order to use something, such as information representation because this type of information manipulation could be applied only in some ways to represent information. However computers, as we speak, have machine knowledge knowledge . The computer may not do it by machine knowledge but it does make use of the information it receives from it</p>
----------------	---

Tabla 11

GENERACIÓN DE PREGUNTAS Y RESPUESTAS

Nombre: Question generator [3]

Enlace: https://github.com/AMontgomerie/question_generator

5 preguntas generadas a partir de un texto (.txt) generado en el algoritmo GPT2 Hugging Face

PRUEBA 1	
Entrada	<p>Computer science is the study of algorithmic processes, vernacular verbs. From ancient Roman Rome, classical mathematics was the study of the laws of physics, applied to all fields of human psychology and human behavior, since it is the research of those sciences to study the natural world. The classical philosophers and theologians were all familiar with the development of mathematical sciences over a period of centuries, as has been mentioned by Bocci, G.G.D., S.B. and others. All these economists understood that each of their mathematical sciences had an associated goal: to develop and test ideas in the natural sciences, in nature, in philosophy and psychology, or in theory. The problem of making general progress in mathematics is then solved.</p> <p>According to modern mathematical theories, natural numbers and mathematical formulas are not simply words or rules but rather represent the human and machine experience. According to mathematical models of man, they can be described as propositions which are logically true</p>
Pregunta y respuesta 1	<p>1) Q: Who were the classical philosophers and theologians? A: The classical philosophers and theologians were all familiar with the development of mathematical sciences over a period of centuries, as has been mentioned by Bocci, G.</p>
Pregunta y respuesta 2	<p>2) Q: What is the definition of computer science? A: Computer science is the study of algorithmic processes, vernacular verbs.</p>
Pregunta y respuesta 3	<p>3) Q: What is the meaning of the term "mathematical"? A: From ancient Roman Rome, classical mathematics was the study of the laws of physics, applied to all fields of human psychology and human behavior, since it is the research of those sciences to study the natural world.</p>
Pregunta y respuesta 4	<p>4) Q: What is the purpose of the study of mathematical sciences? A: All these economists understood that each of their mathematical sciences had an associated goal: to develop and test ideas in the natural sciences, in nature, in philosophy and psychology, or in theory.</p>
Pregunta y respuesta 5	<p>5) Q: What is the meaning of the word "math"? A: According to modern mathematical theories, natural numbers and mathematical formulas are not simply words or rules but rather represent the human and machine experience.</p>

Tabla 12

5 preguntas generadas a partir de un texto (.txt) escogido de Wikipedia

PRUEBA 2	
Entrada	<p>Google LLC is an American multinational technology company that specializes in Internet-related services and products, which include online advertising technologies, a search engine, cloud computing, software, and hardware. It is considered one of the Big Five technology companies in the U.S. information technology industry, alongside Amazon, Facebook, Apple, and Microsoft.</p> <p>Google was founded in September 1998 by Larry Page and Sergey Brin while they were Ph.D. students at Stanford University in California. Together they own about 14 percent of its shares and control 56 percent of the stockholder voting power through supervoting stock. They incorporated Google as a California privately held company on September 4, 1998, in California. Google was then reincorporated in Delaware on October 22, 2002.[12] An initial public offering (IPO) took place on August 19, 2004, and Google moved to its headquarters in Mountain View, California, nicknamed the Googleplex. In August 2015, Google announced plans to reorganize its various interests as a conglomerate called Alphabet Inc.</p>
Pregunta y respuesta 1	<p>1) Q: What is Google?</p> <p>A: Google LLC is an American multinational technology company that specializes in Internet-related services and products, which include online advertising technologies, a search engine, cloud computing, software, and hardware.</p>
Pregunta y respuesta 2	<p>2) Q: Where did they get their Ph.D.?</p> <p>A: students at Stanford University in California.</p>
Pregunta y respuesta 3	<p>3) Q: When was Google founded?</p> <p>A: </s> Google was founded in September 1998 by Larry Page and Sergey Brin while they were Ph.</p>
Pregunta y respuesta 4	<p>4) Q: How many months ago did Google announce plans to reorganize its various interests?</p> <p>A: 1. September 1998 2. August 2015 (correct) 3. August 19, 2004 4. September 4, 1998</p>
Pregunta y respuesta 5	<p>5) Q: What state did Google incorporate as a privately held company?</p> <p>A: 1. 56 percent 2. California (correct) 3. Delaware 4. Mountain View</p>

Tabla 13

10 preguntas generadas a partir de un texto (.txt) escogido de Wikipedia

PRUEBA 3	
Entrada	<p>Google LLC is an American multinational technology company that specializes in Internet-related services and products, which include online advertising technologies, a search engine, cloud computing, software, and hardware. It is considered one of the Big Five technology companies in the U.S. information technology industry, alongside Amazon, Facebook, Apple, and Microsoft.</p> <p>Google was founded in September 1998 by Larry Page and Sergey Brin while they were Ph.D. students at Stanford University in California. Together they own about 14 percent of its shares and control 56 percent of the stockholder voting power through supervoting stock. They incorporated Google as a California privately held company on September 4, 1998, in California. Google was then reincorporated in Delaware on October 22, 2002.[12] An initial public offering (IPO) took place on August 19, 2004, and Google moved to its headquarters in Mountain View, California, nicknamed the Googleplex. In August 2015, Google announced plans to reorganize its various interests as a conglomerate called Alphabet Inc.</p>
Pregunta y respuesta 1	<p>1) Q: What is Google?</p> <p>A: Google LLC is an American multinational technology company that specializes in Internet-related services and products, which include online advertising technologies, a search engine, cloud computing, software, and hardware.</p>
Pregunta y respuesta 2	<p>2) Q: Where did they get their Ph.D.?</p> <p>A: students at Stanford University in California.</p>
Pregunta y respuesta 3	<p>3) Q: When was Google founded?</p> <p>A: </s> Google was founded in September 1998 by Larry Page and Sergey Brin while they were Ph.</p>
Pregunta y respuesta 4	<p>4) Q: How many months ago did Google announce plans to reorganize its various interests?</p> <p>A: 1. September 1998 2. August 19, 2004 3. October 22, 2002 4. August 2015 (correct)</p>
Pregunta y respuesta 5	<p>5) Q: What state did Google incorporate as a privately held company?</p> <p>A: 1. Delaware 2. California (correct) 3. Mountain View 4. August 19, 2004</p>
Pregunta y respuesta 6	<p>6) Q: What state did Google incorporate as a privately held company?</p> <p>A: 1. Mountain View 2. California (correct) 3. American 4. Delaware</p>

Pregunta y respuesta 7	7) Q: How many percent of the stockholders vote for Google? A: Together they own about 14 percent of its shares and control 56 percent of the stockholder voting power through supervoting stock.
Pregunta y respuesta 8	8) Q: how many percent of the stockholders vote? A: 1. California 2. U. 3. 56 percent (correct) 4. about 14 percent
Pregunta y respuesta 9	9) Q: When was Google founded? A: 1. September 1998 (correct) 2. September 4, 1998 3. October 22, 2002 4. August 2015
Pregunta y respuesta 10	10) Q: What is the name of the company? A: It is considered one of the Big Five technology companies in the U.

Tabla 14

Nombre: Question Generation using transformers [4]

Enlace: https://github.com/AMontgomerie/question_generator

5 preguntas generadas a partir de un texto (.txt) escogido de Wikipedia

PRUEBA 1	
Entrada	<p>Cloud computing is the on-demand availability of computer system resources, especially data storage (cloud storage) and computing power, without direct active management by the user. The term is generally used to describe data centers available to many users over the Internet. Large clouds, predominant today, often have functions distributed over multiple locations from central servers. If the connection to the user is relatively close, it may be designated an edge server.</p> <p>Clouds may be limited to a single organization (enterprise clouds), or be available to multiple organizations (public cloud).</p> <p>Cloud computing relies on sharing of resources to achieve coherence and economies of scale.</p> <p>Advocates of public and hybrid clouds note that cloud computing allows companies to avoid or minimize up-front IT infrastructure costs. Proponents also claim that cloud computing allows enterprises to get their applications up and running faster, with improved manageability and less maintenance, and that it enables IT teams to more rapidly adjust resources to meet fluctuating and unpredictable demand, providing the</p>

	burst computing capability: high computing power at certain periods of peak demand.
Pregunta y respuesta 1	1) What is another name for data storage? A: cloud storage
Pregunta y respuesta 2	2) What is the term cloud computing used for? A: data centers available to many users over the Internet
Pregunta y respuesta 3	3) What type of cloud has functions distributed over multiple locations? A: Large clouds
Pregunta y respuesta 4	4) If the connection to the user is relatively close, what may be designated as a cloud? A: edge server
Pregunta y respuesta 5	5) What does cloud computing depend on to achieve coherence and economies of scale? A: sharing of resources
Pregunta y respuesta 6	6) What types of cloud computing allow companies to avoid or minimize up-front IT infrastructure costs? A: public and hybrid clouds
Pregunta y respuesta 7	7) What is a burst computing capability? A: high computing power

PRUEBA 2	
Entrada	<p>Google LLC is an American multinational technology company that specializes in Internet-related services and products, which include online advertising technologies, a search engine, cloud computing, software, and hardware. It is considered one of the Big Five technology companies in the U.S. information technology industry, alongside Amazon, Facebook, Apple, and Microsoft.</p> <p>Google was founded in September 1998 by Larry Page and Sergey Brin while they were Ph.D. students at Stanford University in California. Together they own about 14 percent of its shares and control 56 percent of the stockholder voting power through supervoting stock. They incorporated Google as a California privately held company on September 4, 1998, in California. Google was then reincorporated in Delaware on October 22, 2002.[12] An initial public offering (IPO) took place on August 19, 2004, and Google moved to its headquarters in Mountain View, California, nicknamed the Googleplex. In August 2015, Google announced plans to reorganize its various interests as a conglomerate called Alphabet Inc.</p>

Pregunta y respuesta 1	1) What is the name of the American multinational technology company that specializes in Internet-related services and products? A: Google LLC
Pregunta y respuesta 2	2) What is Google considered one of in the U.S. information technology industry? A: Big Five technology
Pregunta y respuesta 3	3) Who founded Google in 1998? A: Larry Page and Sergey Brin
Pregunta y respuesta 4	4) How many percent of Google's stockholder voting power does Larry Page and Sergey Brin control? A: 56
Pregunta y respuesta 5	5) What type of company did Larry Page and Sergey Brin incorporated Google as? A: California privately held

PRUEBA 3	
Entrada	Computer science is the study of algorithmic processes, and has the potential to improve human behavior. It involves understanding how the algorithms of algorithms, including the algorithms of social networks, affect human behavior. The current research into how algorithms affect human behavior is aimed at understanding how algorithms affect human behavior. The research is intended to be used to develop a practical, cost-effective model of human behavior. The model should take into account the natural tendency to choose the best choices to maximize social benefits in the pursuit of their own personal goals. If this is the case, then the model should provide a way for human behavior to be measured and understood as a whole. The next part of the research will explore how algorithms affect human behavior and how they affect human behavior in general.
Pregunta y respuesta 1	1) What is the study of algorithmic processes? A: Computer science
Pregunta y respuesta 2	2) What is the current research into how algorithms affect human behavior aimed at? A: understanding how algorithms affect human behavior
Pregunta y respuesta 3	3) What is the research intended to develop? A: practical, cost-effective model of human behavior
Pregunta y respuesta 4	4) What should the model take into account? A: natural tendency to choose the best choices to maximize social benefits in the pursuit of their own personal goals

Pregunta y respuesta 5	5) What should the model provide if the natural tendency to choose the best choices to maximize social benefits in the pursuit of their own personal goals is the case? A: the model should provide a way for human behavior to be measured and understood as a whole
-------------------------------	---

Conclusiones

Generación de texto

Para el algoritmo de *generación de texto* se seleccionó GPT-2 de Hugging Face [2] principalmente por la diferencia de rendimiento en comparación con GPT2-Pytorch. Dicha diferencia representa casi el doble de tiempo de ejecución en igualdad de condiciones (entradas). Dado que el algoritmo de GPT2-Pytorch está claramente basado en GPT-2 implica que los textos generados sean similares al modelo original y que la diferencia se defina por el rendimiento de generación de texto.

El algoritmo proporcionado por Hugging Face es oficial respecto a GPT2-Pytorch que es una modificación de GPT2 hecha por un desarrollador, de allí las diferencias en algunos aspectos como la calidad de los resultados, esto se debe a que, Hugging Face es un proveedor de código abierto reconocido en el campo de Procesamiento de Lenguaje Natural (PLN).

Rendimiento

Prueba	Entrada	Cantidad de ejemplos	Longitud de oración	Tiempo de ejecución (s)	
				GPT2 Hugging Face	GPT2 Pytorch
1	"Computer science is the study of algorithmic processes, "	20	200	53,969	103,718
2	"Computer science is the study of algorithmic processes, "	20	100	23,298	52,566
3	"Google is a company "	20	200	54,516	97,795
4	"Google is a company "	20	100	24,455	51,413
5	"Wikipedia is a free, "	20	200	53,723	98,302
6	"Wikipedia is a free, "	20	100	23,866	52,140
Promedio				38,971	75,989

Tabla 15. Comparativa de rendimiento. Pruebas realizadas en Google Colab con GPU

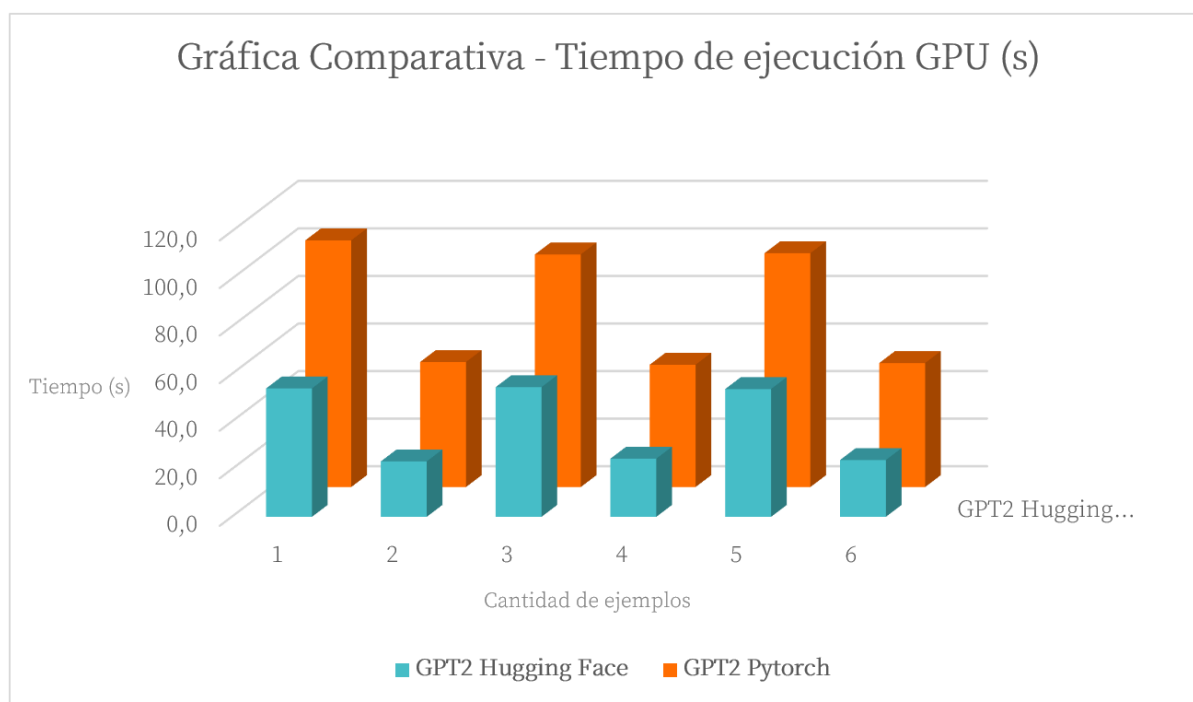


Fig. 1

El algoritmo de Hugging Face en relación con el tiempo de ejecución rinde aproximadamente la mitad de tiempo de lo que rinde GPT-Pytorch (ver Fig. 1).

Evaluation results

The model achieves the following results without any fine-tuning (zero-shot):

Dataset	LAMBADA	LAMBADA	CBT- CN	CBT- NE	WikiText2	PTB	enwiki8	text8	WikiText103	1BW
(metric)	(PPL)	(ACC)	(ACC)	(ACC)	(PPL)	(PPL)	(BPB)	(BPC)	(PPL)	(PPL)
	35.13	45.99	87.65	83.4	29.41	65.85	1.16	1,17	37.50	75.20

Fig. 2 [2] Métricas de evaluación y resultados de GPT2 Hugging Face

Documentación

En cuanto a la documentación, en la página de HuggingFace se encuentra una cantidad abundante de información donde se explican aspectos que van desde la creación, entrenamiento y evaluación hasta la implementación de tuberías (pipelines) que facilitan su uso, esta documentación está seccionada por modelos, entre ellos GPT2.

Por el lado del algoritmo GPT2-Pytorch, a pesar de que posee documentación de uso (poca), no es detallada en comparación a GPT2 de Hugging Face.

Limitaciones

Ambos algoritmos de generación de texto poseen las mismas limitaciones debido a que su base es GPT2, en la tarjeta de modelo de GPT2 de OpenIA señalan las siguientes limitaciones:

“Debido a que los modelos de lenguaje a gran escala como GPT-2 no distinguen la realidad de la ficción, no admitimos casos de uso que requiera que el texto generado sea verdadero.

Además, los modelos de lenguaje como GPT-2 reflejan los sesgos inherentes a los sistemas en los que fueron entrenados, por lo que no recomendamos que se implementen en sistemas que interactúan con humanos> a menos que los implementadores primero lleven a cabo un estudio de sesgos relevantes para los caso de uso. No encontramos diferencias estadísticamente significativas en las sondas de sesgo de género, raza y religión entre 774M y 1.5B, lo que implica que todas las versiones de GPT-2 deben abordarse con niveles similares de precaución en los casos de uso que son sensibles a los sesgos en torno a los atributos humanos.” [2]

Es importante mencionar que el modelo de Hugging Face está “perfeccionado” y disponible en el centro de modelos de la página web, listo para ser utilizado en la tarea de generación de texto, a pesar de esto, se debe tener precaución con el texto generado ya que puede tener sesgos como señala OpenIA.

Encuesta realizada

Se realizó una encuesta con el fin de obtener información acerca de la calidad de los textos generados por ambos algoritmos, esta encuesta fue aplicada a un Docente de Inglés, quién tuvo en cuenta diferentes aspectos para la calificación, tales como: coherencia, gramática, interés del tema y secuencia.

La encuesta consistía en diez textos generados (cinco de cada algoritmo), y se le pedía al docente que proporcionará de acuerdo a su conocimiento una calificación en un rango de uno (1) a diez(10), siendo uno (1) el valor más bajo y diez (10) el valor más alto.

En los resultados obtenidos se puede observar que el algoritmo GPT2 Hugging Face obtiene una mejor calificación, un promedio de 7,4 a diferencia de 6,6 obtenido por el algoritmo GTP2-Pytorch.

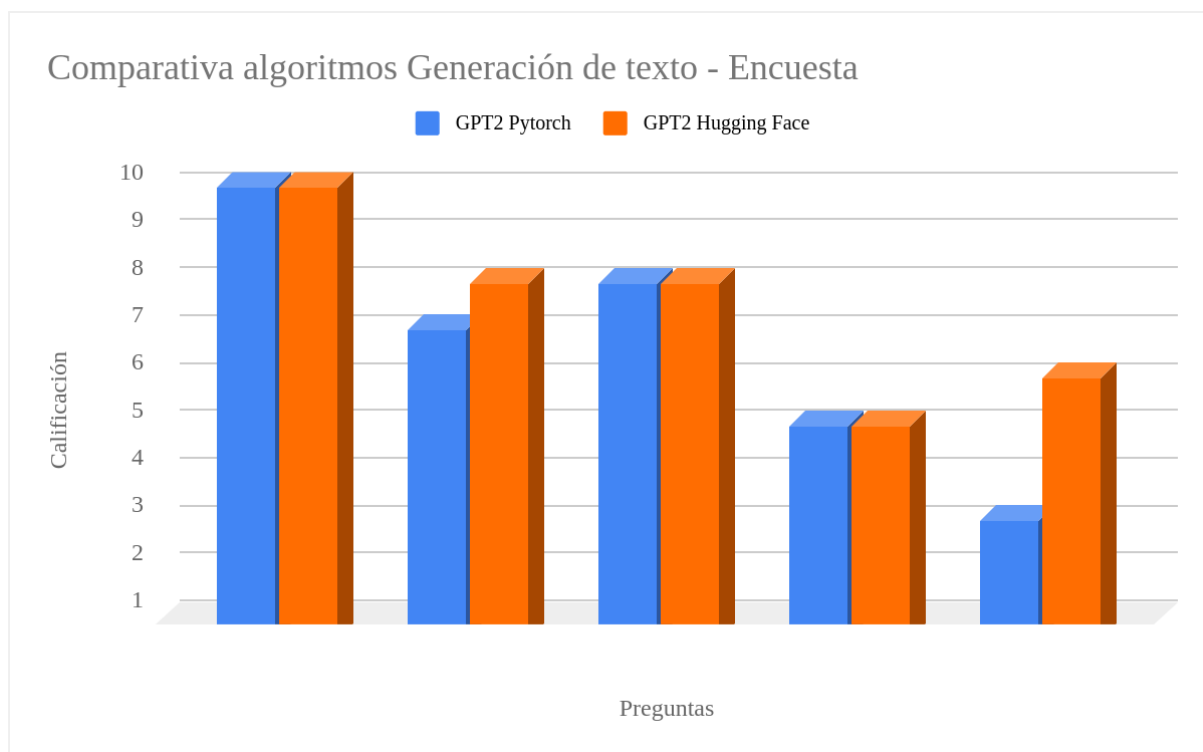


Fig 3.

Evidencia encuesta: Encuesta 2 Trabajo de grado - Formularios de Google.pdf

Generación de preguntas y respuestas

Algoritmo 1: Question generator [3]

Algoritmo 2: Question Generation using transformers [4]

Para la selección del algoritmo de **generación de preguntas y respuestas**, se compararon dos algoritmos de código abierto para esta tarea, sin embargo, desde un inicio uno de los algoritmos presentaba mayor potencial de acuerdo al propósito del proyecto, ya que, de los recolectados es el único algoritmo capaz de generar preguntas y respuestas de opción múltiple y de oraciones (preguntas abiertas), es decir dos tipos de preguntas.

Los otros algoritmos que se recolectaron para esta tarea, incluyendo el algoritmo 2 [4], generan un solo tipo de pregunta (preguntas abiertas). El tipo de preguntas es un factor importante para el propósito del proyecto, ya que, a mayor tipo de preguntas más posibilidades de generar exámenes variados para los estudiantes.

Cabe resaltar que estos algoritmos de respuesta a preguntas van perdiendo calidad en la generación de preguntas cuando la cantidad de estas aumenta, sin embargo, moderando la cantidad de preguntas y proporcionando un texto de entrada completo y extenso en información se puede obtener preguntas y respuestas aceptables. La pérdida de calidad a medida que aumenta la cantidad de preguntas es un factor que está presente en la mayoría de algoritmos dedicados a esta tarea.

La facilidad de uso también es un factor importante en la elección de este algoritmo, se puede ejecutar fácilmente en Google Colab o Kaggle, y aprovechar directamente las ventajas que ofrecen estas plataformas.

De acuerdo a las pruebas y posterior análisis realizado entre ambos algoritmos inicialmente seleccionados, se decidió seleccionar finalmente el algoritmo 1 [3], debido a todas las ventajas obtenidas respecto a el propósito del proyecto expuestas anteriormente.

Referencias

- [1] Tae Hwan Jung, GPT2-Pytorch with Text-Generator (Feb. de 2019), disponible en:
<https://github.com/graykode/gpt-2-Pytorch>
- [2] Julien Chaumond, GPT2 Hugging Face (Feb. de 2019), disponible en:
<https://huggingface.co/gpt2>
- [3] Adam Montgomerie, Question generator (Jul. de 2020), disponible en:
https://github.com/AMontgomerie/question_generator
- [4] Suraj Patil, Neeraj Varshney, Priyank Soni, Manuel Romero, Tom Wilde. Question Generation using transformers (Jul. de 2020), disponible en:
https://github.com/patil-suraj/question_generation