.UBAderecho

IALAB

Impact Assessment of Generative Artificial Intelligence on skills in the workplace

Support by:

Fondazione Adriano Olivetti

UNESCO — Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura

uniTwin — Cátedra UNESCO "Agua y Educación para el Desarrollo Sostenible" UNIVERSIDAD NACIONAL DEL LITORAL FACULTAD DE INGENIERÍA Y CIENCIAS HÍDRICAS

UNIVERSIDAD NACIONAL DEL SUR — UNS BAHÍA BLANCA

LIDeSIA

Cámara Argentina de Comercio y Servicios — 100 años

Thomson Reuters™

Start Coding ia_

puzzle.

Research

Impact Assessment of Generative Artificial Intelligence on skills in the workplace

April 2024

Team Work

Directors

Juan Gustavo **Corvalán**

PhD in Legal Sciences. Master in Artificial Intelligence. Director of the Innovation and Artificial Intelligence Laboratory of the Faculty of Law of the University of Buenos Aires (UBA IALAB).

Enzo María **Le Fevre Cervini**

PhD in Public Policies. Master in Artificial Intelligence for public services. Head of sector of the European Commission.

The opinions expressed by the author are strictly personal and do not necessarily reflect those of the European Commission.

Soledad **Guilera**

Master in International Studies. Professor at the University of California at Berkeley (USA) and at the Torcuato Di Tella University (Argentina). Team Leader of the AI Policy Clinic of the Center for Artificial Intelligence and Digital Policy (USA).

Laura **Díaz Dávila**

Engineer and PhD in Public Policies. LIDESIA Director of the Faculty of Exact, Physical and Natural Sciences of the National University of Córdoba and of the Research, Development and Deployment Group of Applied Artificial Intelligence of UBA IALAB.

.UBAderecho   IALAB

Research Leaders
  Mariana **Sanchez Caparros**
  Giselle **Heleg**
  Melisa **Raban**
  Lautaro **Vasser**

Specialist Consultants
  Belen **Carreira**
  Milagros **Etcheberry Le Fort**
  Pamela **Tolosa**

Research Team
  Carina **Mariel Papini**
  Florencia **Croci**
  Margarita **Gonzalez**
  Sol **Roschkavan**

Graphic Design
  Paula C. **Petroni**
  Victoria **Mafud**

Index

Executive Summary

This document provides a detailed analysis of the implementation of Generative Artificial Intelligence (GenAI) tools in a wide spectrum of tasks, with the aim of evaluating its impact in key sectors such as justice, public administration, law firms, legal departments, translation, education and research[1]. This study takes on special importance, since it provides valuable information on how GenAI can transform and improve various areas that are fundamental for the development and well-being of today's society.

The investigation was carried out between August and December 2023, and focused on analyzing the execution of a set of specific tasks in two different instances. First, the task was performed without using GenAI and then the task was executed again using GenAI. In both cases, the times necessary for its completion were measured. For all tasks, a categorization was carried out regarding their efficiency in relation to the level of complexity, degree of automation, level of repetitiveness, required human judgment and incidence of GenAI. The results of the research arise from the analysis of the times required to execute the tasks with and without the help of GenAI, and the degree of efficiency achieved with respect to the indicated dimensions.

Through the study of **83 tasks**, a significant improvement is observed in the times required to complete them, with an **average reduction of 77%** in the time necessary to complete them, thanks to the use of GenAI tools. Furthermore, a **maximum efficiency level of 99%** is identified for the preparation of a draft translation of 15,000 words.

The **high efficiency (81%)** dof GenAI in tasks that require a high level of human judgment demonstrates its potential to assist or complement human work in complex tasks. This is a clear indication of how GenAI has moved forward to address more sophisticated tasks that require a significant degree of insight or decision. On the other hand, the incorporation of GenAI as support for the performance of a task can give rise to subtasks that did not exist before but are necessary to achieve real optimization of the chosen task.

Furthermore, the analysis highlights that GenAI tools have a high potential for automating routine, well-defined, structured or medium- and high-complexity tasks, while their benefit is lower in simpler tasks. This may be because not all tasks are equally suitable for automation with GenAI. This varies from one institution to another, and requires a specific analysis of each of the processes to determine how and when to implement GenAI tools.

---

1   In order to analyze the impact, tests were carried out in seven judicial branches of Argentina, in two areas of the Public Administration and in five legal firms and legal departmentss of companies.

When working on use cases within an organization, improvements can be presented in both quantitative and qualitative terms. The quantitative improvements are those that we describe, essentially and mainly, in this report and that are reflected in reductions in the time necessary to complete the task. Qualitative improvements are those that imply an improvement in the quality of the result expected from a certain task. The latter may or may not mean saving time in carrying out the task, but they are usually very useful when it comes to increasing the competitiveness of the organization.

The results obtained show the presence of a new paradigm on how to understand and apply GenAI in tasks and jobs.

Tests on 83 tasks

Average time optimization with GenAI

Based on the tests conducted, the average time optimization from performing tasks with **GenAI tools is 77%.**

Maximum **efficiency** achieved by the tool

The maximum efficiency achieved by the **GenAI tool was 99%** for the production of a first draft translation of a 15,000-word text.

Complexity

High: **24 tasks**

Medium: **45 tasks**

Low: **14 tasks**

Efficiency according to complexity level

High complexity: **73%**

Medium complexity: **81%**

Low complexity: **52%**

Degree of automation

Automatable: **21 tasks**

Semi-automatable: **32 tasks**

Non-automatable: **30 tasks**

Efficiency according to automation possibility

Automatable: **94%**

Semi-automatable: **40%**

Non-automatable: **69%**

Level of repetitiveness

High: **24 tasks**

Medium: **45 tasks**

Low: **14 tasks**

Efficiency according to degree of repetitiveness

High: **86%**

Medium: **88%**

Low: **43%**

Human judgment required

High: **42 tasks**

Medium: **22 tasks**

Low: **10 tasks**

Efficiency according to the level of human judgment required

High: **81%**

Medium: **74%**

Low: **44%**

Impact of GenAI

Assistant: **49 tasks**

Complement: **16 tasks**

Substitute: **10 tasks**

Displacement: **1 task**

Efficiency according to the level of impact of GenAI

Assistant: **94%**

Complement: **68%**

Substitute: **45%**

Displacement: **27%**

1 CONTEXT, OPPORTUNITY AND JUSTIFICATION

Context, Opportunity and Justification

This work takes place at a revolutionary moment for a technology industry that impacts all areas of our society, but at the same time is critical in terms of its influence on the global labor market.

Historically, concerns, challenges and opportunities have arisen around the convergence of these two domains that have given rise to more or less intense debates. All of these proposals, today, have been re-presented with greater force and intensity due to certain particularities that characterize our time. Next, we will begin by reviewing these conditions.

a. Trends and outlook in the world of work

Labor markets face significant challenges. With high rates of unemployment and underemployment, finding a job is often difficult. A significant skills mismatch also takes place, where there is a mismatch between the skills of the workforce and the needs of employers. The challenges of automation and digitalization transform the available jobs and require new skills; increasing job insecurity, with an increase in temporary or contractual jobs that offer less security and benefits; pronounced income inequalities between different sectors, regions and demographic groups; unsafe or exploitative working conditions in certain sectors or regions; the effects of demographic changes, such as an aging population, which impact labor supply and statement of claim; barriers to labor mobility are some of the challenges facing the current labor market.

In addition to this, the COVID-19 crisis has contributed, on the one hand, to accelerating certain transformations that were already evident, such as the large-scale adoption of teleworking and the proliferation of digital platforms. But, simultaneously, they introduced new forms of control on workers' performance and productivity derived from technological innovations[2].

On the other hand, the pandemic deepened existing structural problems, such as unemployment, labor informality and gender gaps[3]. Beginning with the start of the recovery process in 2021, the International Labor Organization (ILO) notes that employment growth

---

2   Novick Marta, Andrade Eliana, Ballatore Ferretti Lucila, Estrella María Josefina, Malpede Agustín, Munilla Diana, Núñez Vanesa, Parrilla SantGenAlo, Pieckeinstain Lucía, Pomar Tomás, *"Technology and Digitalization: The Trade Union Challenge"*, SinDigital, May 2022

3   Ripani Laura, *"Impact of COVID-19 on the labor market: What has happened since the economic crisis, and what is next?"* IDB, May 2022, available at: https://blogs.iadb.org/tralow/es/el-mercado-laboral-desde-el-covid-19/ (accessed March 4, 2024).

was solid thanks to the resumption of activity in key sectors of the economy, and that labor market conditions continued improving in 2022. However, by 2023, a slowdown in said growth was already expected, and the persistence of significant decent work deficits around the world that violate social justice[4].

The conditions, the quality of jobs, and the undervaluation of workers do not present a favorable outlook either[5]. Likewise, economic factors such as slowing productivity growth and supply shortages due to geopolitical tensions such as the prolonged conflict in Ukraine have caused high and persistent uncertainty around the world that impacts the challenges that labor markets[6] are facing.

b. The Fourth Industrial Revolution and the rise of intelligent natural language processing systems

According to Klaus Schwab - businessman and economist, president of the World Economic Forum -, humanity has gone through and is going through a total of four industrial revolutions[7]. The first occurred approximately between 1760 and 1840, and was triggered by the construction of the railroad and the invention of the steam engine, inaugurating mechanical production systems. The second one took place between the end of the 19th century and the beginning of the 20th, and was characterized by mass production, thanks to the incorporation of electricity and the assembly line. The third industrial revolution began in 1960 with the development of semiconductors and computing using mainframe servers[8]. And then, personal computing (1970s and 1980s) and the Internet (1990s) follow, leading to the Digital Revolution.

Currently, we are going through the Fourth Industrial Revolution or Industry 4.0[9], focused on the possibilities of artificial intelligence and intelligent machines. But it is not only that, the sophistication and integration of computer networks and systems also constitute key elements for the transformation of the economy and society. In particular, in addition to AI, the following technologies and developments are considered drivers of this revolution: internet of things (IoT), robotics, drones, autonomous cars, 3D printing, cloud computing, nanotechnology and blockchain.

---

4   *"Social and employment perspectives in the world. Trends 2023",* World Labor Organization, Reference Report, Executive Summary, available at: https://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/---publ/documents/publication/wcms_865368.pdf (accessed March 4, 2024).

5   *"Social and employment perspectives in the world. The value of essential work",* World Labor Organization, Reference Report, Executive Summary, available at: https://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/---publ/documents/publication/wcms_871018.pdf (accessed March 4, 2024).

6   *"Social and employment perspectives in the world. Trends 2023",* World Labor Organization, Reference Report, Executive Summary, available at: https://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/---publ/documents/publication/wcms_865368.pdf (accessed March 4, 2024).

7   https://hipermediumciones.com/2018/06/10/la-4o-revolucion-industrial/ (accessed March 4, 2024)

8   Mainframe systems are high-performance computers with large amounts of memory and processors that process billions of simple calculations and transactions in real time. The mainframe is essential for business databases, transaction servers, and applications that require security and agility. See IBM, "What is a mainframe, available at:: https://www.ibm.com/mx-es/topics/mainframe (accessed 03/2/2024)

9   https://www2.deloitte.com/content/dam/Deloitte/uy/Documents/human-capital/Revoluci%C3%B3n%204.0.pdf (accessed 03/04/2024)

This latest revolution has impacted our way of living, working and even relating, establishing the principle of instantaneousness and high speed[10]. In this new economic model, efficiency prevails, that is, achieving an objective or fulfilling a function using the smallest amount of resources possible, such as time and/or materials. It is a paradigm in constant expansion that at the same time penetrates and solidifies in all areas of our lives.

Within this framework, in recent years specifically, we have witnessed an exponential advance of certain artificial intelligence systems, especially emerging in the field of natural language processing, but then taken and exploited by applications in other disciplines and for other purposes. Let's take a look.

Until 2020, existing conversational agents could not avoid for more than a couple of lines, spouting inconsistencies, reproducing discriminatory biases, and suddenly changing topics to skip awkward chit-chat. Replika[11], Kuki[12] and Dynabench[13] performance fell significantly short of expectations when it came to simulating natural language understanding as a human does[14]. EEven so, they managed to not be a complete disaster unlike what happened with other attempts, such as Microsoft's Tay chatbot[15] that had to be deactivated hours after its launch for publishing Nazi slogans, among other atrocities.

As techniques based on deep learning - neural networks - became more sophisticated and optimized, they caused an exponential jump in the performance of language models, and made it possible for AI systems to solve increasingly sophisticated feats.

This is how in June 2020, OpenAI presented the first version of GPT-3, 100 times larger than its previous version, GPT-2 (2019). From supposedly therapeutic babbling and experiments in manual labeling of hate speech[16], users began to interact with a predictive chatbot, in which a few lines could automatically lead to a newspaper article, a poem, guitar chords, computer codes and even a text summary. And although it was far from perfect, it seemed that the natural language processing revolution had begun.

If GPT-3 was fed with 410,000 million texts[17], this trend has only been on the rise during

---

10   Borges Blázquez Raquel, *"Artificial Intelligence and Criminal Procedure"* , Thomson Reuters Aranzadi 2021, pp. 31.
11   https://replika.com/ (accessed on 03/04/2024)
12   https://www.kuki.ai/ (accessed on 03/04/2024)
13   https://dynabench.org/ (accessed on 03/04/2024)
14   At IALAB we carry out tests on the conversational agents Kuki, Dynabench and Replika, evaluating aspects such as discriminatory biases, their understanding of causal relationships, the coherence of their responses, among others. The tests are available at: Corvalán Juan G., Carro María Victoria (collaborator), *"The limits of Artificial Intelligence. Correlations, Causality, Shakira, GPT-3 and Alice in Wonderland"*, in Corvalán Juan G. *"Treatise on Artificial Intelligence and Law"*, Thompson Reuters La Ley, Volume I, 2021. Also, see *"The bias in conversational agents. Again on coded prejudice"*, IALAB, available at: https://ialab.com.ar/wp-content/uploads/2022/09/El-sesgo-de-los-Agentes-Conversacionales.pdf?utm_source=email_marketing&utm_admin=151772&utm_medium=email&utm_campaign=Novedades_UBA_IALAB_septiembre.
15   "Perspectivas sociales y del empleo en el mundo. Tendencias 2023", Organización Mundial del Tralow, Informe de Referencia, Resumen Ejecutivo, disponible en: https://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/---publ/documents/publication/wcms_865368.pdf ((accessed on 03/04/2024)
16   Heaven Will, Douglas, *"The almost impossible challenge of creating a chatbot that is not racist and sexist"*, MIT Technology Review, November 4, 2020, available at: https://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/---publ/documents/publication/wcms_865368.pdf (accessed March 4, 2024).
17   Heaven, Will Douglas, "OpenAI's new language generator GPT-3 is shockingly good—and completely mindless ," MIT Technology Review, July 20, 2020, available at: https://www.technologyreview.com/2020/07/20/1005454/openai-machine-learning-language-generator-gpt-3-nlp/ (accessed March 4, 2024)

2021, to the point of considering it as the year of giant language models[18]. Thus, while Big Tech 's achievments in understanding human speech focused on absorbing more and more online text, the problems associated with its use, such as the reproduction of unfair biases and a lack of understanding of language, persisted. The same methodology was replicated over and over again, but with greater scale and scope, which was extremely costly and complex, and resulted in its exclusive use and benefit being concentrated in the hands of a few creative companies.

In **2022**, innovations leaned towards models that were capable of not only processing text, but also images. Experts realized that if they could make systems "see," they could correct certain common sense problems in machines, to imitate a learning process that was more similar to that of a human brain. This is how multimodal models were born. Examples of these have been DALL-E 2[19] and CLIP[20] from OpenAI or Stable Diffusion[21] from Stability AI.

c. ChatGPT and the keys to its success

But when it seemed that the AI news of 2022 had come to an end, OpenAI announced the launch of ChatGPT[22], uone of the most important technological milestones in recent times. Two months after its launch, it had become the fastest growing platform in the entire history of the Internet, reaching 100 million users in just 2 months, until Meta's latest social network, Threads, obtained more than 100 million registered users in just five days[23]. The arrival of ChatGPT has had an impact on all professions and fields of knowledge, such as education, programming, law, art and politics.

Why ChatGPT? Not even OpenAI itself imagined that its product would become a mega viral success[24]. However, now that a few months have passed and we can examine it retrospectively, we believe that its mass adoption can be explained by several factors.

First of all, **performance**. We have summarized a few lines ago how language models have become rapidly sophisticated in a couple of years, increasingly imitating language as if it were generated by a human. Although ChatGPT is not perfect, it has considerably increased the quality of its responses compared to its predecessor GPT-3, as we have demonstrated

---

18   Heaven Will Douglas *"2021: The year of giant artificial intelligence models"* MIT Technology Review, December 27, 2021, available at: https://www.technologyreview.es/s/13901/2021-el-ano-de-los-modelos-de-inteligencia-artificial-gigantes (consultado el 4 de marzo de 2024)..
19   https://openai.com/dall-e-2 (accessed March 4, 2024)
20   https://openai.com/research/clip (accessed March 4, 2024)
21   https://stability.ai/ (accessed March 4, 2024)
22   OpenAI announced the release of ChatGPT on November 30, 2022. ChatGPT is built on top of OpenAI's GPT-3.5 and later, and was designed to provide a more interactive and conversational user experience.
23   https://www.forbesargentina.com/innovacion/threads-ya-record-convierte-app-crecimiento-mas-rapido-historia-n36788#:~:text=en%20dos%20meses.-,Al%20lograr%20el%20hito%20en%20menos%20de%20cinco%20d%C3%ADas%2C%20Threads,100%20millones%20en%20dos%20meses.
24   Heaven Will Douglas, *"This exclusive conversation with the creators of ChatGPT shows the surprise of its success"* MIT Technology Review, March 2023, available at: https://www.technologyreview.es/s/15168/esta-conversacion-exclusiva-con-los-creadores-de-chatgpt-evidencia-la-sorpresa-de-su-exito (accessed March 4, 2024)

in a series of experiments we carry out at UBA IALAB[25].

These improvements are not only observed in terms of the coherence of the responses, but also in relation to certain ethical standards that had been proving problematic for language models and that currently remain almost impossible to resolve and completely eradicate. This is the reproduction of unfair or discriminatory biases or the generation of hate speech[26].

ChatGPT has been attractive to users in that sense because when a direct attempt is made to make it fall into these assumptions, it is capable of evading the response, demanding greater efforts from whoever intends to circumvent the rules. While the aforementioned Tay in 2016, after a few hours became a fan of Hitler, ChatGPT puts creativity to work for whoever wants to make it throw answers of this style, and when these people succeed, the developers update the model and correct the flaw.

On the other hand, if ChatGPT makes a mistake and the user points it out, it recognizes their mistake and changes the response. This, regardless of whether the person introduced absurdities or not. In this sense, the chatbot materializes the old slogan attributed to a London merchant at the beginning of the 20th century: "the customer is always right." That is, the model is superior in terms of its **flexibility and learning capacity**.

All of these components in relation to the content and quality of the responses are part of a holistic human-centered AI approach that OpenAI has adopted to guide certain aspects of ChatGPT, which is also evident from the reinforcement learning technique based on human feedback. Placing the user at the center implies, among other things, paying attention to their experience and satisfaction, which necessarily leads to more attractive and captivating products.

Secondly, it must be recognized that the platform is **simple and intuitivea**, even easier to manage than many of the massive social networks we know. Before ChatGPT, language models like GPT-3 were accessible through a waiting list managed by the creating company, in which users had to enter a lot of data to join. Additionally, once enabled, the interface was a bit more complex, with a long list of tasks that one could select and certain settings that could be adjusted by those with technical knowledge.

Now ChatGPT allows you to log in with just an email and a password. And to use it, simply write a message in a chat line, just as you do when you talk to your friends on WhatsApp, Telegram or WeChat, platforms with which we are all already familiar. From there, you can

---

25   *"ChatGPT vs. GPT-4 Imperfect by Design?"* UBA IALAB, available at: https://ialab.com.ar/wp-content/uploads/2023/03/Libro-ChatGPT-vs-GPT-4.-UBA-Thomson-Reuters-La-Ley.pdf (accessed on March 4, 2024)

26   In different research carried out at UBA IALAB, we have highlighted the need to know and understand the intrinsic and external limitations of large language models such as ChatGPT, which cause the models to hallucinate; generate harmful content and/or provide biased responses that reproduce or reflect gender stereotypes, prejudices, beliefs or negative social values. Regarding the point, it is suggested to expand and see some examples based on the tests carried out to analyze the possibility of these models yielding biased, incoherent or fantasized responses: *"ChatGPT vs. GPT-4: imperfect by design?"*, UBA IALAB - Thomson Reuters La Ley, March 2023, p.15-16, available at: https://ialab.com.ar/wp-content/uploads/2023/ 03/Libro-ChatGPT-vs-GPT-4.-UBA-Thomson-Reuters-La-Ley.pdf (accessed on 03/05/2024)

ask it to do all sorts of natural language feats, like writing a poem, translating a text, or telling a story.

Based on these facilities, the contrast with other competing companies has become more evident regarding their tendency to keep their innovations in the field away from the public's reach. There is consensus among experts that the latest OpenAI models are neither the first nor the only ones to develop such advanced technologies. There is no reason to think that Google or Meta are below them or even that they do not surpass them. The difference is that OpenAI has been following a policy of openness, launching defective products so that society adapts and collaborates in correcting these defects, in the case of ChatGPT indicating, for example, which response option has been most useful to the user.

Thirdly, ChatGPT is extremely **useful** for users in performing all kinds of natural language based tasks. You can answer emails, organize travel itineraries, assist in solving academic work and even correct computer code. In this sense, it has the potential to facilitate a significant percentage of the tasks we carry out daily.

But the usefulness is not only important for the average user, but also for developers. Rather, we are referring to scalability, when, being a pre-trained model (as indicated by its acronym Generative, Pre-trained, Transformer), it allows it to be taken, calibrated or reused to become a specialist in specific fields of knowledge. Since the arrival of this model, many other projects that have it as a base have proliferated, exploiting and maximizing its functionalities such as Bing from Microsoft, Gemini from Google or LlaMa from Meta[27].

Almost seven hours is the average daily time that people around the world spend connected to the Internet[28]. This virtual space, in which we inhabit much of our day, is governed by a few technological giants that have launched a race to capture the software market .

Microsoft has acquired the rights to use ChatGPT and the rest of the OpenAI tools, becoming its main shareholder. In this way, it integrated these powerful AI systems into its Bing search engine, as well as into all the tools that are part of the Office package (Word, Excel and Power Point). In early 2024, Google launched Gemini, which seeks to be the successor to PaLM, the language model currently used by Bard. The idea is that gradually PaLM will be replaced by Gemini within Bard. Gemini stands out for outperforming all its rivals in the main tests, ahead of OpenAI's GPT-4[29].

---

27   For some examples see, Fernández Yúbal, *"26 projects based on ChatGPT: Artificial Intelligence alternatives to create texts, websites, role-playing games and more"*, Xataka, February 2023, available at: https://www.xataka.com/basics/26-proyectos-basados-chatgpt-alternativas-inteligencia-artificial-para-crear-textos-webs-partidas-rol (accessed March 4, 2024)

28   Pasquali Marina, *"How many hours a day do we spend connected to the internet?"*, Statista, February 2023, available at: https://es.statista.com/grafico/22701/tiempo-medium-de-uso-diario-de-internet/#:~:text=Seis%20horas%20y%2037%20minutos,Overview%20Report%22%20publicado%20por%20DataReportal (accessed March 4, 2024)

29   https://www.xataka.com/basics/google-gemini-que-como-funciona-diferencias-gpt-cuando-podras-usar-este-modelo-inteligencia-artificial (accessed on 02/26/2024)

This flow of innovation has transformed and continues to transform the way companies and individuals perform tasks and make decisions. The ongoing technological revolution has generated enthusiastic expectations but also significant concerns. On the one hand, there is unprecedented potential to improve people's efficiency, productivity and quality of life through the automation of routine and repetitive tasks. On the other hand, concerns arise about the impact on work, economic inequality and the skills required in the future, which this research aims to address.

2 RESEARCH PURPOSES AND PROBLEMS

Research Purposes and Problems

In 2019, at UBA IALAB we promoted research that was published under the title *"Artificial Intelligence and Work. Building a new employment paradigm"*[30], written by Luis Cevasco, Juan Gustavo Corvalán and Enzo María Le Fevre Cervini, and prefaced by the Head of the Department of Macroeconomic Policies of the ILO, Dr. Ekkerhard Ernst. This research aimed to evaluate the impact of automation on work, and one of the main strategies used was to do so based on the Prometea intelligent system, driven by the same team, in the preparation of legal opinions in justice.

As a result, and based on the data produced by other measurements[31], some of the most relevant conclusions at that time were the following:

> » Automation does not necessarily tend to reduce work, but in different sectors or specific areas it can also increase it.
>
> » AI will advance mechanical and routine tasks, and staff will be reassigned to more creative and productive tasks. Therefore, it is key to educate and accompany the transition so that workers can acquire new skills.
>
> » The opportunities and benefits of this new paradigm must be taken advantage of by the countries' economies to become more competitive.

Although only five years have passed since the publication of that book, the abrupt changes that have been evident since then force us to review and test these conclusions. As we have explained when developing the context, since 2019 our society has gone through and overcome a global pandemic that implied various modifications in the labor market. On the other hand, at a technological level, the ChatGPT boom was the starting point for the popularization of GenAI, which is now available to everyone.

When we took Prometea in the previous work, we were able to evaluate its specific impact with respect to the specific area where we had implemented it and for which it had been designed. In this way, it is difficult to make generalizations about the impact that a specific intelligent system, custom-designed for one environment, can have on a different

---

30   The book is available at: https://ialab.com.ar/wp-content/uploads/2023/03/ia_y_tralow.pdf
31   Corvalán, Juan G. "Prometea. Artificial intelligence to transform public organizations", Editorial Astrea, DPI and IMODEV, year 2019, available at: https://ialab.com.ar/wp-content/uploads/2023/03/Libro-Prometea.pdf (accessed on 02/26/2024)

organization. In contrast, the advantage that ChatGPT presents is that its massive adoption allows us to consider its consequences in the workplace of any task in any organization, whether in the field of law or in other fields such as education or health.

In conclusion, in line with the purpose of reviewing the previous works and results of UBA IALAB and in response to the context developed, the two main questions to be answered by this work are: What is the impact of ChatGPT and other similar models of generative AI languages in the workplace? And more specifically, what level of impact does GenAI have on the tasks and subtasks carried out by people within the work processes?

From both, the following research problems emerge:

- » **First. Do ChatGPT and GenAI models tend to reduce the time needed to perform certain tasks or activities??**

- » **Second. Do ChatGPT and GenAI models lead to the reconfiguration of certain tasks in different jobs?**

- » **Third. Do ChatGPT and GenAI models require workers to develop new skills? If the answer is yes, what are these new skills and which are and will be the most valued?**

3
METHODOLOGY

Research methodology

Selection of:

- Sectors included in the investigation
- Tasks, subtasks and microtasks to analyze
- Composition of each of them

Indicators for analyzing impact

Quantitative

Time required to the completion of the task, subtask and microtask

- with GenIA
- without GenIA

This includes:

- Time for creation and adaptation of the prompt
- Time for post-human control
- Level of task repetitiveness: **High / Medium / Low**

Qualitative

Task complexity:
- High
- Medium
- Low

Human judgment required for completition:
- High
- Medium
- Low

Degree of automation of the task:
- Automatable
- Semi-automatable
- Non-automatable

Required skills

Level of impact of GenAI in the tasks, subtasks and microtasks:
- Assistance
- Complement
- Substitution
- Displacement

Degree of familiarity with the GenAI tool

Conceptual Framework and Methodology

Key concepts for this work will be those related to artificial intelligence that have emerged, become popularized and spread during the last period of growth and accelerated sophistication of the natural language processing models that we have developed in the contextual framework. Likewise, other relevant concepts will be those that allow us to classify and categorize work tasks to apply the research methodology and obtain conclusions about the impact of ChatGPT and GenAI on them.

a. Concepts related to artificial intelligence

Artificial intelligence. An AI system is a machine-based system that, for explicit or implicit goals, infers, from the input it receives, how to generate results such as predictions, content, recommendations or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptability after deployment[32].

Generative Artificial Intelligence (GenAI). This is a subfield of AI that is used to create new and original content, or to modify or improve existing content, such as text, images, music, and videos. GenAI relies on machine learning to identify patterns in data and then uses those patterns to generate new content[33].

The deep learning models on which GenAI is based can take raw data and learn to generate statistically probable results when requested. They encode a simplified representation of your training data and extract it to generate output similar, but not identical, to the original data[34]. In essence, GenAI can create new data, unlike traditional AI that tries to label or categorize data[35].

---

32   OECD, *"Explanatory memorandum on the updated OECD definition of an AI system"*, March 5, 2024, available at: https://www.oecd-ilibrary.org/science-and-technology/explanatory-memorandum-on-the -updated-oecd-definition-of-an-ai-system_623da898-en (accessed 03/05/2024)

33   OECD Artificial Intelligence Papers, Initial Policy Considerations for Generative Artificial Intelligence, September 2023, available at: https://www.oecd-ilibrary.org/deliver/fae2d1e6-en.pdf?itemId=/content/paper/fae2d1e6-en&mimeType =pdf (accessed 02/28/2024)

34   See more in *What is artificial intelligence (AI)?*, IBM, available in: What is Artificial Intelligence (AI)? | IBM (accessed on 02/28/2024)
       See also Greg Pavlik, *What is Generative AI? How does it work?*, Oracle, September 15, 2023, available at: What is generative artificial intelligence? | Oracle Argentina (accessed on 02/28/2024).

35   See more in Elena Astorga, *Generative artificial intelligence is going to change what it means to be human*, MIT Technology Review, August 9, 2023, available at: "Generative artificial intelligence is going to change what it means to be human" | MIT Technology Review in Spanish (accessed 02/28/2024)

Impact[36]. "*...it implies changes in people's lives. This could include changes in the knowledge, skill, behavior, health or living conditions of children, adults, families or communities. Such changes are long-term positive or negative effects on identifiable population groups produced by a development intervention, directly or indirectly, intended or unintended. These effects may be economic, sociocultural, institutional, environmental, technological or other types*"[37].

Prompt. A prompt is an instruction, question or text that is used to interact with artificial intelligence systems. We could say that it is like a command through which a user asks the system to perform a specific task[38].

Large Language Models. Large Language Models (LLM) or large linguistic models, are artificial intelligence systems based on the neural network technique that are trained with enormous amounts of text[39]. In this way, they become capable of processing and generating language as if they were human beings. They are used in various applications, including chatbots, virtual assistants, language translation, content creation, and scientific research. As examples, we can mention GPT-3, GLaM, Jurassic-1, Switch-Transformer and Gopher[40].

Foundation Models. Base models or foundation models are those models trained on broad data that can be adapted to a wide range of downstream tasks[41].

b. Methodology and objectives

The application of GenAI introduces a new paradigm in understanding and using AI in a wide variety of tasks and jobs. Tests carried out on 83 tasks reveal that large language models open a universe of new possibilities to improve integration, collaboration, automation and accessibility in the workplace. This innovation allows for unprecedented synergy and cooperation. Automation takes on a new dimension with GenAI tools capable of performing tasks with a high level of autonomy, assisting or complementing human workers in their completion. Likewise, accessibility and scalability are seeing a notable boost, as the democratization of GenAI opens up possibilities for a broader spectrum of people and organizations to take advantage of its advanced capabilities. This progress not only improves existing processes, but also enables the creation of new tasks, jobs and business models, redefining the future of work and human-machine interaction.

---

36   For the purposes of this research, the definition provided by the United Nations Sustainable Development Group was taken as a reference.

37   *"Results Based Management"*, United Nations Development Group, October 2011, p.7, available at: https://unsdg.un.org/sites/default/files/UNDG-RBM-Handbook-2012.pdf (accessed 04 /03/2024)

38   https://www.google.com/amp/s/www.xataka.com/basics/que-prompt-que-importantes-para-usar-inteligencia-artificial/amp

39   Lancaster Anthony, *"Beyond Chatbots: The Rise Of Large Language Models"*, Forbes, March 2023, available at: https://www.forbes.com/sites/forbestechcouncil/2023/03/20/beyond-chatbots-the-rise -of-large-language-models/?sh=4c9f9a2e2319 (accessed March 4, 2024)

40   Heaven Will Douglas, *"2021: the year of giant artificial intelligence models"*, MIT Technology Review, December 2021, available at: https://www.technologyreview.es//s/13901/2021-el-ano-de -the-giant-artificial-intelligence-models# (accessed March 4, 2024)

41   Rishi Bommasani, Percy Liang, *"Reflections on Foundation Models"*, Stanford University, Human-Centered Artificial Intelligence, October 2021, available at: https://hai.stanford.edu/news/reflections-foundation-models (Accessed March 4, 2024).

The research team −according to the parameters explained below− selected the following strategic sectors to measure the impact of GenAI on its internal processes: i) Justice; ii) Public Administration; iii) Legal departments or law firms; iv) Translation; v) Education; and, vi) Research.

The selection of the sectors was made taking into consideration the specialized knowledge and extensive experience of UBA IALAB in the fields of justice, public administration, education and research. Specifically, special focus was placed on analyzing what was happening in this translation sector, based on an August 2023 study, carried out by OpenAI, OpenResearch and the University of Pennsylvania[42], which considered that translation and interpretation were among the jobs most likely to be affected by exposure to large language models[43].

As provided for in the methodology that lays the foundations of this research[44], test teams were formed. These teams were in charge of selecting the processes and analyzing the tasks and subtasks that compose them, and applied said methodology to document the step-by-step measurements obtained.

Mainly, the objective of the research was to collect data to analyze the effect that generative artificial intelligence -GenAI- has on the work carried out in different organizations. Specifically, in the execution of certain tasks specific to the sector.

---
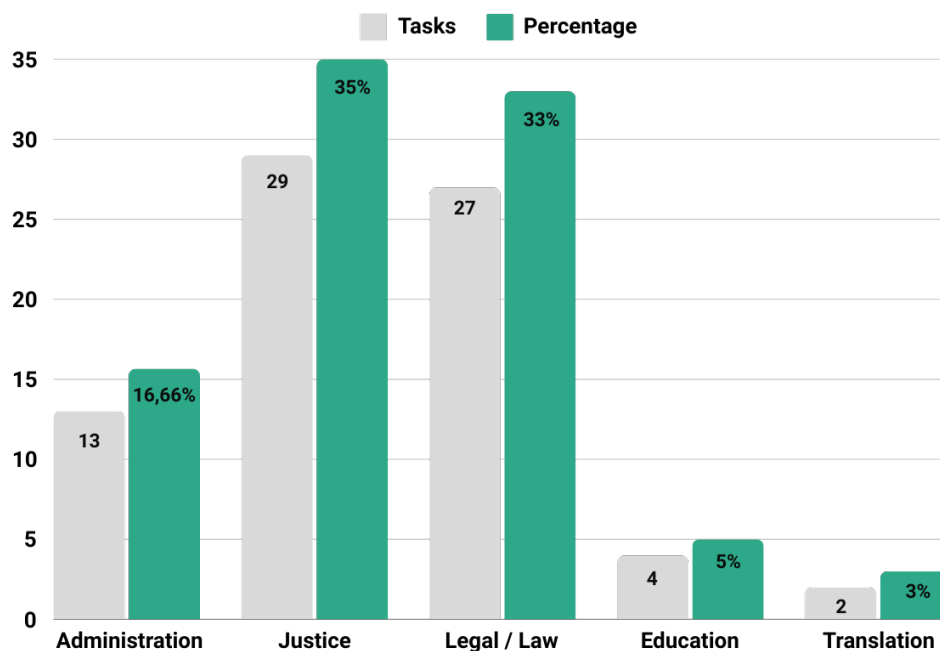
42   Eloundou, Tyna - Manning, Sam - Mishkin, Sam - Rock, Daniel, *"GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models",* available at: https://arxiv.org/pdf/2303.10130 .pdf (accessed 03/07/2024)
43   In a second stage, IALAB will convene references from other vectors such as health, human resources and finance, in order to understand the impact on other sectors.
44   IALAB, Methodological guide for evaluating the impact of generative artificial intelligence on employment and organizations. Identification of new job skills, August 2023, available at: https://ialab.com.ar/wp-content/uploads/2024/02/Guia-metodologica-1.pdf (accessed on 02/26/2024)

This involved analyzing:

1) The immediumte and obvious results that GenAI causes in these tasks, what we call "direct impacts";
2) The side effects and subtle influences that could arise, the "indirect aspects;
3) Changes that are consciously sought, as well as those that may occur in an unplanned manner - intentional or not.

In this way, the evaluation focuses on obtaining a complete and holistic understanding of the impact of GenAI on all these dimensions, to make informed and effective decisions on how to address it.

To analyze how GenAI impacts jobs, it is necessary to identify:

    a. What sectors and industries will be included in the investigation;
    b. What tasks, subtasks and microtasks of those sectors and industries will be analyzed;
    c. How each of them is formed.

Likewise, to evaluate the impact that GenAI has in the sectors included in this research, indicators were used:

    i. **Quantitative.** Specifically: time to complete the task, subtask and microtask; time for the construction of the prompts ; time for human supervision of the result provided by GenAI.
    ii. **Qualitative.** Specifically: degree of automation of the task; task complexity; prompt used and considerations for its construction; level of incidence of GenAI in the tasks, subtasks and microtasks; identification and differentiation of different tasks that GenAI performs to obtain the desired result; tool adaptability; satisfaction and user experience.

To measure the impact of GenAI on tasks and its correlation with work, the following metrics were considered: a) efficiency; b) degree of automation possible according to task segmentation; c) effectiveness that also involves the impact of GenAI on the tasks in terms of complementation, substitution, displacement and assistance.

These metrics have been chosen because they provide a comprehensive view of GenAI's impact on work. They also allow us to understand the impact on tasks, subtasks and microtasks according to their degree of automation, and, in addition, consider other factors such as the adaptability of the tool and the worker's experience when using it.

The chosen concepts and metrics are described below:

a. Efficiency

- With GenAI: measures the execution time of tasks before and after the implementation of GenAI in the process. This includes measurements on the completion time of the task, subtask and/or microtask.

- Without GenAI: here the time to complete the task, subtask and/or microtask is measured according to the commonly used internal procedures and tools approved by the organization

b. Degree of automation of the task

Degree of automation possible according to task segmentation

This identifies which tasks may be suitable for automation and which are more appropriate to be performed by humans.

The research team agreed on the terms of the approved methodology to carry out the tests. It was a methodology based on a rubric[45] and parameters in matrix format that helped visualize and determine which tasks could be suitable for automation and which could be more appropriate to be performed by humans.

41 people were invited to participate in the research, including lawyers, business administrators, economists, translators, from 17 areas, in organizations belonging to the 5 sectors mentioned above. Test teams were formed to conduct testing of the activities using and without using GenAI, and they were trained so that they could capture the data consistently in the forms provided for this purpose. Calibration meetings were held, based on the first tests, doubts were resolved and the measurement tools were adjusted.

The test teams selected and evaluated the tasks based on two key criteria: repetitiviness and human judgment required, and in turn weighted them according to the following scale: high, medium, low .

Let's briefly review the concepts used as a pillar for the definition of the selected tasks and/or processes:

45   Eloundou, Tyna - Manning, Sam - Mishkin, Pamela - Rock, Daniel, *"GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models"*, March 27, 2023, available at: https://arxiv .org/pdf/2303.10130.pdf

Repetitiveness

Refers to the repetitive or recurring nature of a task or activity. In the work and technological context, repetitiveness is related to the constant performance of similar or identical tasks. These tasks generally involve a series of steps or actions that are executed over and over again without significant changes to the process.

Repetitiveness can be a characteristic of jobs or processes that can be automated or systematized, since they do not require complex decision making or constant adaptation to changing situations. It can be high, medium or low.

- **High repetitiveness.** Repetitiveness will be high when the procedures and actions to follow to carry out a task remain uniform, similar or identical over time.

- **Medium repetitiveness.** The repetitiviness of a task will be medium or average when certain actions can be modified due to the presence of factual hypotheses, which produce variability in the execution of a task. In other words, the task is not performed identically each time, but rather allows for a certain degree of adaptation or change in its actions, depending on the factual assumptions that are manifested.

- **Low repetitiveness.** Repetitiveness will be low when each time a task is performed the steps to follow change or are modified. Creative solutions and critical thinking are usually required here. These tasks are typical in roles that demand innovation or complex problem solving.

Human judgment

Refers to people's ability to evaluate, analyze, make decisions and solve problems based on their experience, knowledge, values and reasoning. It involves the ability to consider multiple factors, contexts, and perspectives when making informed decisions.

- **High human judgment.** The human judgment required is high in complex situations, which usually require analyzing multiple variables and making informed decisions. Above all, these are situations of uncertainty where the person's human judgment is key to defining a route of action, anticipating possible consequences, risks or damages.

- **Medium human judgment.** These are situations in which the skills of workers are important to evaluate and make decisions based on their experience. Such as: the ability to analyze, integrate, and observe.

- **Low human judgment.** Tasks that require a low level of human judgment are those that are routine and predictable, which usually have clearly defined procedures. In them, a deep evaluation, analysis or decision-making based on vast experience or knowledge is not necessary.

	REPETITIVENESS	HUMAN JUDGMENT
Writing Reports	Low	High
Replying to emails	High	Medium
Scheduling Meetings	High	Low
Database management	High	Low
Strategic decision making	Low	High

Now, if we interpret the rubric in the following way, we find:

- **AUTOMATABLE TASKS:** tasks with high repetitiveness and low human judgment required are highly automatable. In this case, " scheduling meetings" and "database management" could be considered for automation.

- **SEMI-AUTOMATABLE TASKS:** Tasks with high repetitiveness but a medium degree of human judgment required could be partially automatable, as they could require some supervision or human intervention. In this example, "replying to emails" could fall into this category.

- **NON-AUTOMATABLE TASKS:** Tasks with low repetitiveness and high human judgment required are less susceptible to automation. These tasks typically require more complex skills and the ability to make decisions based on human judgment and experience. In this example, "writing reports" and "strategic decision making" could be considered non-automatable tasks.

c) Effectiveness

Incidence of GenAI on the execution of the task

In order to understand the impact of GenAI on work tasks, it is first necessary to determine what is meant by effectiveness in this context. For research purposes, effectiveness refers to GenAI's ability to meet its proposed objectives by improving or facilitating the performance of specific tasks. This concept serves as a starting point to analyze how the incorporation of GenAI affects the work environment.

From this perspective, the incidence of GenAI on tasks is examined based on whether its integration results in complement, assistance, substitution or displacement. This approach allows for a more nuanced understanding of how GenAI can coexist with human talent, enhancing or changing the current workforce landscape. For the purposes of this investigation it is understood as:

» **Complement:** phenomenon by which artificial intelligence or GenAI systems automate, through various techniques, the performance of a main part of a human task, which results in the remaining part of the task being performed humanly. In this way, it is possible to save time that was previously required to complete the task. The complement may entail modification in the performance of a complex task. This is what we previously called inclusive cobotization[46]. In certain areas that are difficult to replace, there is the advantage of using technology in favor of workers, who can use it as an instrument to create and generate value.

» **Assistance:** provides greater input for the completion of a task, but does not imply an absorption of the human task. In general, the human knowledge base increases and/or the performance of an accessory part of the task is optimized. For example, if you want to write an article or academic work and you want to consult sources of information, it is useful to copy and paste certain doctrine into ChatGPT to determine the most important points of said article, specifically in relation to the topics one wants to investigate. In this example, the task of producing the academic work or article continues to be human; ChatGPT assists in processing information and increasing the knowledge base.

» **Substitution:** Phenomenon generated from the complete absorption of a task by artificial intelligence, which generates a replacement of human activity by the activity of artificial intelligence. For example, if GPT-4 is requested to prepare a letter document, the system returns the complete model as a result.

» **Displacement:** It is a consequence of the substitution and, in certain cases, of the complement. Displacement is understood as the dedication of humans to new tasks or part of certain tasks. In cases of displacement, it is understood that the original task or subtask was absorbed by the implementation of AI in the process. (For example, the task of reviewing texts and modifying them based on errors in grammar, syntax or searching for synonyms for words that are frequently repeated in the paragraphs of an administrative act/resolution prior to signing).

---

46 See more in Cevasco, Luis. - Corvalán, Juan G., - Le Fevre, Enzo M., *"Artificial Intelligence and work"*, DPI Cuántico, IMODEV, year 2019, page 41, available at: https://ialab.com.ar/wp-content/ uploads/2023/03/ia_y_trabajo.pdf (accessed on 03/04/2024)

Procedure to link tasks, subtasks and microtasks with new skills

One of the points that we sought to analyze with this research is the identification of new skills required in the world of work based on the impact of GenAI on tasks. To do this, some categories of skills approved by international organizations were selected, such as the International Labor Organization (ILO) and the Organization for Economic Cooperation and Development (OECD)[47].

Previously, in the process of categorizing the tasks and subtasks, the type of skills required to carry them out was linked. Likewise, in order to evaluate whether the required skills – with the use of GenAI in carrying out the task – have undergone modifications, this was also recorded in the tests carried out by the teams.

For the purposes of analyzing the required skills, we follow the classification of employment skills established by the ILO and the OECD.

---

47   Skills for employment, *"International Labor Organization"*, May 18, 2015, available at: https://www.ilo.org/skills/pubs/WCMS_371815/lang--es/index.
htm (accessed 4/ 03/2024)
See also OECD, *"Skills for the Digital Transition: Assessing Recent Trends Using Big Data"*, October 19, 2022, available at: https://www.oecd.org/
employment/skills-for-the-digital-transition- 38c36777-en.htm (accessed 03/04/2024)

CATEGORY OF SKILLS ACCORDING TO GENERAL CATEGORIES OR PROFESSION	KEY COMPETENCIES/SKILLS
Learn to learn	• Be willing to learn • Use learning techniques to acquire and apply new knowledge and skills • Work safely • Pursue independent learning • Take responsibility for one's own learning • Think abstractly • Organize, process and retain information • Interpret and communicate information • Conduct systematic inquiries, with follow-up to find answers • Use time effectively and efficiently without sacrificing quality • Choose the best way to approach tasks • Start tasks, continue them and complete them • Be adaptable
Communication	• Read competently • Understand the language in which the organization is managed • Know how to write to satisfy the needs of the target group • Listen and communicate effectively • Know how to listen to understand and learn • Use numeracy effectively • Know how to express one's ideas and visions
Teamwork	• Work in teams or groups • Respect the thoughts, culture and opinions of other group members • Understand and contribute to the organization's goals • Plan and make decisions with other people and support the results • Take responsibility for actions • Form alliances and coordinate diverse experiences • Work towards group consensus in decision making • Value the contributions of others • Accept feedback • Resolve conflicts • Offer guidance, mentor, give feedback • Lead effectively • Mobilize groups to achieve high performance

CATEGORY OF SKILLS ACCORDING TO GENERAL CATEGORIES OR PROFESSION	KEY COMPETENCIES/SKILLS
Problem solving	• Identify problems • Take into account the context of the data and the circumstances • Identify and suggest new ideas to accomplish the job (initiative) • Collect, analyze and organize information (planning and organization) • Plan and manage time, money and other resources to achieve goals
Data scientists and analysts	• Data visualization • Cybersecurity Framework • Network security • Information security • IT management • Data analysis • Database management • Data processing and storage • Machine learning • Artificial intelligence • Big Data
Software developers, programmers and engineers	• Cloud Computing • Web development • Java • Distributed computing • Software quality assurance • IT management • Simulation • Performance management • Tensorflow

CATEGORY OF SKILLS ACCORDING TO GENERAL CATEGORIES OR PROFESSION	KEY COMPETENCIES/SKILLS
UX/UI Designer	• Design thinking • User research • Web development • Software quality assurance
ICT and HR managers / marketing specialists	• Advertising • Marketing management • Business solutions • IT management • Business strategy • Employee training • Strategic thinking • Artificial intelligence

Clarifications on the sample chosen for the 83 tasks

» The selected tests correspond to a random sample. The people called to carry out the tests were free to select the tasks and worked with those related to their daily work, with the condition of specifying the details of their composition, such as their degree of automation and complexity.

» Some tests did not reach the expected levels of efficiency, which denotes the need for people to adapt to the use of GenAI tools and, in other cases, shows that GenAI is not the most appropriate tool to make more efficient the chosen task (for example, to calculate procedural deadlines).

» Some tests were carried out by people with no prior knowledge of technology and no relevant experience in using ChatGPT in their work, while others were carried out by students of the Postgraduate Course in Artificial Intelligence and Law at the UBA who, over 9 months, from the workshops and practical workshops, they acquired the necessary tools to apply GenAI in their daily work. This shows, at least a priori , that optimization can grow if people have specific training in using GenAI tools, such as ChatGPT.

» The study allows us to affirm that even without prior knowledge about ChatGPT, with a careful selection of the task to which to apply said technology, it is possible to optimize the tasks in the different jobs.

» Carrying out the tests allowed us to observe that the user needs a reasonable amount of time to generate detailed, contextual and useful prompts to perform the chosen task.

» This document constitutes the first step of the research that is being carried out by UBA IALAB. The results presented here could be modified. Mainly, because the tests were not repeated by those responsible, which may imply a variation in the results obtained here.

4 RESEARCH RESULTS

Application of generative artificial intelligence in the workplace

This infographic reflects the main results of the research carried out by UBA IALAB to analyze the impact of GenAI and ChatGPT in reducing time and optimizing tasks
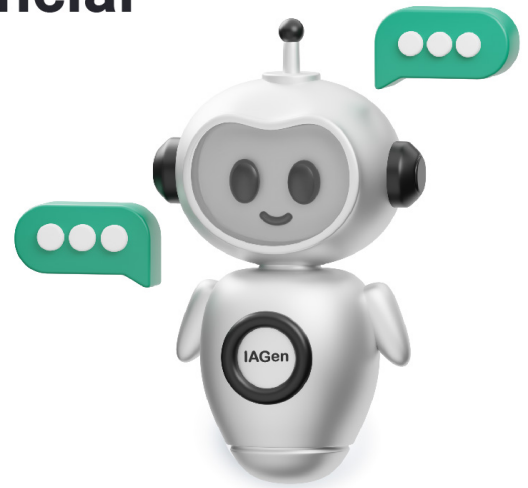
Tests on 83 tasks:
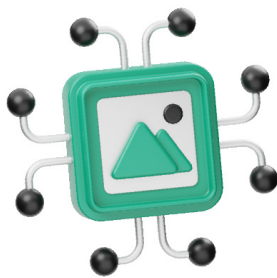
5 areas including Public Administration, Justice, Legal Departments and Law Firms, Translation, Education

77% average time optimizataion

Maximum efficiency achieved

99,96% Draft translation of a 15,000-word text

Efficiency increase in various tasks

97% Drafting of an administrative contract

95% Making a modified copy of a lease agreement

90% Identification of grievances in the unconstitutionality appeal

71% Planning of a practical activity in education

Incidence of GenAI in tasks:

—

Assistant **59.03%**

Complement **19.27%**

Substitute **12.04%**

Efficiency according to complexity level:

—

High complexity **73%**

Medium complexity **81%**

Low complexity **52%**

GenAI hypothesis

Increases efficiency and optimizes subtasks on a large scale

Changes the paradigm in the way of working

Most disruptive tool ever invented

Findings

Tasks of medium complexity are the optimum point for the implementation of GenAI with an average reduction of times by 81%

Certain low-complexity tasks do not benefit as much from automation with AI. Average time reduction with AI: 52%. AI assists or complements human work in complex tasks, achieving high efficiency of 73% in tasks requiring high human judgment.

GenAI is evolving to address more sophisticated tasks that require a significant degree of discernment or decision

The high efficiency (86% and 88%) in tasks with high and medium repetitiveness highlights the utility of applying GenAI to automate and optimize processes

Prioritize automation efforts in those clearly defined and standardized tasks

Detailed and contextualized initial prompt for greater optimization

Optimization of work tasks of high and medium complexity mainly, even without prior knowledge of the operator

GenAI means quantitative and/or qualitative improvements

Increase in efficiency with repeated tests on new cases

IAGen

Tasks compared

	with GenAI	**without GenIA**
Drafting of an administrative contract	**6** minutes	**60** minutes
Making a copy of a lease agreement with modifications	**6** minutes	**120** minutes
Identification of grievances in the appeal of cassation	**30** minutes	**15** minutes
Planning of a practical activity in education	**10** minutes	**35** minutes

Lessons learned

- Detection of the specific subtasks for the use of GenAI
- Training of workers for optimal use of the tool
- Adaptation of prompts to the reality of each organization and subtask
- Specialization of a person from the organization in the use of GenAI
- Measurement of the time taken by the task with and without GenAI to calculate real optimization

Next steps

Confirm or refute the hypothesis with repetition of use cases and optimization of prompts

Prompts as templates

Triggers for next steps
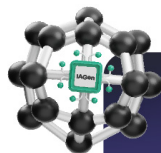
01 Is it possible to develop and refine different types of prompts to function as models or "templates" to further optimize the execution of various tasks or subtasks?

02 Can prompts as templates be extrapolated to other processes or tasks?

03 Are people required to be trained in using generative IA optimally?

04 Do prompts as templates imply a radical improvement in terms of quantitative and/or qualitative optimization in the medium and long term?

05 Is it advisable for a team to constantly teach, adjust and interact with generative AI strategically within the organization?

Preliminary results of standardized prompts as templates

- Greater optimization in work tasks
- Require periodic review and adjustments
- Shortcuts in GenAI
- Specific for each subject and subtask

Research Results.
Analysis of the Data Obtained.

Next, we present the results of the research. The organization of the content is distributed into several sections to facilitate its understanding: a) Main findings; b) Sectors and areas included in the research; c) Time optimization; d) Complexity; e) Degree of automation; f) Repetitiveness of tasks; g) Human judgment required; h) Incidence of GenAI on the task; i) Efficiency.

The analysis of the results, segmented by areas and sectors, allows the identification of specific patterns and trends in each domain and by virtue of each of the metrics selected to measure the impact. Finally, an evaluation of the results obtained according to the complexity, level of repetitiveness and human judgment required, offers a detailed perspective on the data collected.

a. Main findings

Based on the analysis of the results obtained by virtue of the tests carried out, the following findings are identified:

GenAI repetitiveness and efficiency level

- **High and medium repetitive tasks:** Particularly high efficiency (**86% and 88%**) in tasks with high and medium repetitiveness underlines the importance and usefulness of applying GenAI to automate and optimize processes. These tasks represent the optimal spot for GenAI implementation and highlight the importance of prioritizing automation efforts into clearly defined and structured processes.

- **Low-repetitive tasks:** The lower efficiency in low-repetitive tasks (**43%**) highlights the need to reevaluate how GenAI is used, and consider alternatives or complements, such as more direct human intervention to improve efficiency.

- By virtue of the above, the implementation of GenAI should preferably focus on highly repetitive or medium repetitive tasks, where significant improvements in terms of efficiency are more likely to be achieved.

Complexity of GenAI tasks and application

- **Medium complexity:** Efficiency in tasks of medium complexity (**81%**) indicates that GenAI is particularly useful for optimizing tasks that are neither too simple nor excessively complex. This means that tasks of medium complexity are the optimal spot for GenAI implementation.

- **Low complexity:** Lower efficiency on low-level complexity tasks (**52%**) is an indicator that certain simple tasks do not benefit as much from automation, possibly due to the unique or variable nature of these tasks, making human adaptability necessary.

Human judgment and role of GenAI

- **High Human Judgment:** The high efficiency of **73%** on tasks requiring a high level of human judgment shows GenAI's ability to assist or complement human work in complex activities. This challenges the common notion that GenAI is only suitable for repetitive, low-complexity tasks, and shows how GenAI is evolving to address more sophisticated tasks that require a significant degree of insight or decisiveness.

- **Minimum displacement by GenAI:** The majority of tasks (**59,03%**) benefit from GenAI assistance, which implies a collaborative approach between humans and AI, rather than a complete replacement.

- For **more complex tasks**, which require a high level of human judgment, GenAI acts more as a complement than a substitute. These tasks often require advanced skills, including critical thinking, analysis and creativity, where human intervention remains essential.

Implications for skill development

- The difficulty in automating certain tasks and the high human judgment required for some of them underscore the need for specialized human skills. Training and skills development should simultaneously focus on those areas where GenAI cannot make significant contributions, as well as those that enhance human-machine interaction.

Degree of automation of tasks

- The high efficiency in automatable tasks (**94%**) indicates that automation efforts should be prioritized in those tasks that are clearly defined and structured. For semi-automatable tasks, it may be beneficial to re-evaluate how GenAI is implemented and look for ways to increase its efficiency.

In conclusion, the vast majority of tasks benefit from an GenAI assist effect, rather than a complete replacement. For this reason, it is of fundamental importance to foster a culture of collaboration between humans and AI.

b. Time optimization

Below, we present the average efficiency achieved from the use of GenAI in each of the sectors included in the research. Efficiency was analyzed taking into consideration the time required to execute the task with and without GenAI.

.
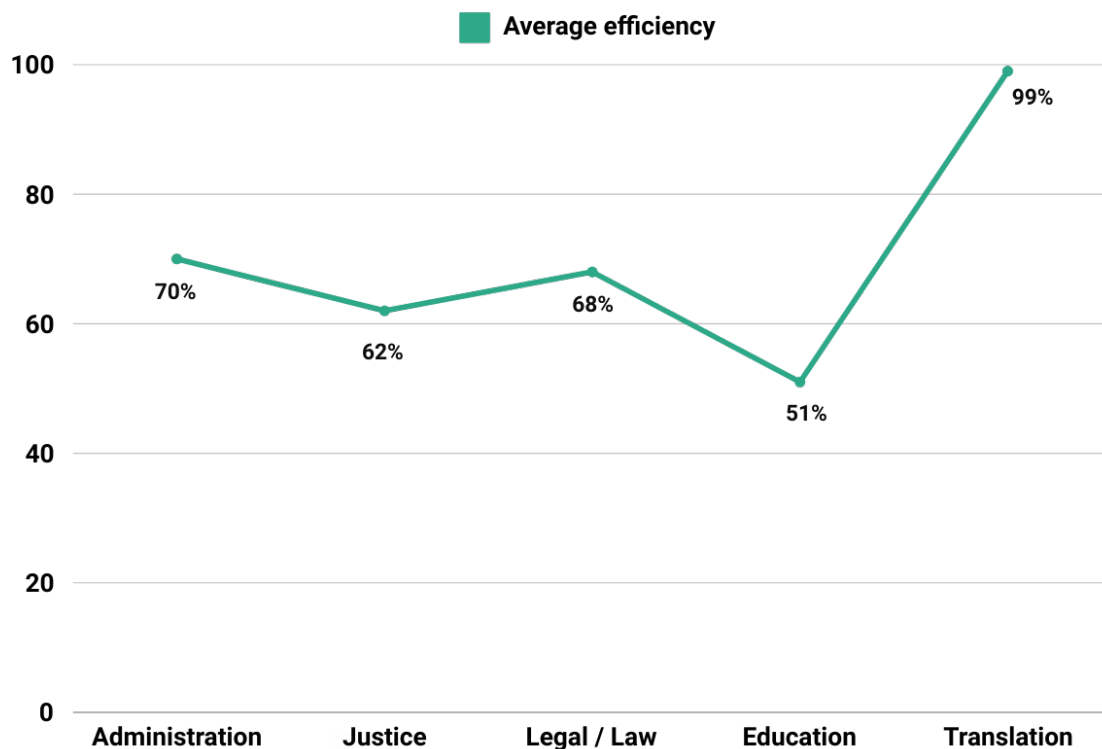
c. Complexity

Regarding the level of complexity of the tasks:

- 24 tasks have **high** complexity
- 45 tasks have **medium** complexity
- 14 tasks have **low** complexity

COMPLEXITY HIGH	COMPLEXITY MEDIUM	COMPLEXITY LOW
28 %	**54 %**	**17 %**
Tasks: 24	Tasks: 45	Tasks: 14
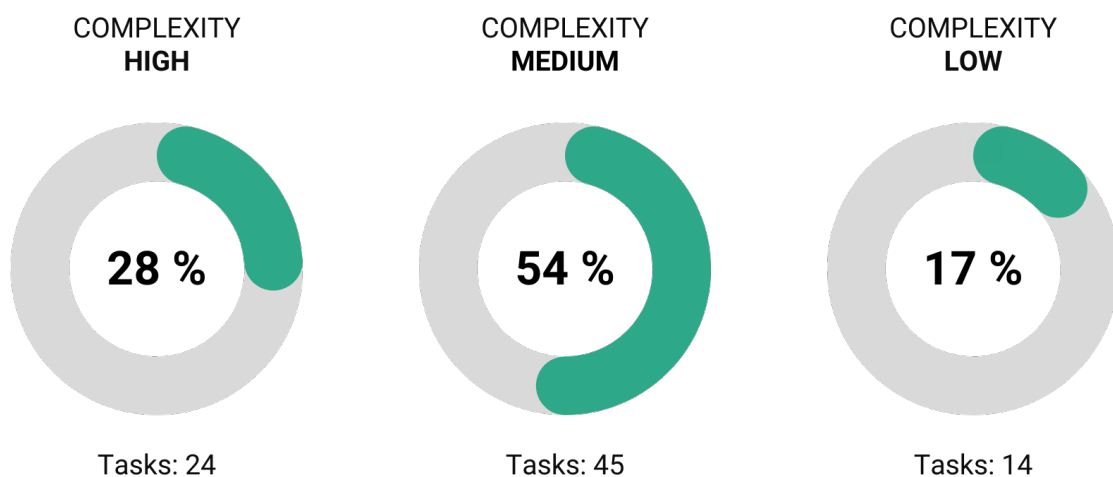
c. Degree of automation

Regarding the degree of automation of tasks:

- 21 tasks were identified as **automatable**
- 32 tasks as semi-automatable
- 30 tasks were considered as non automatable

AUTOMATABLE	SEMI-AUTOMATABLE	NON AUTOMATABLE
25,3 %	**38,5 %**	**36,1 %**
Tasks: 21	Tasks: 32	Tasks: 30
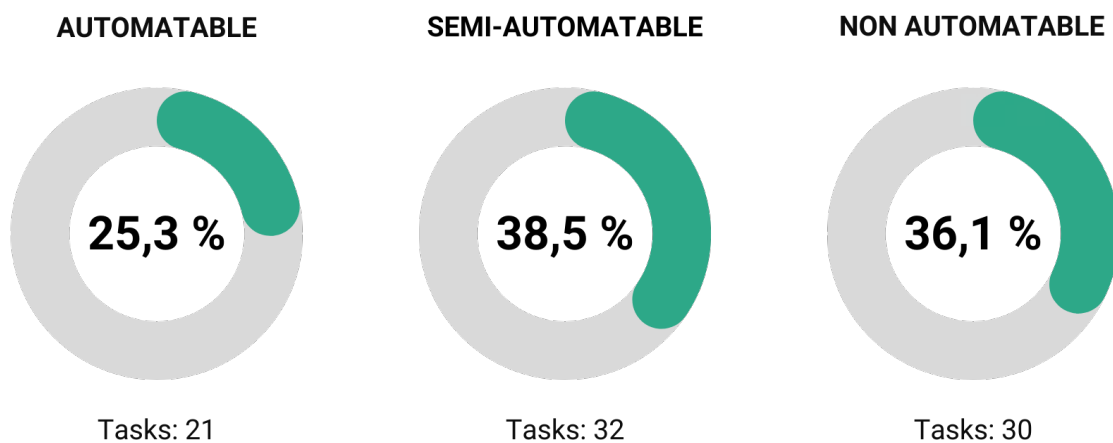
At this point, it is important to note that the classification of a task as "automatable", "semi-automatable", and "non-automatable" was assigned by the person responsible for performing the test.

e. Task repetitiveness

Regarding the level of repetitiveness of the tasks:

- **35 tasks** have **high** repetitiveness

- **17 tasks** have **medium** repetitiveness

- **29 tasks** have **low** repetitiveness

REPETITIVENESS **HIGH**	REPETITIVENESS **MEDIUM**	REPETITIVENESS **LOW**
42,2 %	**20,5 %**	**34,9 %**
Tasks: 35	Tasks: 17	Tasks: 29
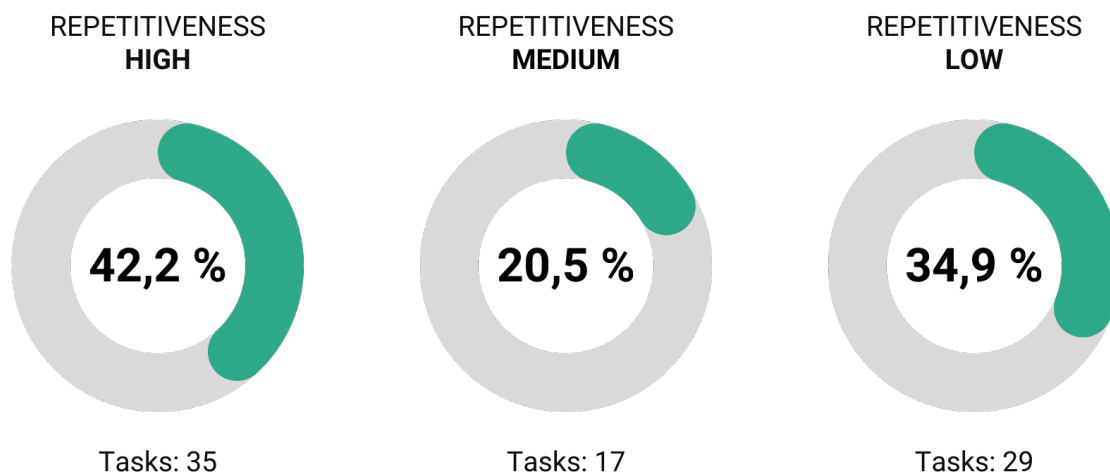
f. Human judgment required

Regarding the human judgment required:

- **45** tasks require a **high** level of human judgment[48]

- **25** a **medium**[49] level

- **11** a **low**[50] level

---

48   For example, drafting a claim for breach of contract within the framework of an identification consumer relationship; the preparation of a labor demand; the individualization of grievances in extraordinary provincial appeals; identification of the object of the claim; identification of the object in an unconstitutionality appeal; the drafting of a claim for damages for contractual liability in the context of gender violence and the analysis of a case of astreintes, among others. See annex context of the tests.

49   For example, putting together a glossary for a document translation; drafting a judicial document letter; the review of bidding documents and conditions in a public procurement and the application of clear language in the vote on a sentence, among others. See annex context of the evidence.

50   For example, preparing an email in "plain language"; drafting the award resolution in a public procurement process; the preparation of a judicial document and the comparison of documents, among others. See annex context of the evidence.

**HUMAN JUDGMENT
HIGH**

54,2 %

Tasks: 45

**HUMAN JUDGMENT
MEDIUM**

30,1 %

Tasks: 25

**HUMAN JUDGMENT
LOW**

13,3 %

Tasks: 11
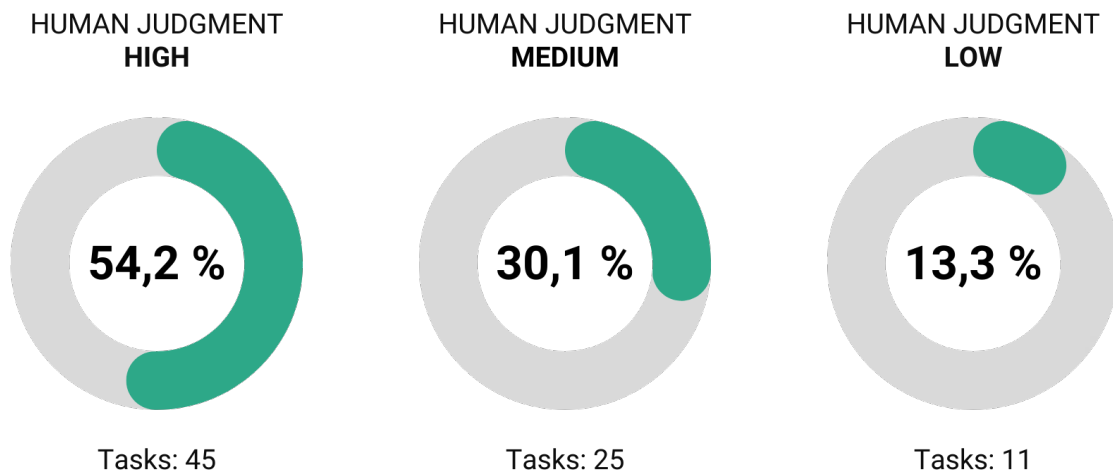
g. Incidence of GenAI on the task

Regarding the level of incidence of GenAI tools in carrying out the task:

- In **49 tasks,** the GenAI tool **provided** assistance in carrying them out

- In **16 tasks,** GenAI **complemented** its implementation

- In **10 tasks, it replaced** the human person in carrying out the task

- In **1 task** the effect caused by the tool was **displacement**

Number of tasks   Percentage

	Number of tasks	Percentage
Assistance	49	59,03%
Complement	16	19,27%
Substitution	10	12,04%
Displacement	1	1,2%
Not assigned	7	8,43%
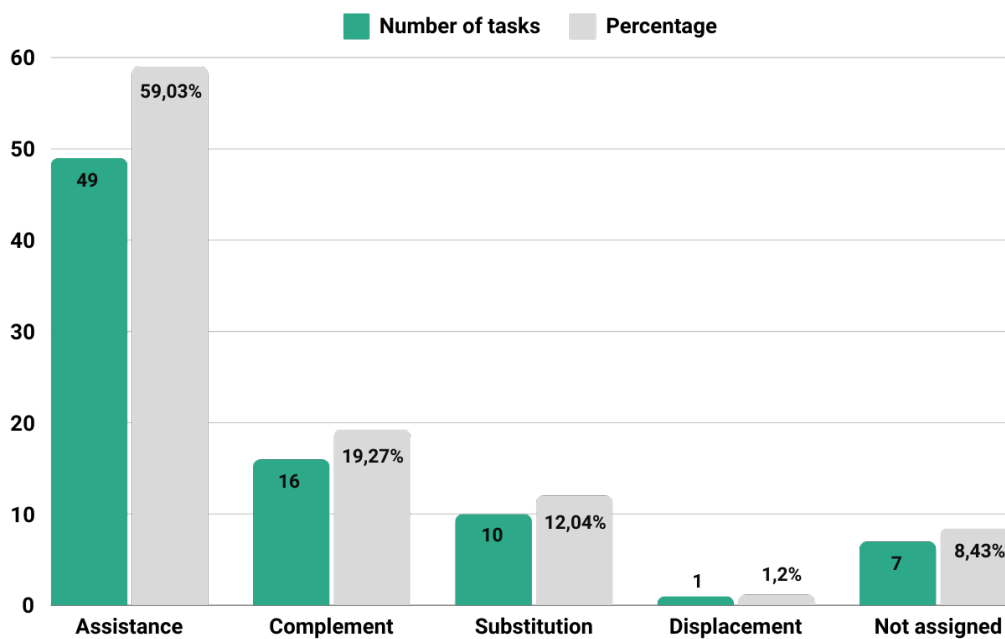
h. Efficiency

1. Efficiency according to the level of complexity of the task

LOW
52

HIGH
73

%
EFFICIENCY
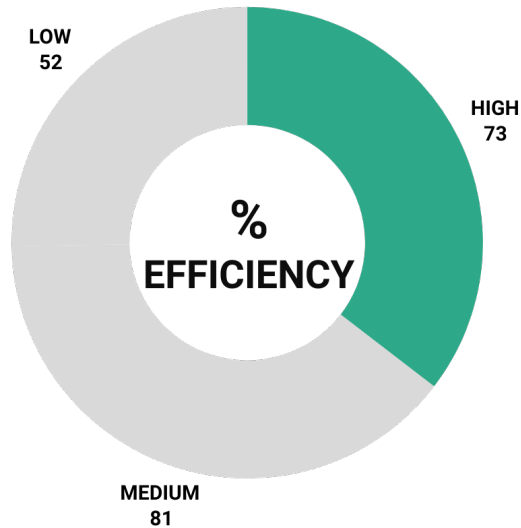
MEDIUM
81

This suggests that GenAI is more effective on medium and high complexity tasks compared to low complexity tasks.

2. Efficiency according to the level of human judgment required[51]

LOW
43

HIGH
86

%
EFFICIENCY
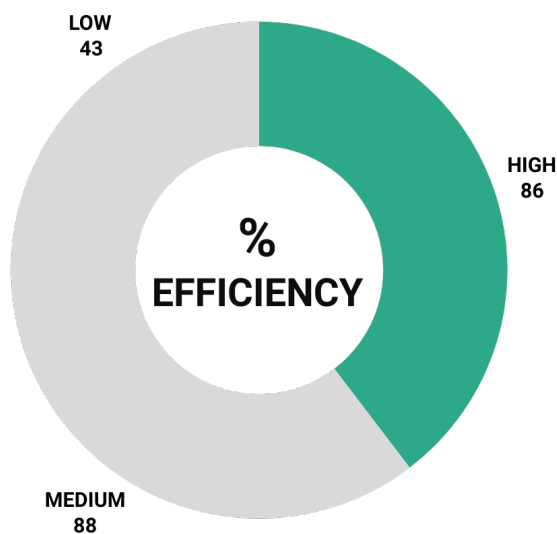
MEDIUM
88

This indicates that GenAI is particularly effective in tasks that require a high and medium level of human judgment.

---

51   Note that the total number of trials performed in low human judgment tasks was substantially lower than in medium or high human judgment tasks. To better understand the performance of GenAI in low human judgment tasks, it is necessary to perform more tests on new tasks and analyze the achieved efficiency..

3. Efficiency according to the degree of repetitiveness of the task

LOW
44

HIGH
81

%
EFFICIENCY
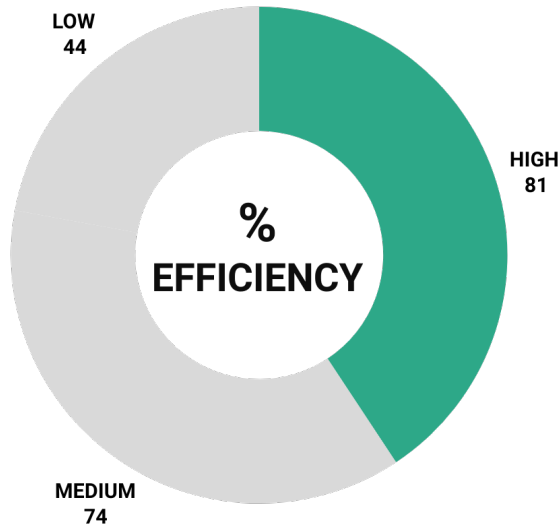
MEDIUM
74

GenAI is more efficient in tasks of medium and high repetitiveness, which is consistent with the idea that automation and AI are effective in routine tasks.

4. Efficiency according to the degree of automation of tasks

NON AUTOMATABLE
69

AUTOMATABLE
94

%
EFFICIENCY

SEMI-AUTOMATABLE
40
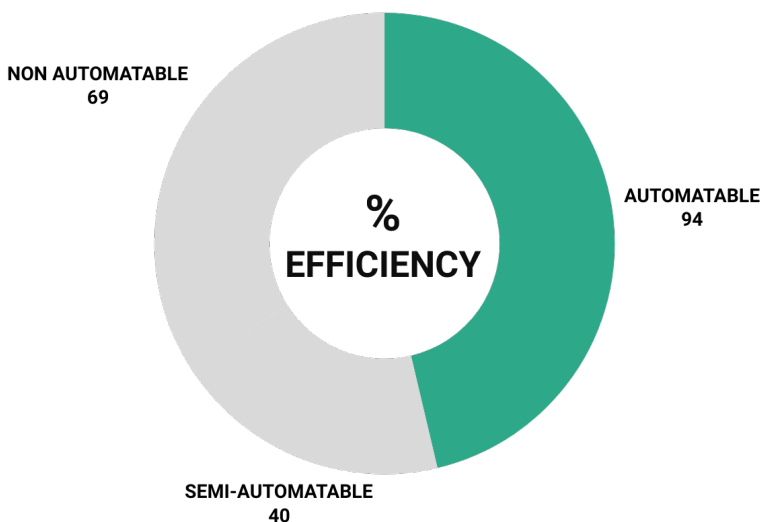
*The **94% efficiency** in automatable tasks indicates that GenAI is highly effective in clearly defined and structured tasks, which shows that GenAI can generate significant benefits in terms of efficiency and productivity in tasks that lend themselves to complete automation.*

Clarification: "Non-automatable" tasks are those that were considered that way before the impact of GenAI. In 69% of cases, these tasks become at least partially automatable with GenAI.

The efficiency of GenAI in non-automatable tasks demonstrates the potential of this technology for tasks that involve greater complexity, in terms of time commitment, precision, knowledge, analysis and sophisticated skills. In this context, it is necessary to highlight that the human must carefully and to a greater extent control the results of GenAI and, where appropriate, correct, expand, complement or directly take advantage of the result. In addition, special focus should be placed on the details, since more complex tasks usually require addressing a greater number of related subtasks.

The efficiency of GenAI in non-automatable tasks is notable, since this changes the paradigm where the highest percentage of tasks that were absorbed by AI were standardized, repetitive or automatable tasks

5. Efficiency according to the level of incidence of GenAI on the task

COMPLEMENT	REPLACEMENT	ASSISTANCE	SUBSTITUTION
94 %	**68 %**	**45 %**	**27 %**
OF EFFICIENCY	OF EFFICIENCY	OF EFFICIENCY	OF EFFICIENCY
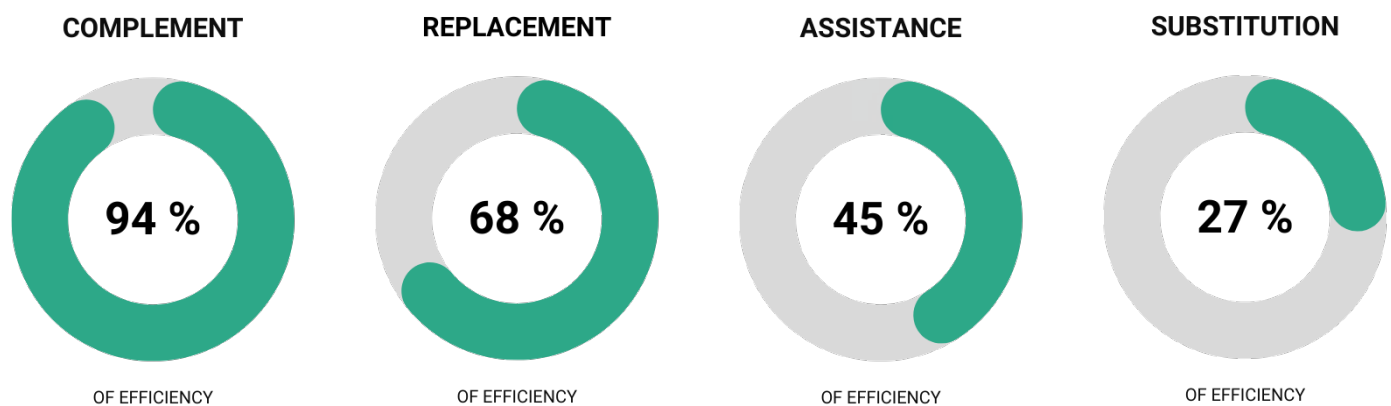
In tasks where GenAI acts as a complement, high efficiency is observed (94%), suggesting that GenAI can improve or enhance human capabilities in certain tasks.

i. Examples of GenAI efficiency in specific tasks

1. Tasks of medium complexity and high repetitiveness with high performance in terms of efficiency

- Example: Drafting an administrative agreement[52]

97 %
EFFICIENCY
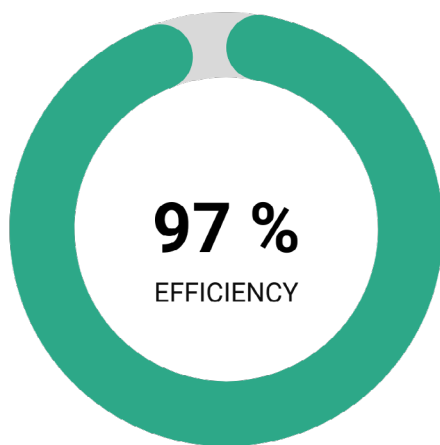
- Complexity: **MEDIUM**

- Repetitiveness: **HIGH**

- Human judgment required: **MEDIUM**

This task, being highly repetitive and of medium complexity, shows a noticeable improvement in efficiency thanks to the implementation of GenAI, which indicates a significant potential in reducing the time required to complete it.

2. Tasks of high complexity and medium repetitiveness with high performance in terms of efficiency

- Example: Drafting the facts of a claim for damages[53]

---

52   The test was carried out in a department of the Public Administration of the Autonomous City of Buenos Aires. Specifically in the Ministry of Production and Economic Development, (Legal Management) whose main task is the preparation and review of legal documents that are processed before State agencies. For these purposes, an Administrative Contract is understood as: An administrative contract is a legal agreement between a governmental or administrative entity and a private contractor or a company for the performance of works, the provision of services, the acquisition of goods or any other type of activity. related to public interests. These contracts can cover a wide range of activities, from the construction of public infrastructure to the provision of health, education, security services, among others. Its primary objective is to ensure that public resources are used efficiently and that the needs of the broader community are met.

53   The test was carried out within the Alfaro Abogados law firm. For this, the subject, data of the defendant, date, place, vehicle, among others, were provided.

74 %
EFFICIENCY
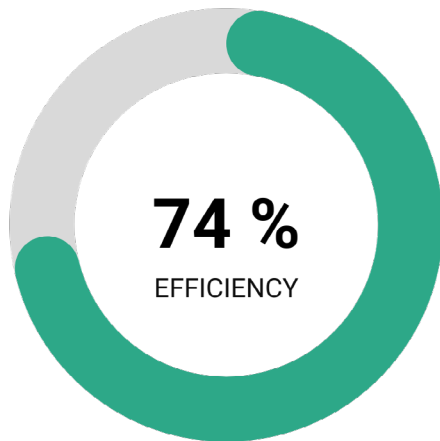
- Complexity: **HIGH**
- Repetitiveness: **MEDIUM**
- Human judgment required: **HIGH**

This test reflects the significant impact that GenAI can have on complex legal tasks, by improving efficiency in the drafting and structuring of legal documents, while on more routine and less complex tasks, its impact can be even greater.

3. Tasks of low complexity and high repetitiveness with high performance in terms of efficiency

- Example: Drafting an award resolution in a public procurement process[54]

90 %
EFFICIENCY
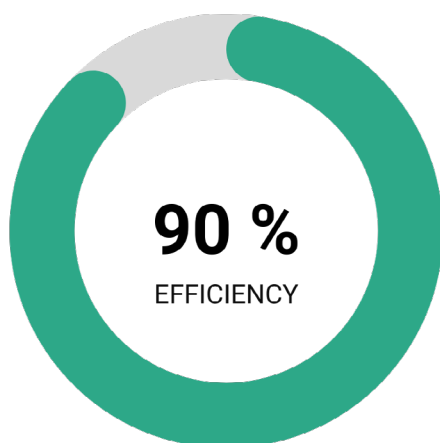
- Complexity: **LOW**
- Repetitiveness: **HIGH**
- Human judgment required: **MEDIUM**

In this case, despite GenAI's potential to automate routine, low-complexity tasks with a high level of efficiency, the medium level of human judgment required underscores the need for a balance between technology and human oversigh.

---

54   The test was carried out in a Purchasing and Contracting Department in the Public Administration of the Autonomous City of Buenos Aires. Public procurement, also known as public bidding, is the process through which public entities (governments, state-owned companies, etc.) acquire goods, services and works necessary for the functioning of their operations. Specifically public bidding. It is the most common public procurement procedure. It is based on the presentation of offers by the companies interested in the contract.

4. Balance between repetitiveness, complexity and human judgment

- Example: Offer Analysis[55]

88 %
EFFICIENCY
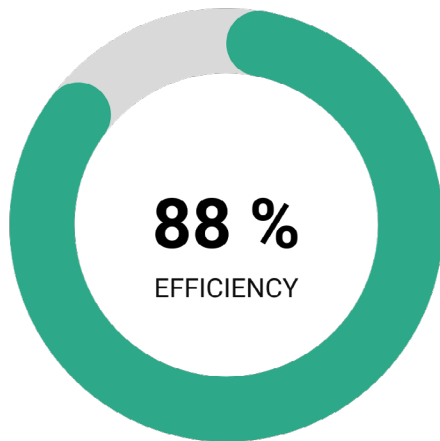
- Complexity: **MEDIUM**
- Repetitiveness: **MEDIUM**
- Human judgment required: **MEDIUM**

*The **88% efficiency** highlights GenAI's ability in tasks that require a balance between automation and human skills. This underlines its potential to improve productivity, precision and efficiency in tasks of medium complexity, reconfiguring the way these activities are approached.*

j. Skills

As explained in the methodology section, the study also covered the analysis and combination of certain skills required to carry out the tasks. To do this, the test teams selected (according to pre-established categories) those skills that predominate in each of the tasks that were tested with GenAI. Some of the results obtained are briefly presented below.

Among the skills that showed the best results were:

---

55   The test was carried out in a Purchasing and Contracting Department in the Public Administration of the Autonomous City of Buenos Aires, on a public bidding process governed by Purchasing and Contracting Law No. 2095. Specifically, the use of ChatGPT was evaluated in the analysis of offers presented in said administrative process.

Organization and processing of information

Skills related to organizing and processing information, such as choosing the best way to approach tasks, reading competently, and understanding language, have an average efficiency o **99.94%**.

Effective use of time

Skills focused on effective use of time show an efficiency of **95%**. This indicates the potential of GenAI to optimize times and improve productivity.

Analysis and organization of information

Information collection, analysis, and organization skills (planning and organization) yielded an efficiency of **90%**, suggesting the potential of GenAI for tasks involving data analysis.

Creative and abstract thinking

The skills of thinking creatively and abstractly reach an efficiency of **86.67%**. This highlights the usefulness of GenAI in supporting tasks that require innovative approaches and unconventional solutions.

Text analysis and writing skills

Skills including analysis, text writing, execution, control and management have an efficiency of **85.83%**. This indicates that GenAI is effective in tasks that combine analytical and administrative aspect.

k. Lessons learned

Regarding efficiency:

» Some tests did not reach the expected efficiency levels, which highlights the need to get used to working with GenAI tools in specific contexts.

» It is useful to measure the time taken by the task with GenAI to calculate the real optimization it means within the daily, weekly or annual workload, and define usage strategies.

» The quantitative improvements are those that we describe, essentially and mainly, in this report and that are reflected in reductions in the time necessary to complete the task.

» Qualitative improvements are those that imply an improvement in the quality of the result expected from a certain task. The latter may or may not mean saving time in carrying out the task, but they are usually very useful when it comes to increasing the competitiveness of the organization.

For example, if ChatGPT is used within a study to discover new arguments that complement existing models or templates, the time taken by the task will probably increase, but, without a doubt, the quality of the result for the client will increase.

» The incorporation of GenAI as support for carrying out a task can give rise to subtasks that did not exist before but are necessary to achieve real optimization of the chosen task. This is the case of the subtask of anonymization or pseudonymization of personal data contained in work documents. It is necessary to optimize the performance of these subtasks to achieve an improvement in total times.

» Automation should not be seen as a replacement for human judgment, but as a complement that improves efficiency and effectiveness.

Regarding the skills required:

» Challenges are observed in implementing GenAI for tasks that require complex analysis, creative thinking, and management skills, indicating current limitations in GenAI's ability to handle highly creative or critical tasks.

» The effectiveness of GenAI depends, to a large extent, on the ability of users to interact and complement these tools.

Regarding the use of GenAI:

» Not all tasks are equally suitable for automation with GenAI. A careful selection of activities is required. The tasks on which you want to apply GenAI must be selected within each institution and in relation to each of the specific processes to determine how and when to implement GenAI tools.

» The prompts must be made as "tailored suits" adapted to the particular needs of each organization.

» It is important to detect, within each task, the subtasks and microtasks that make it up, and determine those in which GenAI can be useful. Also, confirm the hypotheses with tests and measurements on them.

» GenAI can provide useful results to optimize the performance of tasks, even when used by people without technological knowledge, but with great knowledge and management of their daily work.

» It is essential to allocate time to prepare the initial prompt to describe the context and to clearly state the orders that the system must execute.

» In the medium term, if institutions intend to apply GenAI to certain use cases, it is advisable to have at least one expert in the use of these large generative language models. It is important that this profile knows the skills, discovers opportunities and is able to take certain shortcuts to reach the desired results, which includes the ability to generate prompts as templates to reuse.

» When working on use cases within an organization, improvements can be presented in both quantitative and qualitative terms.

Regarding the use of GenAI in the legal sector:

Overall, GenAI can be a useful tool for automating simple legal tasks, such as identifying grievances, applying plain language, and comparing texts. However, it always requires subsequent human supervision to ensure accuracy and legal correctness.

The quality of the product obtained depends on the prompt, but also on the complexity of the file and the quality of the text that is provided to GenAI to analyze. Consequently, in complex cases or with disorganized or poorly drafted documents, GenAI may have difficulties identifying relevant elements to achieve the desired objective (e.g. extracting grievances from a judicial resource).

Tasks that require an in-depth analysis of the file, such as preparing a tax opinion or resolving a dispute, may be more difficult to automate with GenAI due to their complexity and the

knowledge base of national and local law necessary to decide the cases.

Limitations were found in its use due to the need to anonymize data to protect the information of third parties involved in the judicial process, which reflects a limitation in this sector, when the privacy terms of the tool do not allow negotiating adjustments to extent.

The need for human supervision and editing of the result achieved is noted in most cases. Despite the good results, the importance of human control is highlighted to avoid errors or adapt responses to specific needs.

Initially, GenAI tools are not useful for performing tasks related to the calculation of procedural deadlines (for example, an expiration period in an administrative dispute), since they cannot properly execute the prompt because they make errors when performing calculation operations of deadlines.

Skills required to use GenAI tools

In most of the tests carried out in the Justice sector, the following skills were identified as relevant to be able to operate them effectively:

- Be willing to learn
- Work safely
- Use time effectively and efficiently without sacrificing quality
- Create efficient and effective prompts
- Be adaptable
- Organize, process and retain information

Finally, based on the tests carried out and results achieved, we propose to answer the three questions raised in the research purposes and problems.

First. Do ChatGPT and GenAI models tend to reduce the time needed to perform certain tasks or activities?

Evidence shows that large language models can significantly reduce the time needed to perform a wide range of tasks. The average efficiency of 77% in the times required to perform 83 tasks highlights the ability of these tools to optimize processes and increase operational efficiency in various work contexts.

Second. Do ChatGPT and GenAI models lead to the reconfiguration of certain tasks in different jobs?

The use of ChatGPT and similar GenAI tools promotes the reconfiguration of tasks in the work environment. The integration of these technologies not only facilitates the execution of existing tasks but also introduces new subtasks that did not exist before, but which now become essential for the effective optimization of work processes, such as the anonymization of confidential data or the optimization and development of prompts as templates, subtasks that appear from the implementation of large language models.

This transformation requires a change in the organization of work, with new strategies and approaches, where the adaptation and integration of GenAI become key factors for development and work efficiency. This change requires a careful selection of tasks and processes on which to apply GenAI and the adaptation of usage strategies to the specific needs of each organization.

Third. Do ChatGPT and GenAI models require workers to develop new skills? If the answer is yes, what are these new skills and which are and will be the most valued?

The implementation of ChatGPT and GenAI implies a clear requirement for workers to develop new skills. Among the most valued competencies are effective interaction with these technologies, the creation of precise and standardized prompts such as templates, supervision, critical judgment to evaluate the results generated and the adaptability to incorporate these tools into the work routine. Education plays a fundamental role in the acquisition of these skills, highlighting the importance of continuous learning and updating knowledge to make the most of the opportunities and benefits offered by new technologies in general, and GenAI in particular. Finally, the effectiveness of these tools depends largely on the ability of users to interact and complement them, underscoring the need for a collaborative approach between human and artificial intelligence.

5 NEXT STEPS

Next Steps. Presentation of Prompts As Templates

From the Innovation and Artificial Intelligence Laboratory of the Faculty of Law, UBA, we will continue with our research on the impact of GenAI at work.

The next steps consist of testing prompts as templates in different development areas. We consider that prompts that are created to optimize certain tasks and subtasks can become templates, applicable to upcoming similar tasks. The objective is to obtain standardized prompts that can be used quickly, in similar and repetitive tasks, which avoids the human activity of designing a new prompt each time a similar task must be performed.

Just as models or templates of certain standardized writings are made, it is possible to create prompt templates to save time in the execution of tasks.

From the tests carried out, it emerges that prompts as templates have the potential to increase efficiency even further. Observe the results achieved in three tasks.

Task 1 **Drafting arguments to answer a writing**	Annual time savings with prompt as a template: **170,5 hours → 29 minutes per document**	
	Monthly time savings with prompt as a template: **15,5 hours**	
	Task completion time: • **With an optimized prompt as a template: 29 minutes** • **With a non-standardized prompt: 40 minutes** • **Human activity time without GenAI: 1 hour**	Optimization: **51,67%**
	Cost to reach the result: **4 hours**	
	Recovery of time invested in work: **4 days**	

Task 2 **Improvement of arguments presented in the draft document**	Annual time savings with prompt as a template: **155 hours**	
	Monthly time savings with prompt as a template: **5 hours**	
	Task completion time: • **With optimized prompt as a template: 20 minutes** • **With unstandardized prompt: no time improvements.** • **Human activity time without GenAI: 30 minutes**	Optimization: **33,33%**
	Cost to reach the result: **3 hours 45 minutes**	
	Recovery of time invested in work: **10 days**	

Task 3 **Evaluation based on satisfaction surveys**	Monthly time savings with prompt as a template: **293 hours**	
	Task completion time: • **With optimized prompt as a template: 4 minutes** • **Human activity time without GenAI: 30 minutes**	Optimization: **86,66%**
	Cost to reach the result: **2 hours**	
	Recovery of time invested in work: **Less than a day**	

a. Drafting arguments to answer a writing

Human activity take **1 hour** of work.

- By using optimized prompts as templates, **31 minutes were saved.**

- With the first newly designed prompt test, **20 minutes were saved.**

- The efficiency with the prompt as a template was **51.67%.**

i. Costs until reaching the desired result:

- 2 tests were carried out until a use case was discovered with a prompt that allowed us to achieve an optimal result.

- 6 tests were carried out to convert the optimal prompt into a template. This work took 3 hours.

- 16 minutes and 30 seconds were allocated to reach the prompt optimized as a template.

Total time

4
HOURS
OF WORK

ii. Optimization projection:

- If it takes 1 hour without the use of ChatGPT and there are 30 demand responses made per month, 30 hours of human work are required to prepare the arguments.

- If it takes 29 minutes to prepare the arguments with ChatGPT (human activity + use of ChatGPT) and 30 demand responses are made per month, 14.5 hours are required.

In a month of work

15
HOURS
ARE
SAVED

The sacrifice involved in the 4 hours of work with ChatGPT is recovered in around 4 days.

- If it takes 1 hour without the use of ChatGPT and 330 demand responses are made per year, 330 hours of human work are required to develop the arguments.

- If it takes 29 minutes to develop the basics with ChatGPT (human activity + use of ChatGPT) and 330 demand responses are made per year, 159.5 hours are required.

In one year

170
HOURS ARE SAVED

b. Improvement of arguments presented in the draft document:

Human activity takes **30 minutes** of work.

- By using optimized prompts as templates, **10 minutes were saved.**

- With the first newly designed prompt test, there were no time improvements.

- The efficiency with the prompt as a template was **33.33%.**

i. Costs until reaching the desired result:

- 2 tests were carried out until a use case was discovered with a prompt that allowed us to achieve an optimal result.

- 5 tests were carried out to convert the optimal prompt into a template. This task took 2 hours and 45 minutes.

- 6 minutes and 50 seconds were allocated to reach the prompt optimized as a template.

Total time

3 HOURS & 45 MIN.

ii. Optimization projection:

Monthly

- If it takes 30 minutes without the use of ChatGPT and 30 demand responses are made per month that require the improvement of the arguments, it requires 15 hours of human work to prepare the arguments.

- If it takes 20 minutes to improve arguments with ChatGPT (human activity + use of ChatGPT) and 30 demand responses are made per month, it takes 10 hours..

In a month of work

5 HOURS ARE SAVED

The sacrifice involved in the 3 hours and 45 minutes of working with ChatGPT is recovered in around 10 days.

Annual

- If it takes 30 minutes without the use of ChatGPT and 330 demand responses requiring improvement of the arguments are made per year, 165 hours of human work are required to prepare the foundations.

- If it takes 20 minutes to improve the arguments with ChatGPT (human activity + use of ChatGPT) and 330 demand responses are made per year, it takes 110.5 hours.

In one year

55

HOURS ARE SAVED

c. Evaluation based on satisfaction surveys

The optimization is notable if you compare the time it takes for humans to perform the task and to perform the task using ChatGPT, through an optimized prompt. For example, in the task consisting of the evaluation of satisfaction surveys on the leadership exercised by the directors of certain areas, within an organization, with the use of ChatGPT with an optimized prompt, 26 minutes are saved. Note:

- Reading the survey and drawing up the conclusion with the human evaluation of the responses takes around **30 minutes.**

- Completing the task with ChatGPT involves 4 minutes, which includes the conversation time with the system, with the optimized prompt, the preparation of the report with conclusion and evaluation with the system and the time involved in the human subtask of copying and pasting the conclusion drawn up in the corresponding report.

- **26 minutes** are saved per survey.

- The efficiency with the prompt as a template was **86.66%.**

i. Costs until reaching the desired result:

- 4 tests were carried out until a use case was discovered with a prompt that allowed us to reach an optimal result and convert it into a template. This task lasted 2 hours.

- 10 minutes and 30 seconds were allocated to reach the prompt optimized as a template.

Total time

2

HOURS
OF WORK

ii. Optimization projection:

- f it takes 30 minutes without the use of ChatGPT and 100 evaluation conclusions are made per month, it requires 300 hours of human work.

- If it takes 4 minutes with ChatGPT (human activity + use of ChatGPT) and 100 evaluation conclusions are made per month, it requires 7 hours of human work.

In a month of work

293

HOURS ARE SAVED

The sacrifice involved in the 4 hours of work with ChatGPT is recovered in less than a day.

*Therefore, our hypothesis is that **the time invested in obtaining model prompts or templates saves time in completing future tasks**. With improved prompts, in the long and medium term, it will be possible to obtain better results in less time, which increases productivity.*

Findings and lessons learned

It is clarified that the research phase of prompts as templates has recently begun. Therefore, the findings and lessons learned that we have so far may vary as the amount of evidence increases. However, we consider that to date and, in order for more and more people to take advantage of the advantages that GenAI can bring, it is necessary to make them known.

With the tests already carried out in this new phase of the investigation, we have reached the following findings:

» It is necessary to select tasks, subtasks and microtasks that are repetitive within the organization. They can be of high, medium or low complexity.

» It is necessary to detect use cases where ChatGPT can increase efficiency.

» Once a prompt is designed, it is necessary to carry out new tests in order to improve the writing to achieve the best result.

» It is necessary to take the time to evaluate the cost-benefit involved in the sacrifice of improving the prompts that can be converted into templates, since the activity usually takes time. To do this, long-term optimization must be calculated. For example, if 20 minutes are saved in correcting the wording of a statement of claim, the calculation of time savings in completing the task must be made within a period of one month, six months and one year. Impact measurements must be carried out in the most precise

way possible, in relation to the time of use of ChatGPT and the human time that must be allocated to complete the task. The procedure that we propose is the one found in the three tables that were presented at the beginning of this item in relation to three tasks, where the time projection is shown.

» It is advisable to document the time it takes to reach the prompt that generates greater optimization.

» Converting optimal prompts into templates allows you to further increase efficiency in completing tasks. Conversation time with the conversational agent is saved and the activity, in certain complex, creative and arduous cases, of designing the prompt is saved.

» In organizations that have a large volume of work, prompts as templates function as shortcuts for the use of artificial intelligence, since they save creative work that, ultimately, can be carried out by experts in the use of ChatGPT.

» Prompts as templates allow you to reuse the work done, which led to good results. In this way, the use of ChatGPT within the organization is debureaucratized.

» Prompts as templates must be designed as tailored suits. To achieve optimization, it is necessary to design prompts as templates applicable to certain tasks, subtasks or specific microtasks, within the same category and theme. It is not recommended, at the moment, to try to reuse a prompt - template designed for one task within a certain subject, to another. In any case, it is advisable to carry out the corresponding tests in this regard.

» Prompts as templates are useful for those people who do not want to be trained in the use of ChatGPT, since for reuse, it is enough to copy and paste said template in the conversation and continue with the indicated iterations. It is recommended to prepare user guides that indicate the steps to follow for using prompts as templates in ChatGPT or other GenAI tools.

» The development of prompts as templates requires human dedication. In addition to the work of creating specific, detailed and contextualized prompts, there is the task of converting them into templates, testing them in new similar tasks and continually improving them.

» In the long or medium term, it is considered that it will be useful for organizations to have a person who specializes in the use of ChatGPT and in the development and continuous improvement of the prompts - templates that each area requires depending on its tasks. The tasks of discovering use cases, developing optimized prompts, designing templates, continuous improvement and developing use guides require experience and specialization.

» Prompts as templates tend to reduce the possibilities of ChatGPT delusions and generate the conditions to try to reach similar results in new cases. However, it is necessary to evaluate the performance of the prompt as a template periodically and, if necessary, make the corresponding adjustments.

» Prompts as templates should be used in both use cases in which ChatGPT has the potential to achieve qualitative and quantitative improvements. Qualitative improvements can occur in two ways. On the one hand, they impact the quality of the task being performed and improve the result, which is more detailed, complete and precise than what is achieved humanly. On the other hand, qualitative improvements imply, in many cases, an increase in the tasks performed by the conversational agent when releasing the result, which were not performed by humans. For example, in the use case corresponding to the preparation of the conclusion in satisfaction surveys, the result that ChatGPT produced consisted of a subtitled report, where: a general summary was made, the strengths that emerged from the responses to the survey, areas of improvement, suggestions for improvements and a short closing. The conclusions drawn by humans were limited to reiterating issues raised in the responses, without subtitles or further development. Quantitative improvements are those that reduce the time it takes to complete the task. It is possible that in a case there will be both quantitative improvements and qualitative improvements. Our hypothesis, at the moment, is that, in the long term, qualitative improvements also generate quantitative improvements..

ANNEX.
RESULTS

ANNEX I. Results

In this Annex we present in detail the results of our research, categorizing according to the different sectors where the measurements were carried out.

We specify, for each of the 83 analyzed tasks, the time necessary for its execution with and without GenAI.

We also describe the level of complexity, the degree of human judgment required, the repetitiveness, and the degree of automation of these 83 tasks.

a. Sector and areas involved in the research

1. Public Administration

The tests below were carried out within the scope of the Legal and Technical Direction of the Ministry of Economic Development of the City of Buenos Aires, in the legal, sales and contracting, and human resources departments.

Total tasks analyzed	**13**
Tasks on which measurements were made	• Drafting of an administrative contract • Offer analysis • Review of Particular Bases and Conditions Documents • Legal control of an agreement • Drafting of the administrative act of approval of a competition • Selection process with analysis of 50 CVs for Senior Legal Advisor position • Supervision of the administrative act draft for retroactive dismissal of the gross income tax • Legal opinion • Drafting of the award resolution in a public contracting process of the National Telecommunications Corporation (CNT EP) • Drafting of award resolution in a public procurement process
Efficiency	**70%**
Complexity level	High: 2 tasks Medium: 9 tasks Low: 2 tasks
Degree of automation	Assistance: 1 tasks Semi-automatable: 12 tasks
Human judgment required	High: 3 Medium: 10
Repetitiveness	High: 7 tasks Medium: 3 tasks Low: 3 tasks
Incidence of GenAI	Assistance: 11 tasks Unassigned: 2 tasks
Efficiency according to the level of complexity	High: 75% Medium: 74% Low: 87%
Efficiency according to the level of human judgment required	High: 31% Medium: 76%
Efficiency according to the level of repetitiveness	High: 87% Medium: 75% Low: 16%

2. Justice

The tests below were carried out by different judicial branches of Argentina. Specifically by teams made up of judges, officials from the provinces of Mendoza, San Juan, Tierra del Fuego, Misiones, Tucumán, Buenos Aires and the Autonomous City of Buenos Aires.

Total tasks analyzed	29
Tasks on which measurements were made	• Analysis and comparison of writings • Identification of grievances in an appeal • Identification of grievances in an extraordinary provincial appeal presented against rulings of the Labor Chamber • Identify grievances in an unconstitutionality appeal • First decree presenting an extraordinary provincial appeal • Apply clear language in a vote • Apply clear language in a sentence • Preparation of trades from a model • Confrontation of trades • Extraction of relevant information from a claim • Preparation of an email in clear language that informs the internal procedure of the Specialized Consumer Relations Unit and explains the ordinary process of the Procedural Code of Justice in Consumer Relations that will be followed • Preparation of a tax ruling on severance • Preparation of a tax opinion on periodic penalty payments • Identify the object of a claim • Order the facts chronologically and express them in legal language • Propose arguments to respond to the transfer of a response to the claim • Computation of expiration period in a contentious action • Extraction of relevant information from a claim • Drafting a claim for damages for state responsibility in the context of gender violence • Analysis of a case of periodic penalty payment • Analysis of the prosecutor's opening argument in a criminal trial • Analysis of examination and cross-examination of a forensic doctor within the framework of a jury trial • Summary of a sentence and generation of relevant information table • Extraction of legal reasoning from a section of another sentence to incorporate into a project • Extraction of specific information from testimonial statements

Efficiency	**62%**
Complexity level	High: 11 tasks Medium: 14 tasks Low: 4 tasks
Degree of automation	Assistance: 10 Semi-automatable: 8 No Assistance: 1
Human judgment required	High: 17 Medium: 2 Low: 8 Unassigned: 2
Repetitiveness	High: 16 tasks Medium: 7 tasks Low: 4 tasks Unassigned: 2 tasks
Incidence of GenAI	Assistance: 19 tasks Complement: 3 tasks Sustitución: 3 tasks Unassigned: 4 tasks
Efficiency according to the level of complexity	High: 77% Medium: 38% Low: 28%
Efficiency according to the level of human judgment required	High: 65% Medium: 57% Low: 44%
Efficiency according to the level of repetitiveness	High: 45% Medium: 71% Low: 67%

3. Law firms / legal departments of companies

Total tasks analyzed	**27**
Tasks on which measurements were made	• Control of reductions according to attribution matrix • Sanitation control with corresponding approval level and justification • Disposal analysis • Writing of the disposal report • Writing an administrative release • Copy with amendments of a lease contract • Writing of the agreement payment's receipt • Think about the test / Test offer / Position statement • Preparation of a response to a labor statement of claim • Production of a testimony • Drafting a claim for breach of contract within the framework of a consumer relationship (Automotive Savings Plans) • Statement of positions for drafting a claim for breach of contract within the framework of a consumer relationship (Automotive Savings Plans) • Email writing. (Content: Information on the status of the case for the client.) Drafting of a claim for breach of contract within the framework of a consumer relationship (Automotive Savings Plans) • Review of Conflict of Interest procedure • Identify controls to be carried out to verify compliance with service level agreements in contracts with suppliers • Preliminary extraction of obligations from a Contract • Creation of a standard contractual letter (notify an Event, generate a contractual reservation) • Analysis of the Report Received for a Tax Violation - Comparison between two similar reports (same violation) but that dealt with different facts • Analysis of the tax resolution received • Preparation of judicial document • Preparation of lease agreement • Writing a labor document letter • Preparation of payment order • Preparation of facts of a claim (damages) • Labour testimonial evidence. Preparation of document for witness • Testimony evidence: preparation of documents for the benefit of litigating without expenses
Efficiency	**68%**

Complexity level	High: 9 tasks Medium: 12 tasks Low: 6 tasks
Degree of automation	Assistance: 8 Semi-automatable: 11 No Assistance: 8
Human judgment required	High: 15 Medium: 9 Low: 3
Repetitiveness	High: 11 tasks Medium: 5 tasks Low: 11 tasks
Incidence of GenAI	Assistance: 10 tasks Complement: 9 tasks Sustitución: 7 tasks Unassigned: 1 tasks
Efficiency according to the level of complexity	High: 73% Medium: 62% Low: 50%
Efficiency according to the level of human judgment required	High: 71% Medium: 57% Low: 44%
Efficiency according to the level of repetitiveness	High: 55% Medium: 63% Low: 72%

4. Translation

Total tasks analyzed	2
Tasks on which measurements were made	• Assembly of the glossary • First draft translation of a text of 15,000 words
Efficiency	**99,9%**

Complexity level	Medium: 2 tasks
Degree of automation	Assistance: 2 tasks
Human judgment required	High: 1 Medium: 1
Repetitiveness	High: 1 task Medium: 1 task
Incidence of GenAI	Complement: 2 tasks

5. Education

Total tasks analyzed	**4**
Tasks on which measurements were made	• Selection of the Program Unit to Work • Propose practical activities for students of the selected unit • Select the appropriate practical activity • Plan the selected practical activity
Efficiency	**51%**
Complexity level	High: 1 task Medium: 3 tasks
Degree of automation	No Assistance: 4 tasks
Human judgment required	High: 4 tasks
Repetitiveness	Low: 4 tasks
Incidence of GenAI	Assistance Complement: 2 tasks

b. Time optimization

The time required to complete the task with and without the use of GenAI tools is analyzed below. It also shows the level of efficiency achieved in each case.

1. Public Administration

Tasks	Task completion time without GenAI	Task completion time with GenAI	Efficiency
Offer Analysis	249	30	88%
Review of Particular Bases and Conditions Documents	95	55	42%
Legal control of agreement	61	38	38%
Drafting of the administrative act of approval of a competition	34	49	-44%
Selection process with Analysis of 50 CVs for Senior Legal Advisor position	30	18	40%
Supervision of the administrative act draft for retroactive dismissal of the gross income tax	20	7	65%
Legal Opinion	15	17	-13%
Drafting of the award resolution in a public contracting process of the National Telecommunications Corporation(CNT EP)	30	5	83%
Drafting an award resolution in a public procurement process	30	3	90%
Drafting an administrative agreement	60	6	90%
Drafting an administrative agreement	60	6	90%
Drafting an administrative agreement	60	3	95%
Drafting an administrative agreement	60	2	97%

2. Justice

Tasks	Task completion time without GenAI	Task completion time with GenAI	Efficiency
Analysis and comparison of writing	20	25	-25%
Identification of grievances in an appeal (second test)	30	15	50%
Individualization of grievances in extraordinary provincial appeal presented against rulings of the Labor Chamber	120	35	71%
First decree presentation of Extraordinary Provincial Appeal	30	15	50%
Identification of grievances in an appeal (second test)	30	15	50%
Apply clear language in a vote (second test)	45	15	67%
Preparation of trades from a model	5	10	-100%
Confrontation of trades (second test)	10	10	0%
Extraction of relevant information from a claim	20	20	0%
Preparation of an email in clear language that informs the internal procedure of the UERC and explains the ordinary process of the CPJRC that will be followed	80	45	44%
Apply clear language in the sentence	60	10	83%
Order the facts chronologically and express them in legal language	40	15	63%
Propose arguments to respond to the transfer of a response to the claim	60	20	67%
Calculation of expiration period in contentious administrative action	20	30	-50%
Identification of grievances in an appeal (second test)	30	15	50%

Apply clear language in a vote (second test)	45	15	67%
Confrontation of trades (second test)	10	10	0%
Extraction of relevant information from a claim	20	20	0%
Preparation of tax ruling on severance	240	36	85%
Preparation of tax opinion on periodic penalty payments	240	32	87%
Identify the object of the claim	2	3	-50%
Identify grievances in the unconstitutionality appeal	25	2.5	90%
Drafting a claim for damages for state responsibility in the context of gender violence	60	23.25	61%
Analysis of a case of periodic penalty payments	90	23.75	74%
Analysis of the prosecutor's opening argument in a criminal trial	34	12.75	63%
Analysis of examination and cross-examination of a forensic doctor in the framework of a jury trial	31	12.25	60%
Sentence summary and generation of relevant information table	10	5	50%
Extracting legal reasoning from a section of another ruling to incorporate into a project	60	90	-50%
Extraction of specific information from testimonial statements	120	30	75%

3. Law firms / legal departments

Tasks	Task completion time without GenAI	Task completion time with GenAI	Efficiency
Preparation of judicial document	22	9	59%
Preparation of lease agreement	30	17	43%
Drafting a labor document letter	20	9	55%
Preparation of payment order	15	7	53%
Preparation of statement of claim	35	9	74%
Labor testimonial evidence. Preparation of documents for witnesses	22	8	64%
Testimonial evidence: Preparation of documents for a Benefit of Litigation without Expenses	25	4	84%
Analysis of Reports Received for a Tax Violation - Comparison between two similar reports (same violation) but that dealt with different facts	30	11	63%
Analysis of Tax Resolution Received	60	20	67%
Review of Conflict of Interest procedure	30	6	80%
Identify controls to be carried out to verify compliance with service level agreements in contracts with suppliers	240	70	71%
Preliminary extraction of obligations from a contract	250	120	52%
Preparation of a standard contractual letter (notify an Event, generate a contractual reservation)	30	2	93%
Choose names for a product	120	16	87%
Drafting an administrative release	5	45	-800%
Copy with amendments of a lease agreement	120	6	95%
Drafting the agreement payment's receipt	15	3.5	77%

Think about the test / Test offer / Position statement	45	50	-11%
Preparation of a response to a labor statement of claim	62	36	42%
Making a testimonial	35	11	69%
Drafting a claim for breach of contract within the framework of a consumer relationship (Automotive Savings Plans)	960	136	86%
Statement of positions for drafting a claim for breach of contract within the framework of a consumer relationship (Automotive Savings Plans)	90	40	56%
Email writing. (Content: Information on the status of the case for the client.) Drafting of a claim for breach of contract within the framework of a consumer relationship (Automotive Savings Plans)	15	45	73%
Control of reductions according to attribution matrix	45	33	27%
Sanitation control with corresponding approval level and justification	60	20	67%
Analysis of reductions	60	50	17%
Drafting the reductions report	90	65	28%

4. Translation

Tasks	Task completion time without GenAI	Task completion time with GenAI	Efficiency
Assembling the glossary	2400	2	100%
First draft translation of a 15,000-word text	2400	1	43%

5. Education

Tasks	Task completion time without GenAI	Task completion time with GenAI	Efficiency
Selection of the Program Unit to Work	7	8	-14%
Propose practical activities for students of the selected unit	17	10	41%
Select the appropriate practical activity	2	2	0%
Plan the selected practical activity	35	10	71%

6. Investigation

Tasks	Task completion time without GenAI	Task completion time with GenAI	Efficiency
Definition of a research project	80	570	-613%

c. Complexity level de las tasks

1. Highly complex tasks

The highly complex tasks on which measurements were made are listed below:

Area	Tasks	Complexity level
Public Administration	Legal control of agreement	High
	Legal Opinion	
Education	Plan the selected practical activity	High
Law firms / legal departments	Analysis of reductions	High
	Preparation of a response to a labor statement of claim	
	Drafting a claim for breach of contract within the framework of a consumer relationship (Automotive Savings Plans)	
	Review of Conflict of Interest procedure	
	Identify controls to be carried out to verify compliance with service level agreements in contracts with suppliers	
	Preliminary extraction of obligations from a Contract	
	Analysis of Tax Resolution Received	
	Preparation of a leasing agreement	
	Preparation of a statement of claim	

Justice	Identification of grievances in an appeal (second test)	High
	Individualization of grievances in extraordinary provincial appeal presented against rulings of the Labor Chamber	
	Identification of grievances in an appeal (second test)	
	Identification of grievances in an appeal (second test)	
	Preparation of tax ruling on severance	
	Preparation of tax opinion on periodic penalty payments	
	Identify the object of the claim	
	Identify grievances in the unconstitutionality appeal	
	Drafting a claim for damages for state responsibility in the context of gender violence	
	Analysis of a case of periodic penalty payment	
	Extraction of specific information from testimonial statements	

2. Medium complexity tasks

The tasks of medium complexity on which the measurements were made are listed below:

Area	Tasks	Complexity level
Public Administration	Offer Analysis	Medium
	Review of Particular Bases and Conditions Documents	
	Drafting of the administrative act of approval of the competition	
	Selection process with Analysis of 50 CVs for Senior Legal Advisor position	
	Supervision of the draft administrative act for retroactive cessation of the gross income tax	
	Drafting an administrative agreement	
	Drafting an administrative agreement	
	Drafting an administrative agreement	
	Drafting an administrative agreement	
Education	Selection of the Program Unit to Work	Medium
	Propose practical activities for students of the selected unit	
	Select the appropriate practical activity	

Law firms / legal departments	Making a testimonial	Medium
	Preparation of a standard contractual letter (notify an Event, generate a contractual reservation)	
	Choose names for a product	
	Analysis of reports Received for a Tax Violation - Comparison between two similar reports (same violation) but that dealt with different facts	
	Preparation of a judicial document	
	Preparation of payment order	
	Labor testimonial evidence. Preparation of document for witness	
	Testimonial evidence: Preparation of documents for a Benefit of Litigation without Expenses	
	Control of reductions according to attribution matrix	
	Sanitation control with corresponding approval level and justification	
	Drafting the contract payment receipt	
	Think about the test / Test offer / Position statement	
Investigation	Definition of a research project	Medium

Justice	Analysis and comparison of writing	**Medium**
	First decree presentation of Extraordinary Provincial Appeal	
	Apply clear language in a vote (second test)	
	Extract relevant information from a claim	
	Preparation of an email in clear language that informs the internal procedure of the UERC and explains the ordinary process of the CPJRC that will be followed	
	Apply clear language in the sentence	
	Order the facts chronologically and express them in legal language	
	Propose arguments to respond to the transfer of a response to the claim	
	Calculation of expiration period in contentious administrative action	
	Apply clear language in a vote (second test)	
	Extract relevant information from a claim	**Medium**
	Analysis of examination and cross-examination of a forensic doctor in the framework of a jury trial	
	Sentence summary and generation of relevant information table	
	Extracting legal reasoning from a section of another ruling to incorporate into a project	
Translation	Assembling the glossary	**Medium**
	First draft translation of a 15,000-word text	

3. Low complexity tasks

The low complexity tasks on which the measurements were made are listed below:

Area	Tasks	Complexity level
Public Administration	Drafting of the award resolution in a public contracting process of the National Telecommunications Corporation (CNT EP)	Low
	Drafting an award resolution in a public procurement process	
Law firms / legal departments	Drafting a labor document letter	Low
	Drafting the disposal report	Low
	Drafting an administrative release	
	Copy with modifications of a lease contract	
	Statement of positions for drafting a claim for breach of contract within the framework of a consumer relationship (Automotive Savings Plans)	
	Email writing. (Content: Information on the status of the case for the client.) Drafting of a claim for breach of contract within the framework of a consumer relationship (Automotive Savings Plans)	
Justice	Preparation of trades from a model	Low
	Confrontation of trades (second test)	
	Confrontation of trades (second test)	
	Analysis of the prosecutor's opening argument in a criminal trial	

d. Level of human judgment required

1. High human judgment

The high-level human judgment tasks on which measurements were made are listed below:

Area	Tasks	Human judgment required
Public Administration	Legal control of agreement	High
	Drafting of the award resolution in a public contracting process of the National Telecommunications Corporation (CNT EP)	
	Drafting an award resolution in a public procurement process	
Education	Plan the selected practical activity	High
	Selection of the Program Unit to Work	
	Propose practical activities for students of the selected unit	
	Select the appropriate practical activity	
Law firms / legal departments	Analysis of reductions	High
	Preparation of a standard contractual letter (notify an Event, generate a contractual reservation)	
	Choose names for a product	
	Labor testimonial evidence. Preparation of document for witness	
	Drafting a claim for breach of contract within the framework of a consumer relationship (Automotive Savings Plans)	

Law firms / legal departments	Testimonial evidence: Preparation of documents for a Benefit of Litigation without Expenses	High
	Review of Conflict of Interest procedure	
	Identify controls to be carried out to verify compliance with service level agreements in contracts with suppliers	
	Preliminary extraction of obligations from a Contract	
	Analysis of Tax Resolution Received	
	Preparation of a statement of claim	
	Drafting a release report	
	Statement of positions for drafting a claim for breach of contract within the framework of a consumer relationship (Automotive Savings Plans	
	Drafting the contract payment receipt	
	Think about the test / Test offer / Position statement	
Justice	Analysis and comparison of writings	High
	Extraction of relevant information from a claim	
	Identification of grievances in an appeal (second test)	
	Propose arguments to respond to the transfer of a response to the claim	
	Individualization of grievances in extraordinary provincial appeal presented against rulings of the Labor Chamber	
	Identification of grievances in an appeal (second test)	
	Identification of grievances in an appeal (second test)	
	Preparation of tax ruling on on severance	

Justice	Preparation of tax opinion on periodic penalty payments	High
	Identify the object of the claim	
	Extraction of relevant information from a claim	
	Identify grievances in the unconstitutionality appeal	
	Drafting a claim for damages for state responsibility in the context of gender violence	
	Sentence summary and generation of relevant information table	
	Analysis of a case of periodic penalty payments	
	Extracting legal reasoning from a section of another ruling to incorporate into a project	
	Extraction of specific information from testimonial statements	
Translate	First draft translation of a 15,000-word text	High

2. Medium human judgment

The medium-level human judgment tasks on which the measurements were made are listed below:

Area	Tasks	Human judgment required
Public Administration	Offer Analysis	Medium
	Review of Particular Bases and Conditions Documents	
	Drafting of the administrative act of approval of a competition	
	Supervision of the administrative act draft for retroactive cessation of the gross income tax	
	Drafting an administrative agreement	
	Drafting of the award resolution in a public contracting process of the National Telecommunications Corporation (CNT EP)	
	Drafting an award resolution in a public procurement process	
	Drafting an administrative agreement	
	Drafting an administrative agreement	
	Drafting an administrative agreement	
Law firms / legal departments	Making a testimonial	Medium
	Preparation of a response to a labor demand	
	Drafting a labor document letter	
	Analysis of reports Received for a Tax Violation - Comparison between two similar reports (same violation) but that dealt with different facts	

Law firms / legal departments	Drafting an administrative release	Medium
	Preparation of payment order	
	Copy with modifications of a lease contract	
	Sanitation control with corresponding approval level and justification	
	Preparation of lease agreement	
Investigation	Definition of a research project	Medium
Justice	First decree presentation of Extraordinary Provincial Appeal	Medium
	Order the facts chronologically and express them in legal language	
Translate	Assembling the glossary	Medium

3. Low human judgment

The low-level human judgment tasks on which the measurements were made are listed below:

Area	Tasks	Human judgment required
Law firms / legal departments	Preparation of a judicial document	Low
	Control of reductions according to attribution matrix	
	Email writing. (Content: Information on the status of the case for the client.) Drafting of a claim for breach of contract within the framework of a consumer relationship (Automotive Savings Plans)	
Justice	Preparation of trades from a model	Low
	Preparation of an email in clear language that informs the internal procedure of the UERC and explains the ordinary process of the CPJRC that will be followed	
	Apply clear language in the sentence	
	Confrontation of trades (second test)	
	Confrontation of trades (second test)	
	Calculation of expiration period in contentious administrative action	
	Analysis of examination and cross-examination of a forensic doctor in the framework of a jury trial	
	Analysis of the prosecutor's opening argument in a criminal trial	

e. Level of task repetitiveness

1. High repetitiveness

The highly repetitive tasks on which the measurements were made are listed below:

Area	Tasks	Level of repetitiveness
Public Administration	Legal Opinion	High
	Drafting an administrative agreement	
	Drafting of the award resolution in a public contracting process of the National Telecommunications Corporation (CNT EP)	
	Drafting an award resolution in a public procurement process	
	Drafting an administrative agreement	
	Drafting an administrative agreement	
	Drafting an administrative agreement	
Law firms / legal departments	Making a testimonial	High
	Preparation of a labor statement of claim response	
	Analysis of reports received for a Tax Violation - Comparison between two similar reports (same violation) but that dealt with different facts	
	Drafting a labor document letter	

Law firms / legal departments	Preparation of a judicial document	High
	Preparation of payment order	
	Analysis of Tax Resolution Received	
	Sanitation control with corresponding approval level and justification	
	Email writing. (Content: Information on the status of the case for the client.) Drafting of a claim for breach of contract within the framework of a consumer relationship (Automotive Savings Plans)	
Justice	Preparation of trades from a model	High
	Analysis and comparison of writing	
	First decree presentation of Extraordinary Provincial Appeal	
	Extraction of relevant information from a claim	
	Preparation of an email in clear language that informs the internal procedure of the UERC and explains the ordinary process of the CPJRC that will be followed	
	Apply clear language in the sentence	
	Identification of grievances in an appeal (second test)	
	Order the facts chronologically and express them in legal language	

Justice	Confrontation of trades (second test)	High
	Individualization of grievances in extraordinary provincial appeal presented against rulings of the Labor Chamber	
	Identification of grievances in an appeal (second test)	
	Identification of grievances in an appeal (second test)	
	Confrontation of trades (second test)	
	Calculation of expiration period in contentious administrative action	
	Extraction of relevant information from a claim	
	Sentence summary and generation of relevant information table	
Translate	First draft translation of a 15,000-word text	High

2. Medium repetitiveness

The medium repetitiveness tasks on which the measurements were made are listed below:

Area	Tasks	Level of repetitiveness
Public Administration	Offer Analysis	Medium
	Review of Particular Bases and Conditions Documents	
	Supervision of the draft administrative act for retroactive cessation of the gross income tax	
	Preparation of a standard contractual letter (notify an Event, generate a contractual reservation)	
Law firms / legal departments	Labor testimonial evidence. Preparation of document for witness	Medium
	Identify controls to be carried out to verify compliance with service level agreements in contracts with suppliers	
	Preparation of a statement of claim	
	Drafting the release report	

Justice	Propose arguments to respond to the transfer of a response to the claim	Medium
	Preparation of tax ruling on severance	
	Preparation of tax opinion on periodic penalty payment	
	Identify the object of the claim	
	Identify grievances in the unconstitutionality appeal	
	Extracting legal reasoning from a section of another ruling to incorporate into a project	
	Extraction of specific information from testimonial statements	
Translation	Assembling the glossary	Medium

3. Low repetitiveness

The low repetitiveness of tasks on which the measurements were made are listed below:

Area	Tasks	Level of repetitiveness
Public Administration	Legal control of an agreement	Low
	Drafting of the administrative act of approval of the competition	
	Selection process with Analysis of 50 CVs for Senior Legal Advisor position	
Education	Plan the selected practical activity	Low
	Selection of the Program Unit to Work	
	Propose practical activities for students of the selected unit	
	Select the appropriate practical activity	
Law firms / legal departments	Analysis of removals	Low
	Choose names for a product	
	Drafting a claim for breach of contract within the framework of a consumer relationship (Automotive Savings Plans)	
	Testimonial evidence: Preparation of documents for a Benefit of Litigation without Expenses	
	Review of Conflict of Interest procedure	
	Preliminary extraction of obligations from a Contract	

Law firms / legal departments de empresas	Drafting an administrative release	Low
	Copy with modifications of a lease agreement	
	Statement of positions for drafting a claim for breach of contract within the framework of a consumer relationship (Automotive Savings Plans)	
	Drafting the contract payment receipt	
	Think about the test / Test offer / Position statement	
Investigation	Definition of a research project	Low
Justice	Analysis of examination and cross-examination of a forensic doctor in the framework of a jury trial	Low
	Drafting a claim for damages for state responsibility in the context of gender violence	
	Analysis of a case of periodic penalty payments	
	Analysis of alleged fiscal opening in criminal trial	

ANNEX

ANNEX II

Next, we describe in detail the procedure carried out to measure the tasks on which the impact of GenAI tools was analyzed.

Public Administration

After the UBA IALAB team defined the objectives and scope of the research, within the scope of the General and Technical Administrative and Legal Directorate of the Ministry of Economic Development and Production of the City of Buenos Aires, an interdisciplinary team of volunteer collaborators was convened (7 people in total).

The areas of expertise on which the tests were carried out were: Purchasing and Contracting, Legal and Human Resources. The selected tasks were: "Analysis of Offers", "Legal Control of Agreements" and "Selection Process with Analysis of CVs for a given position".

Specific tasks were assigned to the collaborators of the Legal and Technical Administrative Direction (according to their area of expertise) and according to their skills and knowledge. Some were in charge of pre-selection of tasks relevant to the topic, while others focused on time measurement and content review, all supervised by a general coordination link.

Deadlines were established and partial deliveries agreed to maintain a constant flow of information and minimize delays.

Throughout the process of preparing the research and work tests, periodic meetings were held to discuss progress, solve doubts and make adjustments according to the identified needs. These meetings also allowed the team to have visibility of the progress and work collaboratively to improve the quality of the report.

The volunteers, between 25 and 40 years old, were initially enthusiastic about the project. As the first tests were carried out, the first questions arose about how to interact more efficiently with ChatGPT: "Am I doing it right?" ; fears about the scope of the task performed: "Are you going to replace me? " and a certain "reluctance" to continue moving forward with the project linked to those fears.

At an operational level, the delay in obtaining the desired result was questioned (need to adapt the prompt) "I waste a lot of time until it "understands" what I want, I do it directly", being inevitable to assign " personality" characteristics to the tool.

With the first results, the volunteers began to understand that it could be a collaborative or support tool when developing their tasks, becoming alert in cases in which the tool replaces or substitutes the human task.

They detected that the challenge is to work on the knowledge of tools of this type to add them to their professional development as well as to strengthen the development of soft skills.

The extent of the tests carried out focused on deepening specific aspects of the topic discussed, providing a detailed analysis and greater contextualization for the analyst. We worked on an Excel matrix that was improved in order to facilitate the analysis of the results presented.

As a final stage, after a first return to the IALAB team, an exhaustive review of the matrix was carried out. The coherence of the information was verified and errors of interpretation were corrected in general terms. Additionally, the accuracy and consistency of the data presented, as well as the quality of the content in general, was ensured.

Justice

The tests to identify the object of a claim for periodic penalty payments and the identification of grievances in an appeal for unconstitutionality of a severance case were carried out in the Deputy Attorney General's Office for Contentious, Administrative, Tax and Consumer Relations of the Public Prosecutor's Office of the Autonomous City of Buenos Aires.

They were in charge of a judicial employee with the position of administrative assistant, who specializes in preparing different types of opinions on the matter before the Superior Court of Justice of the Autonomous City of Buenos Aires.

The tests consisted on asking ChatGPT through different prompts to carry out an exhaustive analysis of a statement of claim, in one case, and to identify its object in a detailed and precise way. In another case, the tool was asked to analyze an unconstitutionality appeal and identify its grievances.

In both cases, the tests carried out yielded successful results, meeting the expected objectives and without having to redo any of the initial prompts.

To use only one of the two cases as an example, the time it takes to carry out the human reading of an unconstitutionality appeal to correctly identify the grievances raised is

approximately 25 minutes. With this GenAI tool, that time was shortened to 2 minutes and 30 seconds, that is, a 90% time optimization.

Translate

The glossary assembly tests and the draft translation of a 15,000-word text were carried out by an English translator, who collaborated with the Laboratory for these purposes.

The tests consisted of asking both MateCat[56] and ChatGPT—in this case, through different prompts—to perform an analysis of a text, in one case, and extract different terms to integrate a glossary. In another case, the tools were asked to analyze a text and translate it.

After a few iterations, in both cases, the tests performed produced successful results.

For the two tests carried out, about 2,400 minutes (40 hours) were taken as the estimated time to complete the tasks. With these generative AI tools, that time was shortened to 2 minutes, that is, reaching almost 100% time optimization. So we can conclude that they are tools that serve as a complement.

---

56   MateCat is free translation software based on the limitation of segments within the translation project and the maximum optimization of word repetition.